



HHS Public Access

Author manuscript

Behav Neurosci. Author manuscript; available in PMC 2015 June 02.

Published in final edited form as:

Behav Neurosci. 2015 June ; 129(3): 257–268. doi:10.1037/bne0000055.

STABILITY OF AGE-RELATED DEFICITS IN THE MNEMONIC SIMILARITY TASK ACROSS TASK VARIATIONS

Shauna M. Stark¹, Rebecca Stevenson¹, Claudia Wu¹, Samantha Rutledge¹, and Craig E. L. Stark^{1,2}

¹Department of Neurobiology and Behavior, University of California

Abstract

Several studies in our lab and others have demonstrated age-related declines in mnemonic discrimination during a recognition memory paradigm using repeated items, similar lures, and novel foils. In particular, older adults exhibit a shift in lure discriminability, identifying similar lures as old items at a greater rate than young adults. This shift likely reflects deficits in pattern separation processing as a result of underlying changes in the dentate gyrus of the hippocampus. Here, we explored whether alterations in the task design could rescue the age-related impairment or whether it was ubiquitous as one might expect if the neurobiological mechanisms were truly disturbed by typical aging. Despite overt instructions to study item details during encoding, we replicated the age-related deficit in mnemonic discrimination. We established reliable effects with short lists of stimuli and with repeated testing. Altering the task design from a study/test to a continuous recognition paradigm replicated the age-related shift in lure discrimination as well. Modifying the task to an old/new response (rather than old/similar/new) showed the same effect and a d' analysis showed that lure items were more akin to target items in older adults. Finally, we varied the test instructions in order to promote gist or veridical responses in the old/new task. Even these overt, veridical test instructions did not ameliorate older adults' lure discrimination problems. Together, these findings demonstrate the robust nature of this age-related deficit and support the hypothesis that typical aging results in neurobiological changes that underlie this impairment.

Keywords

Pattern separation; lure discrimination; aging; hippocampus

Introduction

Age-related memory impairments affect various types of memory, such as impairments in episodic and source memory - the knowledge of where or when information was encoded (Johnson et al., 1993; Schacter et al., 1997). Older adults are more prone to false recollections, miscombining features of different events that are confidently held as true (Koutstaal et al., 2001; Lyle et al., 2006). Likewise, there are reports that older adults rely on

²To whom correspondence should be addressed at: 320 Qureshey Research Laboratory, University of California, Irvine, Irvine CA, 92697-3800, Tel (949) 824-4230, Fax (949) 824-2447, cestark@uci.edu.

memory for the general features or gist of studied items, but lose the specific details of individual items (Kensinger & Schacter, 1999).

Previously, we developed a task that is sensitive to these types of age-related declines in mnemonic discrimination utilizing a recognition memory paradigm consisting of repeated items, similar lures, and novel foils (Kirwan & Stark, 2007; Stark, Yassa, Lacy, & Stark, 2013; Yassa et al., 2010a; Yassa, Mattfeld, Stark, & Stark, 2011). Using this Mnemonic Similarity Task¹, we reported a decline across the aging spectrum for identifying lures as “similar”, with no corresponding decrease in recognition for identifying repeat items as “old”. This finding has also been reported by other groups (Toner, Pirogovsky, Kirwan, & Gilbert, 2009) and is sensitive to individual differences in memory ability (Holden, Toner, Pirogovsky, Kirwan, & Gilbert, 2013; Kirwan et al., 2012). Further, when older adults are separated into memory impaired and unimpaired (though still within the normal range for their age) based on their delayed word recall performance, the impaired adults perform particularly poorly on lure discrimination, but normally on recognition of old items (Stark et al., 2013).

