Neuropsychologia 119 (2018) 448–463

Contents lists available at ScienceDirect

Neuropsychologia

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

Recognition memory shielded from semantic but not perceptual interference in normal aging

D. Merika Wilson*, Kevin W. Potter, Rosemary A. Cowell*

Department of Psychological and Brain Sciences, University of Massachusetts, Tobin Hall, 135 Hicks Way, Amherst, MA 01003, United States

A R T I C L E I N F O

Keywords:
Age-related memory deficits
Deese–Roediger–McDermott paradigm
Signal detection theory
Recognition memory

A B S T R A C T

Normal aging impairs long-term declarative memory, and evidence suggests that this impairment may be driven partly by structural or functional changes in the medial temporal lobe (MTL). Theories of MTL memory function therefore make predictions for age-related memory loss. One theory – the Representational-Hierarchical account – makes two specific predictions. First, recognition memory performance in older participants should be impaired by feature-level interference, in which studied items contain many shared, and thus repeatedly appearing, perceptual features. Second, if the interference in a recognition memory task – i.e., the information that repeats across items – resides at a higher level of complexity than simple perceptual features, such as semantic gist, older adults should be less impacted by such interference than young adults. We tested these predictions using the Deese-Roediger-McDermott paradigm, by creating feature-level (i.e., perceptual) interference with phonemically/orthographically related word categories, and higher-level associative interference with semantically related word categories. We manipulated category size in order to compare the effect of less versus more interference (i.e., small versus large category size), which served to (1) avoid potential item confounds arising from systematic differences between words belonging to perceptually- versus semantically-related categories, and (2) ensure that any effect of interference was due to information encoded at study, rather than pre-experimentally. Further, we used signal detection theory (SDT) to interpret our data, rather than examining false alarm (FA) rates in isolation. The d′ measure derived from SDT avoids contamination of the memory measure by response bias, and lies on an interval scale, allowing memory performance in different conditions to be compared without violating assumptions of the statistical tests. Older participants were relatively more impaired by perceptual interference and less impaired by semantic interference than young adults. This pattern is at odds with many current theories of age-related memory loss, but is in line with the Representational-Hierarchical account.

1. Introduction

Long-term declarative memory function declines in old age. Deficits in older adults have been reported for several forms of long-term declarative memory including episodic, semantic, and spatial memory (Bäckman and Nilsson, 1996; Newman and Kaszniak, 2000; Rönnlund et al., 2005; Schaie, 2005). In particular, age-related impairments are frequently found in tasks that require the retrieval of information via associations, such as in contextual, associative, and source memory tasks (Chalfonte and Johnson, 1996; Duarte et al., 2008; Naveh-Benjamin, 2000; Schacter et al., 1991, 1994; Silver et al., 2012; Spencer and Raz, 1995; but see Campbell et al., 2010). Regarding the neuroanatomical changes associated with age-related memory decline, research has revealed disproportionate deterioration of medial temporal lobe (MTL) structures (Devitt and Schacter, 2016; Raz et al., 1998, 2004; Wang et al., 2002). Given the cognitive effects of aging, this finding fits with the well-documented role of the MTL in long-term declarative memory (Scoville and Milner, 1957; Squire and Wixted, 2011; Squire and Zola-morgan, 1991) and, more specifically, in associative or relational memory (e.g., Cohen et al., 1999; Eichenbaum et al., 1994; Giovanello et al., 2004).

One theory of the role of MTL structures in cognition – the Representational-Hierarchical (R-H) account (Bussey and Saksida, 2002; Cowell et al., 2006; Sadil and Cowell, 2017) – has begun to amass empirical support for its explanation of mnemonic and perceptual deficits following brain damage (Barense et al., 2005, 2007, 2012a; Bartko et al., 2010, 2007b, 2007a; Lee et al., 2005a, 2005b, 2007; McTighe et al., 2010). However, to our knowledge this theory has not been applied to the effects of aging on memory (but see Scheerer and Marrone, 2014). And yet it makes clear predictions for memory

* Corresponding authors.
E-mail addresses: dmerikawilso@umass.edu (D.M. Wilson), rcowell@umass.edu (R.A. Cowell).

https://doi.org/10.1016/j.neuropsychologia.2018.07.031
Received 13 October 2017; Received in revised form 2 July 2018; Accepted 28 July 2018
Available online 30 July 2018
0028-3932/

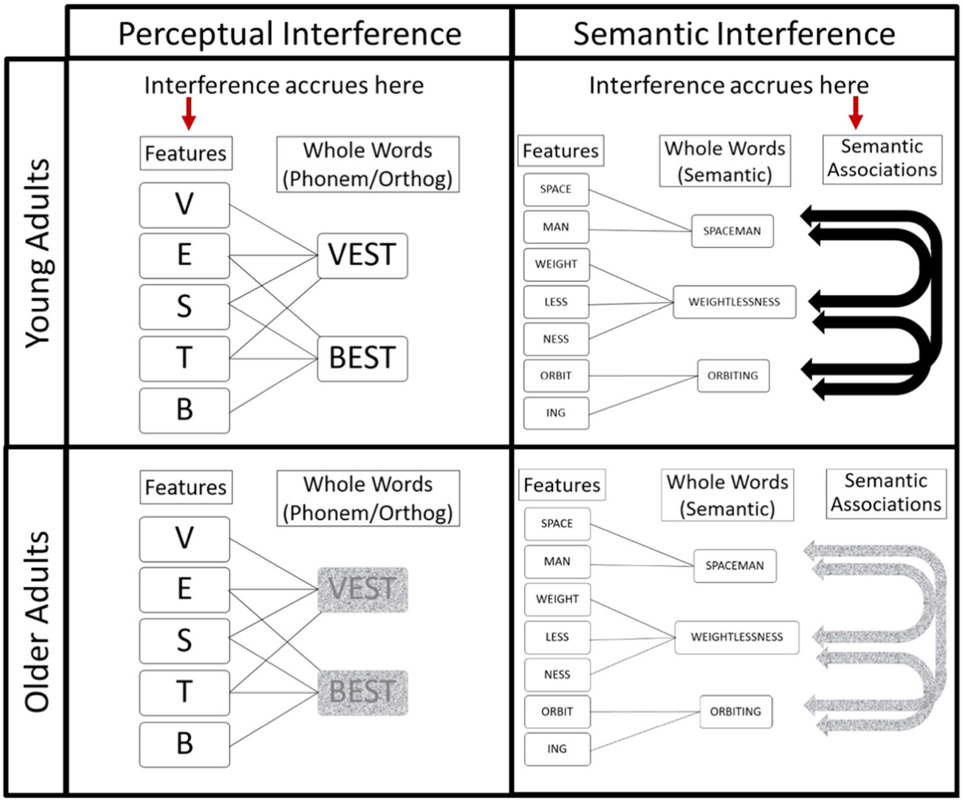

Fig. 1. Age-Dependent Predictions of the R-H Account for Perceptual and Semantic Interference. Two proposed representational hierarchies for words and their corresponding interference: a phonemic/orthographic hierarchy affected by feature-level perceptual interference (left column) and a semantic hierarchy affected by higher-level semantic interference (right column). Each hierarchy contains levels of representations that increase in conjunctive complexity from left to right. The phonemic/orthographic feature level is depicted as letters but may contain any sub-lexical orthographic or phonemic units. The semantic feature level comprises morphemes, which possess elements of meaning but might not be whole words. In line with models of word recognition, the semantic feature level must receive inputs in the form of orthographic/phonemic features (not shown). The two hierarchies' whole word levels are proposed to co-exist in parallel. **Left**: Perceptual interference arises when phonemic/orthographic features occur repeatedly (in perceptually-related categories), causing representations of those perceptual features to appear familiar even when they are part of a novel word. For perceptually-related categories there is no strong semantic relation between the words (e.g., "vest" and "best") such that interference is confined mainly to the feature-level of the phonemic/orthographic hierarchy. **Right:** Semantic interference arises when a semantic theme occurs repeatedly (in semantically-related categories), causing unstudied whole-word representations to appear familiar via inter-item semantic associative activation. For words in a semantically-related category there is little phonemic/orthographic feature overlap such that interference is confined mainly to the association-level of the semantic hierarchy. **Top Row:** In young adults, all representations are intact. **Bottom Row:** In older adults, incipient damage appears (grey shading) for word-level phonemic/orthographic information (left) and for the semantic associations between words (right).

performance that distinguish it from other theories of age-related memory loss (e.g., Brainerd et al., 2009; Naveh-Benjamin, 2000; Yassa and Stark, 2011). In this article, we outline those predictions and report an empirical study that tested them.

Briefly, the R-H account assumes that the ventral visual pathway contains a hierarchically organized system of representations that culminates in the MTL. At the start of the pathway, in visual cortex, the simple features of visual stimuli (e.g., color, orientation) are represented individually, and those features are brought together into cohesive conjunctions of increasing complexity in progressively more anterior regions. MTL structures contain conjunctive representations of whole stimuli such as objects, scenes and episodic events. These representations determine the role of MTL structures in cognition: whenever a cognitive task requires discrimination between stimuli or events containing shared features, conjunctive MTL representations are needed to resolve the feature ambiguity (Bussey and Saksida, 2002; Bussey et al., 2002). This prediction has been confirmed using tasks that are traditional tests of memory (e.g., Bartko et al., 2010, 2007b; McTighe et al., 2010; Yeung et al., 2013) as well as tasks that tap perception (e.g., Barense et al., 2005; Barense et al., 2012a; Barense et al., 2012b; Bartko et al., 2007a, 2007b; Lee, 2006; Lee et al., 2005b; Lee et al., 2006; Lee et al., 2008), lending support to the proposal that the role of MTL structures in cognition is determined by representational content, rather than by specialization for a particular cognitive process.

If normal aging compromises MTL structures and the representations contained therein, then the R-H account makes predictions for age-related changes in memory. Although cognitive aging is a complex phenomenon that likely stems from widespread changes in brain structure and function (Cabeza et al., 2016; Fjell and Walhovd, 2010), to the extent that MTL structures are disproportionately affected by aging, the R-H account predicts that age-related impairments in recognition memory should share some properties with the impairments seen after MTL damage. In this study we test whether the R-H account can explain age-related memory changes by focusing on one of its key predictions for memory function, regarding the resolution of different types of mnemonic interference. Specifically, the R-H account makes clear predictions for "feature-level interference" – a kind of perceptual interference in which features possessed by the to-be-discriminated test stimuli appear repeatedly during a task. According to the R-H account, such interference should impair the discrimination of old and new items when there is damage to conjunctive representations in MTL, because these representations are necessary for uniquely specifying an object (or event) and distinguishing it from similar items with shared features (Cowell et al., 2006). In support of this, studies have shown that feature-level interference causes impairments in recognition memory in individuals with MTL damage (Bartko et al., 2010, 2007b; McTighe et al., 2010; Yeung et al., 2013). To the extent that aging damages the representations and processes normally carried out in MTL, the R-H account predicts the same interference effect in older relative to younger adults. However, since this prediction derives from the claim that conjunctive representations in MTL are compromised, a second, more counterintuitive prediction for interference also follows from the R-H account. Consider a scenario in which the repeatedly appearing (i.e., interfering) attributes of the to-be-discriminated stimuli reside at the level of complex conjunctions, or high-level associations – effectively creating "conjunction-level" interference. In this case, older adults should be paradoxically shielded from interference, because they are less able to encode that interference than young adults.

We operationalized these predictions for healthy older adults using a version of the Deese-Roediger-McDermott (DRM) recognition memory paradigm (Deese, 1959; Roediger and McDermott, 1995). In this paradigm, participants are presented with categories of related items at study, then tested with both studied items and unstudied lures, with some of the lures related in theme to the studied items. Participants of all ages tend to endorse related lures as studied with a higher probability than unrelated lures, a memory impairment that could be explained in terms of interference between similar items (see Gallo, 2006 for a review). Our goal was to use this paradigm to create both feature-level and conjunction-level interference.

Word stimuli constitute the largest and most easily manipulated source of items for use with the DRM paradigm. Given that we had previously applied the R-H account only to visual stimuli, in order to use words, we needed to specify the format of word representations within the assumed representational hierarchy. Popular models of visual word recognition and models of language production assume that humans possess two separate representational hierarchies for words – one stream for phonemic/orthographic features and the other for semantic content (Coltheart et al., 2001; Dell, 1986). This assumption is justified intuitively by considering that two very orthographically similar words such as TIDE and TILE have extremely dissimilar and separable semantic content. We therefore took our inspiration from these prominent models of word representation and blended their assumption with those of the R-H account, yielding a model with two separate streams for word representations that each contain a hierarchical continuum of increasingly complex representations (Fig. 1). We propose that the phonemic/orthographic hierarchy culminates in a conjunctive representation of the combination of phonemic/orthographic features that uniquely defines a whole word, assuming the word has no homonyms (Fig. 1, left column). In contrast, the semantic hierarchy culminates at the level of semantic themes or "gist", which is represented via the semantic associations between individual words, rather than being contained within individual word representations themselves (Fig. 1, right column). We further suggest that older adults have compromised representations at the top level of both hierarchies: aging renders whole-word level representations of the conjunctions of phonemic/orthographic features less precise, and renders representations of the semantic associations between words less robust (Fig. 1, bottom row).

