Journal of Memory and Language 142 (2025) 104625

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Investigating the cognitive correlates of semantic and perceptual false memory in older and younger adults: A multi-group latent variable approach☆

John T. West[a,*], Rebecca L. Wagner[a], Ashley Steinkrauss[b], Nancy A. Dennis[a]

[a] Department of Psychology, Pennsylvania State University, University Park, PA, United States
[b] Department of Psychology, Boston College, Chestnut Hill, MA, United States

ARTICLE INFO

Keywords:
False Memory
Aging
Individual Differences

ABSTRACT

Falsely remembering information can have negative consequences for day-to-day functioning and can be especially problematic for older adults who often experience higher rates of false memory. Because there is considerable variability between older adults in memory and cognition, it is essential that we understand the factors that place older individuals at risk for developing false memories. Whereas lower frontal functioning has previously been related to false memory in general, prior research suggests that there may also be domain-specificity in the factors associated with false memory. To test this possibility, 211 young adults and 152 older adults completed tasks measuring semantic false memory, perceptual false memory, frontal functioning, semantic discrimination, and perceptual discrimination. Factor analyses revealed that – contrary to predictions – individual differences in semantic and perceptual false memory were best represented by a single, overarching false memory factor. Although cognitive abilities were generally not related to false memory when assessed together, semantic discrimination, perceptual discrimination, and frontal functioning were all negatively associated with false memory in isolation, and jointly predicted 37% of the variance in younger adults and 40% in older adults. Importantly, the extent to which these cognitive abilities protected against false memory did not differ between older and younger adults. Results suggest that for both older and younger adults, individual differences in the tendency to falsely remember information are captured by a single overarching construct that has negative (yet redundant) associations with various cognitive abilities.

Because human memory is imperfect, memory errors are a reality of everyday life. One type of memory error that has received considerable empirical attention is false memory, where individuals remember events differently than they occurred or remember events that never occurred (Roediger & McDermott, 2000). Although false memory is a concern for individuals of all ages, they are both increasingly common and problematic for older adults (McCabe et al., 2009). Even in the absence of pathological disorders of aging such as Alzheimer's disease, false memories may pose a critical threat to otherwise healthy older adults' independence, safety, and well-being. For instance, older adults may fall victim to the harmful effects of false memory when attempting to remember a loved one's name, or whether they have taken an important medication. Given these concerns, it is essential that memory researchers understand the factors associated with false memory formation in older adults.

Another important consideration is that older adults are not equally susceptible to false memory. In light of these individual differences, as well as research demonstrating that older adults are more variable in terms of cognition generally speaking (Hertzog, 1985; Lindenberger & Oertzen, 2006; Morse, 1993; Nyberg et al., 2012), there is a pressing need to identify the factors that make certain older individuals vulnerable to false memories. Although prior research has revealed much about the neural and cognitive processes underlying false memory in older adults (Dennis et al., 2008, 2014; for a review, see Devitt & Schacter, 2016), relatively little is known about how aging affects the individual-level factors associated with false memory. Critically, it is also unknown whether the cognitive factors associated with false memory in older adults are the same or different for semantic and

☆ This article is part of a special issue entitled: 'Individual differences in memory' published in Journal of Memory and Language.

* Corresponding author at: Pennsylvania State University, Moore Building, University Park, PA 16802, United States.
E-mail address: johnwest@psu.edu (J.T. West).

https://doi.org/10.1016/j.jml.2025.104625
Received 1 April 2024; Received in revised form 8 February 2025; Accepted 10 February 2025
0749-596X/Published by Elsevier Inc.

perceptual content. The goal of the current study was to address this gap by investigating the cognitive correlates of false memory across a variety of memory tasks and stimulus domains in both older and younger adults.

To date, memory researchers interested in understanding the factors associated with individual variation in long-term memory have successfully applied a *cognitive correlates* approach (Pellegrino & Glaser, 1979). Using this approach, researchers collect measures of cognitive abilities and memory performance and conduct analyses aimed at uncovering sources of individual variability in latent memory constructs. Such research has revealed much about the cognitive correlates of long-term memory performance (for a review, see Unsworth, 2019). For example, it is now well-established that different tasks measuring accurate (i.e., true) memory load onto a common higher-order memory factor (Malmi et al., 1979; Nyberg, 1994; Underwood et al., 1978; Unsworth, 2010; Unsworth & Brewer, 2009). In terms of cognitive correlates, a best-evidence synthesis of the extant literature by Unsworth (2019) revealed that memory accuracy is positively associated with cognitive abilities such as working memory capacity, fluid intelligence, and attentional control (e.g., Brewer & Unsworth, 2012; Robison et al., 2023; Unsworth et al., 2009; Unsworth & Spillers, 2010).

Although much is known about the factors associated with individual differences in true memory, relatively little is known about individual differences in false memory. Some prior research suggests that – as with true memory – false memory rates on different tasks tend to be reliable within individuals (Blair et al., 2002; Ost et al., 2013; Salthouse & Siedlecki, 2007) and load onto a single latent false memory factor (Ball et al., 2022; Lövdén, 2003; Unsworth & Brewer, 2010a). For example, Lövdén (2003) found that false memory rates on tasks of category cued recall, recall of omitted critical words belonging to semantically related lists (i.e., the DRM paradigm, Roediger & McDermott, 1995), and recognition of semantically related images loaded onto a single false memory factor. This study also found that individuals high in false memory tended to be low in long-term memory ability (see also, Zhu et al., 2010) and inhibition. Similar research has shown that individual differences in false memory are negatively associated with factors implicated in frontal functioning (i.e., working memory capacity and attentional control, Ball et al., 2022; Butler et al., 2004; Chan & McDermott, 2007; Colombel et al., 2016; Gerrie & Garry, 2007; Leding, 2012; Lövdén, 2003; McCabe et al., 2009; Peters et al., 2006, 2007; Unsworth & Brewer, 2010a, 2010b; Watson et al., 2005; for a review, see Festini & Katz, 2021), source monitoring ability (Ball et al., 2022; Unsworth & Brewer, 2010a), and perceptual ability (Zhu et al., 2010).

Although such research suggests that individual differences in false memory are associated with various cognitive abilities, this is not always found. For example, Salthouse and Siedlecki (2007, Experiment 2) measured false recognition for words, faces, and dot formations, as well as cognitive abilities such as fluid intelligence, memory ability, and perceptual speed in an adult lifespan sample. The authors found that false memory rates were not correlated between tasks and were generally unrelated to cognitive abilities. These results stand in contrast to the results reviewed above, including studies linking false memory to cognitive abilities specifically within older adults (Butler et al., 2004; Chan & McDermott, 2007; Colombel et al., 2016; Lövdén, 2003; McCabe et al., 2009).

What might explain these discrepant results? One possibility is that the factors associated with false memory may be partially domain-specific, with differences arising between false memories for semantic versus perceptual content. This possibility is supported by a neuroimaging *meta*-analysis of young adult data which demonstrated that, regardless of memoranda, false memories were associated with activation in frontal-parietal regions, including the ventromedial prefrontal cortex (vmPCF; Kurkela & Dennis, 2016). This result suggests that processes mediated by frontal regions such as failures of retrieval monitoring (McDonough et al., 2013; Thomas & McDaniel, 2013) may underly false memory formation regardless of the stimulus domain. Critically, this *meta*-analysis also found evidence of activation that was content-specific, such that false memories for semantic content were associated with activation in areas including the left inferior and middle frontal gyri, whereas false memories for perceptual content were associated with activation in the left inferior and middle occipital gyri. Because false memory is thought to be due in part to failures to use item-specific (i.e., verbatim) information to differentiate between studied and novel items during retrieval (Brainerd & Reyna, 2002), it is possible that these regions play an important role in the distinctive representation of semantic and perceptual items in memory, respectively. Thus, in addition to domain-general processes such as retrieval monitoring, domain-specific processes such as semantic and perceptual discrimination may protect against false memory.

Another basis for predicting that false memory is partially domain-specific comes from prior research demonstrating dissociations between semantic and perceptual false memories for lexical stimuli using the DRM paradigm. As reviewed by Chang and Brainerd (2021), many prior studies have investigated the question of whether different processes underly false memories on DRM tasks in which lures are semantically related to targets (e.g., *doctor* as a lure for *nurse*, *sick*, *hospital*, etc.) vs DRM tasks in which lures are orthographically (i.e., perceptually) similar to targets (e.g., *beat* as a lure for *heat*, *bet*, *beat*, etc.). In support of the proposition that false memories on semantic and perceptual materials tasks are mediated by different retrieval processes, Chang and Brainerd highlight evidence that these tasks exhibit different developmental trajectories (for review, see Brainerd & Reyna, 2012), and are differentially influenced by factors such as depth of processing (Chan et al., 2005; Thapar & McDermott, 2001), presentation duration (Ballardini et al., 2008; McDermott & Watson, 2001), emotional valence (Howe et al., 2010; Pesta et al., 2001), and presentation modality (Gallo et al., 2001; Smith & Hunt, 1998). Taken together, such research suggests that semantic and perceptual false memories as measured using lexical DRM tasks are governed by different principles.

These results have several implications related to individual differences in false memory. First, to the extent that semantic and perceptual false memory depend in part on separate processes, it may be that individual differences in semantic and perceptual false memory are best understood as two related – but distinct – constructs rather than as a single overarching construct. If true, then one's tendency to experience false memories, as well as the factors related to this tendency, may depend on the stimulus domain in question. Such a possibility would explain Salthouse and Siedlecki's (2007, Experiment 2) finding that false memory rates from a variety of stimulus categories did not correlate with one another, as well as other research which has found that false memory rates for critical words from the DRM paradigm are unrelated to the tendency to incorporate suggestive misinformation into visually encoded memories (Calvillo & Parong, 2016; Ost et al., 2013; Patihis et al., 2018). From a practical standpoint, determining whether individual differences in semantic false memory are distinct from individual differences in perceptual false memory is important in that such knowledge could be used to inform targeted interventions for individuals deficient in factors that protect against false memories for a specific type of stimuli. The current study aimed to test these possibilities by identifying the factors associated with semantic and perceptual false memory.

Another important consideration is that it is currently unclear whether the cognitive correlates of false memory differ across the adult lifespan. Given research demonstrating age differences in the neural correlates of false memory (Bowman et al., 2019; Bowman & Dennis, 2015; Chamberlain et al., 2022; Dennis et al., 2007, 2008, 2014; Devitt & Schacter, 2016; Tsukiura, 2014) as well as the finding that cognitive abilities tend to be less differentiated in older adults (La Fleur et al., 2018, but see Koen & Rugg, 2019), one might wonder whether the factors associated with false memory are the same or different in older and younger adults. Investigating this issue is both theoretically and practically meaningful, as understanding whether there are age differences in the extent to which certain cognitive abilities are protective against false memory is critical in determining which deficits should be prioritized as intervention targets when attempting to improve cognitive functioning in older adults.

What little research exists regarding the cognitive correlates of false memory in older adults suggests that – as with younger adults – older adults with lower frontal functioning (Butler et al., 2004; LaVoie et al., 2005; McCabe et al., 2009; Meade et al., 2012; for a review, see Devitt & Schacter, 2016) and inhibitory capacity (Colombel et al., 2016; Lövdén, 2003) are more susceptible to false memory. Such research suggests consistency in the factors associated with false memory between age groups, but does not speak to whether the relationship between these factors and false memory differs depending on age. In the only study to our knowledge to directly test whether the relationship between false memory formation and its cognitive correlates differs between older and younger adults, Chan and McDermott (2007) found that age did not moderate the relationship between frontal functioning and decreased false memory. Although this result suggests similarity between older and younger adults with respect to the factors that protect against false memory, because this study did not differentiate between semantic and perceptual false memory, it remains possible that the factors associated with false memory in older adults differ from those in younger adults in domain-specific ways.

The goal of the current study was to identify the cognitive factors associated with individual differences in semantic and perceptual false memory in older and younger adults. To this end, participants from both age groups completed multiple tasks assessing semantic and perceptual false memory, as well as frontal functioning, semantic discrimination, and perceptual discrimination. Based on prior research demonstrating differences in the neural bases of semantic and perceptual false memory (Kurkela & Dennis, 2016) as well as behavioral research demonstrating that false memories across different stimulus domains are sometimes uncorrelated (Calvillo & Parong, 2016; Ost et al., 2013; Patihis et al., 2018; Salthouse & Siedlecki, 2007), we predicted that these constructs would be captured by related yet distinct individual difference factors. Additionally, we predicted that semantic and perceptual false memory would have domain-specific, negative relationships with semantic and perceptual discrimination, respectively, which would remain significant after controlling for the effects of domain-general processes related to frontal functioning. In line with research linking false memory to low levels of frontal functioning (Ball et al., 2022; Butler et al., 2004; Chan & McDermott, 2007; Colombel et al., 2016; Festini & Katz, 2021; Gerrie & Garry, 2007; Leding, 2012; Lövdén, 2003; McCabe et al., 2009; Peters et al., 2006, 2007; Unsworth & Brewer, 2010a, 2010b; Watson et al., 2005) as well as neuroimaging research implicating frontally-mediated processes in the formation of false memories regardless of stimulus domain (Kurkela & Dennis, 2016), we predicted that both semantic and perceptual false memory would be negatively related to frontal functioning above and beyond the effects of semantic and perceptual discrimination. We also tested whether the relationship between false memory and its cognitive correlates differed between older and younger adults. Because older adults are more susceptible to false memory (McCabe et al., 2009) and more variable in their cognition compared to younger adults (Hertzog, 1985; Lindenberger & Oertzen, 2006; Morse, 1993; Nyberg et al., 2012), we predicted that the protective effects of these cognitive abilities on false memory would be stronger in older adults compared to younger adults.

As a secondary issue, we also investigated age differences in the variability and differentiation of our false memory factors. Prior research suggests that, compared to younger adults, older adults exhibit greater interindividual variability in cognitive performance (Hertzog, 1985; Lindenberger & Oertzen, 2006; Morse, 1993; Nyberg et al., 2012). For example, Morse et al. (1993), synthesizing individual differences data, found that older adults were more variable than younger adults with respect to their reaction times, memory, and fluid intelligence. Based on such findings, we predicted that older adults would exhibit greater variability than younger adults in semantic and perceptual false memory. Additionally, some prior research suggests that that certain cognitive abilities tend to become more intercorrelated (i.e., dedifferentiated) in later life. For example, La Fleur et al. (2018) analyzed data from 11 cognitive batteries and found that the positive association between constructs such as processing speed and mental rotation ability increased linearly with age across the adult lifespan (but see Koen & Rugg, 2019). Given these results, we planned to investigate age-related dedifferentiation within the context of semantic and perceptual false memory, and predicted that these constructs would be more strongly correlated in older compared to younger adults.

Methods

Data availability

The preregistration, materials, data, analysis code, and rendered analysis output for the current study are available at https://osf.io/x45ju/.

Participants

Two-hundred and thirty-four young adults between the ages of 18 and 24 were recruited from the psychology department's participant pool at The Pennsylvania State University. Additionally, 165 older adults between the ages of 60 and 89 were recruited from the communities surrounding the Pennsylvania State University. Participants were excluded if they reported using English for less than half their life (6 younger adults, 2 older adults) or if they did not complete both testing visits (17 younger adults). The final dataset included 211 younger adult participants and 152 older adult participants (see Table 1 for demographic information). All participants had normal or corrected-to-normal vision. All participants provided written informed consent as approved by The Pennsylvania State Institutional Review Board. Young adult participants were compensated for their participation with course credit and older adults were paid $45.

As indicated in the preregistration, we initially planned to collect 200 participants per age group. This sample size was chosen based on a simulation using pilot data from 58 older adults in which the *simsem* R package was used to estimate 80 % power based on 10,000 Monte Carlo simulations assuming a single false memory factor with conservative factor loadings (Pornprasertmanit et al., 2016). However, due to logistical issues with data collection related to the COVID-19 pandemic, we fell short of this goal for the older adult age group. Regardless, our final sample size of 152 older adults and 211 younger adults meets

Table 1
Participant Demographics by Age Group.

	Young Adults ($N = 211$)	Older Adults ($N = 152$)
Age; Mean (SD)	18.89 (1.09), *Range:* 18–24	72.91 (6.00), *Range:* 60–89
Education (years); Mean (SD)	12.61 (0.96)	16.72 (2.61)
MMSE	–	29.59 (0.71)
Sex		
Female	143	101
Male	67	51
Non-binary	1	0
Race		
Asian	34	0
Black or African American	9	2
White	148	144
More than one race	15	1
Other, unknown, or prefer not to answer	5	5
Ethnicity		
Hispanic or Latino	20	1
Not Hispanic or Latino	186	144
Other, unknown, or prefer not to answer	5	7

Note. MMSE = Mini Mental State Examination recommendations regarding the minimum number of participants needed for structural equation modeling in that our multiple-groups analyses well exceed 100 participants per group and both groups have more than 140 participants (i.e., 10 participants for every indicator variable; Wang & Wang, 2019).