We designed this task to tax the process of pattern separation. Computational models have defined pattern separation as the orthogonalization of similar inputs into distinct, non-overlapping representations (McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; Treves & Rolls, 1994). These models suggest that it is a vital component of episodic memory and other complex, multi-dimensional forms of memory by allowing new memories to be stored without inducing large amounts of interference from related episodes. The models propose that, by virtue of its unique anatomical properties and functional organization, the dentate gyrus (DG) subfield of the hippocampus is responsible for reducing the similarity of inputs associated with pattern separation. Support for this idea consists of the sparse activity in the DG, with only a small number of cells active at one time (Chawla et al., 2005). In addition, these cells alter their firing in response to small changes in input that are insufficient to alter firing properties elsewhere in the hippocampus (Leutgeb, Leutgeb, Moser, & Moser, 2007; Neuneubel & Knierim, 2014). Likewise, functional neuroimaging in humans has shown the DG to be more responsive to small changes in input than other hippocampal subfields (Bakker, Kirwan, Miller, & Stark, 2008; Lacy, Yassa, Stark, Muftuler, & Stark, 2011), consistent with its putative role in pattern separation (see Yassa & Stark, 2011 for review).

The MST attempts to evaluate the efficacy of this pattern separation process by assessing responses to highly-similar lure items (Kirwan & Stark, 2007). By including highly-similar lure items (that have a range of “mnemonic similarity”, operationalized as distribution of false alarm rates – see Yassa et al., 2010a) we can assess the ability of the system to preserve unique, detailed memories that, in these computational models, would rely upon pattern separation. Previous research has identified age-related changes in lure discrimination performance for individuals aged 60 and older (Holden & Gilbert, 2012;

¹In Stark et al. (2013), we coined this task the “Behavioral Pattern Separation Task – Objects (BPS-O)”. Upon much discussion and reflection, we have reverted back to referring to this task design as the “Mnemonic Similarity Task (MST)”. While we still suggest that the discrimination of similar lures in this task likely reflects underlying pattern separation processes, the use of the name MST more accurately reflects the demands of the task.

Stark, Yassa, & Stark, 2010; Toner et al., 2009). Such results are consistent with deficits in pattern separation and have been correlated with age-related changes in the functional activity from the DG/CA3 subregion of the hippocampus (Yassa et al., 2010a; Yassa et al., 2011) and with perforant path integrity (the input to the DG from the entorhinal cortex) (Yassa et al., 2011).

While there is clear evidence to support the role of pattern separation in the dentate gyrus and while the MST, and its age-related lure discrimination deficits have been correlated with hippocampal signals and integrity, such links are, by necessity, indirect assessments of causality. Here, we addressed the possibility that other, unrelated cognitive deficits may be contributing to this effect. Perhaps older adults are more “efficient” learners, only encoding the gist and not the details that are subsequently tested. To address this concern, instead of a surprise recognition test, we instructed participants during the study task that they would later be tested on these images using an old/similar/new test. Despite these overt instructions to study details, we replicated the age-related deficit in mnemonic discrimination and no age-related change in recognition of repeated items. In a related vein, we evaluated the effect of study-test set size and repeated testing on lure discrimination performance in young adults and found constant performance in the face of both manipulations. These data are important for the use of this task in intervention paradigms, such as clinical trials, or repeat testing to assess change over time, common in aging research. Next, we examined if limits in the processing speed in older adults might bias them towards using the old response instead of the similar response via a self-paced paradigm that was also adapted to a continuous recognition design. Again we found an age-related impairment in lure discrimination, despite these design modifications. Finally, we evaluated the role of the decision-threshold and task instructions by modifying the task to an old/new response only in order to calculate a d-prime measure of recognition memory performance. Consistent with our predictions, older adults’ shift in d-prime scores indicated that lure items were more akin to target items in older adults. Manipulating the task instructions to bias responses towards gist (responding “old” to lures) or veridical (responding “new” to lures) representations attenuated the age-related bias in lure discrimination, but did not eliminate it. Together, these findings demonstrate the robust nature of this age-related deficit in mnemonic discrimination as measured by the mnemonic similarity task, despite variations in task design. They therefore support the hypothesis that the MST is indexing an age-related neurobiological change – most likely an alteration of hippocampal function and a concomitant alteration in pattern separation.