The notion that the MTL and nearby temporal cortex are involved in representing individual words phonemically/orthographically, particularly when those words are similar, and that aging can affect these representations, was inspired by a number of empirically-based claims: that perirhinal cortex is situated at the top of the visual word processing hierarchy (Liuzzi et al., 2015); that perirhinal cortex is engaged specifically when abstract visual stimuli share a high degree of feature-overlap (Barense et al., 2012a, 2010); and that older adults suffer increased interference, or competition, between phonological neighbors (Abrams et al., 2007; Gordon and Kurczek, 2013).[1] The notion that semantic properties and semantic associations depend upon MTL structures is supported by a considerable body of research (Barense et al., 2010; Bruffaerts et al., 2013; Chadwick et al., 2016; Clarke and Tyler, 2014; Manns et al., 2003; Warrington and Shallice, 1984).

By using the DRM paradigm to create interference that affects these two putative processing streams at different levels, we created the conditions of feature-level and conjunction-level interference necessary for testing our predictions. Specifically, we created categories of items possessing shared phonemic/orthographic features but few semantic associations (e.g., BEST, LEST, NEST, PEST, REST, TEST, VEST, WELT, WENT, WEPT), and categories of items that shared a semantic theme but few phonemic-orthographic features (e.g., ASSAILANT, BANDIT, BOOTY, BURGLARY, HOLDUP, MUGGING, STEALING, STICKUP, THEFT, WALLET). In the phonemic/orthographic categories, the shared information across items that provides the source of interference resides at the *sub-word level*: the letter- and phoneme- components appear repeatedly. This causes the representations of simple perceptual features to appear familiar, even for novel words. In young adults (Fig. 1, top row), intact phonemic/orthographic representations at the whole-word level can resolve this interference because the unique conjunction corresponding to an individual stimulus is not repeated and does not become familiar for novel items. But in older adults the conjunctive word-level representations of phonemes/orthography are compromised (Fig. 1, bottom row), forcing reliance upon feature-level representations, thereby increasing the effects of interference. Thus, the phonemic/orthographic categories provide a "feature-level interference" condition in which older adults should be less able to resolve the interference than young adults. In semantic categories, the shared information across items resides at the *supra-word level*: inter-item semantic associations. When a semantic theme occurs repeatedly, word-level semantic representations for words that were unstudied but related to the theme begin to accrue familiarity owing to inter-item semantic associative activation. That is, the effects of semantic interference on novel test items are *mediated* by the higher-level semantic associations. Because these associations are assumed to be weaker in older adults than in young adults, this interference is less potent for older adults. Thus, the semantic categories provide a "conjunction-level interference" condition in which older adults should be paradoxically shielded from interference, relative to young adults.[2]

Despite a number of prior studies investigating age-related changes in semantic and perceptual memory using the DRM paradigm, these predictions have not yet been tested. Difficulty in drawing relevant inferences from the literature stems from two sources. First, some prior DRM studies of memory in older adults are subject to limitations in task design or data analysis that allow for non-mnemonic explanations of the data, either in terms of differences in response bias between groups and conditions, or in terms of item confounds. Second, perhaps because of the variety of task designs and analyses employed, findings from prior DRM studies comparing perceptual and semantic interference in older adults are conflicting: some have reported age-related deficits for perceptual but not semantic interference (Ly et al., 2013), others have observed greater deficits for semantic than perceptual (Koutstaal et al., 2003), and others still have found deficits for both (Budson et al., 2003; Pidgeon and Morcom, 2014).

Thus, the goal of the present study was to test whether the R-H account of MTL amnesia can explain how the effects of interference on memory change in normal aging, with an empirical study that avoids some of the confounds that complicate the interpretation of prior, related research. Specifically, we included three important features in our experimental design and analysis, as follows.

First, we used signal detection theory (SDT) to extract a measure of discriminability, d′. We use d′ because in a standard memory task like

[1] We acknowledge the evidence for a role of the 'visual word form area' (VWFA, in the fusiform gyrus) in the representation of orthographic properties of words (e.g., Cohen and Dehaene, 2004; Glezer et al., 2015) and for a role of auditory cortex in the representation of phonemic properties (see Hickok and Poeppel, 2007 for a review). We therefore remain cautious about tying the word-level of the phonemic/orthographic hierarchy to a single cortical location, or strictly to anterior temporal regions, but we suggest that it is nonetheless plausible to assume aging causes incipient damage to these conjunctive phonemic/orthographic representations.

[2] We assume that whole-word level semantic representations are intact to make the most conservative claim regarding the effects of aging. One could assume mildly impoverished whole-word semantic representations and the R-H account would make similar predictions. This is because the semantic interference condition does not cause interference (i.e., repeated presentation of mnemonic material) at the level of whole-words, but rather at the *supra*-word level of semantic associations, i.e., the words are much more often associates (e.g., "gambler" and "poker") than synonyms (e.g., "gambler" and "bettor"). Thus, the words most often do not share internal, semantic features but instead are related via association with a higher semantic theme. Because "gambler" and "poker" do not share internal, semantic features, their semantic whole-word representations could be rendered less precise in old age without causing them to become indiscriminable.

the DRM paradigm the R-H account makes predictions for recognition memory performance, rather than for false memory (i.e., false alarm (FA) rates). In standard recognition memory tasks with a yes/no test format, response bias plays a role, therefore assumptions about decision criteria are required to make *a priori* predictions for hit and FA rates. The R-H account makes no such assumptions and so, for the DRM paradigm, it makes predictions only for memory sensitivity, as indexed by a dependent variable such as d′. (But see McTighe et al. (2010) and Yeung et al. (2013) for incidental recognition memory tasks in rats and humans, respectively. In these tasks, decision criteria had little or no influence on memory performance – which was indexed by exploration time or eye fixations – and hence the R-H account was able to make predictions specifically concerning false memory for these studies). Further, the d′ measure confers two key advantages over hit and FA rates. The first advantage is that d′, given the assumptions of SDT, lies on an interval scale. The inferential test that we planned to use – analysis of variance (ANOVA) – assumes that the dependent variable is normally distributed with homogeneous variance; these properties are frequently violated by proportion scores, such as hit and FA rates, but much less often violated by an interval scale measurement such as d′. Furthermore, the use of an interval scale was critical because our key prediction was an interaction between the experimental manipulations (small/large category size, perceptual/semantic interference type) and participant group (older/young age). While main effects are testable regardless of the measurement scale, a non-crossover interaction observed with a dependent variable that does not lie on an interval scale (such as hit or FA rate) may be "removable" and hence uninterpretable (Loftus, 1978; Wagenmakers et al., 2012). The second advantage of d′ is that, unlike FA rate, it is unconfounded by possible differences in response bias between groups or conditions. In particular, we were concerned that participants may be able to maintain different criteria for different categories, even when items from all categories are interleaved at test. For example, the presence of systematically differing word properties across semantic versus phonemic/orthographic categories (e.g., word frequency, discussed further below) might differentially affect response criteria for the two conditions.[3] Moreover, the propensity to use distinct criteria for different categories might differ between young and older participants, such that any group differences in the hit or FA rate reflect differences in response bias rather than memory. Given these important advantages of d′, we report hit and FA rates for completeness only. To draw conclusions about memory, we focus exclusively on d′, which tests the specific predictions of our theoretical account, conforms to the assumptions of our statistical tests, and provides an index of memory performance that is independent from response bias.

Second, similar to many prior DRM studies, we interleaved the presentation of items from all categories in both study and test phases. This served two main purposes: (1) it decreased the salience of the shared themes within each category, increasing the chance that older adults with putatively compromised semantic/associative representations would fail to apprehend those associations; and (2) it helped avoid any confounding effects of fatigue on performance for words drawn from different category sizes (category size was a critical manipulation in our experimental design, as discussed below). That is, if words from small and large sized categories are tested in a blocked format, poorer performance on larger-sized categories could be induced either by additional interference from the category or by fatigue from being tested on a longer test list. In addition, interleaved presentation may have reduced participants' tendencies to use different response criteria across different category types at test (Shiffrin et al., 1995), but we are doubtful that it can eliminate this tendency entirely, as noted above.

Third, and most importantly, we manipulated category size, using categories of 2 and 8 items at study, in order to provide a measure of mnemonic interference for each condition that is not subject to item confounds. In a task design comparing perceptual with semantic interference, the experimenter's goal would typically be to observe an effect of interference type, allowing the conclusion that participants are differentially susceptible to the two types of interference (i.e., the build-up of interfering information in memory during the study phase). However, if only one category size is employed (e.g., Ly et al., 2013), any effect of interference type might be equally well explained by differences in inherent properties of words in the two interference types, such as frequency, pre-experimental familiarity, word length, and so on. That is, because words cannot be counterbalanced across the two conditions, their inherent properties are not controlled for when only a single category size is used. Instead, if each condition includes two category sizes (across which inherent word properties remain constant, if the assignment of particular words to category sizes is counterbalanced across participants) and we examine the difference between d′ for 2-item and 8-item categories, we can measure the influence of study-related *interference* on recognition performance for each condition, in the absence of item confounds. Thus, our key dependent variable was an interference measure – the *difference* between d′ for 2-item and d′ for 8-item categories. (The importance of using a dependent variable that lies on an interval scale is further illustrated in this context: in order for the proposed difference score to be interpretable, we must be confident that the psychological distance between values of 1 and 2 on the dependent variable is the same as the psychological distance between 2 and 3. This is theoretically true for d′, but not true for probability scores such as hit or FA rates. This issue is the same as the point originally made by Loftus (1978; and expanded on by Wagenmakers et al., 2012) when he argued that a statistical interaction – a difference of differences – obtained using response probabilities as a dependent variable is often uninteresting. Only a subset of such interactions are interpretable, because any non-crossover interaction may be removed or augmented depending on exactly how the response probability non-linearly transforms the psychological variable of interest. Hits and FAs non-linearly transform the underlying "memory strength" variable of interest, whereas d′ provides a linear measure.)

Our predictions can be summarized as follows. The R-H account assumes that stimulus representations become more complex with progression through the visual and auditory processing pathways toward anterior and MTL, giving rise to representations of whole words and the semantic relations between them. It also assumes that the more anterior, more complex representations (both word-level and semantic-level) are compromised in older adults, but intact in the young. The R-H account thus predicts an interaction for recognition memory performance between age (i.e., compromised versus intact anterior representations) and interference type (i.e., perceptual versus semantic). In the case of perceptual interference, low-level features of the words appear repeatedly. When anterior representations are compromised (in older but not young participants), participants must rely on posterior feature-based representations; this reduces the discriminability of old from new items, because even new items contain familiar features. This effect is greater in large than in small perceptual categories, because large categories entail more feature repetition. In the case of semantic interference, the compromised anterior representations (in older but not young) have the opposite effect. The build-up of interference across semantic categories occurs at a higher level of representation – the

[3] Prior research has examined whether participants make trial-by-trial adjustments in their decision criterion, in particular when words from different categories are presented in an interleaved fashion in a mixed test list. However, the findings are ambiguous. Some studies have found that participants do not adjust (Stretch and Wixted, 1998; Wixted and Stretch, 2000), while other studies have found that they do (e.g., Singer et al., 2002; Singer and Wixted, 2006). The critical factor appears to be the types of categories to which words in the mixed test list belong (e.g., categories defined by item strength and cued by font color, versus categories defined by study-test delay). To our knowledge, no prior research has examined whether participants shift their decision criterion under the specific conditions used in our task, in particular the perceptual versus semantic distinction.