Materials, stimuli, & task procedures

Recognition memory tasks

There were four recognition memory tasks, two of which were predominantly reliant on semantic processing and two which were predominantly reliant on perceptual processing. All tasks were created using Psychopy (Peirce et al., 2019) and administered using Pavlovia.org. Stimuli across all tasks were standardized for size and presented sequentially in the center of the screen. The order in which stimuli were presented within all encoding phases was randomized prior to data collection, then held constant across all participants. During the encoding phases, participants were informed that they were to view items and remember them for a later memory test. For the categorical false memory task, object similarity task, and abstract images task, the encoding phase was separated into three blocks in which to-be-remembered items were presented for two seconds with a one second interstimulus interval (ISI). In these tasks, each encoding block was followed by a recognition block. For the schematic memory task, the encoding phase was separated into two sequential blocks in which presented scenes containing to-be-remembered items were shown for 10 s with a one second ISI, followed by a single recognition phase.

During the recognition phases, participants were shown items from the encoding phase (*targets*), items that were either semantically or perceptually related to items from encoding (*lures*), and unrelated novel items (*foils*). Participants were made aware of the presence of lures and foils prior to each recognition phase. Across the four memory tasks, items presented during the recognition phases were pseudorandomized to ensure that no more than four items of a given trial type were presented sequentially. Participants were instructed to indicate whether a given item was old or new and indicate to their confidence using the following four-response scale: definitely old, probably old, probably new, or definitely new. All items during recognition were presented for five seconds with a one second ISI. The key indicator of false memory for our individual differences analyses in these tasks was false alarm rates for semantically or perceptually related lures as opposed to false memory rates for unrelated foils.

Semantic Memory Tasks. The *categorical false memory task* is a modified DRM task which consisted of 54 categorically-related word lists from Dennis et al. (2007). Each list consisted of six words, four of which were presented during the encoding phase and the remaining two during recognition as related lures (see Fig. 1A). Each block of encoding consisted of 72 words separated into 18 categories. Prior to viewing items from each category, participants were shown a prompt (e.g., “Types of Fruits”) for one second. Following this prompt, the four words associated with the respective category were presented one-at-a-time for two seconds (e.g., “banana”, “strawberry”, “apple”, “pear”). Presentation order of both the categories and the words within each category were fixed across participants. The recognition phase consisted of 270 words, with each category having two targets and two lures. The remaining 42 words were novel foils which were not related to any presented category.

The *schematic memory task* was modified from Webb et al. (2016). The encoding phase of this task consisted of drawings of 26 complex scenes, split into two sequential blocks, with 13 scenes presented in each block. Scenes were presented for 10 s each, allowing the participants to freely study the objects present within the scenes. Objects within the scenes were either exclusive or non-exclusive to the scene’s general schema (see Fig. 1B). An example of a schematic (i.e., schema-exclusive) object would be an image of a toilet within a bathroom, as toilets appear only within bathrooms but not within other settings. An example of a non-schematic (i.e., schema non-exclusive) object would be an image of a mirror within a bathroom scene, as mirrors appear in bathrooms but also appear in other settings as well (see Webb et al., 2016 for development and norming of the stimuli). The recognition phase was separated into 3 blocks. In total, 96 objects were presented during the recognition phase, including 26 schematically related targets, 26 non-schematic targets, and 26 schematic lures. The remaining 18 objects consisted of foils, which were unrelated to all presented scenes. In addition to ensuring that no more than three objects of a given trial type (e.g., targets, lures, and foils) were presented sequentially, items during the recognition phase were pseudorandomized to ensure that no two objects associated with a given scene appeared consecutively.

Perceptual Memory Tasks. The *object similarity task* consisted of 216 images of everyday objects from Dennis et al. (2014). Images were separated into 54 image categories, with four object exemplars per category (see Fig. 1C). The encoding phase was separated into three blocks, with each block consisting of 72 object images across 18 categories. Prior to viewing a given category, participants were provided with a one-second category prompt (e.g., “Types of Bouquets”) that was followed by the sequential presentation of the four associated exemplar images. The presentation order of both the categories and exemplars within each category was fixed across all participants. A recognition phase followed each encoding phase, and consisted of 12 targets, 12 related lures, and 6 unrelated foils. Across all three recognition phases, 36 targets, 36 related lures, and 17[1] novel foils were presented.

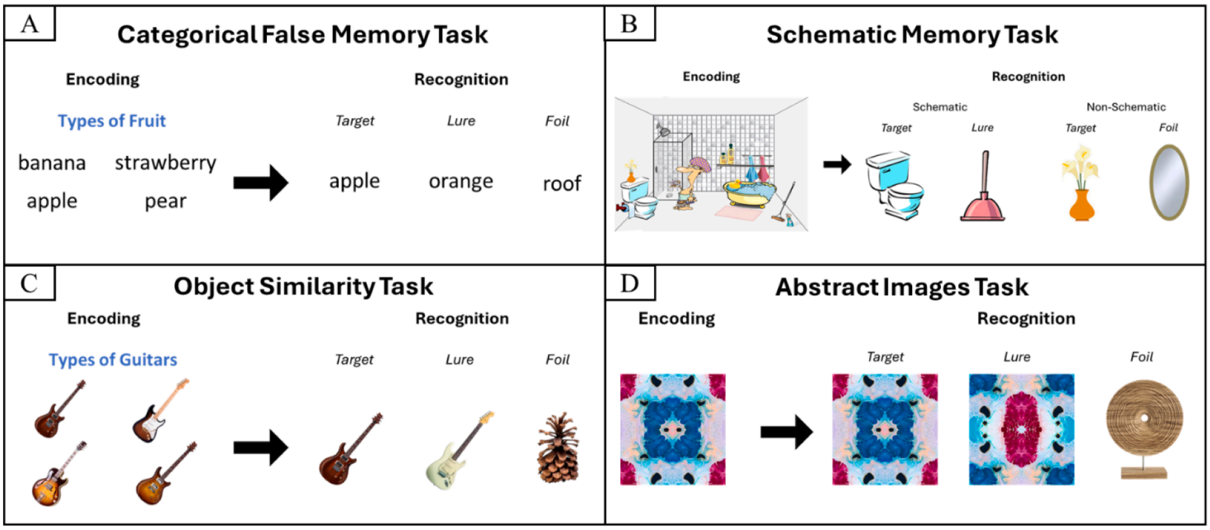
The *abstract images task* consisted of 270 images of abstract art pieces collected from Internet searches (Fig. 1D). These searches were limited to only include paintings and sculptures that had perceptually similar counterparts that could serve as lures. To norm selected images, eight participants were individually shown the image pairs and were asked to rate each pair for relatedness to one another, image quality, and abstractness using five-point Likert scales. For the image relatedness, a response of one indicated that the images were nothing alike whereas a response of five indicated that the images were similar to one another ($M = 4.55$, $SD = 0.74$). For image quality, a response of one indicated that image quality was poor, whereas a response of five indicated that image quality was good ($M = 4.51$, $SD = 0.86$). For image abstractness, a response of one indicated that the participant could generate a label for the image given pair, whereas a response of five indicated that they could not generate a label for the image pair ($M = 4.67$, $SD = 0.86$). If an image pair fell below an average of three across all three scales, the pair was replaced. The encoding phase of the abstract images task consisted of 216 images separated into three blocks of 72 images. Following each encoding phase, there was a recognition phase consisting of 36 targets, 36 perceptually related lures, and 18 foils. Across all three recognition phases all 270 images were presented.

No category from a given memory task was used in a different memory task. For example, if the category “farm animals” was used in the categorical false memory task, then items related to the concept of “farm” were not used in any other memory task. Prior to analyses, key responses made during the memory tasks were evaluated to address accidental shifts of participants’ hands on the keyboard. Key presses were evaluated by two raters to identify responses where a participant shifted within one key of the keys associated with the four response options. When both raters agreed that a participant shifted their hand, their responses were corrected.

Cognitive assessments

The administration of all cognitive assessments adhered to their respective administration guidelines, including provided instructions, scoring guidelines, and adaptive task termination rules when applicable.

[1] Due to experimenter error, one novel foil was presented twice during the retrieval phase of the object similarity task. The second presentation of this item was removed and was not included in any final analyses.



Note. For the abstract images task, we have provided both example target and lure items which correspond to the abstract image presented during encoding for demonstration purposes only. In reality, a given image presented during encoding for this task was only ever matched to either a target repetition or a perceptually similar lure during the recognition phase.

Fig. 1. *Examples of the Encoding and Recognition Items for the Categorical False Memory (A), Schematic Memory (B), Object Similarity (C), and Abstract Images (D) Tasks. Note.* For the abstract images task, we have provided both example target and lure items which correspond to the abstract image presented during encoding for demonstration purposes only. In reality, a given image presented during encoding for this task was only ever matched to either a target repetition or a perceptually similar lure during the recognition phase.

These procedures are briefly reported below.

Frontal Functioning Tasks. Four cognitive assessments were chosen to evaluate frontal functioning: the digital, short form of the Berg Card-Sorting Test-64 (pBCST-64; Fox et al., 2013) and the arithmetic, backwards digit span, and letter-numbering sequencing from the Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2008). Administration of the pBCST-64 (Piper et al., 2012) and the three measures from the WAIS-IV (Wechsler, 2008) have been validated to capture changes in executive function across the lifespan.

The pBCST-64 was administered via computer using The Psychology Experimental Building Language (PEBL; Mueller & Piper, 2014) following standardized task procedure. Participants were required to sort a virtual deck of 128 cards according to the color, shape, or quantity printed on the card (Berg, 1948). The “sorting rule” was unknown to participants throughout the duration of the task and would change after participants correctly sorted 10 cards (i.e., one “set”). Participants had to rely solely on trial-by-trial feedback to determine the sorting rule. The task continued until the participant either 1) sorted all 128 cards regardless of accuracy, or 2) completed nine sets of the task. As such, the minimum number of trials a participant could experience if they did not make any errors was 90 trials and the maximum number of trials was 128. In the current study, younger adult participants completed an average of 122.70 trials ($SD = 7.15$) and older adult participants completed an average of 125.40 trials ($SD = 5.72$). Participant’s performance on the pBCST-64 was calculated as the total number of sorting errors made following a rule switch (i.e., “perseverative errors”), with lower values indicating better performance. Due to experimenter error, data from seven younger adults and eight older adults were lost.

The arithmetic task consisted of 22 timed mental arithmetic problems. Answers to problems were only considered correct if the answer matched the expected answer and were given before the time limit was reached. If a participant did not provide an answer before the time limit, their mental calculations were not cut off by researchers, but their answers were counted as incorrect. The task was terminated once a participant either produced four consecutive incorrect responses or answered all 22 problems. Performance was calculated by totaling the number of correct responses. Thus, possible scores on the arithmetic task ranged from 0 to 22. Due to experimenter error, data from one younger adult were lost.

The backwards digit span consisted of seven sets of digit sequences that were read aloud to participants by the researcher. Following each sequence, participants were instructed to repeat the digits in backwards order. Sequences were separated into sets of two trials. The task was terminated once a participant either incorrectly repeated both sequences within a given set or provided answers to all seven sets. Performance was calculated by totaling the number of correct responses. Thus, scores on this task ranged from 0 to 14.

The letter-number sequencing task consisted of seven sets of letter-number sequences that were read aloud to participants. At the end of each sequence, participants were instructed to repeat the sequence back in a particular order. First, participants were instructed to repeat the digits in numerical order, starting with the lowest number. Next, participants were instructed to repeat the letters in alphabetical order. Sequences were separated into sets of three trials. The task was terminated once a participant either incorrectly repeated all three sequences within a given set or provided answers to all seven sets. Performance was calculated by totaling the number of correct responses. Thus, scores for the letter-number sequencing task ranged from 0 to 21.

Semantic Discrimination Tasks. The PPVT-4 tested participant’s receptive vocabulary knowledge (Dunn & Dunn, 2007). This task consisted of 228 trials separated into sets of 12 trials. Participants aged 18 started the task on trial 145, while participants ages 19 and older started on trial 157. On each trial, the researcher read a vocabulary word and presented four pictures and participants were instructed to choose the picture that best described the vocabulary word. The test began by first identifying a participant's basal score, which was determined based on the first set where a participant was able to answer at least 11 of the 12 trials correctly. If the participant answered two or more trials incorrectly within a given set, the participant moved to a less difficult set until they answered at least 11 trials correctly. Once a participant's basal score was established, the participant advanced through increasingly difficult sets until they incorrectly answered eight trials within a set or reached the end of the task. Performance on the PPVT-4 was calculated by subtracting the total number of incorrect trials from the last trial number a participant received before the task was terminated. Thus, scores on this task ranged from 0 to 228. Due to experimenter error, data from five younger adults was lost.

The PPT assessed participants' ability to derive semantic meaning and relationships from pictures (Howard & Patterson, 1992). The 52-trial task was adapted to be administered via computer using Qualtrics (Qualtrics, https://www.qualtrics.com). On each trial, participants were provided a test image (e.g., a pyramid) and were asked to match the image to one of two image responses: a conceptually-related target image (e.g., a palm tree) or a conceptually-unrelated distractor image (e.g., a fir tree). All images in the task were simple, black-and-white line drawings. Participants were asked to provide an answer to every trial but were able to indicate that they did not know the answer to a particular question by selecting "Not Sure". Participants received all 52 trials regardless of task performance. All correct responses were awarded one point and "Not Sure" responses were awarded a half of a point. Thus, scores on this task ranged from 0 to 52.

Perceptual Discriminability Assessments. The MVPT-4 tested participant's visual discrimination, spatial relationship discrimination, visual memory, figure-ground discriminability, and visual closure perception (Colarusso & Hammill, 2015).The task consisted of 45 test trials and eight practice trials. On each trial, participants were shown a page with an item and four response options. On each trial, researchers read participants a question and participants were expected to choose which response option best answered the given question. Participants were shown all 45 test trials regardless of task accuracy. Performance on the MVPT was calculated by totaling the number of accurate responses to test trials. Thus, scores on this task ranged from 0 to 45.

The (Odd-One-Out) OOO was adapted from four tasks from Barense and colleagues' (2007) perceptual discriminability task: the low ambiguity familiar objects task, high ambiguity familiar objects task, the size task, and the color task. These tasks were selected given their reliance on perceptual feature discrimination processes (Barense et al., 2007). Tasks were presented through Psychopy (Peirce et al., 2019) and Pavlovia.org. The four tasks were presented in the following order across all participants: low ambiguity, high ambiguity, size, color task. Prior to receiving a given task, participants received three practice trials to familiarize themselves with the specific requirements of a given task. For all four OOO tasks, participants were shown four images and were instructed to select the image that was "different" from the others. In the low and high ambiguity familiar objects tasks, four pictures of everyday objects were presented per trial, with the differing object varying in its perceptual details compared to the other objects but not in its size or orientation (i.e., the "odd" object would not be the exact same object presented smaller or angled differently than the other three objects, it would be a new, yet similar object). Ten trials of the low ambiguity task were used solely to familiarize participants with the general procedures of the OOO tasks and data for this task were not used in our analyses. The high ambiguity task consisted of 35 trials where the differing object was highly similar to the three distractor objects. In the size task, four rectangles were presented, with the "odd" rectangle differing in size from the others. Lastly, in the color task, four colored squares were presented with the differing square appearing in a slightly different hue. Unlike the high ambiguity task, the size and color OOO tasks consisted of 17 trials. Participants were allowed unlimited time to respond to each trial. Performance on the high ambiguity task, the size task, and color OOO task was calculated as the proportion of correct responses.

Experimental procedures

The current study took place over two sessions, occurring within one week of each other. All testing procedures were conducted in-person via computer or paper-and-pencil. During the first session, participants provided consent, self-reported demographics, and completed the Short Form Health Survey (SF-12;Ware et al., 1996). Older adult participants also completed the Mini-Mental State Examination (MMSE; Folstein et al., 1975). Next, participants completed a general memory practice test that included example stimuli from the four memory tasks they were to complete over the course of the two visits. This practice was provided to allow participants to familiarize themselves with the memory task procedures that were common across the four tasks, including presentation time, task phases, and recognition response options. Following this practice task, participants completed two false memory tasks and four cognitive assessments during their first session. During the second session, participants completed the remaining two false memory tasks and cognitive assessments.

The order in which the four false memory tasks and cognitive assessments were administered was counterbalanced both between and within sessions. The four false memory tasks were distributed evenly between the two sessions, with one perceptual and one semantic memory task in each session. Administration of the cognitive assessments was also segmented, with two frontal functioning assessments occurring during each session. Administration of the domain-specific cognitive tasks was grouped, such that semantic and perceptual discrimination tasks occurred on separate days.

Data analysis

Exclusions and missing data

As pre-registered, a participant's data for a given false memory task were excluded from analyses based three criteria: consecutive fast responses, low hit rate, and high rates of non-responses (see Table 2). During each recognition test, if a participant made six or more consecutive responses in 400 ms or less, their data were excluded for that task.

Table 2
Data Removed from Each False Memory Task by Age Group.

	Young Adult Excluded	Older Adult Excluded
	Categorical False Memory Task	
Consecutive Fast Responses	2	0
Low Hit Rate	6	1
High Non-Response Rate	0	1
Total	**8**	**2**
	Schematic Memory Task	
Consecutive Fast Responses	3	0
Low Hit Rate	53	42
High Non-Response Rate	0	4
Total	**56**	**46**
	Object Similarity Task	
Consecutive Fast Responses	1	0
Low Hit Rate	4	1
High Non-Response Rate	0	1
Total	**5**	**2**
	Abstract Images Task	
Consecutive Fast Responses	1	0
Low Hit Rate	7	1
High Non-Response Rate	1	0
Total	**9**	**1**

If a participant's hit rate was at or below chance (i.e., 50 %) on a specific memory task, their data were excluded for that task. Finally, data for a given task were excluded if a participant failed to respond to over 10 % of the recognition trials.