General Methods

Participants

Several different groups of cognitively intact healthy adults participated in these tasks. With the exception of Experiment 5 (which employed a within-subjects design between 2 different sets of task instructions) and Experiment 2 (which employed a mixed within- and between-subjects design for each version length), each experiment employed a between-subjects design. Young adults (18–26 years old) were recruited through the University of California at Irvine undergraduate pool and participated for course credit. Older adults (60–

89 years old) were recruited through advertisements and word-of-mouth and participated for \$15 per hour of testing. Older adults were screened for cognitive impairments and all scored within the normal range on the Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975) or the Telephone Interview for Cognitive Status (Brandt, Spencer, & Folstein, 1988). All participants signed consent forms and the study was conducted in compliance with the Institutional Review Board of the University of California at Irvine.

Task Design

The MST (Figure 1) typically consisted of a series of 192 color photographs of everyday objects on a white background, downloaded from the Internet (see Yassa et al., 2010a for more details). In the first phase, participants engaged in an indoor/outdoor judgment for each picture (based on their opinion with no right or wrong answer) via a button press (128 items total, 2s each, 0.5s ISI). Immediately following the encoding task, participants engaged in a recognition memory test in which they identified each item as “Old”, “Similar”, or “New” via button press (192 items total – 64 repeated items, 64 lure items, and 64 foil items; 2.5s each, 0.5s ISI). One-third of the images in the test phase were exact repetitions of images presented in the encoding phase (targets or repeats); one-third of the images were new images not previously seen (foils); and one-third of the images were similar to the those seen during the encoding phase, but not identical (lures). These trial types were randomly intermixed during the test. As in our prior work (e.g., Stark et al., 2013), the Lure Discrimination Index (LDI) was calculated as the difference between the rate of “Similar” responses given to the lure items minus “Similar” responses given to the foils. Recognition (REC) for repeat items was calculated as the difference between the rate of “Old” responses given to repeat items minus “Old” responses given to foils. These scores correct for any response bias on a per-subject basis.

Experiment 1: Overt Study/Test

Typically, we have administered this task with an incidental encoding session followed by a surprise recognition test. Here, we informed the participants prior to the encoding of the items that they would later be tested on those items in addition to similar items and asked to determine which items were old, similar, and new. We were interested in the possibility that if older adults were informed about the task design, they might engage attentional resources (Naveh-Benjamin, Guez, & Shulman, 2004) or intentional strategies (Jennings & Jacoby, 1993) that could overcome the previously observed over-generalization of the similar items. We reasoned that by informing participants that they would later be tested on similar items, they might approach the encoding task differently. In addition, we were interested in the utility of using this task as a repeated measure with the goal of detecting change over time, such as a dependent measure in clinical trials or repeated testing associated with neurodegenerative disease. Thus, we were interested in ascertaining the effect on performance based on prior knowledge regarding the task.

Methods

Participants included 28 Aging (mean age: 72.9, range: 60–86; 19F/9M) and 24 Young (mean age: 20.1, range: 18–26; 20F/4M) adults (2 young and 1 older additional participants

were removed for using the wrong response keys). Prior to the study phase they were informed that they would be studying a set of pictures for a later old/similar/new recognition memory test. They were given example stimuli to demonstrate the difference between a repeated item and a similar lure item. Participants were informed to guess if they were unsure of the correct answer.

We also compared LDI performance following the overt instructions to a set of participants from a separate group who received incidental instructions. 27 Aging (mean age: 72.4; 16F/11M) and 16 Young (mean age: 23.4; 10F/6M) participants were selected from our Stark et al. (2013) study to be age-matched to the overt sample. These participants performed the same indoor/outdoor study task followed by a surprise old/similar/new recognition memory test. The instructions were exactly the same between groups. Only the timing of when the old/similar/new test instructions were presented differed (before the study session or after it).