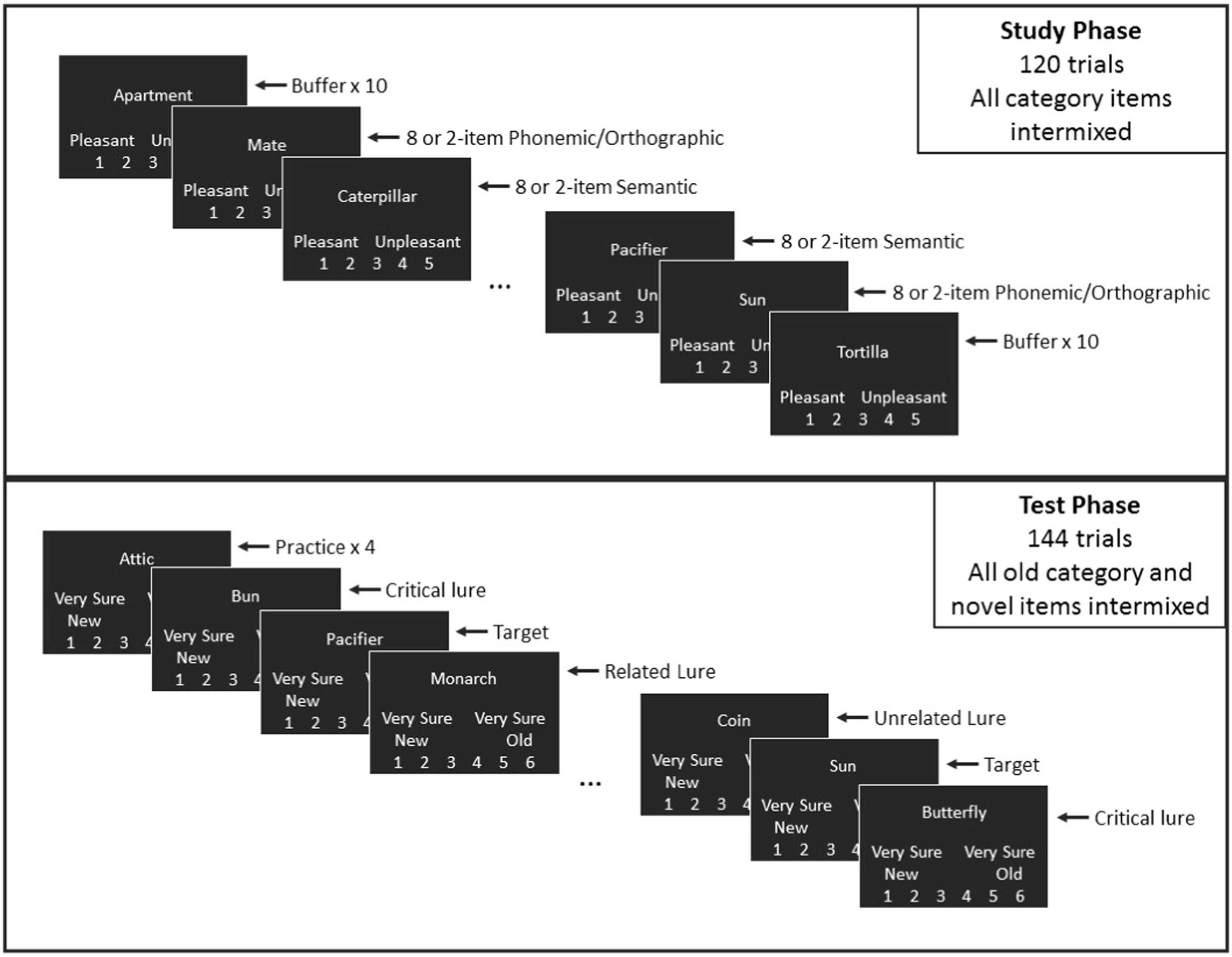

Fig. 2. Illustration of Experimental Paradigm. The study phase comprised 120 trials, 20 of which served as "buffer" trials, 10 at the beginning and 10 at the end of the study list, to prevent recency and primacy effects. The remaining 100 study trials sampled words from 10 semantic categories and 10 phonemic/orthographic categories (drawing 2 items from 5 of the semantic categories and 8 items from the other 5 semantic categories; likewise for phonemic/orthographic categories). Words sampled from these categories were intermixed in the study phase and appeared as one long list. The test phase comprised 144 trials, including 4 practice trials at the beginning of the test list. Of the remaining 140 trials, sampled words were either targets (studied category exemplar words), critical lures (non-studied category prototype words), related lures (non-studied category exemplar words), or unrelated lures (non-studied words unrelated to any category). Words from these categories were intermixed in the test phase, appearing as one long list. During the study phase, participants were asked to respond during the 3000 ms presentation of each word; during the test phase, participants could take as long as they needed to respond.

semantic relations between different words. Now, older participants are paradoxically shielded, because the high-level associations mediating the semantic interference are less robust than in young adults. Therefore, using the d′2-d′8 difference measure, our critical prediction was an interaction between age group and interference type: older adults, relative to young adults, should suffer greater interference for phonemic/orthographic than for semantic categories.

2. Methods

2.1. Participants

A total of 120 participants were recruited from the University of Massachusetts-Amherst and the local community, including 40 older adults and 80 young adults. One older adult was excluded before analysis because MATLAB quit unexpectedly during the study phase. A further seven (three older and four younger) subjects were excluded during analysis (see Section 2.4.1. Signal detection model). Of the remaining participants, older participants were between the ages of 60 and 92 years old (M = 71.4, SD = 7.4) and young adults were between the ages of 18 and 30 (M = 20.7, SD = 2.7). All participants spoke English fluently; had normal or corrected-to-normal vision; and were in general good health with no history of psychiatric or neurological conditions. Participants gave written informed consent after being told about the procedures of the experiment. Undergraduate students were compensated one extra credit that could be applied to an undergraduate psychology class and non-students were compensated $10 per hour of participation.

2.2. Materials

We adapted study materials from Shiffrin et al. (1995). Stimuli consisted of 25 categories that were composed of 10 exemplar words and one prototype word (see Appendices A, B, and C for the complete stimulus set). Fifteen of these categories contained words that were all semantically similar within the category and tended to be relatively long (5–14 letters) with relatively low natural language frequency. The other ten categories contained words that were either phonemically or orthographically related to the prototype word and were either three-letter or four-letter monosyllabic words with a high natural language frequency. In addition to the words within the semantic and phonemic/orthographic categories, there were 44 extra words that met the same criteria as the words in semantic categories (i.e., language frequency and word length), and 20 extra words that met the same criteria as the words in the phonemic/orthographic categories; these words were not exemplars to any of the prototypes (i.e., they were unrelated extra words) and were used as buffer or distracter words (see Section 2.3. Procedures). No words were offensive, emotionally loaded, or otherwise provocative.

Clearly, the two category types (phonemic/orthographic and semantic) were not matched in word language frequencies nor in word length. Regarding word frequency, several prior studies have revealed no effect of word frequency on memory performance in DRM-like paradigms (Madigan and Neuse, 2004; Roediger et al., 2001; Sherman and Jordan, 2011), while others have found an effect on false recognition limited to the cases when conscious recollection is involved (Dewhurst et al., 1998; Kinoshita, 1995). Although the influence of word frequency on recognition memory thus remains unclear, this property does not represent a confound for our study because we manipulated category size. That is, our critical measure was one of memory *interference* – the d′2-d′8 difference score – rather than absolute levels of d′, and any effects of word frequency on absolute levels of recognition memory can be presumed removed because word frequency is the same across the 2-item and 8-item categories within each category type. Regarding word length, it has been found to affect false recognition only when the length of critical lures differs from the length of within-category words (Madigan and Neuse, 2004), which was not the case in this paradigm.

2.3. Procedures

The experiment consisted of two phases: a study phase with 120 words and a test phase with 144 words (see Fig. 2). Both phases were completed on a desktop computer with a separate monitor or on a laptop computer. During the study phase, words were presented for three seconds each, with the list of 120 words comprising eight or two exemplars from each of the ten semantic and ten phonemic/orthographic categories (100 in total), plus ten extra semantic buffer words at the beginning of the study list and ten more at the end, to prevent recency and primacy effects for the items of interest (i.e., the 100 items drawn from the thematic categories). These buffer items were not presented during the test phase. The selection of ten semantic categories out of a total possible 15 semantic categories, for use in the study phase, was counterbalanced across participants. Additionally, we randomized for each participant: the assignment of which categories contributed eight or two exemplars (with the constraint that five semantic and five phonemic/orthographic contributed eight exemplars, with the remaining categories contributing two); the specific exemplar words presented; and the order of presentation of all words during the study phase (with the constraint that the semantic buffer words remained in the first ten and last ten list positions). Consequently, semantic and phonemic/orthographic stimuli were intermixed. All words were presented in a white font on a black background.

Once the study phase was completed, the test phase began without any delay. The test list contained, in a randomized order for each participant, twenty extra semantic distracter words, twenty extra phonemic/orthographic distracter words and, from each of the 20 study categories, two studied exemplar words (targets; counterbalanced across subjects), two non-studied exemplar words (related lures; counterbalanced across subjects), and one prototype word (critical lure). Again, semantic and phonemic/orthographic stimuli were intermixed. The four remaining extra semantic buffer words were used in four practice trials that were presented at the start of the test phase, before any trials containing study category words. Data from these trials were discarded. The other extra words inserted into the test phase (twenty semantic and twenty phonemic/orthographic) were included as distractors (unrelated lures) to measure false memory for words that did not belong to any category and had no interference. Test trials were self-paced, with the participant's response cueing the presentation of the next word.

2.3.1. Task

At the beginning of the experiment, participants were informed that some words presented during the study phase would be repeated during the test phase and that there would be a final recognition memory test. Additionally, participants were informed that many of the words they were to see would seem similar or related. During the study phase participants were asked to rate the pleasantness of the presented word by pressing keys numbered 1–5. On this 5-point scale, 1 was considered very pleasant, 5 was very unpleasant, and 3 was neutral. In the test phase, participants were prompted to give a rating of their confidence that the current word was seen before. This confidence was measured on a 6-point scale where pressing 1 meant that he/she was very sure the item was a new word and pressing 6 meant that he/she was very sure it was an old word.

2.3.2. Neuropsychological tests

After completion of the experimental study, all older adult participants completed an additional one-hour neuropsychological battery, to assess cognitive abilities and confirm healthy cognitive status. These tests assessed memory, thinking, language, and visual perception and included Wechsler Memory Scale-IV Logical Memory I & II (Wechsler, 2009); Trails Making Test Parts A and B (Lezak et al., 2004); Wechsler Adult Intelligence Scale-IV Digit Span (Wechsler, 2008); Mini-Mental State Examination (Folstein et al., 1975); and Visual Object Space Perception Silhouettes (Warrington and James, 1991). Young adult participants did not complete the neuropsychological tests.

2.4. Statistical analysis

2.4.1. Signal detection model

SDT was used to analyze the data. SDT models assume that a participant's response is based on a combination of the degree of discriminability – in this context, the ability to detect whether a word was previously studied – and a criterion (k) value, which governs the participant's decision rule (Macmillan and Creelman, 2005). In this model, on any given trial a participant must make an old-new recognition decision based on a sampled ‘memory strength’ value for the current word. Memory strength values are assumed to vary from trial to trial according to a normal distribution (see Fig. 3). We assumed differing distributions of memory strength for each type of word (e.g., target, lure, etc.). The likelihood of a particular response (e.g., a miss or a hit) is determined from the area under the curve to the left or right of the k (criterion) value.

Typically, targets (i.e., previously seen words) have a greater mean memory strength than lures and so the memory strength distribution for targets is shifted to the right; however, there exist some items with ambiguous memory strength values, where the target and lure distributions overlap. The discriminability measure d′ is the difference between the mean memory strengths for the target and lure distributions, taking into account the variance of those distributions. A greater d′ reflects less overlap between the two distributions. In psychological experiments, it is unlikely that the two distributions do not overlap at all (corresponding to a participant with perfect recognition performance).

In order to make a recognition decision, participants must employ a criterion, k. Any memory strength value above k elicits an “old” response and any value below k elicits a “new” response. This leads to

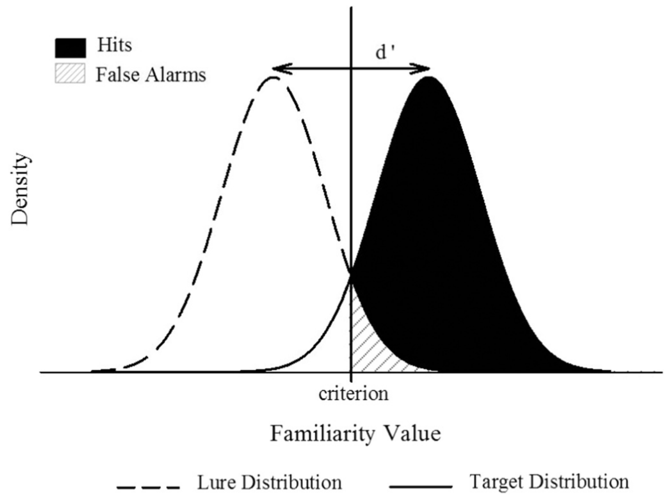

Fig. 3. Illustration of SDT Measurements. The discriminability measure d′ is calculated by converting the hit and FA rate to z scores and finding the difference, d′ = z[Hit rate] – z[FA rate]. Hits are determined from the area under the target distribution to the right of the criterion (i.e., when the word is old and the participant's response is old). FAs are determined from the area under the lure distribution to the right of the criterion (i.e., when the word is new and the participant's response is old).