Modeling approach

For all latent variable models, we report several model fit indices. Based on existing recommendations (Hu & Bentler, 1999; Kenny, 2020; MacCallum et al., 1996), acceptable model fit is indicated by a non-significant chi-squared test,[2] TLI and CFI values greater than 0.90, and RMSEA and SRMR values below 0.08.

Prior to testing our hypotheses about the structural relationships between our latent factors, it was necessary to first establish the appropriate measurement model given our data within each age group. We did this by fitting a series of stratified *Confirmatory Factor Analysis* (CFA) models within to the older and younger adult data. Decisions about CFA model selection were made primarily on the basis of chi-squared difference tests, which indicate whether the fit of a more restrictive model which has been trimmed by eliminating free parameter (s) is significantly worse than the fit of a less restrictive nested model which retains said parameter(s). Given that chi-squared difference tests evaluate the significance of decrements in model fit due to model trimming, a significant chi-squared test indicates a preference for the less restrictive model whereas a non-significant chi-squared test indicates a preference for the more restrictive and parsimonious model (Kline, 2016). However, additional information was also considered as part of the model selection process including inter-factor correlations, overall model fit, and suggested model modifications (i.e., modification indices).

After establishing the appropriate measurement model for each group using CFA, group differences in the relationship between false memory and cognitive ability variables between older adults and younger adults were tested using multiple-groups structural equation modeling (MG-SEM, Hirschfeld & Brachel, 2019).[3] Provided that both groups have the same general factor structure, this analysis allows one to test whether various model parameters are equivalent between groups (i.e., *measurement invariance*) by constraining these parameters to be equal between groups and testing whether fit decreases substantially as a result. Following the recommendations of Hirschfeld and Brachel (2019), our decisions about whether to retain the more constrained model were informed by whether the nested likelihood ratio test comparing model fit was significant and whether the change in the CFI fit index was greater than 0.01, with both of these outcomes indicating worse fit in the more constrained model.

We assessed three levels of measurement invariance: *configural invariance* (where the same general factor structure but none of the specific parameters are the same between groups), *weak invariance* (where the configural invariance model is constrained by forcing item loadings to be equal), and *strong invariance* (where the weak invariance model is further constrained by forcing intercepts to be equal).[4] Importantly, establishing weak invariance allows one to test for group differences in latent covariances and regression paths, and establishing strong invariance allows one to test for group differences in latent means.

All analyses were conducted using RStudio (Posit team, 2024) and were rendered into reproducible research reports using Quarto which are available on the OSF repository (Allaire et al., 2024). Structural equation modeling and confirmatory factor analysis were carried out using the *lavaan* R package (Rosseel, 2012). Participants missing data on indicators were retained through maximum-likelihood estimation, which enables participants with missing data to inform model estimates for indicators for which they have data provided that the missing data is assumed to be missing at random (Allison, 2003). Additionally, as in Robison et al. (2023) we supplemented several of our frequentist model comparisons with exploratory Bayesian analyses. For these Bayesian analyses, Bayes factors were computer using the bic_to_bf function in the *bayestestR* R package (Makowski et al., 2019), which computes Bayes factors based on differences between models in the Bayesian information criterion (BIC). These Bayes factors can be interpreted as the likelihood of one model relative to another given their observed BIC values.

Results

Descriptive statistics and reliability for each task are presented in Tables 3 and 4 for younger and older adults, respectively. For each task, we computed odd–even reliability and applied a Spearman-Brown correction. False memory reliability was computed for related lure only, and was computed following the exclusions described in Table 2.

Correlations for our measured (i.e., manifest) variables are presented in Figs. 2–4. Generally, the observed pattern of correlations aligns with our hypothesized factor structure (Fig. 5) in that variables predicted to measure the same latent construct tend to be intercorrelated. An exception to this is seen in the correlations between the four false memory tasks, which – despite being correlated with one another – do

Table 3
Descriptive Statistics and Reliability in the Younger Adult Sample.

Measure	*n*	*Mean*	*SD*	*Skew*	*Kurtosis*	*Reliability*
	False Memory Tasks					
Categorical False Memory Task	203	0.27	0.19	0.82	0.02	0.911
Schematic Memory Task	152	0.48	0.15	−0.04	−0.48	0.500
Object Similarity Task	205	0.38	0.20	0.43	−0.33	0.884
Abstract Image Task	201	0.52	0.15	−0.15	−0.44	0.777
	Frontal Functioning Tasks					
Letter-Number Sequencing	211	10.70	2.93	−0.35	1.30	0.797
Backwards Digit Span	211	7.29	2.26	0.54	0.01	0.817
Arithmetic	210	9.09	3.54	−0.15	−0.96	0.848
pBCST perseverative errors	206	15.71	9.29	1.32	2.05	0.765
	Semantic Discrimination Tasks					
PPVT-4	206	195.51	12.67	−0.41	−0.19	0.914
PPT	211	47.07	2.83	1.02	0.19	0.640
	Perceptual Discrimination Tasks					
MVPT						
OOO (High)	211	0.63	0.19	−0.52	−0.49	0.845
OOO (Color)	211	0.80	0.16	−1.22	1.89	0.682
OOO (Size)	211	0.58	0.21	−0.10	−0.78	0.730

Note. pBCST = PEBL Berg Card Sort Task, PPVT-4 = Peabody Picture Vocabulary Task (4th Ed.), PPT = Pyramids and Palm Trees Task, MVPT = Motor-free Visual Perception Task, OOO = Odd One Out task (high discriminability, color, and size components). Descriptive statistics for the false memory tasks reflect false recognition of related lures.

[2] It should be noted that, being a significance test, the chi-squared test can be significant despite good fit at larger sample sizes, as in the current study (Bentler & Bonett, 1980).

[3] Our use of MG-SEM was not preregistered. However, given that MG-SEM allows one to address important questions about differences in latent parameters between groups, we ultimately decided that such analyses were more appropriate given our research questions of interest. The decision to use MG-SEM as opposed to the more standard SEM occurred prior to data analysis.

[4] In addition to configural, weak, and strong measurement invariance, it is also possible to test for *strict invariance* (where the strong invariance model is further constrained by forcing residual variances to be equal between groups). Because strong invariance was not established in the current study, we did not pursue strict invariance in any of our analyses.

Table 4
Descriptive Statistics and Reliability in the Older Adult Sample.

Measure	*n*	*Mean*	*SD*	*Skew*	*Kurtosis*	*Reliability*
	False Memory Tasks					
Categorical False Memory Task	148	0.23	0.16	1.01	0.56	0.884
Schematic Memory Task	101	0.50	0.16	0.17	−0.01	0.535
Object Similarity Task	147	0.39	0.17	0.51	−0.15	0.849
Abstract Image Task	150	0.59	0.14	−0.20	−0.37	0.780
	Frontal Functioning Tasks					
Letter-Number Sequencing	152	10.10	2.93	−0.10	1.30	0.792
Backwards Digit Span	152	7.61	2.28	0.10	−0.25	0.815
Arithmetic	152	10.88	3.37	−0.42	−0.83	0.883
pBCST perseverative errors	144	19.87	11.61	0.62	−0.05	0.914
	Semantic Discrimination Tasks					
PPVT-4	152	219.75	5.63	−1.36	2.19	0.856
PPT	152	50.38	1.40	−1.60	4.83	0.465
	Perceptual Discrimination Tasks					
MVPT						
OOO (High)	152	0.54	0.19	−0.17	−0.69	0.840
OOO (Color)	152	0.68	0.19	−0.22	−0.87	0.712
OOO (Size)	152	0.63	0.65	−0.08	−0.51	0.624

Note. pBCST = PEBL Berg Card Sort Task, PPVT-4 = Peabody Picture Vocabulary Task (4th Ed.), PPT = Pyramids and Palm Trees Task, MVPT = Motor-free Visual Perception Task, OOO = Odd One Out task (high discriminability, color, and size components). Descriptive statistics for the false memory tasks reflect false recognition of related lures.

not show specificity with respect to our distinction between semantic and perceptual false memory tasks. If it were the case that semantic and perceptual false memories were captured by two distinct constructs, one would expect semantic false memory tasks to be more highly correlated with one another than with perceptual false memory tasks, and vice versa. This prediction is not supported by the observed correlations; the highest correlation for our false memory measures in both age groups is between the perceptual Object Similarity and semantic Categorical False Memory tasks, and Object Similarity false memory is also moderately correlated with semantic Schematic false memory.

Following Robison et al. (2023), we conducted exploratory tests of age differences for our observed correlations by applying a Fisher's *r*-to-*z* transformation to the correlations in each age group and testing group differences using Fisher's method (1925). Results of this analysis are reported in our rendered analysis output: https://osf.io/q46p2. Of the 91 pair-wise correlations in Figs. 2–4, 8 differed significantly between older adults and younger adults (smallest $p = 0.005$). In all cases, these age differences reflected stronger correlations between measures in younger adults compared to older adults. However, it is important to note that none of these age differences were significant at a Bonferroni-corrected alpha threshold of 0.05/91.

Age differences in cognitive assessments

For the domain-general frontal functioning assessments, there was a significant difference in performance between the age groups on the pBCST-64, $t(210.35) = 3.87$, $p < 0.001$, $d = 0.44$, with older adults having significantly more perseverative errors than younger adults. There was also a significant difference in performance on the arithmetic task, $t(334.23) = 4.87$, $p < 0.001$, $d = 0.52$, with older adults performing better than younger adults. There was no difference in performance between the age groups on the backwards digit span task, $t(323.13) = 1.32$, $p = 0.19$, $d = 0.14$. Lastly, older adults performed significantly worse on the letter-number sequencing task compared to younger adults, $t(346.53) = -2.06$, $p = 0.04$, $d = 0.20$.

For the domain-specific discriminability assessments, older adults performed significantly better than younger adults in the semantic domain, as captured by the PPVT-4, $t(300.25) = 24.39$, $p < 0.001$, $d = 2.47$, and the PPT, $t(325.08) = 14.67$, $p < 0.001$, $d = 1.48$. In contrast, in the perceptual domain assessments younger adults outperformed the older adults on nearly all assessments. In the Odd One Out subtasks, younger adults performed significantly better than older adults on the *High* discriminability subtest, $t(326.90) = -4.10$, $p < 0.001$, $d = 0.47$, and on the *Color* subtest, $t(283.35) = -6.21$, $p < 0.001$, $d = 0.68$. On the *Size* subtest, older adults performed significantly better than younger adults, $t(359.06) = 2.12$, $p = 0.03$, $d = 0.27$. Finally, on the MVPT, younger adults, again, performed significantly better than the older adults, $t(331.25) = -3.49$, $p < 0.001$, $d = 0.37$.

Age differences in memory discriminability

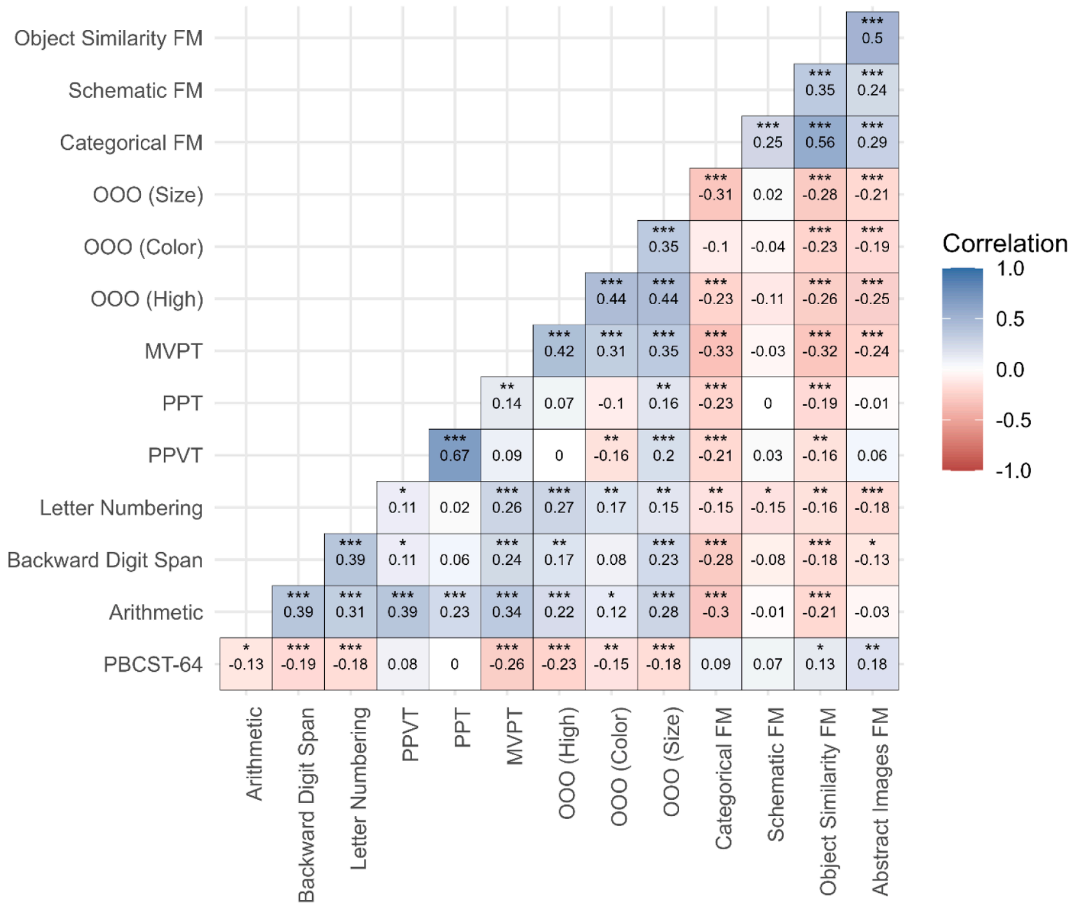
To investigate age differences in memory discriminability (Table 5), d' was calculated for each memory task as *z*(Hits) – *z*(Related False Alarms). Hit rates of 1 were corrected by subtracting 0.5 to the hit rate numerator and false alarm rates of 0 were corrected by adding 0.5 to the false alarm rate numerator. For the categorical false memory task, older adults had better discriminability ($M = 2.09$, $SD = 0.80$) than younger adults ($M = 1.90$, $SD = 0.95$), $t(342.48) = 2.05$, $p = 0.04$, $d = 0.22$. For the schematic memory task, younger adults had better discriminability ($M = 0.28$, $SD = 0.43$) than older adults ($M = 0.17$, $SD = 0.40$), $t(225.70) = -2.01$, $p = 0.045$, $d = -0.26$. In the object similarity task, older adults had significantly better discriminability ($M = 1.77$, $SD = 0.58$) than younger adults ($M = 1.51$, $SD = 0.89$), $t(347.93) = 3.29$, $p = 0.001$, $d = 0.34$. Finally, in the abstract images task, there was no significant difference in discriminability between the two age groups, $t(348.65) = 0.98$, $p = 0.33$, $d = 0.10$ (younger adults: $M = 0.81$, $SD = 0.66$; older adults: $M = 0.87$, $SD = 0.48$).

Age differences in false memory

Four 2 (Age: young adult, older adult) x 2 (Item type: lures, foils) analysis of variance (ANOVAs) were conducted to evaluate the effects of age group and item type on the production of false memories across the four memory tasks (Table 5). Including item type as a factor in these analyses allowed us to verify that false memory rate differed between lures and foils, which was the case for all four memory tests.

Semantic Memory Tasks. For the categorical false memory task, a significant main effect of age group was found, $F(1, 349) = 8.13$, $p = 0.005$, $\eta_p^2 = 0.023$, with older adults' false memory rates ($M = 0.12$, $SD = 0.16$) being lower than younger adults' false memory rates ($M = 0.16$, $SD = 0.19$). There was also a significant main effect of item type, $F(1, 349) = 780.488$, $p < 0.001$, $\eta_p^2 = 0.691$, with false memory rates for lures ($M = 0.26$, $SD = 0.18$) being greater than false memory rates for novel foils ($M = 0.03$, $SD = 0.08$). The interaction effect between age group and item type was not significant, $F(1, 349) = 0.914$, $p = 0.340$, $\eta_p^2 = 0.003$.