Results

On the trials used for traditional recognition memory (REC: actual repetitions and novel foils), younger and older adults scored similarly (see Supplemental Table 2 for response rates in each condition), whereas a marked difference was observed in lure discrimination (Figure 2A). These data were entered in to a Repeated-Measures 2×2 ANOVA, using age group (Young and Aging) and memory test type (LDI and REC). We found a main effect of age group ($F(1,50) = 11.5, p < .01$), a main effect of memory test type ($F(1,50) = 454.2, p < .001$), and an interaction ($F(1,50) = 9.3, p < .01$). Consistent with our previous findings, post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a decrease in the LDI for aging adults compared to young ($t(100) = 4.5, 95\% \text{ CI } [.09, .29]$), but no difference in REC between groups ($t(100) = 1.1, 95\% \text{ CI } [-.05, .14]$).

To determine whether the overt instructions improved LDI performance compared to the incidental instructions (Figure 2B), we entered the LDI data into a 2×2 ANOVA, using age group (Young and Aging) and instruction type (Overt and Incidental). We found a main effect of age group ($F(1,91) = 35.7, p < .0001$), but no effect of instruction type ($F(1,91) = .03, p = .40$) and no interaction ($F(1,91) = .46, p = .50$). Post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) indicated that there was no difference in LDI between the two sets of instructions for Young ($t(91) = .99, 95\% \text{ CI } [-.07, .18]$) or Aging ($t(91) = .13, 95\% \text{ CI } [-.10, .11]$).

Discussion

Simply making the study task overt and describing exactly what will happen at test did not alleviate the age-related impairment in lure discrimination. Older adults were not able to employ attentional resources or intentional strategies based on the overt knowledge of how they would be later tested that were strong enough to overcome their bias to endorsing lures as old items. Further, lure discrimination performance was no different following the overt instructions than the incidental instructions with a surprise recognition test.

These results indicated that prior knowledge of the recognition task did not have a restorative effect on mnemonic similarity performance. While direct comparisons with our

prior work are challenging (owing to slight differences in age, recruitment methods, etc.) the performance in this experiment is on par with (or slightly below) our prior work using incidental study instructions (e.g., Stark et al., 2013), suggesting that overt study is not having a significant, positive effect. This finding gives hope that the MST can be utilized repeatedly within a person to ascertain change over time as a surprise test is not required.

In a supplemental set of experiments, we tested the effect of repeated testing and whether performance is significantly impacted by set size. In many repeated-measures designs utilized in clinical trials or other interventions, numerous measures are collected at each time point, making both practice effects and the total time to complete a task are critical factors. Therefore, we explored several set sizes: 16, 20, 32, and 64 items per condition in a mixed between-subjects design in young adults (Supplement 1). We found that reducing the set size significantly had no effect on LDI performance, despite substantially reducing the time and number of intervening items between study and test. In addition, we did not observe significant practice effects in short-term repeated testing on the LDI. The second run of the task immediately followed the first run, likely maximizing any potential for practice effects and for interference between sets of objects. Thus, in more typical study designs evaluating change over time, the subsequent exposure to the task using different object sets would be further reduced due to a longer delay.

Experiment 2: Continuous, self-paced

Typically when exploring behavioral performance in this paradigm, participants engaged in separate study and test sessions. However, in studies exploring functional magnetic resonance imaging (fMRI) brain activity associated with this task (Bakker et al., 2008; Kirwan & Stark, 2007; Lacy et al., 2011; Yassa et al., 2010a), a continuous design has been used. While it appears from Experiment 1 that having overt knowledge of the test prior to the encoding task did not eliminate the difference in LDI for older adults, the continuous task design may encourage a change in encoding strategy that could influence this discrimination. In particular, in a continuous version the notion that their memory will be tested (and here with both similar and novel foils) is continuously present.