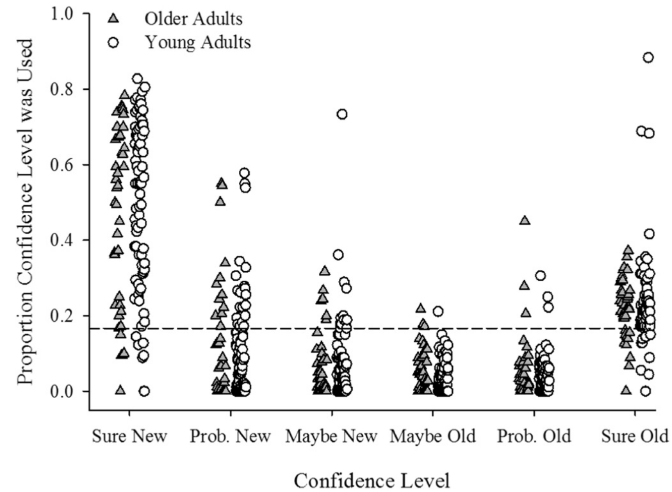

Fig. 4. Proportion of Times a Confidence Level was used by Older and Young Subjects. Each point shows the proportion of responses for one subject averaged across conditions. The dotted line shows proportions corresponding to equal use of each level.

four possible response types: 1) hits when the word is old and the response is old; 2) FAs when the word is new and the response is old; 3) misses when the word is old and the response is new; and 4) correct rejections when the word is new and the response is new. To calculate d′, it is sufficient to know the frequencies of hits and FAs, along with total number of target and lure trials. The proportion of misses and correct rejections is simply one minus the proportion of hits and FAs, respectively. The predicted rates of hits and FAs can be calculated by the proportion of area under the target distribution curve to the right of k and the area under a given lure distribution curve to the left of k, respectively. Therefore, d′ and k provide a complete description of a participant's recognition performance, including their hits and FAs. Assuming k is fixed to zero, a d′ value of zero would produce 50% hits and FAs (i.e., chance performance). As d′ increases, the percentage of hits increases, and the percentage of FAs decreases.

We applied a signal detection model to derive estimates of the discriminability between targets and lures for each condition, defined by lure type (critical, related), interference type (phonemic/orthographic, semantic), and category size (large, small). We did not examine the discriminability of targets from unrelated lures (which differed neither by interference type nor category size), since this was of no a priori theoretical interest. Discriminability, measured as d′, was derived using the algebraic formula, d′ = z[Hit rate] – z[FA rate], separately for each participant in each condition. The six confidence ratings given by participants were collapsed into a dichotomous response (old/new, in which responses 1–3 were mapped to 'new' and responses 4–6 were mapped to 'old') because an insufficient number of participants used the full scale when responding (see Fig. 4). Consequently, the model was constrained to assume variance equal to one for both lures and targets, corresponding to an equal-variance signal–detection (EVSD) model. Therefore, a total of eight values were calculated for each participant: two d′ values, namely d′ between targets and critical lures and d′ between targets and related lures, for each of the four possible combinations of category size (2-item or 8-item) and interference type (phonemic/orthographic or semantic).

Ratcliff et al. (1992) noted that recognition memory experiments analyzed using SDT typically produce results in which the variance of the old-item (target) distribution is greater than that of the new-item (lure) distribution. This can be demonstrated by plotting Receiver Operating Characteristic (ROC) curves for z[Hit rate] against z[FA rate], since it can be shown mathematically that the slope of a z-ROC curve is equal to the ratio of the lure: target distribution variances. The value found for this ratio across empirical studies is typically ∼ 0.8 but can be as low as 0.7 (Glanzer et al., 1999; Ratcliff et al., 1992; Rotello et al., 2004). As noted above, the distribution of responses provided by our participants was dichotomous, which precluded plotting ROC curves (since we did not have data at a range of bias levels) and necessitated instead an assumption of equal variance. Assuming an EVSD model when the ground truth is unequal variance (as in the unequal-variance signal-detection (UVSD) model) can produce misleading results under some circumstances. Specifically, if the use of an EVSD model to analyze data from two experimental conditions indicates a difference in both d′ and response criterion across those conditions, but in reality the underlying memory strength distributions have unequal variance, then ostensible differences in d′ may in fact reflect changes only in response bias (Dougal and Rotello, 2007). Because we were unable to plot z-ROC curves to derive an estimate of the target variance in our data, we performed a simplified UVSD analysis to check for the possibility that our EVSD analysis produced misleading d′ values. In this analysis, we calculated d′ scores from the hit and FA rates directly by simply assuming a lure distribution variance, σ_{lure}, of 1 and a target distribution variance, σ_{targ}, of 1.25 or 1.4 (corresponding to lure:target ratios of 0.8 and 0.71). That is,

$$d'_{UVSD} = \frac{z(Hit\ rate) - z(FA\ rate)}{\sqrt{\frac{(\sigma_{targ}^2 - \sigma_{lure}^2)}{2}}}$$

in which the cumulative normal distribution used to calculate z[Hit rate] had a variance of σ_{targ}, whereas the cumulative normal distribution for z[FA rate] had a variance of σ_{lure}. If any differences in d′ across conditions produced under an assumption of equal variance were spurious effects of that assumption, then this simple UVSD analysis should reduce or eliminate those differences.

Before deriving the values of the signal detection model we prescreened the data in three ways. First, we identified subjects whose average accuracy across conditions was significantly lower than chance (0.41; n = 4) and relabeled their old/new responses to the opposite mapping, assuming that these participants had accidentally switched the keys used for indicating 'old' and 'new' during the experiment. Second, any subject whose average accuracy fell within a 95% confidence interval around chance performance (0.41–0.59) was removed from the analysis (n = 7). Third, we applied the Hautus (1995) correction for extreme proportions and their biasing effect on d′, by adding

0.5 to the response count in every cell (i.e., the total response count per condition per participant) and 1 to the total number of trials in each condition; this method has been found to produce less biased estimates of d′ than the 1/(2 N) method that corrects only extreme proportion values of 0 or 1 (Snodgrass and Corwin, 1988).

As discussed above, the category size condition was critical in providing an index of the extent to which any observed false memory for related lures is induced by mnemonic interference (i.e., by studying a category of related items) rather than by inherent properties of the words that are used to form related categories in the DRM paradigm. For example, there may be systematic differences between semantic and phonemic/orthographic words in terms of backward association strength, base frequency, or pre-experimental familiarity, making one class of lure more prone to be falsely remembered, or one class of target more difficult to remember after study. However, these inherent word properties should be the same regardless of category size (in this study, 2-item versus 8-item), provided that the assignment of items to category sizes is counterbalanced across participants. Therefore, in order to control for the effects of inherent word properties we calculated the difference between the d′ scores for 8-item and 2-item categories, for each participant within each interference type (phonemic/orthographic, semantic). This d′ difference score (d′2 – d′8) not only addresses potential confounds caused by inherent word-category properties, but can also be considered a measure of study-related interference (i.e., the effect of seeing more related lures) on recognition memory performance.

In drawing conclusions from the data we focused primarily on the discriminability of targets from related lures, rather than targets from critical lures. This is because related lures have a key advantage over critical lures: related lures can be counterbalanced across participants, removing any effect of inherent word properties. In DRM word categories, the word that serves as the critical lure is fixed and (similar to potential differences between perceptual versus semantic category words) may have certain properties of frequency, pre-experimental familiarity, and backward association strength that differ from all other words in the category, making it more prone to be falsely remembered. Moreover, these distinctive properties of the critical lure may be more pronounced for one condition (semantic versus perceptual) than the other, complicating the interpretation of any condition effects. In contrast, related lures are drawn randomly from the unstudied members of a category, which allows counterbalancing across participants, thereby removing any potential item confounds caused by properties specific to critical lures.

2.4.2. Statistical analyses of recognition memory and interference measures

All ANOVA's were calculated using the ‘ez’ R package (Version 4.2; Lawrence, 2015) and all *t*-tests were Welch's *t*-tests calculated using the base ‘stats’ R package (Version 3.2.4; R Core Team, 2016). First, d′ scores were analyzed in a four-way (2 × 2 × 2 × 2) ANOVA, with a between-subjects factor of Age (older and young adults); and within-subjects factors of (1) Interference Type (phonemic/orthographic and semantic); (2) Item Type (critical lures and related lures); and (3) Category Size (2-item and 8-item categories).

Next, we assessed the d′ difference score (d′2 – d′8), to measure study-related interference. We analyzed d′ difference scores using a three-way (2 × 2 × 2) ANOVA, with factors Age Group, Interference Type and Item Type, as above. According to the R-H account, we expected older adults’ memory performance to be affected less by study-related interference for semantic words, but suffer from more interference for phonemic/orthographic words, relative to young adults. This prediction should manifest as an interaction between Age and Interference Type; specifically, older adults should have smaller d′2 – d′8 than young adults in the semantic condition, but a larger d′2 – d′8 in the phonemic/orthographic condition. The cleanest test of this prediction was provided by related lures, which, unlike critical lures, were counterbalanced to eliminate item confounds.

Table 1
Older adult average raw scores for neuropsychological battery.

Test (Maximum score)	M (SD)
MMSE (/30)	29.0 (1.1)
WMS-IV LM Immediate Recall (/50)	23.6 (5.4)
WMS-IV LM Delayed Recall (/50, 20-min delay)	19.6 (6.3)
WMS-IV LM Recognition (/30)	22.2 (3.9)
Trails A	24.0 s (6.3 s)
Trails B	64.2 s (37.2 s)
WAIS-IV Digit Span Forward (/9)	6.7 (1.1)
WAIS-IV Digit Span Backward (/8)	4.8 (1.1)
VOSP Silhouettes (/30)	19.4 (4.9)

Note: The mean (M) performance on all cognitive tasks was within the normal range relative to established norms or within established passing cutoff scores. MMSE = Mini-Mental State Examination; WMS-IV LM = Wechsler Memory Scale, 4th ed., Logical Memory subtest; WAIS-IV = Wechsler Adult Intelligence Scale, 4th ed.; VOSP = Visual Object Space Perception battery.

3. Results

3.1. Neuropsychological test performance

Results of the neuropsychological battery are shown in Table 1. Older adults demonstrated intact group performance on all cognitive tasks, with average performance within the normal range relative to established norms or within established passing cutoff scores (Crum et al., 1993; Folstein et al., 1975; Tombaugh, 2004; Warrington and James, 1991; Wechsler, 2008, 2009). Individually, all older participants included in the analysis passed the study's inclusion criterion of a score greater than 25 on the Mini-Mental State Examination.

3.2. The effects of semantic and perceptual interference on recognition memory

Hits and FAs for all participants in all conditions are shown in Table 2, separated according to Item Type (Unrelated Lures were not assigned to any category, and thus had no perceptual/semantic or 2-item/8-item status). Because these raw proportion scores are influenced by both memory and response bias and we are interested only in memory, we focused our statistical analyses on the d′ measures derived from them, shown in Figs. 5, 6.

We began by analyzing d′ of targets from critical and related lures. A four-way ANOVA with d′ as the dependent variable revealed main effects of Age Group (older, young), F (1, 110) = 4.19, p = .043, Interference Type (phonemic, semantic), F (1, 110 = 11.44, p = .001, Item Type (critical, related), F (1, 110) = 161.2, p < .001, and Category Size (2-item, 8-item), F (1, 110) = 27.51, p < .001 (see Fig. 5). Additionally, there were significant interactions between Age Group and Interference Type, F (1, 110) = 4.52, p = .036; Interference Type and Item Type, F (1, 110) = 5.91, p = .012; and Item Type and Category Size, F (1, 110) = 19.8, p < .001. Because related lures provided a cleaner test of our predictions than critical lures, and because we revealed a main effect of Item Type, we followed this analysis with two additional 2 × 2 × 2 ANOVA's examining critical lures and related lures separately.

For related lures, we found main effects of Interference Type, F (1, 110) = 18.42, p < .001, with semantic d′ scores greater than phonemic/orthographic d′ scores, and Category Size, F (1, 110) = 5.66, p = .019, with 2-item d′ scores greater than 8-item d′ scores, as expected given the greater interference introduced by 8-item categories. There was no main effect of Age (F (1, 110) = 2.16, p = .144). There was a borderline interaction between Age Group and Interference Type, F (1, 110) = 3.48, p = .065, in which young adults showed a greater d′ advantage for semantic over phonemic/orthographic categories than did older adults. Finally, there was a three-way interaction between Age Group, Interference Type, and Category Size, F (1, 110) = 4.76,

Table 2
True and false recognition responses by age group, item type, interference type and category size.

	True recognition				False recognition								
	Targets				Critical lures				Related lures				Unrelated lures
	Semantic		Perceptual		Semantic		Perceptual		Semantic		Perceptual		
Category Size	2	8	2	8	2	8	2	8	2	8	2	8	
Younger adults													
Mean	0.89	0.89	0.85	0.90	0.17	0.37	0.21	0.40	0.12	0.20	0.20	0.28	0.09
SEM	0.014	0.018	0.017	0.017	0.023	0.031	0.026	0.027	0.013	0.017	0.021	0.023	0.009
Older adults													
Mean	0.76	0.81	0.76	0.76	0.15	0.39	0.17	0.32	0.11	0.18	0.12	0.19	0.06
SEM	0.034	0.033	0.039	0.041	0.027	0.043	0.029	0.043	0.020	0.032	0.025	0.030	0.011

$p = .031$. For older adults, d′ differed numerically less between 2-item and 8-item categories for semantically related categories than for phonemically/orthographically related categories. In contrast, for young adults, d′ differed numerically more between 2-item and 8-item categories for semantically related categories than for phonemically/orthographically related categories. Thus, d′ for older adults tended to be more influenced by phonemic/orthographic interference than semantic, whereas for young adults the reverse was true (Fig. 5). This three-way interaction in the d′ scores is equivalent to a 2-way interaction between Age and Interference Type in the d′ difference scores (i.e., d′2 – d′8), as predicted by the R-H account. Analyses of the d′ difference (i.e., ‘interference’) scores are reported below.