For the schematic memory task, the main effect of age group was not significant, $F(1, 251) = 2.51$, $p = 0.115$, $\eta_p^2 = 0.010$. The main effect of item type was significant, $F(1, 251) = 454.59$, $p < 0.001$, $\eta_p^2 = 0.644$. False memory rates for lures ($M = 0.49$, $SD = 0.15$) were greater than false memory rates for foils ($M = 0.29$, $SD = 0.17$). The interaction effect between age group and item type was significant, $F(1, 251) = 24.93$, $p < 0.001$, $\eta_p^2 = 0.090$. Follow-up simple effects tests revealed that this interaction was driven by group differences in false memory rates for novel foils, $F(1, 251) = 13.57 p < 0.001$, $\eta_p^2 = 0.051$, with younger adults' false memory rates ($M = 0.32$, $SD = 0.17$) being greater than older adults' false memory rates ($M = 0.25$, $SD = 0.15$) for novel foils. There were no significant differences in false memory rates for lure items, *F*(1,



Note. For the PBCST-64, lower values indicate *higher* performance. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Fig. 2. *Correlations of Measured Variables in the Full Sample. Note*. For the PBCST-64, lower values indicate *higher* performance. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

251) = 1.34, $p = 0.287$, $\eta_p^2 = 0.005$, between younger ($M = 0.48$, $SD = 0.15$) and older ($M = 0.50$, $SD = 0.16$) adults.

Perceptual Memory Tasks. For the object similarity task, the main effect of age group was not significant, $F(1, 350) = 0.43$, $p = 0.513$, $\eta_p^2 = 0.001$. The main effect of item type was significant, $F(1, 350) = 1443.78$, $p < 0.001$, $\eta_p^2 = 0.805$, with false memory rates for lures ($M = 0.39$, $SD = 0.19$) being greater than false memory rates for foils ($M = 0.02$, $SD = 0.06$). The interaction between age group and item type was not significant, $F(1, 350) = 2.64$, $p = 0.105$, $\eta_p^2 = 0.007$.

Finally, for the abstract images task, the main effect of age group was marginally significant, $F(1, 349) = 3.80$, $p = 0.052$, $\eta_p^2 = 0.011$, as older adults' false memory rates ($M = 0.38$, $SD = 0.25$) were greater than younger adults' false memory rates ($M = 0.35$, $SD = 0.23$). The main effect of item type was also significant, $F(1, 349) = 2715.46$, $p < 0.001$, $\eta_p^2 = 0.886$, with false memory rates for lures ($M = 0.55$, $SD = 0.15$) being greater than false memory rates for novel foils ($M = 0.17$, $SD = 0.16$). Additionally, the interaction between age group and item type was significant, $F(1, 349) = 28.18$, $p < 0.001$, $\eta_p^2 = 0.075$. Follow-up simple effects tests indicated that this interaction was driven by age differences in false memory rates for the lure items, $F(1, 349) = 18.46$, $p < 0.001$, $\eta_p^2 = 0.050$, with older adults' false memory rates being greater ($M = 0.59$, $SD = 0.14$) than younger adults' false memory rates for lures ($M = 0.52$, $SD = 0.15$). There were no differences in false memory rates for novel foils, $F(1, 349) = 0.42$, $p = 0.515$, $\eta_p^2 = 0.001$, between younger ($M = 0.18$, $SD = 0.16$) and older ($M = 0.16$, $SD = 0.15$) adults. A summary of results regarding age differences in our cognitive assessments and false memory tasks is presented in Table 6.

Confirmatory factor analyses

We began our latent variable analyses by comparing the fit of four potential theory-informed factor structures. Model 1 aligned with our preregistered hypotheses in that this model had two separate factors for semantic and perceptual false memory and three separate factors for cognitive abilities. In addition to testing the fit of this preregistered measurement model, we also tested three competing models based on theory which were not preregistered. Model 2 tested the possibility that false memory is captured by a single, overarching false memory factor

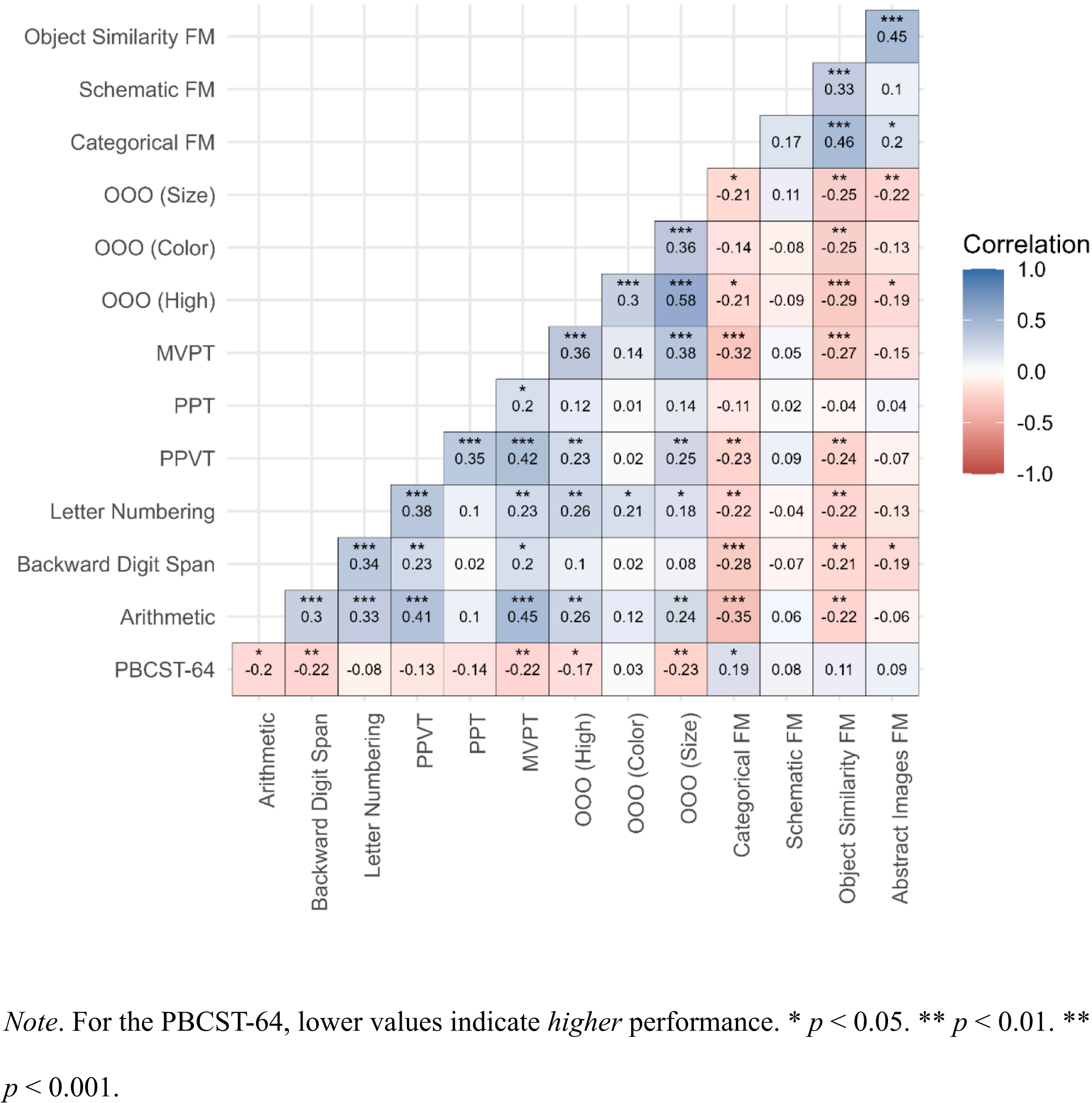
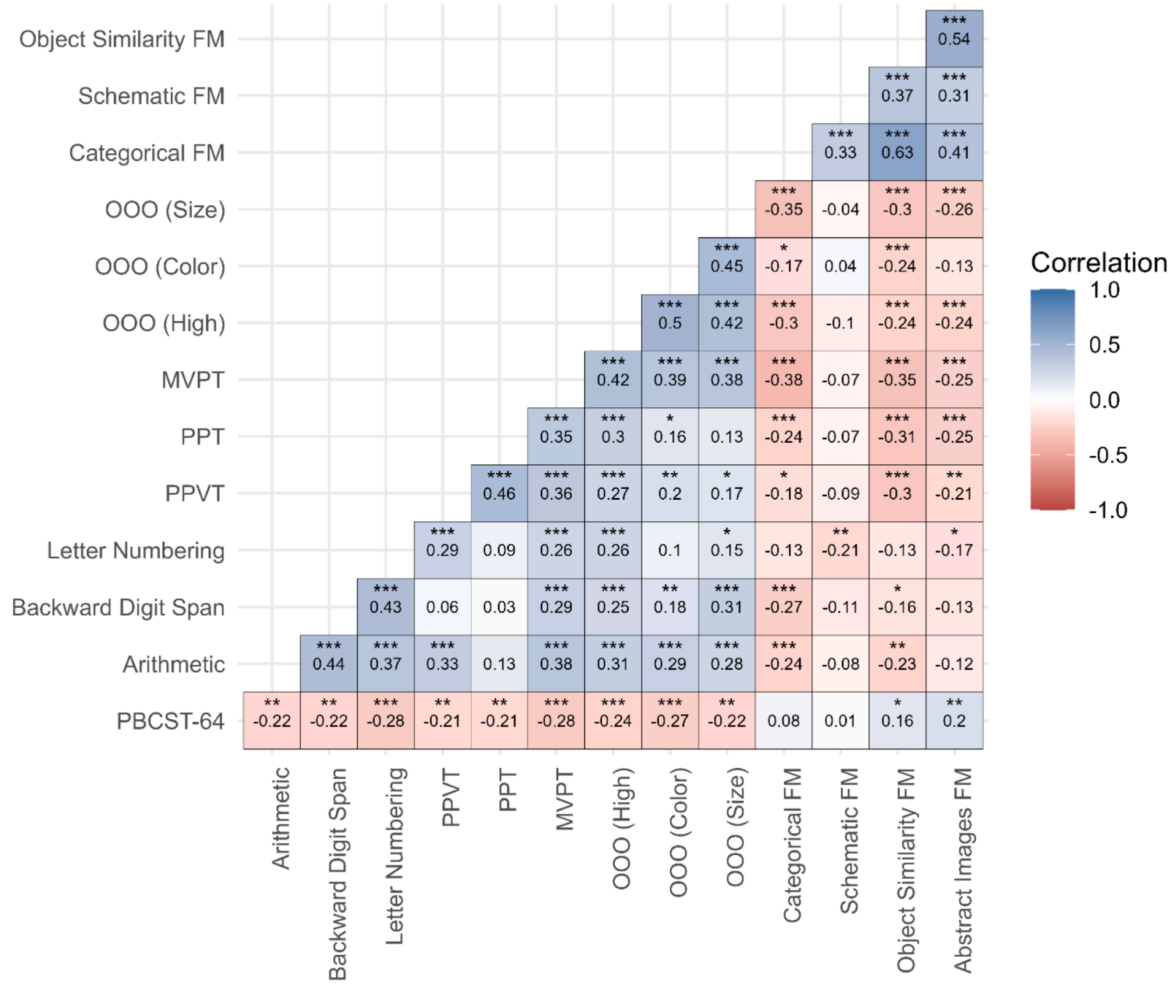


Fig. 3. *Correlations of Measured Variables in Older Adults. Note.* For the PBCST-64, lower values indicate *higher* performance. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

rather than by two distinct factors but retained the three distinct cognitive ability factors. Model 3 tested the possibility that cognitive abilities are best captured by a single factor but retained the separate semantic and perceptual false memory factors from Model 1. At the suggestion of an anonymous reviewer, we also tested the fit of Model 4, which had a single false memory factor and a single cognitive ability factor. Because producing Models 2–4 entails placing various restrictions on Model 1 (i.e., fixing between-factor correlations to be 1), Model 1 is nested within Models 2–4. Models 2 and 3 are nested within Model 4 for similar reasons. Residual covariances were specified for the three measures taken from the odd-one-out to allow these measures to covary with one another above and beyond their shared status as indicators of perceptual discrimination. Models 1 and 3 indicated a small number of negative variance estimates close to zero for the older adult data. As is often done, these variance estimated were fixed to zero (Farooq, 2022). Path diagrams for the CFA models are presented in our rendered output: https://osf.io/erdkf.[5]

Fit indices for the three CFA models are presented in Table 7. For younger adults, Models 1 and 2 fit the data well whereas Models 3 and 4 fit the data more poorly. All models converged without warnings. Chi-squared difference tests indicated that the more restrictive Model 3 fit the data significantly worse than the less restrictive Model 1 [$X^2(7) = 61.158$, $p < 0.001$]. Model 2, however, did not fit the data significantly worse than Model 1 [$X^2(4) = 7.019$, $p = 0.135$], suggesting that Model 2 is more appropriate than Model 1 for the younger adult data. Because

[5] At the suggestion of a colleague, we also conducted exploratory CFAs investigating the fit of bifactor models, where false memory was modeled using a combination of orthogonal domain-general and domain-specific false memory factors (Cucina & Byle, 2017; Markon, 2019). Because these models exhibited poor fit and convergence issues, we do not discuss them here. Nonetheless, the results of these analyses can be found in our OSF repository: https://osf.io/j5rpk.



Note. For the PBCST-64, lower values indicate *higher* performance. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Fig. 4. *Correlations of Measured Variables in Younger Adults. Note.* For the PBCST-64, lower values indicate *higher* performance. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Model 4 fit the data significantly worse than Model 2, Model 2 was retained as the appropriate model [$X^2(5) = 55.696$, $p < 0.001$]. Thus, results suggest that Model 2, with a single false memory factor and three separate cognitive ability factors, is the measurement model that best describes the younger adult data.

All four models fit the older data well. Chi-squared difference tests indicated that the more restrictive Models 2 [$X^2(2) = 9.25$, $p = 0.010$] and 3 [$X^2(6) = 19.00$, $p = 0.004$] both fit the data significantly worse than the less restrictive Model 1. Although these chi-squared tests suggest that Model 1 is the appropriate model for the older adult data, there are several indications that, as was the case with the younger adult data, the more restrictive Model 2 which combines the semantic and perceptual false memory factors may also be a viable measurement model for the older adult data.

First, the semantic and perceptual false memory factors were highly correlated in Model 1 ($r = 0.803$) for the older adult data, suggesting that, even when modeled as separate factors, semantic and perceptual false memory exhibit a high degree of overlap. This high correlation was also present when estimating the SEM model in Fig. 5 for older adults (see https://osf.io/erdkf for model results). This SEM assumes the separate semantic and perceptual false memory factors of Model 1 and is based on our a priori hypotheses in which semantic and perceptual false memory exhibit domain-general and domain-specific relationships with cognitive abilities. In this SEM, the semantic and perceptual false memory factors are highly correlated ($r = 0.806$), again, highlighting the overlap of these factors. Also consistent with the notion that the two false memory factors had considerable overlap, several modification indices suggested making changes to this SEM which hint at overlap between these factors (e.g., allowing schematic and object similarity false memory to covary [$MI = 8.687$], allowing schematic false memory to load onto perceptual false memory [$MI = 6.808$], allowing categorical false memory to load onto perceptual false memory [$MI = 4.787$]). These same modification suggestions were also present for CFA Model 1 (e.g., allowing schematic and object similarity false memory to covary [$MI = 8.847$], allowing schematic false memory to load onto perceptual false memory [$MI = 7.264$], allowing categorical false memory to load

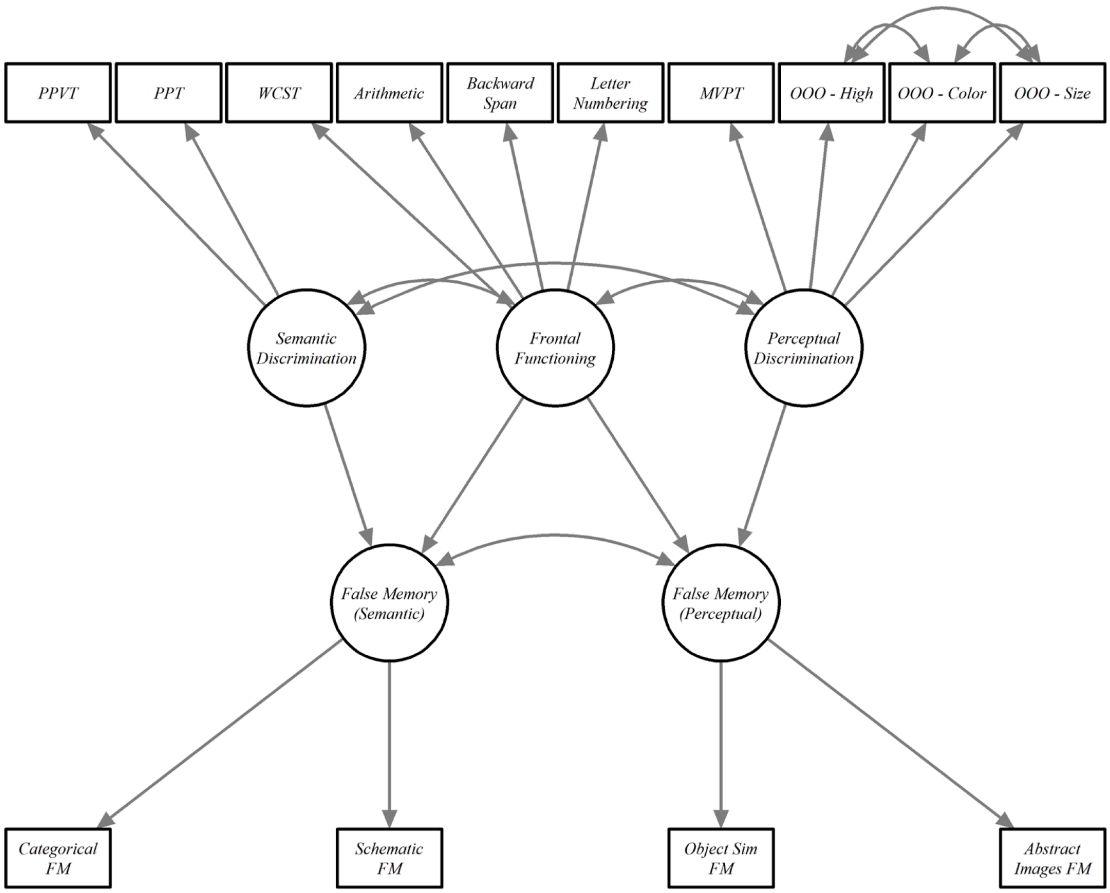


Fig. 5. *Path Diagram Depicting A Priori Hypotheses.*

onto perceptual false memory [$MI = 7.264$]). Finally, the semantic and perceptual false memory factors from Model 1 had similar relationships with the cognitive ability factors for older adults in that both factors were significantly and negatively correlated with the three cognitive abilities (r's between -0.241 and -0.785, p's < 0.001 to 0.016). This result suggests that our semantic and perceptual false memory factors are interchangeable with respect to their cognitive correlates.