In Experiment 1, responses during the old/similar/new task were collected within the 2-second time window that the item was presented on the screen. If participants did not respond in that time, the trial was marked as a no-response trial. It is certainly possible that the timed nature of the test was rushing the responses, particularly in older adults and those with poorer overall memory performance. Although they still responded to more than 90% of the trials, the older adults did have more no-responses to lure trials than younger adults in Experiment 1 (2.12% Young and 6.96% Aging; $t(41) = 3.8, p < .01$). We reasoned that if older adults retained the capacity to perform well at lure discrimination, removing any limitations imposed by our somewhat rapid response pace might allow this to be revealed. However, to be able to compare performance between groups, the display presentation time (encoding time) for each item was held constant.

Methods

There were a total of 480 trials: 96 first presentations with 96 repeats, 96 first presentations with 96 lure items, and 96 novel foils. The number of intervening items between the first presentation and a subsequent target or lure varied from 0 to 180 items. On each trial, items were presented for 2-seconds, followed by a blank screen and participants were asked to determine if the item was old, similar, or new (see General Methods). Participants could make responses either when the item was on the screen or during the blank screen following it. Upon their response, the next item was presented. In this way, no trials were excluded due to responses made outside the 2-second time window and participants had an unlimited period of time for their response. Importantly, by restricting participants to the 2-second viewing window, we prevented any group differences in the amount of time allowed to encode the item. Data from 23 Young (mean age: 20.5, range: 18–28; 17F/6M) and 21 Aging (mean age: 73.8, range: 60–83; 16F/5M) participants (1 young and 2 older additional participants were removed for using the wrong response keys or not complying with task instructions) were used.

Results

The data were entered into a Repeated-Measures 2×2 ANOVA, using age group (Young and Aging) and memory score (LDI and REC). We found a main effect of age group ($F(1,42) = 8.1, p < .01$), a main effect of memory score ($F(1,42) = 268.3, p < .001$), and an interaction ($F(1,42) = 12.6, p < .01$; Figure 2C). Consistent with our previous findings, post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a decrease in the LDI for aging adults compared to young ($t(84) = 4.4, 95\% \text{ CI } [.06, .28]$), but no difference in REC between groups ($t(84) = .29, 95\% \text{ CI } [-.09, .12]$).

We also examined the role of lag (e.g. number of intervening items between the first presentation of an item and its repeat or lure item) (see Supplemental Table 2 for response rates in each condition). We calculated the LDI and the REC for those trials with a short lag (0–9 items), medium lag (20–80 items), and long lag (120–180 items) and entered these data into two separate 2×3 ANOVAs, using age group (Young and Aging) and Lag (Short, Medium, and Long) as factors. For LDI, we found a main effect of age group ($F(1,126) = 31.7, p < .001$), a main effect of lag ($F(2,126) = 10.7, p < .001$), but no interaction ($F(2,126) = .40, p = .67$) (Figure 2D). In contrast, for REC, we found no main effect of age ($F(1,126) = .37, p = .54$), a main effect of lag ($F(2,126) = 10.3, p < .0001$) and no interaction ($F(2,126) = .03, p = .97$). Thus, while the number of intervening items does result in a drop in both lure discrimination and repeated item recognition performance, we did not detect an interaction with age, suggesting that older adults are not disproportionately affected by the interference from intervening items.

Discussion

Despite this change in task design, from a study/test format to a continuous recognition test with no response deadline, the aging impairment on the LDI remained. Inspection of the results from Experiment 1 (study/test) and Experiment 2 (continuous recognition) reveal that the REC scores are consistent across the two designs, but the LDI is higher for both age groups in the continuous recognition task. While there is still an aging impairment, lure

discrimination for both groups is improved in the self-paced, continuous recognition design. Perhaps this effect is due to additional time to make the decision and response, fewer intervening items (Alley, Hussey, Ko, & Molitor, 2013), or less overall time between the presentation of the original item and the subsequent lure. It is worth noting that the 16-item version of the task resulted in an average of 40 intervening items between the first presentation and the lure, comparable to the medium lag (20–80 items), while the 64-item version of the task resulted in an average of 160 items between the first presentation and the lure, comparable to the long lag (120–180 items). However, there was still a greater difference in time between the first presentation and the lure in study/test design because of the delay period during the instructions for the test session. Regardless of these differences, the age-related memory impairment remained.