For critical lures, there were main effects of Age Group, $F(1, 110) = 6.39$, $p = .013$, and Category Size, $F(1, 110) = 40.5$, $p < .001$. Younger adults' d′ scores were greater than older adults' d′ scores, and d′ for 2-item categories was greater than d′ for 8-item categories. There was also a trend towards an interaction between Age Group and Interference Type, $F(1, 110) = 3.08$, $p = .082$, in which – similar to the pattern seen in related lures – young adults showed a greater d′ advantage for semantic over phonemic/orthographic categories than did older adults. In the critical lures d′ scores, the three-way interaction between Age Group, Interference Type and Category Size was not significant.

Next, we analyzed the d′2 – d′8 interference score in a 3-way ANOVA with factors Item Type, Interference Type and Age Group. This was the most important analysis because, as explained above, this dependent variable minimizes the effect of item confounds by exploiting the category size manipulation. We found a main effect of Item Type (critical lures, related lures), $F(1, 110) = 19.78$, $p < .001$ (Fig. 6). To recap, d′2 – d′8 serves as a measure of study-related interference (i.e., the effect of seeing more related items) on recognition memory performance. Because the scores from critical and related lures differed significantly, and because we had a priori reasons to place greater confidence in the data from related lures, we analyzed them separately.

For d′2 – d′8 scores from related lures, we found the predicted 2-way interaction between Age Group and Interference Type, $F(1, 110) = 4.76$, $p = .031$. Numerically, older adults had greater d′2 – d′8 scores than young adults for phonemically/orthographically related categories, whereas young adults had greater d′2 – d′8 scores than older adults for semantically related categories. Within this interaction, there was no simple main effect of Age Group for either Interference Type. However, for semantically related categories, older adults' d′2 – d′8 scores did not differ significantly from zero, Welch's $t(35) = 0.03$, $p = .976$, but young adults' scores did, $t(75) = 2.91$, $p = .005$. That is, older adults were not impaired by an increase in semantic interference, but young adults were. Although these two results cannot be taken as evidence that young and older adults differed from each other within semantic categories, they are nonetheless instructive about the nature of the interaction between Age Group and Interference Type. In contrast, for phonemically/orthographically related categories, neither d′2 – d′8 scores from older adults, $t(35) = 1.55$, $p = .13$, nor young adults, $t(75) = 0.20$, $p = .84$, differed significantly from zero. Again, although the simple main effect of Age Group was not significant, it is informative to note that the numerical pattern seen in phonemically/orthographically related categories was in the opposite direction from that seen in semantic categories, which presumably contributed to the overall interaction that was observed. For completeness, we also examined the complementary simple main effect – Interference Type within each Age Group separately – revealing no difference in older adults, and a borderline significant difference in young adults ($p = .057$) in the direction of greater d′2-d′8 scores for semantic than phonemic/orthographic categories. In sum, the presence of a 2-way interaction between Age and Interference Type for the d′ difference scores provides evidence that – as predicted by the R-H account –

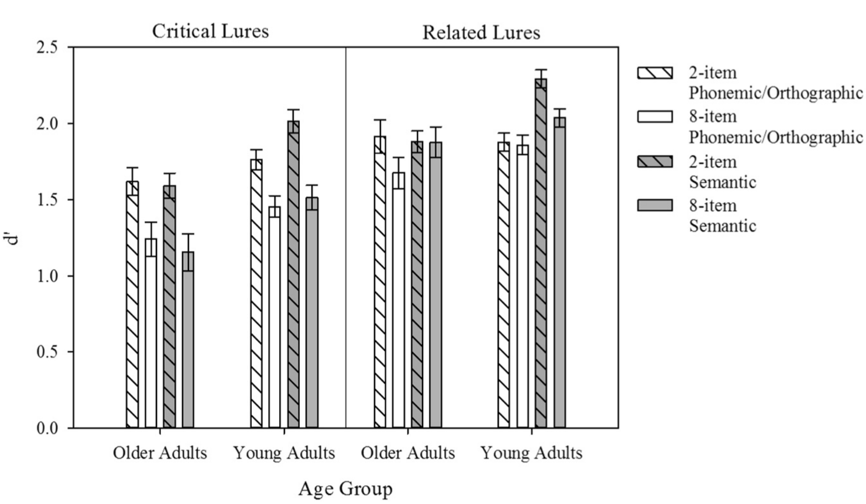

Fig. 5. Analysis of d′ between Target and Critical Lure or Target and Related Lure Distributions across Age Group, Category Size, and Interference Type. A main effect of Item Type ($p < .001$) supported separate analyses for critical and related lures. For related lures, there was a three-way interaction between Age Group, Category Size and Interference Type ($p = .031$), in which older adults' d′ scores differed numerically more between 2-item and 8-item categories for phonemically/orthographically related categories than semantic categories, whereas the opposite pattern held for young adults. This interaction was not significant for critical lures. Error bars are within-subject standard error of the mean.

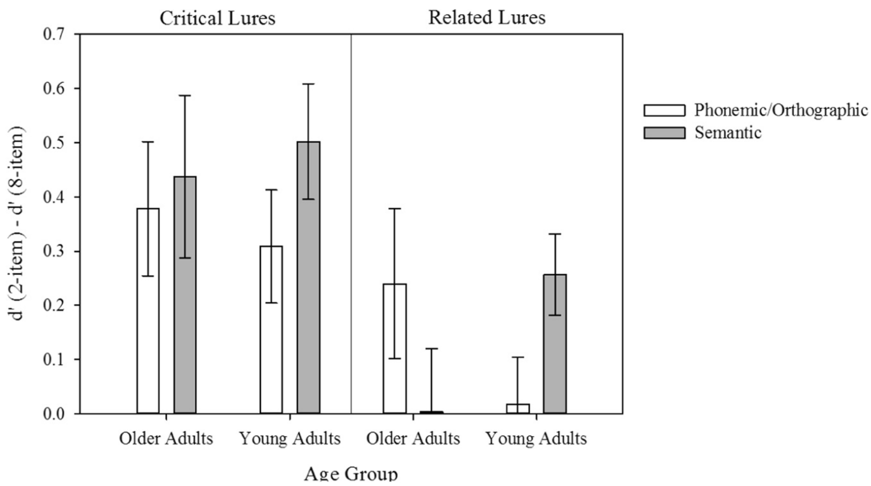

Fig. 6. Analysis of Study-Related Interference on d′ between Target and Critical Lure or Target and Related Lure Distributions across Age Group and Interference Type. Study-related interference was measured by subtracting the d′ score from 8-item categories from the d′ score from 2-item categories (d′ [2-item] – d′ [8-item]) within a given condition (Item Type, Age Group, Interference Type). For related lures, there was an interaction between Age Group and Interference Type (p = .031), such that older adults' d′ scores were impaired less by semantic and more by phonemic/orthographic interference compared to young adults. There were no significant effects for critical lures, though a similar numerical pattern can be seen. Error bars are within-subject standard error of the mean.

recognition memory in older adults, relative to young adults, is less impaired by semantic interference and more impaired by phonemic/orthographic interference.

For critical lures, there were no significant effects in the d′2 – d′8 scores, but general patterns similar to those observed in related lures were seen. That is, older adults produced numerically greater d′ difference scores than young adults for phonemic/orthographic critical lures, while young adults gave numerically greater d′ difference scores than older adults for semantic critical lures.

Finally, to assess the effect of assuming equal variance of the lure and target distributions (since this assumption may produce misleading results, if it is inaccurate), we recalculated the signal detection analysis comparing Targets and Related Lures for all conditions, assuming greater variance in the target distribution than the lure distribution. Specifically, in two separate analyses we assumed a lure distribution variance, σ_{lure}, of 1, and a target distribution variance, σ_{targ}, of either 1.25 or 1.4 (we chose 1.25 to reflect the mean ratio typically found in the empirical literature, and 1.4 to reflect the extreme end of the range of ratios found empirically, following Glanzer et al. (1999), Ratcliff et al. (1992), and Rotello et al. (2004)). When assuming σ_{targ} = 1.25, a three-way ANOVA with d′ as the dependent variable again revealed a three-way interaction between Age Group, Interference Type, and Category Size, F(1, 110) = 1.54, p = .025. As stated above, this three-way interaction in the d′ scores is equivalent to a two-way interaction between Age Group and Interference Type in the d′ difference scores (i.e., d′2 – d′8; F (1, 110) = 3.07, p = .025). Similarly, when assuming σ_{targ} = 1.4, a three-way ANOVA with d′ as the dependent variable again revealed a three-way interaction between Age Group, Interference Type and Category Size, F (1, 110) = 1.61, p = .023, which was also reflected in the two-way ANOVA with d′2 – d′8 as the dependent variable as an interaction between Age Group and Interference Type, F (1, 110) = 3.21, p = .023. Additionally, as in the EVSD analysis reported above, for semantically related categories, older adults' d′2 – d′8 scores did not differ significantly from zero for either the σ_{targ} = 1.25 UVSD analysis (Welch's t (35) = 0.32, p = .753) or the σ_{targ} = 1.4 UVSD analysis (t (35) = 0.49, p = .628), but young adults' d′2 – d′8 scores did differ from zero in the semantic condition, for both the σ_{targ} = 1.25 UVSD analysis (t (75) = 2.43, p = .017) and the σ_{targ} = 1.4 UVSD analysis (t (75) = 2.19, p = .032). Finally, as in the EVSD analysis, both UVSD analyses found that for phonemically/orthographically related categories the d′2 – d′8 scores from neither older nor young adults differed significantly from zero (p > .05). In sum, the pattern of results seen in the d′ scores under the EVSD model persisted under our simplified implementation of the UVSD model, and the critical 3-way interaction between Age Group, Interference Type and Category Size in the d′ dependent measure remained significant (even becoming more so). Thus, it seems unlikely that a potentially inaccurate assumption of equal variance in our original analysis produced misleading results.

4. Discussion

The key result was a differential effect of perceptual and semantic interference on recognition memory performance in older compared with younger adults. Specifically, recognition memory in older adults was impaired relatively more by perceptual interference and less by semantic interference than was the case in young adults. This finding was predicted by the R-H account of cognition, applied here for the first time to memory performance in aging.

Several theories have previously been put forward to explain age-related memory loss. One theory holds that aging causes deficits primarily in associative memory, or the ability to bind information together into a cohesive, complex unit (Chalfonte and Johnson, 1996; Naveh-Benjamin, 2000; Naveh-Benjamin et al., 2003). Other theorists have suggested that while older adults can encode and retrieve memory for the gist of an experience, mnemonic deficits emerge because verbatim traces for individual items are compromised (Fuzzy Trace Theory; Brainerd et al., 2009). Another recent theory hypothesizes a central role for pattern separation in memory, and proposes that this function depends critically on the hippocampus, which is compromised by aging (Holden et al., 2012; Schacter et al., 1997; Stark et al., 2010; Yassa et al., 2011). These three theories make different predictions for memory function in old age, including whether age-related impairments occur to different extents for specific classes of mnemonic content such as perceptual and semantic. The Associative-Deficit Hypothesis (Naveh-Benjamin, 2000) predicts that memory should be impaired by aging whenever the stimulus material is associative (Naveh-Benjamin et al., 2003); a corollary of this prediction is that memory should not be impaired by perceptual interference in which the interfering material is merely similar to, rather than associatively related to, studied material. In contrast, Fuzzy Trace Theory predicts that old age should impair memory for any fine-grained details – whether perceptual or semantic – that cannot be incorporated into the gist of an experience (Brainerd et al., 2009). Finally, hippocampal pattern separation theories make a third prediction: that older adults should be impaired on any mnemonic discrimination involving a high degree of overlap between stimuli, including perceptual overlap at the level of single items (Bakker et al., 2008; Yassa et al., 2011).

The R-H account shares several properties with these theories, for example, that deterioration of MTL structures is critical to age-related cognitive deficits, that compromised associative representations play an important role, and that older adults should exhibit reduced ability to resolve overlap between stimulus representations (Barense et al., 2012a; Cowell, 2012). However, the predictions of the R-H account for the effects of semantic and perceptual interference in aging differ from all three sets of predictions just outlined. In the case of perceptual interference, the R-H account predicts that, if the stimuli contain sufficient feature-overlap to necessitate conjunctive representations for their unique identification, old age should increase susceptibility to such interference. This prediction is similar to those of the hippocampal pattern separation account and Fuzzy Trace Theory, but differs from the Associative Deficit Hypothesis, which predicts no age-related susceptibility to perceptual interference unless the perceptual material is associatively related to the test discriminanda (in the present study, it was not). In the case of semantic interference, the R-H account makes a unique prediction: that older adults should be paradoxically less affected by such interference than young adults, because they are less able to apprehend and encode the interfering associative relations between studied items. In this case, all three alternative theories appear to predict age-related impairments, because semantic interference (1) is associative (Associative Deficit Hypothesis), (2) necessitates memory for verbatim details in order to discriminate accurately between semantically-related targets and lures (Fuzzy Trace Theory), and (3) causes greater overlap between item representations (hippocampal pattern separation account). Thus, the present finding of an interaction between age and interference type, with older adults relatively shielded from semantic interference and impaired by perceptual interference compared to young adults, seems to be compatible with the predictions of the R-H account alone.