Thus, although chi-squared difference tests indicate that the more restrictive Model 2 fit the data worse than the less restrictive Model 1 for the older adult data [$X^2(2) = 9.25$, $p = 0.010$], there is evidence that collapsing the perceptual and semantic false memory factors as in Model 2 is warranted, as these factors are highly corelated and have similar cognitive correlates. Given these considerations, we chose Model 2 as the measurement model for the older adult data. Doing so also had the benefit of allowing us to use MG-SEM analyses to test for age differences given that choosing Model 2 for both age groups means that both groups have the same underlying factor structure. We return the issue of whether the older adult data is best described by one or two false memory factors in the Discussion section.

Multi-group structural equation modeling

Having established the measurement model for our data, we next conducted MG-SEM analyses to examine age differences the structural relationships between our latent factors. In this model, domain-general false memory was predicted as a function of semantic discrimination, perceptual discrimination, and frontal functioning. A path diagram of this model is presented in Fig. 6.

We began by estimating a configural invariance model in which the same general factor structure and regression paths were fit to both age groups simultaneously. Fit indices for this and other MG-SEM models are presented in Table 8. Because this model fit the data well, we concluded configural invariance, indicating that the same general factor structure applies to both older adults and younger adults. Next, we tested for weak invariance by constraining item loadings to be equal between groups. Because this constraint did not significantly reduce fit based on a likelihood ration test [$X^2(10) = 11.245$, $p = 0.339$] and resulted in a minimal decrease to CFI ($CFI_{diff} = 0.001$), we concluded weak invariance. Finally, we tested for strong invariance by constraining both the item loadings and intercepts to be equal between groups. Because this constraint resulted in a significant reduction in fit [$X^2(10) = 160.340$, $p < 0.001$] and a large decrease to CFI ($CFI_{diff} = 0.140$), we did not conclude strong invariance.

Even when the requirements of strong invariance are not met, it is still possible to achieve *partial strong invariance* by releasing the equality constrains for a small number of item intercepts (Hirschfeld & Brachel, 2019). To test for this possibility, we conducted both frequentist and Bayesian analyses to compare changes in model fit after iteratively equating each item intercept. Analyses suggested that suggested that nearly all of the item intercepts differed between age groups, making partial strong invariance untenable (see the rendered output: https://osf.io/ur64k). Ultimately, the current analyses suggest a pattern of weak measurement invariance, whereby the relationship between cognitive abilities and false memory is captured by the same general model in older adults and younger adults, with similar item loadings but not item intercepts between groups.

Having established weak invariance, we turn now to our interpretation of our final MG-SEM model. Path diagrams depicting this model in older adults and younger adults are presented in Figs. 7 and 8, and scatterplots depicting latent covariances and regression paths as estimated by factor scores are presented in Figs. 9 and 10. As seen in Fig. 9, frontal functioning, semantic discrimination, and perceptual

Table 5
Descriptive Statistics for Memory Tasks.

Memory Task	*n*	*Mean*	*SD*	*Skew*	*Kurtosis*
	Older Adults				
Categorical False Memory Task					
Lure False Alarm Rate	148	0.23	0.16	1.01	0.56
Foil False Alarm Rate	148	0.02	0.05	5.12	32.79
Target-Lure Discriminability (*d'*)	148	2.09	0.80	0.27	−0.20
Schematic Memory Task					
Lure False Alarm Rate	101	0.50	0.16	0.17	−0.01
Foil False Alarm Rate	101	0.25	0.15	0.62	0.26
Target-Lure Discriminability (*d'*)	101	0.17	0.40	0.14	0.23
Object Similarity Task					
Lure False Alarm Rate	147	0.39	0.17	0.51	−0.15
Foil False Alarm Rate	147	0	0.02	5.77	40.82
Target-Lure Discriminability (*d'*)	147	1.77	0.59	−0.03	−0.24
Abstract Image Task					
Lure False Alarm Rate	150	0.59	0.14	−0.20	−0.37
Foil False Alarm Rate	150	0.16	0.15	1.14	0.95
Target-Lure Discriminability (*d'*)	150	0.87	0.48	0.55	0.16
	Younger Adults				
Categorical False Memory Task					
Lure False Alarm Rate	203	0.27	0.19	0.82	0.02
Foil False Alarm Rate	203	0.04	0.09	2.99	9.77
Target-Lure Discriminability (*d'*)	203	1.89	0.95	0.61	0.69
Schematic Memory Task					
Lure False Alarm Rate	152	0.48	0.15	−0.04	−0.48
Foil False Alarm Rate	152	0.32	0.17	0.29	−0.28
Target-Lure Discriminability (*d'*)	152	0.28	0.43	0.33	1.01
Object Similarity Task					
Lure False Alarm Rate	205	0.38	0.20	0.43	−0.33
Foil False Alarm Rate	205	0.03	0.07	3.39	12.76
Target-Lure Discriminability (*d'*)	205	1.51	0.88	0.33	−0.40
Abstract Image Task					
Lure False Alarm Rate	201	0.52	0.15	−0.15	−0.44
Foil False Alarm Rate	201	0.17	0.16	1.06	0.84
Target-Lure Discriminability (*d'*)	201	0.81	0.66	0.91	0.77

discrimination were positively and significantly intercorrelated in both older adults and younger adults. Notably, cognitive abilities jointly predicted 37.2 % of the variance in false memory in older adults and 40.0 % of the variance in younger adults. Regarding the specific effects of these cognitive abilities on false memory, only the effect of perceptual discrimination in younger adults was significant, with higher perceptual discrimination predicting decreased false memory when controlling for semantic discrimination and frontal functioning.

We predicted that cognitive abilities would be negatively related to false memories in both older and younger adults but only observed a single significant effect of perceptual discrimination on false memory in younger adults. However, inspection of the factor score scatterplots in Fig. 10 appears to show the expected negative relationship between each of these cognitive abilities and false memory. This suggests that although these cognitive abilities generally do not have *unique* relationships with false memory when these effects are estimated jointly (perhaps due to some degree of multicollinearity, as correlations between the cognitive ability factors ranged from 0.436 to 0.752), we reasoned that these factors might predict false memory in isolation. Consistent with this interpretation, exploratory analyses indicated that when estimated in models not including the other cognitive abilities, frontal functioning (Older adults: $\beta = -0.54$, $z = -3.21$, $p < 0.001$, $R^2 = 0.295$; Younger adults: $\beta = -0.37$, $z = -3.23$, $p = 0.001$, $R^2 = 0.139$),

Table 6
Summary of Age Difference Analyses.

Measure	Outcome of Age Comparison
Categorical False Memory Task	
Lure False Alarm Rate	Younger Adults > Older Adults
Foil False Alarm Rate	Older Adults > Younger Adults
Target-Lure Discriminability (*d'*)	Older Adults > Younger Adults
Schematic Memory Task	
Lure False Alarm Rate	No Age Difference
Foil False Alarm Rate	Older Adults > Younger Adults
Target-Lure Discriminability (*d'*)	Younger Adults > Older Adults
Object Similarity Task	
Lure False Alarm Rate	No Age Difference
Foil False Alarm Rate	Older Adults > Younger Adults
Target-Lure Discriminability (*d'*)	Older Adults > Younger Adults
Abstract Image Task	
Lure False Alarm Rate	Older Adults > Younger Adults
Foil False Alarm Rate	No Age Difference
Target-Lure Discriminability (*d'*)	No Age Difference
Frontal Functioning Tasks	
Letter-Number Sequencing	Younger Adults > Older Adults
Backwards Digit Span	No Age Difference
Arithmetic	Older Adults > Younger Adults
pBCST perseverative errors	Younger Adults > Older Adults
Semantic Discrimination Tasks	
PPVT-4	Older Adults > Younger Adults
PPT	Older Adults > Younger Adults
Perceptual Discrimination Tasks	
MVPT	Younger Adults > Older Adults
OOO (High)	Younger Adults > Older Adults
OOO (Color)	Younger Adults > Older Adults
OOO (Size)	Older Adults > Younger Adults

Note. pBCST = PEBL Berg Card Sort Task, PPVT-4 = Peabody Picture Vocabulary Task (4th Ed.), PPT = Pyramids and Palm Trees Task, MVPT = Motor-free Visual Perception Task, OOO = Odd One Out task (high discriminability, color, and size components).

Table 7
Fit Indices for CFA Models in Older and Younger Adults.

	Model 1	Model 2	Model 3	Model 4
	Older Adults			
Chi-squared test	$X^2(66) = 68.773$, $p = 0.384$	$X^2(68) = 78.022$, $p = 0.190$	$X^2(72) = 87.773$, $p = 0.100$	$X^2(73) = 94.635$, $p = 0.045$
CFI	0.992	0.972	0.956	0.940
TLI	0.989	0.963	0.945	0.925
RMSEA	0.017 [0, 0.052] 90 % *CI*	0.031 [0, 0.059] 90 % *CI*	0.038 [0, 0.063] 90 % *CI*	0.044 [0.007, 0.068] 90 % *CI*
SRMR	0.060	0.064	0.066	0.070
	Younger Adults			
Chi-squared test	$X^2(64) = 100.146$, $p = 0.003$	$X^2(68) = 107.165$, $p = 0.002$	$X^2(71) = 161.304$, $p < 0.001$	$X^2(73) = 162.861$, $p < 0.001$
CFI	0.950	0.945	0.874	0.875
TLI	0.928	0.927	0.838	0.844
RMSEA	0.052 [0.031, 0.071] 90 % *CI*	0.052 [0.032, 0.071] 90 % *CI*	0.078 [0.062, 0.094] 90 % *CI*	0.076 [0.061, 0.092] 90 % *CI*
SRMR	0.053	0.054	0.066	0.066

Note. For RMSEA and SRMR, lower values indicate better fit.

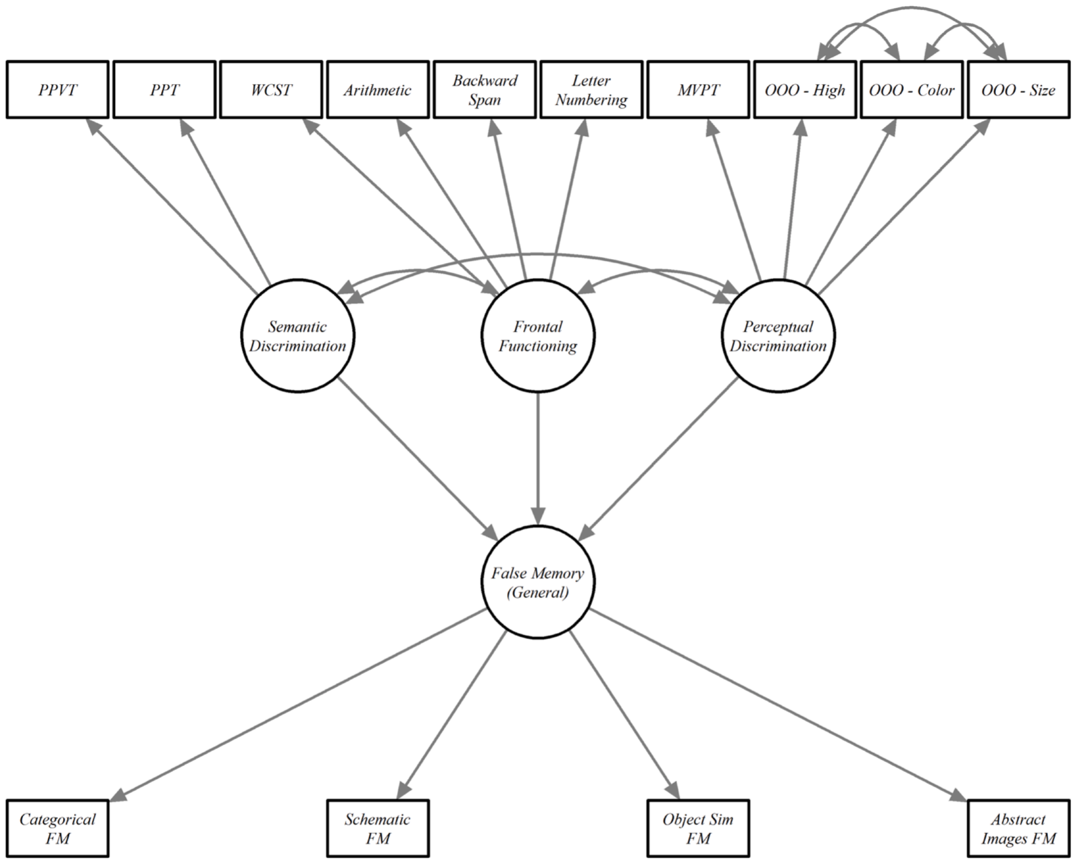


Fig. 6. *Path Diagram of the MG-SEM Model.*

Table 8
Fit Indices for MG-SEM Models.

	Configural Invariance	Weak Invariance	Strong Invariance
Chi-squared test	$X^2(136) =$ 185.187, $p = 0.003$	$X^2(146) =$ 196.432, $p = 0.003$	$X^2(156) =$ 356.771, $p < 0.001$
CFI	0.954	0.953	0.814
TLI	0.939	0.942	0.783
RMSEA	0.045 [0.027, 0.060] 90 % *CI*	0.044 [0.026, 0.059] 90 % *CI*	0.084 [0.073, 0.096] 90 % *CI*
SRMR	0.058	0.061	0.083

Note. For RMSEA and SRMR, lower values indicate better fit.

semantic discrimination (Older adults: $\beta = -0.32$, $z = -2.36$, $p = 0.018$, $R^2 = 0.099$; Younger adults: $\beta = -0.48$, $z = -3.80$, $p < 0.001$, $R^2 = 0.227$), and perceptual discrimination (Older adults: $\beta = -0.60$, $z = -3.97$, $p < 0.001$, $R^2 = 0.362$; Younger adults: $\beta = -0.60$, $z = -5.07$, $p < 0.001$, $R^2 = 0.361$) were each negatively related to false memory in both older adults and younger adults. These results suggest that, although frontal functioning, semantic discrimination, and perceptual discrimination do not uniquely predict false memory, they are negatively related to false memory when considered in isolation of one another.

Next, we conducted analyses to examine whether factor variances and regression paths differed between older adults and younger adults. For these analyses, we tested for moderation by releasing the equality constraint on a single model parameter one at a time and testing whether releasing that constraint resulted in improved fit (see Robison et al., 2023). Frequentist and Bayesian results (i.e., Bayes factors presented as BF_{10}) of these moderation analyses are presented in Table 9.

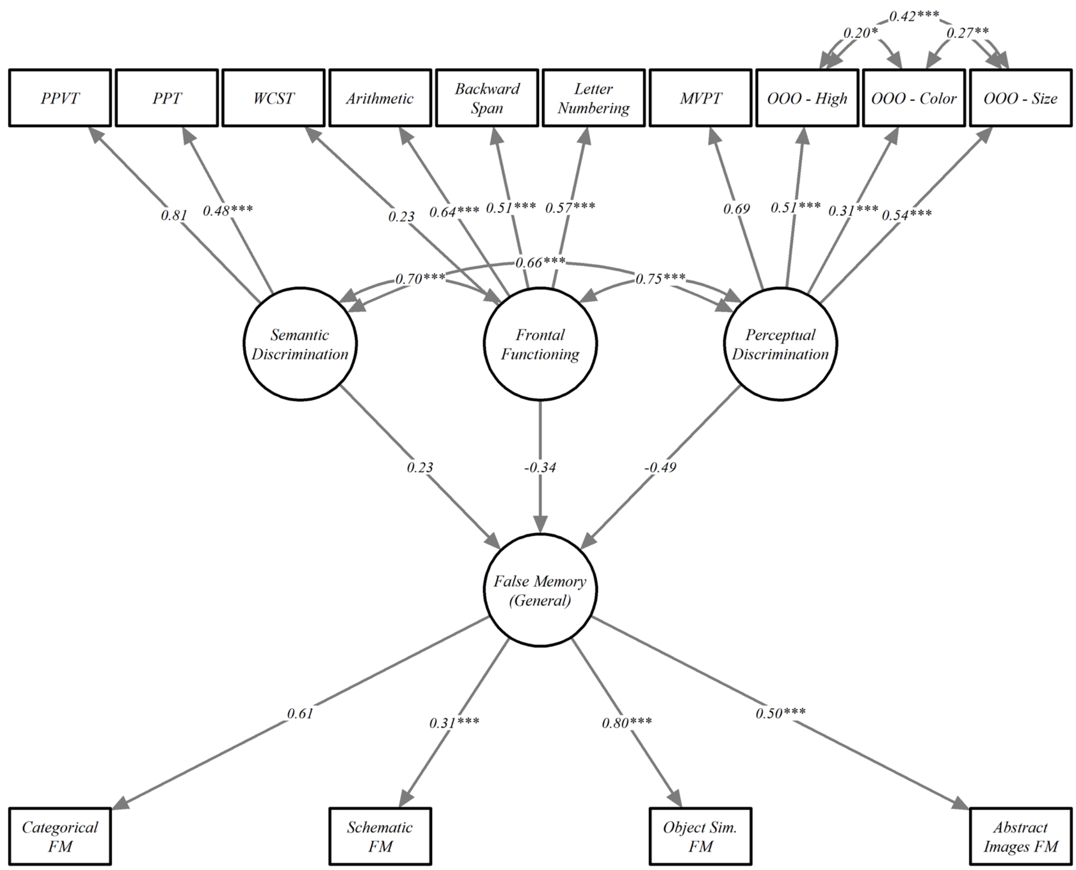
Testing whether factor variances differed between older and younger adults allowed us to test the prediction that older adults would exhibit greater interindividual variability than younger adults (Hertzog, 1985; Lindenberger & Oertzen, 2006; Morse, 1993; Nyberg et al., 2012). Moderation analyses revealed that younger adults were more variable with respect to semantic discrimination than older adults (0.400 vs. 0.080; see the Variance Estimates section of Table 9). Frequentist analyses suggested that younger adults were also more variable with respect to false memory, though Bayesian results for this test were more ambiguous (0.557 vs. 0.312). We did not find evidence for age differences in variability in terms of frontal functioning or perceptual discrimination.

Given these group differences in variance estimates, moderation tests for covariance estimates were not assessed. Instead, we addressed the prediction that older adults would exhibit increased cognitive dedifferentiation by inspecting latent correlations. Results showed that frontal functioning was more strongly related to semantic discrimination in older adults than in younger adults ($r = 0.696$ vs $r = 0.436$). Semantic and perceptual discrimination were similarly correlated in older and younger adults ($r = 0.662$ vs $r = 0.618$), as were frontal functioning and perceptual discrimination ($r = 0.752$ vs $r = 0.726$).

Finally, we tested for group differences in regression coefficients within both our joint and isolated models, which allowed us to determine whether the effects of cognitive abilities on false memory differed between older and younger adults. As seen in Table 9, our analyses did not provide evidence of moderation, suggesting that the joint and isolated effects of cognitive abilities on false memory did not differ between older adults and younger adults.

Discussion

The goal of the current study was to investigate the factor structure



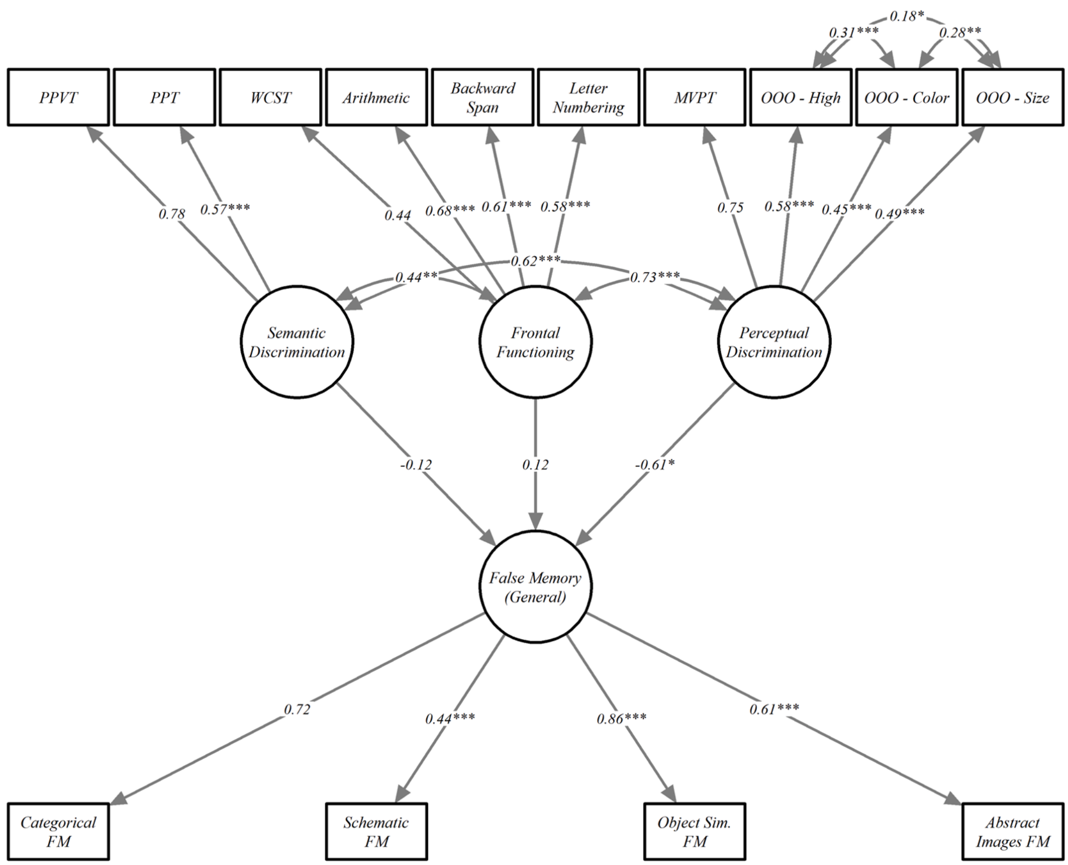
Note. All parameters are standardized. Significance is not calculated for the loading associated with the first indicator of each factor, as these indicators have fixed unstandardized values of 1 for the purpose of model identification. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Fig. 7. *Path Diagram of Final MG-SEM in Older Adults. Note.* All parameters are standardized. Significance is not calculated for the loading associated with the first indicator of each factor, as these indicators have fixed unstandardized values of 1 for the purpose of model identification. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

and cognitive correlates of individual differences in semantic and perceptual false memory in younger and older adults. To this end, participants completed a series of tasks designed to assess semantic false memory, perceptual false memory, frontal functioning, semantic discrimination, and perceptual discrimination. Based on research suggesting that the neural processes associated with semantic false memory differ from those associated with perceptual false memory (Kurkela & Dennis, 2016) as well as research suggesting that false memory tasks from different stimulus modalities are uncorrelated (Calvillo & Parong, 2016; Ost et al., 2013; Patihis et al., 2018; Salthouse & Siedlecki, 2007), we predicted that measures of semantic and perceptual false memory would load onto separate factors which would be negatively related to semantic and perceptual discrimination, respectively. Based on prior research (Ball et al., 2022; Butler et al., 2004; Chan & McDermott, 2007; Colombel et al., 2016; Festini & Katz, 2021; Gerrie & Garry, 2007; Leding, 2012; Lövdén, 2003; McCabe et al., 2009; Peters et al., 2006, 2007; Unsworth & Brewer, 2010a, 2010b; Watson et al., 2005), we predicted that both types of false memory would be negatively related to frontal functioning. In terms of age differences, we predicted that the protective effects of cognitive abilities against false memory production would be stronger in older adults, and that older adults would show both increased variability and decreased differentiation (i.e., higher intercorrelation) for the false memory factors.

Semantic and perceptual false memory are captured by a single overarching factor

Contrary to our predictions, we did not find convincing evidence that tasks of semantic and perceptual false memory loaded onto distinct factors. For younger adults, there was clear evidence that these tasks were captured by a single overarching factor, as the model with a single false memory factor (Model 2) did not fit worse than the less restrictive model with separate semantic and perceptual false memory factors (Model 1). Indeed, in Model 1 these factors correlated at $r = 0.906$. Results regarding the older adult data were more ambiguous, with both Models 1 and 2 fitting the data well (see Table 7). For older adults, although a chi-squared difference test indicated that Model 2 fit worse than Model 1, the high overlap between our semantic and perceptual



Note. All parameters are standardized. Significance is not calculated for the loading associated with the first indicator of each factor, as these indicators have fixed unstandardized values of 1 for the purpose of model identification. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Fig. 8. *Path Diagram of Final MG-SEM in Younger Adults. Note.* All parameters are standardized. Significance is not calculated for the loading associated with the first indicator of each factor, as these indicators have fixed unstandardized values of 1 for the purpose of model identification. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

false memory factors was demonstrated by the high correlation of these factors ($r = 0.803$ in CFA Model 1, $r = 0.806$ in our hypothesized SEM model), modification indices which suggested changes to both models consistent with the overlap of these factors (e.g., allowing semantic and perceptual tasks to covary or load onto both factors), and the similar pattern of relationships between these factors and cognitive abilities.

If one concludes that semantic and perceptual false memory are separate factors based on our older adult data, they must concede that these factors were very highly correlated and had similar negative correlations with cognitive abilities. In our view, these considerations are sufficient to conclude that the two false memory factors are largely redundant with respect to their ability to describe the older adult data. Thus, given the clear preference for Model 2 in the younger adult data and evidence supporting the overlap of semantic and perceptual false memory in older adults, we consider Model 2 to be the most plausible explanation of the current data.

Accordingly, our results suggest that although semantic and perceptual false memories may differ in certain aspects, such differences may not provide the best characterization of individual-level variation in the tendency to falsely remember information. Our results are more consistent with the idea that individuals who are prone to falsely remember information exhibit this tendency regardless of the stimulus domain in question.[6] In light of prior neuroimaging research suggesting that false memory has an important domain-general component (Kurkela & Dennis, 2016), it is possible that the contribution of frontally-mediated domain-general processes to false memory, such as failure of retrieval monitoring (McDonough et al., 2013; Thomas & McDaniel, 2013), outweighs the contribution of more subtle domain-specific processes related to fine-grained stimulus discrimination.

[6] As mentioned in the Introduction, prior research suggests that individual differences in DRM false memory are unrelated to individual differences in the misinformation effect (Calvillo & Parong, 2016; Ost et al., 2013; Patihis et al., 2018). Although such research is suggestive of the possibility that semantic and perceptual false memories are uncorrelated, it is also possible that other differences between the methods of evoking false memories using these two paradigms account for this lack of correlation (for example, the fact that these tasks differ in the role of self-generation vs external suggestion; Ost et al., 2013).

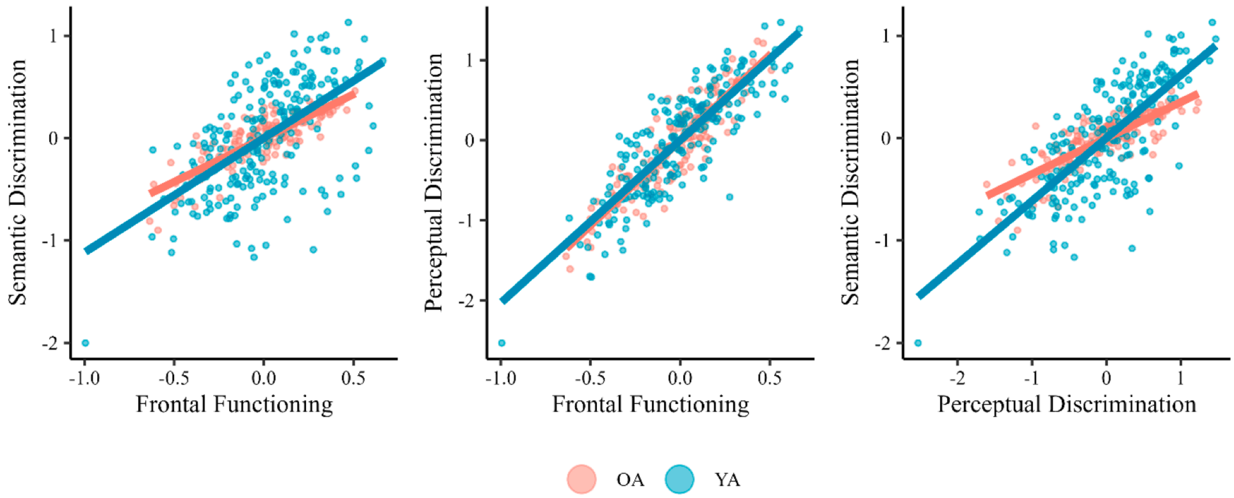


Fig. 9. *Scatterplots of Latent Factor Score Covariances.*

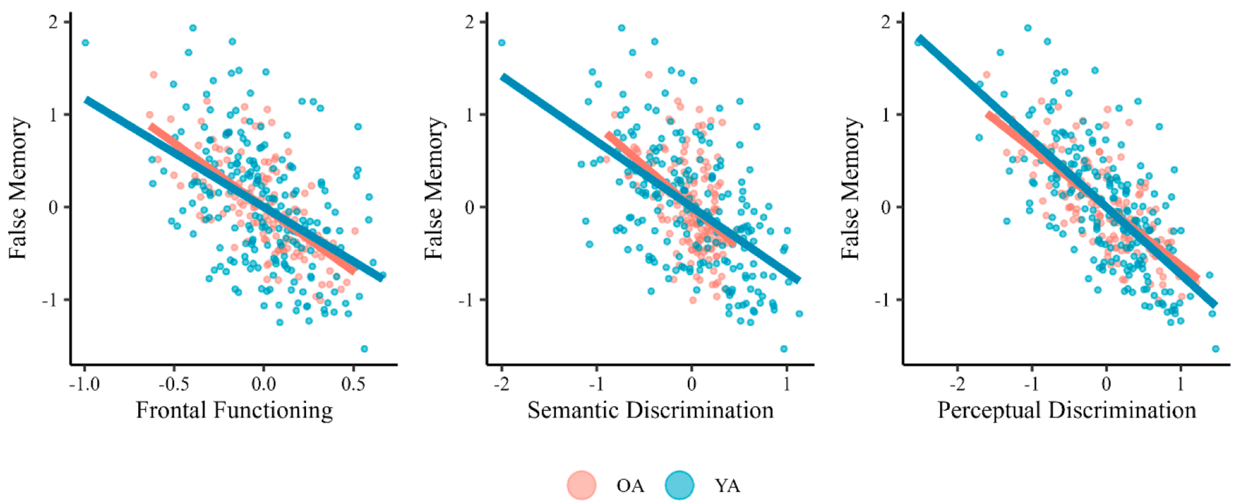


Fig. 10. *Scatterplots of Latent Factor Score Regression Paths.*

Evidence supporting a model with a single false memory factor as opposed to a model with two separate factors was unexpected given prior research suggesting that there are differences in the processes mediating semantic and perceptual false memory. For example, a *meta*-analysis of the neural correlates of false memory formation found evidence of domain-specificity, with semantic false memory being associated with activation in verbal processing regions such as the left middle and inferior frontal gyri and perceptual false memory being associated with activation in regions within the visual cortex (Kurkela & Dennis, 2016). Additionally, Salthouse and Siedlecki (2007, Experiment 2) conducted a large lifespan study ($n = 332$) and found that, despite sufficient task reliability, false memory rates across tasks assessing memory for words, dot formations, and faces did not correlate with one another, suggesting that tasks of semantic and perceptual false memory may not be measuring the same underlying construct.

Differences between our results and Salthouse and Siedlecki's (2007, Experiment 2) may be due to differences in stimuli used between our studies. Specifically, Salthouse and Siedlecki's perceptual stimuli included dot formations and faces, whereas our study used images of everyday objects and abstract art pieces. In addition to differences in the nature of our stimuli, our study also differed from Salthouse and Siedlecki's in how stimuli were presented. In our object similarity task, objects were presented alongside a category label (e.g., "Types of Guitars"). By providing a semantic label for these stimuli, participants in the current study may have been able to organize these perceptual stimuli in a way which emphasized their semantic interrelatedness, thus resulting in a stronger correlation between false memories on this task and our tasks designed to measure semantic false memory (see Figs. 2–4). In contrast, the dot formations and faces used by Salthouse and Siedlecki did not include semantic or categorical labeling, likely making these stimuli more difficult to categorize than our everyday objects. We contend, however, that although the object similarity task does indeed have a semantic component in that the task presents participants with nameable objects, this task measures perceptual memory in that the ultimate decision about whether a given exemplar within a given category (e.g., guitars) is old or new must be made on the basis of the visual features of the exemplar itself and cannot be made based on object name alone (for studies assessing perceptual memory using similar tasks, see Abe et al., 2013; Davidson et al., 2019; Dennis et al., 2012; Gutchess & Schacter, 2012; C. E. L. Stark et al., 2023).