Our interpretation of the key finding is that older adults' pattern of susceptibility to interference is caused by compromised conjunctive representations in the MTL or nearby temporal lobe regions. Specifically, we propose that older adults possess deficits in conjunctive representations in both a phonemic/orthographic hierarchy of word representations, and in a separate semantic hierarchy. In the phonemic/orthographic hierarchy, the conjunctions that bind together the phonemic and/or orthographic features that uniquely identify a word are compromised, reducing the ability to correctly identify a target word as studied by discriminating it from unstudied words that share perceptual features. In the semantic hierarchy, the representations of inter-item semantic associations are compromised, reducing the extent to which unstudied but semantically related words are activated via associations with studied words, thereby reducing the source of interference that in young individuals impairs the discriminability of targets from related lures at test. These compromised conjunctive representations thus have two opposing effects: they tend to increase older adults' susceptibility to interference from phonemically/orthographically-related items, but paradoxically shield older adults from interference caused by semantically-related items. This interpretation is paradoxical in that it explains a tendency for older adults to be *facilitated* relative to young adults in terms of a cognitive *deficit*: compromised associative-semantic representations protect older adults' memory discrimination performance in the face of increasing semantic interference.

A skeptic might reasonably argue, instead, that the present data are better explained by an age-related improvement in associative-semantic representations. However, this interpretation seems less plausible than our preferred interpretation, for two reasons. First, a number of studies have reported a decline in semantic memory function with normal aging (e.g., Bäckman and Nilsson, 1996; Berkowitz, 1953; Kaufman, 2001; Lindenberger and Baltes, 1994; Rönnlund et al., 2005). Indeed, even in theoretical accounts of memory in aging that posit the over-reliance of older adults on semantic gist, it is not assumed that older individuals encode semantic information more deeply than young adults (e.g., Koutstaal et al., 2003). Second, the data from our own study point to the same conclusion: comparing semantic discrimination performance in older and younger adults (by taking d′ of related lures from targets and collapsing across category size) reveals that older adults performed significantly worse, t (65.37) = -2.17, p = .034. In other words, we found no evidence of superior ability to discriminate semantically related items in older adults, as would be expected if semantic representations become more precise and detailed with age. Importantly, it is also not the case that semantic discrimination in older adults was at floor – their d′ scores from semantic categories, collapsed across category size, differed significantly from zero, t (35) = 16.98, $p < .001$ – ruling out the possibility that the shielding from semantic interference seen in this group occurred simply because performance had no room to go lower with additional interference.

It is also possible that our results reflect age-related differences in the strategies that participants used to memorize list items. For example, perhaps older adults in our task devoted more attention to the perceptual details of words and less to their semantic content, relative to young adults, producing the pattern wherein older adults suffer more from perceptual and less from semantic interference (as measured by the d′2-d′8 interference score). If young participants were instructed to encode items in this way (i.e., to focus more on perceptual than semantic properties), memory performance across the groups might be rendered more similar, and such a result would imply a role for strategic factors in the effect we observed. However, such a strategy-based explanation of our results would invoke the *opposite* strategy than has been attributed to older adults in the prior literature. Previous accounts of age-related strategy differences have typically claimed that older adults rely on a gist-based strategy – where "gist" refers to the general meaning of a stimulus without item-specific perceptual details – along with less thorough encoding of item-specific (verbatim) details (Dennis et al., 2007, 2008; Dennis and Turney, 2018). In other words, the strategy used by older adults according to these accounts is the preferential encoding of semantic information over perceptual information. Using such a strategy in our task, older adults should suffer more from semantic than perceptual interference – the opposite pattern to the one we found (Figs. 5, 6). Thus, although a strategy-based explanation cannot be ruled out, it conflicts with prior claims about age-related strategy use in the literature. Future studies in which participants' strategy use is explicitly manipulated via instruction might help to clarify the issue. Of course, our results are not only at odds with existing *strategy-based* accounts of age-related memory deficits; they also contradict similar accounts claiming that older adults are not simply less inclined to, but are less able to, encode perceptual details than semantic gist (Brainerd et al., 2009). Some (but not all) prior studies of recognition memory in aging have supported such accounts. Next, we discuss several potential reasons for the divergence between our results and these prior findings.

A number of prior studies have examined the existence and separability of age-related deficits in perceptual and semantic memory. The results of these studies conflict with each other: one reported greater age-related susceptibility to semantic interference (Koutstaal et al., 2003); another revealed age-related deficits in perceptual rather than semantic mnemonic discrimination (Ly et al., 2013); and two further studies found that older adults are impaired by both perceptual and semantic interference (Budson et al., 2003; Pidgeon and Morcom, 2014). Moreover, the results of the present study align with none of these prior reports. Even the findings from Ly et al. (2013), which appear to yield conclusions similar to our own, in fact differed critically from our results, as explained below. However, the heterogeneity of prior results is perhaps unsurprising in light of the considerable range of task parameters employed by these studies (e.g., category size, stimulus format, between- versus within-subject designs, and means of manipulating semantic and perceptual content), as well as the variety of dependent variables (e.g., d′, FA rate, or baseline-corrected FA rate) that the authors used. Below, we explain two ways in which the task design and data analyses of the present study were optimized to test our a priori predictions and to rule out important confounds that may have applied to prior studies.

First, many DRM studies examining memory in older adults have focused principally on FA rates, which – unlike d′ or d_a measures derived using SDT – do not account for potential differences in response bias between groups (Budson et al., 2003; Koutstaal et al., 2003; Pidgeon and Morcom, 2014). In some analyses of the data, critical lure

FA rates were corrected for baseline levels of false recognition by subtracting the FA rate for unrelated lures (Budson et al., 2003; Koutstaal et al., 2003; Pidgeon and Morcom, 2014). Although there may be some theoretical account of guessing (i.e., a psychological model that offers an explanation of how unrelated lures come to be endorsed as targets) under which subtraction of FA rate for unrelated lures is an appropriate correction, none was provided for these studies. In fact, this correction appears to rest upon an assumption that false memory and guessing are mutually exclusive; this seems implausible, given that one property of such a model is that as guess rate increases false memory must decrease. Thus, it is unclear how to appropriately interpret critical lure FA rates that have been corrected in this way, and it cannot be assumed that such a correction controls for response bias. In contrast, SDT is based upon a well-defined psychological model and provides a measure of mnemonic discriminability, d′, that is uncontaminated by response bias. This increases confidence that the measure of performance provided by d′ is a reliable index of mnemonic processing alone.

Second, some prior DRM studies have employed only one category size (Budson et al., 2003; Ly et al., 2013), or have employed more than one category size but have not placed great import on finding an interaction between category size and other key factors such as interference type and age group (Koutstaal et al., 2003; Pidgeon and Morcom, 2014; but see Koutstaal and Schacter, 1997). A manipulation of category size provides a manipulation of the amount of mnemonic interference from the study phase. As noted above, this manipulation is crucial in ruling out item confounds that may exist when stimuli cannot be counterbalanced between conditions (e.g., perceptual versus semantic categories). Thus, to demonstrate a differential effect of semantic versus perceptual interference unconfounded by inherent item properties, it is necessary to find an interaction of category size and interference type. To demonstrate a differential effect that differs by age, it is necessary to find a three-way interaction of category size, interference type and age group, as reported in this study for the first time.

The manipulation of category size is one way in which our study differed from that of Ly et al. (2013), and consideration of this factor reveals an important difference between the two sets of findings. Because Ly et al. tested only one category size, the most analogous comparison to their data that we can make is an assessment of d′ for related lures collapsed across category size, which in fact reveals the opposite pattern of results from Ly et al.: a marginal interaction between Age Group and Interference Type, F (1, 110) = 3.48, p = .065, in which older adults were impaired by semantic (Welch's t (65.37) = -2.17, p = .034) but not phonemic/orthographic (t (66.28) = -0.52, p = .608) overlap between targets and lures (see Fig. 5). To make the closest possible comparison to Ly et al., who used very small categories of 1 item, we also examined d′ for 2-item categories alone, but again found the opposite pattern from Ly et al. (2013): an interaction between Interference Type and Age Group (F (1, 110) = 9.565, p = .003) in which phonemic/orthographic d′ scores did not differ by age (p = .825) but semantic d′ scores did (p = .005). Thus, unlike in Ly et al., older adults in our study were not better at semantic discrimination and worse at perceptual discrimination *per se*, indeed, we found the opposite pattern. Two further differences between the two studies may have contributed to the discrepant results. First, Ly et al. used very small (1-item) categories and presented items such that interference occurred only at test: each related lure in the test list was related to one item from study. Second, the dependent variable for which Ly et al. reported the key interaction of age (young/older) and task (perceptual/semantic) was d′ between targets and unrelated foils, ignoring hits and FAs for related lures. In combination with 1-item categories, this targets-versus-foils d′ essentially provides a measure of recognition memory in the presence of perceptual or semantic similarity *at the time of test*. In contrast, participants in our study saw multiple items (2 or 8) from each category at study, and d′ was calculated between targets and related lures, providing a measure of recognition memory that is influenced by study-related interference. Thus, the two experimental designs may have engendered quite different representational and cognitive demands (as discussed by Ly et al., 2013). We note that Ly et al. reported another dependent variable that did measure the discriminability of targets from related lures, but the age*task interaction for this measure was not significant, making it difficult to conclude in favor of a differential effect of perceptual versus semantic study-related interference in older compared to young adults.

Finally, we note that the effects of perceptual or semantic interference in isolation are subtle in our data. Specifically, older adults were not significantly impaired by perceptual interference relative to young adults, nor did the effects of perceptual and semantic interference differ in the older adult group. Given that these were healthy older adults with no known neurological disorder, we expected only small effect sizes. Indeed, this pattern is in line with a prior study of recognition memory comparing young adults, healthy older adults and individuals at risk for developing Mild Cognitive Impairment (MCI; Yeung et al., 2013) in which perceptual interference had a robust effect on the at-risk group and an intermediate effect on the healthy older group: performance in healthy older adults was impaired relative to young adults on some but not all comparisons, and was often preserved relative to at-risk individuals. Critically, in the present study, the effect that reached significance was the predicted interaction between age and interference type for the d′2 – d′8 interference score. In addition, in young adults, there was a borderline significant effect of Interference type (perceptual, semantic) for the d′2 – d′8 score, which became significant under the assumption of $\sigma_{targ} > 1$. Since there was no such effect in older adults (and, numerically, the difference was in the opposite direction), these two findings regarding the simple main effects corroborate our interpretation of the overall interaction – that the effects of perceptual and semantic interference differed for older versus young participants.

In sum, in this study we set forth the predictions of the R-H account of cognition for how interference should affect memory in healthy aging, and tested those predictions with a recognition memory experiment employing the DRM paradigm. We found that older adults, relative to young adults, were paradoxically shielded from semantic interference but impaired by perceptual interference. This pattern of results appears at odds with most current theories of memory in aging, but is in line with the predictions of the R-H account for mnemonic interference, which explains the performance of older participants in terms of compromised conjunctive representations in anteromedial temporal lobe regions.

Acknowledgements

This work was supported by the National Science Foundation, United States, Award BCS-1554871. The sponsors had no involvement in the execution or reporting of this work. We thank David Huber and Jeff Starns for helpful discussions.

Appendix A

See Table A1

Table A1
Semantic word categories.

Astronaut	**Butterfly**	**Castle**	**Comedian**	**Diamond**
Atmosphere	Camouflage	Chateau	Buffon	Brilliance
Cosmonaut	Caterpillar	Courtyard	Clown	Carat
Gravity	Cocoon	Dungeon	Comic	Emerald
Orbiting	Dragonfly	Feudal	Humorist	Glittering
Rocket	Flutter	Fortress	Improvisation	Hardness
Satellite	Fragile	Mansion	Joker	Precious
Shuttle	Metamorphosis	Medieval	Lampoon	Priceless
Spaceman	Monarch	Stronghold	Monologue	Rhinestones
Voyager	Slight	Throne	Punster	Rubies
Weightlessness	Wings	Towers	Slapstick	Sparkle
Dinosaur	**Fitness**	**Gambler**	**Infant**	**Lunatic**
Amphibians	Aerobics	Bettor	Babble	Asylum
Artifacts	Barbells	Blackjack	Cradle	Demented
Brontosaurus	Biceps	Bookie	Diapers	Deranged
Extinction	Exertion	Casino	Highchair	Hallucinations
Fossils	Jogging	Jackpot	Lullaby	Insanity
Glaciers	Nutrition	Lottery	Pacifier	Madman
Mammoth	Physique	Poker	Rattle	Manic
Reptiles	Sweating	Roulette	Stork	Psychopath
Skeletons	Toning	Stakes	Stroller	Psychotic
Swamps	Workout	Wager	Teething	Ranting
Magician	**Phantom**	**Pyramid**	**Robbery**	**Tornado**
Conjure	Apparition	Catacombs	Assailant	Cyclone
Enchanted	Beckon	Egyptian	Bandit	Funnel
Hypnotist	Ghost	Embalming	Booty	Gusts
Juggling	Ghoul	Hieroglyphics	Burglary	Sirens
Rabbit	Gloomy	Mummies	Holdup	Spiral
Sorcerer	Goblin	Pharaoh	Mugging	Twister
Spells	Haunting	Tombs	Stealing	Typhoon
Trickster	Paranormal	Triangular	Stickup	Whirling
Vanish	Specter	Underworld	Theft	Whirlwind
Wizard	Spooky	Vault	Wallet	Windstorm

Note: Prototypes appear in bold type.