Importantly, based on a re-analysis of our data suggested by a reviewer, our conclusion that individual differences in false memory are best represented by a single domain-general factor holds regardless of fact that the object similarity task contains nameable categories of objects. That is, even when excluding this task from our CFA analyses and modeling perceptual false memory as a single indicator latent factor (Hayduk & Littvay, 2012; Kline, 2016, 2018) using just the abstract images task, chi-squared difference tests indicate a preference for the

Table 9
Moderation Analyses for Latent Variances and Regression Paths.

	Likelihood Ratio Test	BF_{10}
	Variance Estimates	
False Memory	$X^2(1) = 3.91, p = 0.048$	0.37
Frontal Functioning	$X^2(1) = 1.27, p = 0.261$	0.10
Semantic Discrimination	$X^2(1) = 27.28, p < 0.001$	44,058.69
Perceptual Discrimination	$X^2(1) = 1.00, p = 0.318$	0.09
	Regression Paths (Joint Model)	
Frontal Functioning to False Memory	$X^2(1) = 1.56, p = 0.211$	0.12
Semantic Discrimination to False Memory	$X^2(1) = 1.88, p = 0.171$	0.13
Perceptual Discrimination to False Memory	$X^2(1) = 0.29, p = 0.588$	0.06
	Regression Paths (Isolated Effects)	
Frontal Functioning to False Memory	$X^2(1) = 0.51, p = 0.473$	0.07
Semantic Discrimination to False Memory	$X^2(1) = 0.18, p = 0.668$	0.06
Perceptual Discrimination to False Memory	$X^2(1) = 0.20, p = 0.658$	0.06

Note. BF_{10} values reflect Bayes factors which indicate the degree of support for a moderation effect compared to a lack of moderation given model BIC values. Based on convention standards (Dienes, 2014), a BF_{10} greater than 3 indicates moderate evidence for group moderation, whereas a BF_{10} less than 1/3 indicates moderate evidence of a lack of moderation. BF_{10} values between 3 and 1/3 are interpreted as being ambiguous regarding the presence of moderation.

model with a single false memory factor based on all three tests over separate semantic and perceptual factors for both age groups (see https://osf.io/9rq7c or this analysis). Nonetheless, future research should explore this issue further by assessing the intercorrelations and factor structure of semantic and perceptual false memory tasks across a wider range of stimulus domains while also varying whether explicit category labels are provided during encoding. In this respect, it would be particularly fruitful to replicate our factor analysis results in a sample which not only completes tasks similar to those used by Salthouse and Siedlecki (2007), but also semantic and phonological DRM lists as reviewed in our Introduction (Chang & Brainerd, 2021; Sommers & Lewis, 1999).

Cognitive abilities predict semantic and perceptual false memory

Regarding the cognitive correlates of our false memory factor, our results showed that, when considered together, cognitive abilities predicted a substantial portion of the variability in false memory within both age groups (37 % in younger adults and 40 % in older adults). Such results indicate that adults who score high on tests of cognitive abilities tend to be less susceptible to false memory formation regardless of their age group. In terms of the nature of these relationships, we found that cognitive abilities generally do not have unique associations with false memory (with the exception of a significant, negative effect of perceptual discrimination on false memory for younger but not older adults). Importantly, however, exploratory analyses indicated that each of the three cognitive ability factors was negatively related to false memory when considered in isolation, with these effects being similar in magnitude for both age groups. Taken together, these analyses indicate that cognitive abilities have substantial, negative effects on false memory formation, but that these effects are redundant with one another in that none of these abilities are *uniquely* associated with false memory when controlling for the other ability constructs. Substantively, these results suggest that being high in semantic discrimination, perceptual discrimination, or frontal functioning is protective against false memory formation, but being high in more than one of these constructs is unlikely to provide further additive protection.

The finding that frontal functioning was negatively related to false memory when considered in isolation is consistent with a wealth of prior research demonstrating a negative relationship between frontal functioning constructs such as working memory capacity and false memory (Ball et al., 2022; Butler et al., 2004; Chan & McDermott, 2007; Colombel et al., 2016; Festini & Katz, 2021; Gerrie & Garry, 2007; Leding, 2012; Lövdén, 2003; McCabe et al., 2009; Peters et al., 2006, 2007; Unsworth & Brewer, 2010a, 2010b; Watson et al., 2005) as well as neuroimaging research implicating frontally-mediated processes such as failures of retrieval monitoring in false memory formation (for a *meta*-analysis, see Kurkela & Dennis, 2016). Our results are also consistent with research using the *Mnemonic Similarity Task* (MST), a task in which participants are required to discriminate between previously presented targets and highly similar novel lures (S. M. Stark et al., 2013). Similar to the current results, prior research using the MST has demonstrated that mnemonic discrimination is related individual differences in both frontal functioning (Foster & Giovanello, 2020; Gellersen et al., 2021; Jensen et al., 2023) and to perceptual ability (Davidson et al., 2019; Gellersen et al., 2021). Additionally, our finding that perceptual discrimination was negatively related to false memory is consistent with research demonstrating that false memory is negatively related to perceptual ability (Zhu et al., 2010). The finding that semantic discrimination was negatively related to false memory is also consistent with prior research indicating that having less distinct neural representations within brain regions associated with semantic processing is predictive of false memory (Chadwick et al., 2016). Collectively, such results suggest that false memory formation is due to both domain-general failures of frontally-mediated processes such as retrieval monitoring as well as failures to discriminate between information within domain-specific stimulus categories.

Generally, our results were more consistent with age similarities than age differences in false memory and its cognitive correlates. Because older adults tend to exhibit increased false memory (McCabe et al., 2009) and increased cognitive variability relative to younger adults (Hertzog, 1985; Lindenberger & Oertzen, 2006; Morse, 1993; Nyberg et al., 2012), we predicted that the negative effects of cognitive abilities on false memory would be stronger for older compared to younger adults. However, neither the joint nor the isolated effects of cognitive abilities on false memory differed between older and younger adults. Although the negative effect of perceptual discrimination on false memory was significant for younger but not older adults in the model which also included frontal functioning and semantic discrimination, moderation analyses provided evidence against age differences for this effect. Our finding that the effects of cognitive abilities on false memory were similar for older and younger adults is consistent with the results of Chan and McDermott (2007), who likewise found that age did not moderate the relationship between frontal functioning and false memory. Substantively, such results suggest that the protective effects of factors such as frontal functioning on false memory formation do not differ in magnitude for older and younger adults. This may have meaningful implications for our understanding of the production of false memory across adulthood. Specifically, results suggest that factors which foster cognitive reserve, such as education, may be more influential to false memory than age alone.

Our results regarding the impact of cognitive abilities on false memory in older adults contribute critical knowledge to our understanding of the factors associated with memory errors in later life. Specifically, our factor analysis results suggest that, regardless of age, individuals with an increased susceptibility to false memory exhibit this tendency regardless of stimulus domain. Additionally, the fact that frontal functioning, semantic discrimination, and perceptual discrimination all exhibited negative (albeit redundant) effects on false memory in both age groups suggests that any one of these cognitive abilities may be promising targets for interventions aimed at reducing false memory formation for both older and younger adults. Broadly speaking, the finding that individual differences in cognitive abilities predict a substantial portion of the variance in false memory within older adults highlights the fact that, rather than being interchangeable in their risk for cognitive decline, older adults exhibit important, systematic heterogeneity in their propensity to make memory errors. This conclusion dovetails nicely with prior neuroimaging research, which has shown that individual differences in false memory rates between older adults correlate with neural activity during false memory (Dennis et al., 2014; Dennis & Turney, 2018; Webb & Dennis, 2019).

Age differences in false memory, interindividual variability, and cognitive differentiation

Typically, older adults exhibit higher rates of false memory than younger adults (McCabe et al., 2009). Interestingly, the only task in which older adults had higher rates of false memory for related lures was the abstract images task. This was the only task that did not include any semantic component, nor could participants rely on their prior knowledge during this task. In contrast, the other three memory tasks all allowed for some degree of overt semantic processing (e.g., category labeling and generation). Thus, it may be that the presence of an age-related increase in false memory on the abstract images task is explained by older adults being less able to leverage their relatively intact semantic/prior knowledge during this task to make discriminations between previously seen targets and novel lures. This explanation is consistent with the finding that performance on our semantic discrimination tasks was unrelated to false memory during the abstract images task (see Fig. 2). Even so, this explanation is undermined somewhat by the fact that tasks of semantic discrimination were also unrelated to schematic false memory, which one would assume should rely on semantic knowledge,[7] and the fact that the abstract images task loaded onto the same factor as the other three false memory tasks (*CFA Model 2 younger adult loading* = 0.62, *CFA Model 2 older adult loading* = 0.49).

Though not typical, the absence of age differences in false memory is not without precedent. For example, Salthouse and Siedlecki found that age was unrelated to false memory in two large lifespan samples (2007). Similarly, LaVoie et al. (2005) found that older adults without frontal functioning impairment did not differ from younger adults in false memory. Additionally, previous work from our own lab using the schematic false memory task did not find age differences in false memory (Webb & Dennis, 2019). Importantly, our assessment of false memory using multiple different memory tasks highlights the fact that age-related increases in false memory are not inevitable, but may instead depend critically on task-specific characteristics such as the nature of the studied material and how such materials are presented.

Part of our rationale for predicting that the relationship between cognitive abilities and false memory would be stronger in older adults relates to the finding that older adults typically show increased false memory compared to younger adults. We predicted that, because prior research has found that older adults may be more likely to develop false memories, the protective effects of cognitive abilities might be especially beneficial in this group in comparison to younger adults, for whom false memory may be less of a concern. Because we did not find strong evidence of age-related increases in false memory, an important future direction will be to replicate the current age by cognitive ability moderation results within the context of a sample showing age-related increases in false memory. We note, however, that because older and younger adults often differ in the mechanisms by which they carry out a given cognitive process even when matched on task performance (Daselaar et al., 2006; Reuter-Lorenz & Cappell, 2008; Webb & Dennis, 2019), our investigation of whether the cognitive correlates of false memory are the same or different in older and younger adults is still interesting and informative.

As issues of secondary importance, we also tested for age differences in cognitive variability and cognitive differentiation. Prior aging research emphasizes the fact that older adults tend to be highly variable with respect to their cognition (Hertzog, 1985; Lindenberger & Oertzen, 2006; Morse, 1993; Nyberg et al., 2012). Based on such research, we predicted that older adults would be more variable in terms of false memory than younger adults. Despite this, frequentist analyses demonstrated that younger adults were more variable in false memory than older adults (though Bayesian results were ambiguous). Such results suggest that although it is sometimes the case that older adults exhibit increased heterogeneity compared to younger adults in their cognition, this is not always the case. Future research should continue to examine this issue in order to determine whether the phenomenon of age-related increases in cognitive heterogeneity does or does not generalize to false memory formation.

We also predicted that older adults would show age-related dedifferentiation (La Fleur et al., 2018, but see Koen & Rugg, 2019) with respect to false memory, as indicated by a stronger correlation between their semantic and perceptual false memory factors than younger adults. Although our decision to model false memory using a single latent factor prevented us from addressing this question, we observed some evidence of age-related dedifferentiation with respect to our cognitive ability factors in that frontal functioning was more strongly associated with semantic discrimination in older compared to younger adults. These results suggest that, consistent with prior research, certain cognitive abilities tend to be less differentiated in later life.

Limitations

This study is not without limitations. First, although the reliability of our manifest variables was generally within acceptable ranges, there were two notable exceptions: reliability was lower for the PPT and for the schematic false memory task. Regarding the PPT, we are not the first to report issues related to reliability for this task, as others have reported low reliability both in college students (Klein & Buchanan, 2009) and in a Turkish adult lifespan sample (Bozdemir & Gurvit, 2022). Regarding our false memory tasks, although three of our four tasks exhibited evidence of reliability (*odd–even reliability* between 0.777 and 0.911), the schematic false memory task did not (*younger adult reliability* = 0.500, *older adult reliability* = 0.535). Future research should keep these psychometric limitations in mind when selecting and designing tasks intended to assess individual differences in false memory.

Another aspect of the schematic false memory task that should be kept in mind is that performance on this task was quite low. As such, data from 42 older adults and 53 younger adults were excluded based on our a priori exclusion criterion regarding task performance. Although this left us with a sufficiently large sample of younger adults with data on this task ($n = 152$) based on recommendations for conducting SEM analyses (Wang & Wang, 2019), our analyses included schematic false memory data from only 101 older adults. Although our sample size for this indicator meets Wang and Wang's recommendation of having at least 100 participants per group, it does not meet their recommendation of having 10 participants per indicator ($n = 140$ for the current models). Speculatively, it is possible that the low performance on the schematic false memory task relative to the other three false memory tasks is explained by the fact that this was the only task that required participants to study arrays of multiple items simultaneously, though participants were given additional study time with each array in anticipation of this fact. Future research may refine this test by adjusting parameters such as the number of items in each array in conjunction with each array's presentation time.

An additional limitation of the current study is that, because older adults performed near ceiling on our two semantic discrimination tasks (see https://osf.io/q46p2 for visualizations of these distributions), variability on these tasks within this age group was limited. Such results are unsurprising considering the wealth of research demonstrating that

[7] Though this lack of correlation may also be an artifact of this task having the lowest reliability of the four memory tasks (*younger adult reliability* = 0.500, *older adult reliability* = 0.535).

older adults tend to perform similarly or even outperform younger adults on tests of semantic ability (Hartshorne & Germine, 2015; Park et al., 2002; Salthouse, 2019; Verhaeghen, 2003). A challenge for future research will be to identify appropriate tests which are able to offer more precise measures of semantic discrimination for both older and younger adults while also ensuring that sufficient variability exists within both groups.

A final limitation of the current study is that we only had two indicators for each of our latent false memory factors. It is possible that this fact may explain some of the issues in model estimation seen in CFA Models 1 and 3 which had separate semantic and perceptual false memory factors (i.e., these models had negative variance estimates for some manifest variables). However, this aspect of our design does not explain the fact that our false memory factors exhibited high overlap with one another, which we used as the basis for selecting Model 2 with its single false memory factor for the older adult data. Nonetheless, future research should expand upon this investigation by replicating the current results in a sample which completes a larger number of false memory tests per hypothesized factor.

Conclusions

In conclusion, data from both older and younger adults supports the idea that false memory in both semantic and perceptual tasks is best represented as a single false memory construct. In terms of cognitive correlates, results showed that frontal functioning, semantic discrimination, and perceptual discrimination predicted a substantial portion of the variance in false memory, and each had negative, yet redundant, associations with false memory in both age groups. Importantly, the relationship between cognitive factors and false memory across age groups alongside the absence of age-related increases in false memory in three of the four memory tasks highlights the fact that age-related increases in false memory are not inevitable. Rather, false memory is dependent on cognitive factors that vary within individuals regardless of their age. Collectively, results suggest that these cognitive abilities may be promising targets for interventions intended to reduce false memory formation for both older and younger adults.

Funding and Acknowledgement

This work was supported by a National Science Foundation Grant (BCS1025709) awarded to NAD. JTW was supported by National Institute on Aging Grant T32 AG049676 to The Pennsylvania State University. The authors would like to thank Juliet Palumbo, Andrea John, Desiree Meca, and Kai Ludwig for their help with data collection.

CRediT authorship contribution statement

John T. West: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Rebecca L. Wagner:** Writing – review & editing, Supervision, Project administration, Investigation, Formal analysis, Conceptualization. **Ashley Steinkrauss:** Supervision, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Nancy A. Dennis:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The preregistration, materials, data, analysis code, and rendered analysis output for the current study are available at https://osf.io/x45ju/.

References

Abe, N., Fujii, T., Suzuki, M., Ueno, A., Shigemune, Y., Mugikura, S., Takahashi, S., & Mori, E. (2013). Encoding- and retrieval-related brain activity underlying false recognition. *Neuroscience Research, 76*(4), 240–250. https://doi.org/10.1016/j.neures.2013.05.006

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., & Dervieux, C. (2024). *Quarto* (Version 1.4) [Computer software]. doi: 10.5281/zenodo.5960048.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545–557. https://doi.org/10.1037/0021-843X.112.4.545

Ball, B. H., Robison, M. K., Coulson, A., & Brewer, G. A. (2022). Individual differences in disqualifying monitoring underlie false recognition of associative and conjunction lures. *Memory & Cognition, 50*(4), 751–764. https://doi.org/10.3758/s13421-021-01243-5

Ballardini, N., Yamashita, J. A., & Wallace, W. P. (2008). Presentation duration and false recall for semantic and phonological associates. *Consciousness and Cognition, 17*(1), 64–71. https://doi.org/10.1016/j.concog.2007.01.008

Barense, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia, 45*(13), 2963–2974. https://doi.org/10.1016/j.neuropsychologia.2007.05.023

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology, 39*(1), 15–22. https://doi.org/10.1080/00221309.1948.9918159

Blair, I. V., Lenton, A. P., & Hastie, R. (2002). The reliability of the DRM paradigm as a measure of individual differences in false memories. *Psychonomic Bulletin & Review, 9*(3), 590–596. https://doi.org/10.3758/BF03196317

Bowman, C. R., Chamberlain, J. D., & Dennis, N. A. (2019). Sensory representations supporting memory specificity: Age effects on behavioral and neural discriminability. *Journal of Neuroscience, 39*(12), 2265–2275. https://doi.org/10.1523/JNEUROSCI.2022-18.2019

Bowman, C. R., & Dennis, N. A. (2015). Age differences in the neural correlates of novelty processing: The effects of item-relatedness. *Brain Research, 1612*, 2–15. https://doi.org/10.1016/j.brainres.2014.08.006

Bozdemir, M., & Gurvit, I. H. (2022). Normative data for the Turkish version of the pyramids and palm trees test. *Applied Neuropsychology: Adult, 29*(6), 1571–1577. https://doi.org/10.1080/23279095.2021.1900187

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brainerd, C. J., & Reyna, V. F. (2012). Reliability of children's testimony in the era of developmental reversals. *Developmental Review, 32*(3), 224–267. https://doi.org/10.1016/j.dr.2012.06.008

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language, 66*(3), 407–415. https://doi.org/10.1016/j.jml.2011.12.009

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review, 11*(5), 921–925. https://doi.org/10.3758/BF03196722

Calvillo, D. P., & Parong, J. A. (2016). The misinformation effect is unrelated to the DRM effect with and without a DRM warning. *Memory, 24*(3), 324–333. https://doi.org/10.1080/09658211.2015.1005633

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences, 113*(36), 10180–10185. https://doi.org/10.1073/pnas.1610686113

Chamberlain, J. D., Bowman, C. R., & Dennis, N. A. (2022). Age-related differences in encoding-retrieval similarity and their relationship to false memory. *Neurobiology of Aging, 113*, 15–27. https://doi.org/10.1016/j.neurobiolaging.2022.01.011

Chan, J. C. K., & McDermott, K. B. (2007). The effects of frontal lobe functioning and age on veridical and false recall. *Psychonomic Bulletin & Review, 14*(4), 606–611. https://doi.org/10.3758/BF03196809

Chan, J. C. K., McDermott, K. B., Watson, J. M., & Gallo, D. A. (2005). The importance of material-processing interactions in inducing false memories. *Memory & Cognition, 33*(3), 389–395. https://doi.org/10.3758/BF03193057

Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language, 119*, Article 104210. https://doi.org/10.1016/j.jml.2020.104210

Colarusso, R. P., & Hammill, D. D. (2015). *Motor-Free Visual Perception Test-4*. Academic Therapy Publications Inc.

Colombel, F., Tessoulin, M., Gilet, A.-L., & Corson, Y. (2016). False memories and normal aging: Links between inhibitory capacities and monitoring processes. *Psychology and Aging, 31*(3), 239–248. https://doi.org/10.1037/pag0000086

Cucina, J., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence, 5*(3), 27. https://doi.org/10.3390/jintelligence5030027

Daselaar, S. M., Fleck, M. S., Dobbins, I. G., Madden, D. J., & Cabeza, R. (2006). Effects of healthy aging on hippocampal and rhinal memory functions: An event-related fmri study. *Cerebral Cortex, 16*(12), 1771–1782. https://doi.org/10.1093/cercor/bhj112

Davidson, P. S. R., Vidjen, P., Trincao-Batra, S., & Collin, C. A. (2019). Older adults' lure discrimination difficulties on the mnemonic similarity task are significantly correlated with their visual perception. *The Journals of Gerontology: Series B, 74*(8), 1298–1307. https://doi.org/10.1093/geronb/gby130

Dennis, N. A., Bowman, C. R., & Peterson, K. M. (2014). Age-related differences in the neural correlates mediating false recollection. *Neurobiology of Aging, 35*(2), 395–407. https://doi.org/10.1016/j.neurobiolaging.2013.08.019

Dennis, N. A., Bowman, C. R., & Vandekar, S. N. (2012). True and phantom recollection: An fMRI investigation of similar and distinct neural correlates and connectivity. *NeuroImage, 59*(3), 2982–2993. https://doi.org/10.1016/j.neuroimage.2011.09.079

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. https://doi.org/10.1016/j.neuropsychologia.2007.07.003

Dennis, N. A., Kim, H., & Cabeza, R. (2008). Age-related differences in brain activity during true and false memory retrieval. *Journal of Cognitive Neuroscience, 20*(8), 1390–1402. https://doi.org/10.1162/jocn.2008.20096

Dennis, N. A., & Turney, I. C. (2018). The influence of perceptual similarity and individual differences on false memories in aging. *Neurobiology of Aging, 62*, 221–230. https://doi.org/10.1016/j.neurobiolaging.2017.10.020

Devitt, A. L., & Schacter, D. L. (2016). False memories with age: Neural and cognitive underpinnings. *Neuropsychologia, 91*, 346–359. https://doi.org/10.1016/j.neuropsychologia.2016.08.030

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00781

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test* ((4th ed.).). Pearson.

Farooq, R. (2022). Heywood cases: Possible causes and solutions. *International Journal of Data Analysis Techniques and Strategies, 14*(1), 79–88. https://doi.org/10.1504/IJDATS.2022.121506

Festini, S. B., & Katz, B. (2021). A frontal account of false alarms. *Journal of Cognitive Neuroscience, 33*(9), 1657–1678. https://doi.org/10.1162/jocn_a_01683

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.

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. https://doi.org/10.1016/0022-3956(75)90026-6

Foster, C. M., & Giovanello, K. S. (2020). Domain general processes moderate age-related performance differences on the mnemonic similarity task. *Memory, 28*(4), 528–536. https://doi.org/10.1080/09658211.2020.1743321

Fox, C. J., Mueller, S. T., Gray, H. M., Raber, J., & Piper, B. J. (2013). Evaluation of a short-form of the Berg Card Sorting Test. *PLoS ONE, 8*(5), Article e63885. https://doi.org/10.1371/journal.pone.0063885

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 339–353. https://doi.org/10.1037/0278-7393.27.2.339

Gellersen, H. M., Trelle, A. N., Henson, R. N., & Simons, J. S. (2021). Executive function and high ambiguity perceptual discrimination contribute to individual differences in mnemonic discrimination in older adults. *Cognition, 209*, Article 104556. https://doi.org/10.1016/j.cognition.2020.104556

Gerrie, M. P., & Garry, M. (2007). Individual differences in working memory capacity affect false memories for missing aspects of events. *Memory, 15*(5), 561–571. https://doi.org/10.1080/09658210701391634

Gutchess, A. H., & Schacter, D. L. (2012). The neural correlates of gist-based true and false recognition. *NeuroImage, 59*(4), 3418–3426. https://doi.org/10.1016/j.neuroimage.2011.11.078

Hartshorne, J. K., & Germine, L. T. (2015). When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. *Psychological Science, 26*(4), 433–443. https://doi.org/10.1177/0956797614567339

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology, 12*(1), 159. https://doi.org/10.1186/1471-2288-12-159

Hertzog, C. (1985). An individual differences perspective: Implications for cognitive research in gerontology. *Research on Aging, 7*(1), 7–45. https://doi.org/10.1177/0164027585007001002

Hirschfeld, G., & von Brachel, R. (2019). Improving Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research, and Evaluation, 19*(1). https://doi.org/10.7275/qazy-2946

Howard, D., & Patterson, K. (1992). *The Pyramids and Palm Trees Test*. Pearson.

Howe, M. L., Candel, I., Otgaar, H., Malone, C., & Wimmer, M. C. (2010). Valence and the development of immediate and long-term false memory illusions. *Memory, 18*(1), 58–75. https://doi.org/10.1080/09658210903476514

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jensen, A., Karpov, G., Collin, C. A., & Davidson, P. S. R. (2023). Executive function predicts older adults' lure discrimination difficulties on the mnemonic similarity task. *The Journals of Gerontology: Series B, 78*(10), 1642–1650. https://doi.org/10.1093/geronb/gbad091

Kenny, D. A. (2020). *SEM: Measuring model fit*. https://davidakenny.net/cm/fit.htm.

Klein, L. A., & Buchanan, J. A. (2009). Psychometric properties of the pyramids and palm trees test. *Journal of Clinical and Experimental Neuropsychology, 31*(7), 803–808. https://doi.org/10.1080/13803390802508926

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford Publications.

Kline, R. B. (2018). Response to Leslie Hayduk's review of Principles and Practice of Structural Equation Modeling, 4th edition. *Canadian Studies in Population [ARCHIVES], 45*(3–4), Article 3–4. doi: 10.25336/csp29418.

Koen, J. D., & Rugg, M. D. (2019). Neural dedifferentiation in the aging brain. *Trends in Cognitive Sciences, 23*(7), 547–559. https://doi.org/10.1016/j.tics.2019.04.012

Kurkela, K. A., & Dennis, N. A. (2016). Event-related fMRI studies of false memory: An Activation Likelihood Estimation meta-analysis. *Neuropsychologia, 81*, 149–167. https://doi.org/10.1016/j.neuropsychologia.2015.12.006

La Fleur, C.-G., Meyer, J. M., & Dodson, C. (2018). Exploring dedifferentiation across the adult lifespan. *Psychology and Aging, 33*(5), 855–870. https://doi.org/10.1037/pag.0000274

LaVoie, D. J., Willoughby, L., & Faulkner, K. (2005). Frontal lobe dysfunction and false memory susceptibility in older adults. *Experimental Aging Research, 32*(1), 1–21. https://doi.org/10.1080/01902140500325023

Leding, J. K. (2012). Working memory predicts the rejection of false memories. *Memory, 20*(3), 217–223. https://doi.org/10.1080/09658211.2011.653373

Lindenberger, U., & von Oertzen, T. (2006). Variability in cognitive aging: From taxonomy to theory. In E. Bialystok, & F. I. M. Craik (Eds.), *Lifespan Cognition: Mechanisms of Change* (pp. 297–314). Oxford University Press. https://doi.org/10.1093/acprof:Oso/9780195169539.003.0021.

Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*(2), 268–283. https://doi.org/10.1016/S0749-596X(03)00069-X

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software, 4*(40), 1541. https://doi.org/10.21105/joss.01541

Malmi, R. A., Underwood, B. J., & Carroll, J. B. (1979). The interrelationships among some associative learning tasks. *Bulletin of the Psychonomic Society, 13*(3), 121–123. https://doi.org/10.3758/BF03335032

Markon, K. E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretation. *Annual Review of Clinical Psychology*, *15*(Volume 15, 2019), 51–69. doi: 10.1146/annurev-clinpsy-050718-095522.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., & Balota, D. A. (2009). Aging reduces veridical remembering but increases false remembering: Neuropsychological test correlates of remember–know judgments. *Neuropsychologia, 47*(11), 2164–2173. https://doi.org/10.1016/j.neuropsychologia.2008.11.025

McDermott, K. B., & Watson, J. M. (2001). The Rise and Fall of False Recall: The Impact of Presentation Duration. *Journal of Memory and Language, 45*(1), 160–176. https://doi.org/10.1006/jmla.2000.2771

McDonough, I. M., Wong, J. T., & Gallo, D. A. (2013). Age-related differences in prefrontal cortex activity during retrieval monitoring: Testing the compensation and dysfunction accounts. *Cerebral Cortex, 23*(5), 1049–1060. https://doi.org/10.1093/cercor/bhs064

Meade, M. L., Geraci, L. D., & Roediger, H. L. (2012). Neuropsychological status in older adults influences susceptibility to false memories. *The American Journal of Psychology, 125*(4), 449–467. https://doi.org/10.5406/amerjpsyc.125.4.0449

Morse, C. K. (1993). Does variability increase with age? An archival study of cognitive measures. *Psychology and Aging, 8*(2), 156–164. https://doi.org/10.1037/0882-7974.8.2.156

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods, 222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Nyberg, L. (1994). A structural equation modeling approach to the multiple memory systems question. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 485–491. https://doi.org/10.1037/0278-7393.20.2.485

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences, 16*(5), 292–305. https://doi.org/10.1016/j.tics.2012.04.005

Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K., & Salmon, K. (2013). False memory ≠ false memory: DRM errors are unrelated to the misinformation effect. *PLoS ONE, 8*(4), Article e57939. https://doi.org/10.1371/journal.pone.0057939

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17*(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Patihis, L., Frenda, S. J., & Loftus, E. F. (2018). False memory tasks do not reliably predict other false memories. *Psychology of Consciousness: Theory, Research, and Practice, 5*(2), 140–160. https://doi.org/10.1037/cns0000147

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pellegrino, J. W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. *Intelligence, 3*(3), 187–215. https://doi.org/10.1016/0160-2896(79)90017-5

Pesta, B. J., Murphy, M. D., & Sanders, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 328–338. https://doi.org/10.1037/0278-7393.27.2.328

Peters, M. J. V., Jelicic, M., Haas, N., & Merckelbach, H. (2006). Mild executive dysfunctions in undergraduates are related to recollecting words never presented. *International Journal of Neuroscience, 116*(9), 1065–1077. https://doi.org/10.1080/00207450600808768

Peters, M. J. V., Jelicic, M., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *European Journal of Cognitive Psychology, 19*(2), 213–232. https://doi.org/10.1080/09541440600760396

Piper, B. J., Li, V., Eiwaz, M. A., Kobel, Y. V., Benice, T. S., Chu, A. M., Olsen, R. H. J., Rice, D. Z., Gray, H. M., & Mueller, S. T. (2012). Executive function on the Psychology Experiment Building Language tests. *Behavior Research Methods, 44*(1), 110–123. https://doi.org/10.3758/s13428-011-0096-6

Pornprasertmanit, S., Miller, P., Schoemann, A., Quick, C., Jorgensen, T., & Pornprasertmanit, M. S. (2016). *Package "simsem"* [Computer software]. https://cran.r-project/.

Posit team. (2024). *Rstudio: Integrated development environment for R* [Computer software]. Posit Software, PBC. http://www.posit.co/.

Qualtrics. (2024). *Qualtrics* [Computer software]. Qualtrics. https://www.qualtrics.com/.

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current Directions in Psychological Science, 17*(3), 177–182. https://doi.org/10.1111/j.1467-8721.2008.00570.x

Robison, M. K., Celaya, X., Ball, B. H., & Brewer, G. A. (2023). Task sequencing does not systematically affect the factor structure of cognitive abilities. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02369-0

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), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9*(4), 123–127. https://doi.org/10.1111/1467-8721.00075

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging, 34*(1), 17–24. https://doi.org/10.1037/pag0000288

Salthouse, T. A., & Siedlecki, K. L. (2007). An individual difference analysis of false recognition. *The American Journal of Psychology, 120*(3), 429–458. https://doi.org/10.2307/20445413

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5*(4), 710–715. https://doi.org/10.3758/BF03208850

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. https://doi.org/10.1006/jmla.1998.2614

Stark, C. E. L., Noche, J. A., Ebersberger, J. R., Mayer, L., & Stark, S. M. (2023). Optimizing the mnemonic similarity task for efficient, widespread use. *Frontiers in Behavioral Neuroscience, 17*. https://doi.org/10.3389/fnbeh.2023.1080366

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. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29*(3), 424–432. https://doi.org/10.3758/BF03196393

Thomas, A. K., & McDaniel, M. A. (2013). The interaction between frontal functioning and encoding processes in reducing false memories. *Aging, Neuropsychology, and Cognition, 20*(4), 443–470. https://doi.org/10.1080/13825585.2012.736468

Tsukiura, T. (2014). Neural correlates of remembering false memories in young and older adults: A brief review of fMRI studies. *The Journal of Physical Fitness and Sports Medicine, 3*(2), 155–161. https://doi.org/10.7600/jpfsm.3.155

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General, 107*(4), 393–419. https://doi.org/10.1037/0096-3445.107.4.393

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134*(1), 16–28. https://doi.org/10.1016/j.actpsy.2009.11.010

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin, 145*(1), 79–139. https://doi.org/10.1037/bul0000176

Unsworth, N., & Brewer, G. A. (2009). Examining the relationships among item recognition, source recognition, and recall from an individual differences perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1578–1585. https://doi.org/10.1037/a0017255

Unsworth, N., & Brewer, G. A. (2010a). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62*(1), 19–34. https://doi.org/10.1016/j.jml.2009.08.002

Unsworth, N., & Brewer, G. A. (2010b). Variation in working memory capacity and intrusions: Differences in generation or editing? *European Journal of Cognitive Psychology, 22*(6), 990–1000. https://doi.org/10.1080/09541440903175086

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity—Fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin & Review, 16*(5), 931–937. https://doi.org/10.3758/PBR.16.5.931

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*(4), 392–406. https://doi.org/10.1016/j.jml.2010.02.001

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging, 18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332

Wang, J., & Wang, X. (2019). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

Ware, J., Kosinski, M., & Keller, S. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220–233. https://doi.org/10.1097/00005650-199603000-00003

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 76–85. https://doi.org/10.1037/0278-7393.31.1.76

Webb, C. E., & Dennis, N. A. (2019). Differentiating true and false schematic memories in older adults. *The Journals of Gerontology: Series B, 74*(7), 1111–1120. https://doi.org/10.1093/geronb/gby011

Webb, C. E., Turney, I. C., & Dennis, N. A. (2016). What's the gist? The influence of schemas on the neural correlates underlying true and false memories. *Neuropsychologia, 93*, 61–75. https://doi.org/10.1016/j.neuropsychologia.2016.09.023

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-IV* ((4th ed.).). Pearson.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., Li, H., Xue, G., Lu, Z., & Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory, 18*(5), 543–555. https://doi.org/10.1080/09658211.2010.487051