Appendix B

See Table B1

Table B1
Phonemic/orthographic word categories.

Boon	**Bun**	**Cat**	**Cop**	**Fate**
Boom	Bud	Bat	Bop	Date
Boos	Bum	Cab	Cob	Face
Boot	Bus	Cam	Cod	Fade
Coon	But	Cap	Cog	Fake
Goon	Fun	Fat	Con	Fame
Loon	Gun	Hat	Cot	Gate
Moon	Nun	Mat	Hop	Hate
Noon	Pun	Pat	Mop	Late
Soon	Run	Rat	Pop	Mate
Toon	Sun	Sat	Top	Rate
Mire	**Role**	**Sip**	**Teal**	**West**
Dire	Dole	Dip	Deal	Best
Fire	Hole	Hip	Heal	Lest
Hire	Mole	Lip	Meal	Nest
Mice	Pole	Nip	Peal	Pest
Mike	Robe	Rip	Real	Rest
Mile	Rode	Sin	Seal	Test
Mime	Rope	Sis	Team	Vest
Mine	Rose	Sit	Teas	Welt
Tire	Rote	Six	Teat	Went
Wire	Sole	Tip	Veal	Wept

Note: Prototypes appear in bold type.

Appendix C

See Table C1

Table C1
Extra words.

Semantic		Phonemic/Orthographic
Antiquity	Mooring	Bird
Apartment	Opossum	Book
Attic	Pauper	Bout
Bagel	Piccolo	Bur
Biologist	Podium	Coin
Bison	Promenade	Cow
Canvas	Purple	Foil
Carport	Sable	Fur
Convenience	Schoolyard	Joy
Dachshund	Scissors	Lawn
Gazette	Silhouette	Loud
Gutter	Stationer	Now
Honeycomb	Storeroom	Null
Housecoat	Synopsis	Perk
Industry	Thesaurus	Pull
Inferno	Thicket	Raw
Invitation	Tortilla	Saw
Jargon	Tribesman	Toy
Jasmine	Triplicate	Wood
Linguistics	Undergrowth	Yaw
Mechanical	Unformed	
Monoxide	Warmhearted	

References

Abrams, L., Trunk, D.L., Merrill, L.A., 2007. Why a superman cannot help a tsunami: activation of grammatical class influences resolution of young and older adults' tip-of-the-tongue states. Psychol. Aging 22 (4), 835–845. https://doi.org/10.1037/0882-7974.22.4.835.

Bäckman, L., Nilsson, L.-G., 1996. Semantic memory functioning across the adult life span. Eur. Psychol. 1 (1), 27–33. https://doi.org/10.1027/1016-9040.1.1.27.

Bakker, A., Kirwan, C.B., Miller, M., Stark, C.E.L., 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science (March), 1640–1643.

Barense, M.D., Bussey, T.J., Lee, A.C.H., Rogers, T.T., Davies, R.R., Saksida, L.M., Graham, K.S., 2005. Functional specialization in the human medial temporal lobe. J. Neurosci. 25 (44), 10239–10246. https://doi.org/10.1523/JNEUROSCI.2704-05.2005.

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.

Barense, M.D., Groen, I.I.A., Lee, A.C.H., Yeung, L.K., Brady, S.M., Gregori, M., Henson, R.N.A., 2012a. Intact memory for irrelevant information impairs perception in amnesia. Neuron 75 (1), 157–167. https://doi.org/10.1016/j.neuron.2012.05.014.

Barense, M.D., Ngo, J.K.W., Hung, L.H.T., Peterson, M.A., 2012b. Interactions of memory and perception in amnesia: the figure-ground perspective. Cereb. Cortex 22 (11), 2680–2691. https://doi.org/10.1093/cercor/bhr347.

Barense, M.D., Rogers, T.T., Bussey, T.J., Saksida, L.M., Graham, K.S., 2010. Influence of conceptual knowledge on visual object discrimination: insights from semantic dementia and MTL amnesia. Cereb. Cortex 20 (11), 2568–2582. https://doi.org/10.1093/cercor/bhq004.

Bartko, S.J., Cowell, R.A., Winters, B.D., Bussey, T.J., Saksida, L.M., 2010. Heightened susceptibility to interference in an animal model of amnesia: impairment in encoding, storage, retrieval – or all three? Neuropsychologia 48, 2987–2997. https://doi.org/10.1016/j.neuropsychologia.2010.06.007.

Bartko, S.J., Winters, B.D., Cowell, R.A., Saksida, L.M., Bussey, T.J., 2007a. Perceptual functions of perirhinal cortex in rats: zero-delay object recognition and simultaneous oddity discriminations. J. Neurosci. 27 (10), 2548–2559. https://doi.org/10.1523/JNEUROSCI.5171-06.2007.

Bartko, S.J., Winters, B.D., Cowell, R.A., Saksida, L.M., Bussey, T.J., 2007b. Perirhinal cortex resolves feature ambiguity in configural object recognition and perceptual oddity tasks. Learn. Mem. 14, 821–832. https://doi.org/10.1101/lm.749207.

Berkowitz, B., 1953. The Wechsler-Bellevue performance of white males past age 50. J. Gerontol. 8, 76–80.

Brainerd, C.J., Reyna, V.F., Howe, M.L., 2009. Trichotomous processes in early memory development, aging, and neurocognitive impairment: a unified theory. Psychol. Rev. 116 (4), 783–832. https://doi.org/10.1037/a0016963.

Bruffaerts, R., Dupont, P., Peeters, R., Deyne, S. De, Storms, G., Vandenberghe, R., 2013. Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. J. Neurosci. 33 (47), 18597–18607. https://doi.org/10.1523/JNEUROSCI.1548-13.2013.

Budson, A.E., Sullivan, A.L., Daffner, K.R., Schacter, D.L., 2003. Semantic versus phonological false recognition in aging and Alzheimer's disease. Brain Cogn. 51 (3), 251–261. https://doi.org/10.1016/S0278-2626(03)00030-7.

Bussey, T.J., Saksida, L.M., 2002. The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex. Eur. J. Neurosci. 15, 355–364. https://doi.org/10.1046/j.0953-816x.2001.01850.x.

Bussey, T.J., Saksida, L.M., Murray, E.A., 2002. Perirhinal cortex resolves feature ambiguity in complex visual discriminations. Eur. J. Neurosci. 15, 365–374. https://doi.org/10.1046/j.0953-816x.2001.01851.x.

Cabeza, R., Nyberg, L., Park, D. (Eds.), 2016. Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging. Oxford University Press. https://doi.org/10.1176/appi.ajp.163.3.560.

Campbell, K.L., Hasher, L., Thomas, R.C., 2010. Hyper-binding. Psychol. Sci. 21 (3), 399–405. https://doi.org/10.1177/0956797609359910.

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. Proc. Natl. Acad. Sci. USA 201610686. https://doi.org/10.1073/pnas.1610686113.

Chalfonte, B.L., Johnson, M.K., 1996. Feature memory and binding in young and older adults. Mem. Cogn. 24 (4), 403–416. https://doi.org/10.3758/BF03200930.

Clarke, A., Tyler, L.K., 2014. Object-specific semantic coding in human perirhinal cortex. J. Neurosci. 34 (14), 4766–4775. https://doi.org/10.1523/JNEUROSCI.2828-13.2014.

Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. NeuroImage 22 (1), 466–476. https://doi.org/10.1016/j.neuroimage.2003.12.049.

Cohen, N.J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., Nash, C., 1999. Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. Hippocampus 9, 83–98. https://doi.org/10.1002/(SICI)1098-1063(1999)9:1 < 83::AID-HIPO9 > 3.0.CO;2-7.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J., 2001. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol. Rev. 108 (1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204.

Cowell, R.A., 2012. Computational models of perirhinal cortex function. Hippocampus 22 (10), 1952–1964. https://doi.org/10.1002/hipo.22064.

Cowell, R.A., Bussey, T.J., Saksida, L.M., 2006. Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. J. Neurosci. 26 (47), 12186–12197. https://doi.org/10.1523/JNEUROSCI.2818-06.2006.

Crum, R.M., Anthony, J.C., Bassett, S.S., Folstein, M.F., 1993. Population-based norms for the Mini-Mental State Examination by age and educational level. JAMA 269 (18), 2386–2391.

Deese, J., 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. J. Exp. Psychol. 58 (1), 17–22. https://doi.org/10.1037/h0046671.

Dell, G.S., 1986. A spreading-activation theory of retrieval in sentence production. Psychol. Rev. 93 (3), 283–321.

Dennis, N.A., Kim, H., Cabeza, R., 2007. Effects of aging on true and false memory formation: an fMRI study. Neuropsychologia 45, 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. J. Cogn. Neurosci. 20 (8), 1390–1402. https://doi.org/10.1162/jocn.2008.20096.Age-related.
Dennis, N.A., Turney, I.C., 2018. The influence of perceptual similarity and individual differences on false memories in aging. Neurobiol. 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.
Dewhurst, S.A., Hitch, G., Barry, C., 1998. Separate effects of word frequency and age of acquisition in recogntion and recall. J. Exp. Psychol. 24 (2), 284–298.
Dougal, S., Rotello, C.M., 2007. Remembering" emotional words is based on response bias, not recollection. Psychon. Bull. Rev. 14 (3), 423–429. https://doi.org/10.3758/BF03194083.
Duarte, A., Henson, R.N., Graham, K.S., 2008. The effects of aging on the neural correlates of subjective and objective recollection. Cereb. Cortex 18 (9), 2169–2180. https://doi.org/10.1093/cercor/bhm243.
Eichenbaum, H., Otto, T., Cohen, N.J., 1994. Two functional components of the hippocampal memory system. Behav. Brain Sci. 17 (3), 449–472. https://doi.org/10.1017/S0140525X00035391.
Fjell, A.M., Walhovd, K.B., 2010. Structural brain changes in aging: courses, causes and cognitive consequences. Rev. Neurosci. 21 (3), 187–221. https://doi.org/10.1515/REVNEURO.2010.21.3.187.
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. J. Psychiatr. Res. 12 (3), 189–198.
Gallo, D.A., 2006. Associative Illusions of Memory: False Memory Research in DRM and Related Tasks. Psychology Press, New York, NY.
Giovanello, K.S., Schnyer, D.M., Verfaellie, M., 2004. A critical role of the anterior hippocampus in relational memory: evidence from an fMRI study comparing associative and item recognition. Hippocampus 14, 5–8. https://doi.org/10.1002/hipo.10182.
Glanzer, M., Kim, K., Hilford, A., Adams, J.K., 1999. Slope of the receiver-operating characteristic in recognition memory. J. Exp. Psychol.: Learn. Mem. Cogn. 25 (2), 500–513. https://doi.org/10.1037/0278-7393.25.2.500.
Glezer, L.S., Kim, J., Rule, J., Jiang, X., Riesenhuber, M., 2015. Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. J. Neurosci. 35 (12), 4965–4972. https://doi.org/10.1523/JNEUROSCI.4031-14.2015.
Gordon, J.K., Kurczek, J.C., 2013. The ageing neighbourhood: phonological density in naming. Lang. Cogn. Process. 1–19. https://doi.org/10.1080/01690965.2013.837495.
Hautus, M.J., 1995. Corrections for extreme proportions and their biasing effects on estimated values of d′. Behav. Res. Methods Instrum. Comput. 27 (1), 46–51. https://doi.org/10.3758/BF03203619.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402.
Holden, H.M., Hoebel, C., Loftis, K., Gilbert, P.E., 2012. Spatial pattern separation in cognitively normal young and older adults. Hippocampus 1832, 1826–1832. https://doi.org/10.1002/hipo.22017.
Kaufman, A.S., 2001. WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. Intelligence 29, 131–167. https://doi.org/10.1016/S0160-2896(00)00046-5.
Kinoshita, S., 1995. The word-frequency effect in recognition memory versus repetition priming. Mem. Cogn. 23 (5), 569–580. https://doi.org/10.3758/bf03197259.
Koutstaal, W., Reddy, C., Jackson, E.M., Prince, S., Cendan, D.L., Schacter, D.L., 2003. False recognition of abstract versus common objects in older and younger adults: testing the semantic categorization account. J. Exp. Psychol. Learn. Mem. Cogn. 29 (4), 499–510. https://doi.org/10.1037/0278-7393.29.4.499.
Koutstaal, W., Schacter, D.L., 1997. Gist-based false recognition of pictures in older and younger adults. J. Mem. Lang. 37, 555–583. https://doi.org/10.1006/jmla.1997.2529.
Lawrence, M.A., 2015. ez: Easy Analysis and Visualization of Factorial Experiments. Retrieved from 〈https://cran.r-project.org/package=ez〉.
Lee, A.C.H., 2006. Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond long-term declarative memory: a double dissociation in dementia. J. Neurosci. 26 (19), 5198–5203. https://doi.org/10.1523/JNEUROSCI.3157-05.2006.
Lee, A.C.H., Bandelow, S., Schwarzbauer, C., Henson, R.N.A., Graham, K.S., 2006. Perirhinal cortex activity during visual object discrimination: an event-related fMRI study. NeuroImage 33, 362–373. https://doi.org/10.1016/j.neuroimage.2006.06.021.
Lee, A.C.H., Buckley, M.J., Pegman, S.J., Spiers, H., Scahill, V.L., Gaffan, D., Graham, K.S., 2005a. Specialization in the medial temporal lobe for processing of objects and scenes. Hippocampus 15, 782–797. https://doi.org/10.1002/hipo.20101.
Lee, A.C.H., Bussey, T.J., Murray, E.A., Saksida, L.M., Epstein, R.A., Kapur, N., Graham, K.S., 2005b. Perceptual deficits in amnesia: challenging the medial temporal lobe "mnemonic" view. Neuropsychologia 43 (1), 1–11. https://doi.org/10.1016/j.neuropsychologia.2004.07.017.
Lee, A.C.H., Levi, N., Davies, R.R., Hodges, J.R., Graham, K.S., 2007. Differing profiles of face and scene discrimination deficits in semantic dementia and Alzheimer's disease. Neuropsychologia 45 (9), 2135–2146. https://doi.org/10.1016/j.neuropsychologia.2007.01.010.
Lee, A.C.H., Scahill, V.L., Graham, K.S., 2008. Activating the medial temporal lobe during oddity judgment for faces and scenes. Cereb. Cortex 18, 683–696. https://doi.org/10.1093/cercor/bhm104.
Lezak, M.D., Howieson, D.B., Loring, D.W., 2004. Neuropsychological Assessment, 4th ed. Oxford University Press, New York, NY.
Lindenberger, U., Baltes, P.B., 1994. Sensory functioning and intelligence in old age: a strong connection. Psychol. Aging 9 (3), 339–355. https://doi.org/10.1037/0882-7974.9.3.339.
Liuzzi, A.G., Bruffaerts, R., Dupont, P., Adamczuk, K., Peeters, R., De Deyne, S., Vandenberghe, R., 2015. Left perirhinal cortex codes for similarity in meaning between written words: comparison with auditory word input. Neuropsychologia 76, 4–16. https://doi.org/10.1016/j.neuropsychologia.2015.03.016.
Loftus, G.R., 1978. On interpretation of interactions. Mem. Cogn. 6 (3), 312–319.
Ly, M., Murray, E., Yassa, M.A., 2013. Perceptual versus conceptual interference and pattern separation of verbal stimuli in young and older adults. Hippocampus 23 (6), 425–430. https://doi.org/10.1002/hipo.22110.
Macmillan, N.A., Creelman, C.D., 2005. Detection Theory: A User's Guide, 2nd ed. Lawrence Erlbaum Associates, Mahwah, NJ.
Madigan, S., Neuse, J., 2004. False recognition and word length: a reanalysis of Roediger, Watson, McDermott, and Gallo (2001) and some new data. Psychon. Bull. Rev. 11 (3), 567–573. https://doi.org/10.3758/BF03196612.
Manns, J.R., Hopkins, R.O., Squire, L.R., 2003. Semantic memory and the human hippocampus. Neuron 38 (1), 127–133. https://doi.org/10.1016/S0896-6273(03)00146-6.
McTighe, S.M., Cowell, R.A., Winters, B.D., Bussey, T.J., Saksida, L.M., 2010. Paradoxical false memory for objects after brain damage. Science 330, 1408–1410. https://doi.org/10.1126/science.1194780.
Naveh-Benjamin, M., 2000. Adult age differences in memory performance: tests of an associative deficit hypothesis. J. Exp. Psychol.: Learn. Mem. Cogn. 26 (5), 1170–1187. https://doi.org/10.1037//0278-7393.26.5.1170.
Naveh-Benjamin, M., Hussain, Z., Guez, J., Bar-On, M., 2003. Adult age differences in episodic memory: further support for an associative-deficit hypothesis. J. Exp. Psychol.: Learn. Mem. Cogn. 29 (5), 826–837. https://doi.org/10.1037/0278-7393.29.5.826.
Newman, M.C., Kaszniak, A.W., 2000. Spatial memory and aging: performance on a human analog of the Morris water maze. Aging Neuropsychol. Cogn. 7 (2), 86–93. https://doi.org/10.1076/1382-5585(200006)7:2;1-U;FT086.
Pidgeon, L.M., Morcom, A.M., 2014. Age-related increases in false recognition: the role of perceptual and conceptual similarity. Front. Aging Neurosci. 6 (OCT), 1–17. https://doi.org/10.3389/fnagi.2014.00283.
Ratcliff, R., Sheu, C.-F., Gronlund, S.D., 1992. Testing global memory models using ROC curves. Psychol. Rev. 99 (3), 518–535.
Raz, N., Gunning-Dixon, F.M., Head, D., Dupuis, J.H., Acker, J.D., 1998. Neuroanatomical correlates of cognitive aging: evidence from structural magnetic resonance imaging. Neuropsychology 12 (1), 95–114. https://doi.org/10.1037//0894-4105.12.1.95.
Raz, N., Rodrigue, K.M., Head, D., Kennedy, K.M., Acker, J.D., 2004. Differential aging of the medial temporal lobe: a study of a five-year change. Neurology 62, 433–438. https://doi.org/10.1212/01.WNL.0000106466.09835.46.
Roediger III, H.L., McDermott, K.B., 1995. Creating false memories: remembering words not presented in lists. J. Exp. Psychol.: Learn. Mem. Cogn. 21 (4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803.
Roediger III, H.L., Watson, J.M., McDermott, K.B., Gallo, D.A., 2001. Factors that determine false recall: a multiple regression analysis. Psychon. Bull. Rev. 8 (3), 385–407. https://doi.org/10.3758/BF03196177.
Rönnlund, M., Nyberg, L., Bäckman, L., Nilsson, L.-G., 2005. Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. Psychol. Aging 20 (1), 3–18. https://doi.org/10.1037/0882-7974.20.1.3.
Rotello, C.M., Macmillan, N.A., Reeder, J.A., 2004. Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model. Psychol. Rev. 111 (3), 588–616. https://doi.org/10.1037/0033-295X.111.3.588.
Sadil, P.S., Cowell, R.A., 2017. A computational model of perceptual and mnemonic deficits in medial temporal lobe amnesia. J. Cogn. Neurosci. 29 (6), 1075–1088. https://doi.org/10.1162/jocn.
Schacter, D.L., Kaszniak, A.W., Kihlstrom, J.F., Valdiserri, M., 1991. The relation between source memory and aging. Psychol. Aging 6 (4), 559–568. https://doi.org/10.1037/0882-7974.6.4.559.
Schacter, D.L., Koutstaal, W., Norman, K.A., 1997. False memories and aging. Trends Cogn. Sci. 1 (6), 229–236.
Schacter, D.L., Osowiecki, D., Kaszniak, A.W., Kihlstrom, J.F., Valdiserri, M., 1994. Source memory: extending the boundaries of age-related deficits. Psychol. Aging 9 (1), 81–89. https://doi.org/10.1037/0882-7974.9.1.81.
Schaie, K.W., 2005. What can we learn from longitudinal studies of adult development? Res. Hum. Dev. 2 (3), 133–158. https://doi.org/10.1207/s15427617rhd0203_4.
Scheerer, N., Marrone, D.F., 2014. Age-related deficits in conjunctive representation of complex objects. Curr. Aging. Sci. 214–219. https://doi.org/10.2174/1874609808666150201215549.
Scoville, W., Milner, B., 1957. Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatry 20, 11–21.
Sherman, S.M., Jordan, T.R., 2011. Word-frequency effects in long-term semantic priming and false memory. Br. J. Psychol. 102 (3), 559–568. https://doi.org/10.1111/j.2044-8295.2011.02017.x.
Shiffrin, R., Huber, D.E., Marinelli, K., 1995. Effects of categoy length and strength on familiarity recognition. J. Exp. Psychol.: Learn. Mem. Cogn. 21 (2), 267–287.
Silver, H., Goodman, C., Bilker, W.B., 2012. Impairment in associative memory in healthy aging is distinct from that in other types of episodic memory. Psychiatry Res. 197 (1–2), 135–139. https://doi.org/10.1016/j.psychres.2012.01.025.
Singer, M., Gagnon, N., Richards, E., 2002. Strategies of text retrieval: a criterion shift account. Can. J. Exp. Psychol. 56 (1), 41–57. https://doi.org/10.1037/h0087384.

Singer, M., Wixted, J.T., 2006. Effect of delay on recognition decisions: Evidence for a criterion shift. Mem. Cogn. 34 (1), 125–137. https://doi.org/10.3758/BF03193392.

Snodgrass, J.G., Corwin, J., 1988. Pragmatics of measuring recognition memory: applications to dementia and amnesia. J. Exp. Psychol. Gen. 117 (1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34.

Spencer, W.D., Raz, N., 1995. Differential effects of aging on memory for content and context: a meta-analysis. Psychol. Aging 10 (4), 527–539. https://doi.org/10.1037/0882-7974.10.4.527.

Squire, L.R., Wixted, J.T., 2011. The cognitive neuroscience of human memory since H.M. Annu. Rev. Neurosci. 34 (1), 259–288. https://doi.org/10.1146/annurev-neuro-061010-113720.

Squire, L.R., Zola-morgan, S., 1991. The medial temporal lobe memory system human memory. Science 253 (5026), 1380–1386. https://doi.org/10.1146/annurev.neuro.27.070203.144130.

Stark, S.M., Yassa, M.A., Stark, C.E.L., 2010. Individual differences in spatial pattern separation performance associated with healthy aging in humans. Learn. Mem. 17 (6), 284–288. https://doi.org/10.1101/lm.1768110.

Stretch, V., Wixted, J.T., 1998. On the difference between strength-based and frequency-based mirror effects in recognition memory. J. Exp. Psychol.: Learn. Mem. Cogn. 24 (6), 1379–1396. https://doi.org/10.1037/0278-7393.24.6.1379.

Team, R.C., 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from 〈https://www.r-project.org/〉.

Tombaugh, T.N., 2004. Trail making test A and B: normative data stratified by age and education. Arch. Clin. Neuropsychol. 19, 203–214. https://doi.org/10.1016/S0887-6177(03)00039-8.

Wagenmakers, E.-J., Krypotos, A.-M., Criss, A.H., Iverson, G., 2012. On the interpretation of removable interactions: a survey of the field 33 years after Loftus. Mem. Cogn. 40 (2), 145–160. https://doi.org/10.3758/s13421-011-0158-0.

Wang, D., Chalk, J.B., Rose, S.E., de Zubicaray, G., Cowin, G., Galloway, G.J., Semple, J., 2002. MR image-based measurement of rates of change in volumes of brain structures. Part II: application to a study of Alzheimer's disease and normal aging. Magn. Reson. Imaging 20, 41–48. https://doi.org/10.1016/S0730-725X(02)00472-1.

Warrington, E.K., James, M., 1991. The Visual Object and Space Perception battery. Thames Valley Company, Bury St. Edmunds, UK.

Warrington, E.K., Shallice, T., 1984. Category specific semantic impairments. Brain 107 (3), 829–853. https://doi.org/10.1093/brain/107.3.829.

Wechsler, D., 2008. Wechsler Adult Intelligence Scale, Fourth edition. Pearson, San Antonio, TX.

Wechsler, D., 2009. Wechsler Memory Scale, Fourth edition. Pearson, San Antonio, TX.

Wixted, J.T., Stretch, V., 2000. The case against a criterion shift account of false memory. Psychol. Rev. 107 (2), 368–376.

Yassa, M.A., Lacy, J.W., Stark, S.M., Albert, M.S., Gallagher, M., Stark, C.E.L., 2011. Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. Hippocampus 21, 968–979. https://doi.org/10.1002/hipo.20808.

Yassa, M.A., Stark, C.E.L., 2011. Pattern separation in the hippocampus. Trends Neurosci. 34 (10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006.

Yeung, L.-K., Ryan, J.D., Cowell, R.A., Barense, M.D., 2013. Recognition memory impairments caused by false recognition of novel objects. J. Exp. Psychol.: Gen. 142 (4), 1384–1397. https://doi.org/10.1037/a0034021.