Furthermore, we concluded that providing additional time for responses past the 2-second encoding window did not ameliorate the lure discrimination difference between older and younger adults. While this manipulation allowed for longer time to collect the response, it is worth noting that participants were still limited in the amount of encoding time for each image. It is possible that a longer encoding period could alter lure discrimination performance.

Experiment 3: Agnostic Old/New, Study/Test with confidence ratings

In the previous experiments, recognition memory judgments were made using three choices: old, similar and new. It is possible that the lure discrimination issues observed in older adults are driven, at least in part, by problems in their use of “similar” responses and not in problems with the similar lure items per se. Older participants may set a different threshold for the use of the similar response or be less inclined to use it altogether simply because of the novel nature of the response. Thus, in Experiment 3, we employed a study/test design, but the test instructions departed from the standard old/similar/new task. Instead, participants made an old/new recognition memory judgment for each item. In addition, as we wanted to determine whether issues in confidence levels might be driving our observed impairments and as we wanted to explore the data in a signal detection theory framework (it is not possible to ascribe a clear threshold used for “similar” responses), we added a confidence component to the response.

Methods

The study phase was the same as described in the General Methods, but four possible response choices were given on each test trial: 1) sure old, 2) maybe old, 3) sure new, and 4) maybe new. Neither at study nor at test were participants instructed that there would be similar lures or how to categorize them. We will refer to this as ‘agnostic’ instructions (in contrast to the ‘gist’ and ‘veridical’ instructions in Experiment 5 below). Data from 27 Young (mean age: 20.9, range: 18–26; 21F/6M) and 19 Aging (mean age: 74.3, range: 60–86; 13F/6M) participants (1 young and 3 older additional participants were removed for using the wrong response keys or not complying with task instructions) were used.

Results

The data from Experiment 3 were entered in to a Repeated-Measures 2×3 ANOVA, using group (Young and Aging) and trial type (Repeats, Lures, and Foils) (see Supplemental Table 3 for response rates in each condition). In contrast to the previous experiments, the two-choice response format precludes the calculation of an LDI score and we therefore focus on the “old” responses (Figure 3A). There were relatively few “maybe old” and “maybe new” responses, so the old and new responses were combined here. There was no main effect of age group ($F(1, 132) = 1.4, p=.23$), but there was a main effect of trial type ($F(2,132) = 269.8, p<.001$), and an interaction ($F(2,132) = 5.5, p<.01$). Post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a decrease in old responses to the lures for aging adults compared to young ($t(132) = 2.9, 95\%CI [-.26, -.03]$), but no difference in old responses to repeats or foils ($t(132) = .86, 95\%CI [-.16, .08]$ and $t(132) = 1.7, 95\%CI [-.03, .20]$, respectively). As with all of our previous findings, older adults were more likely to identify lures as old items.

One advantage of using a two-choice response instead of a three-choice design is that a response bias can be mitigated by using a d' score, or sensitivity index. A d' score is used in signal detection theory to provide a separation between the means of the signal and noise distributions compared against the standard deviation of the noise distribution. In this case, there are three distributions (and three possible d' scores) because there are three trials types. Here, we calculated three d' scores: 1) d' (T,F) reflects the difference in the distributions between old responses to repeats and old responses to foils; 2) d' (L,F) reflects the difference in the distributions between old responses to lures and old responses to foils; and 3) d' (T,L) reflects the difference in the distributions between old responses to targets and old responses to lures. We predicted that there would be a shift in the lure distribution for older adults, such that the difference between target and lures would be lower, with a corresponding increase in the difference between lures and foils (Figure 3C). In addition, we calculated a variant of d' , d'_a (described below), to account for possible contributions of unequal variance.

Since we were specifically interested in the shift in the lure distribution, we entered these d' scores into a Repeated-Measures 2×2 ANOVA, using age group (Young and Aging) and d' type (d' (T,L), and d' (L,F)) as factors. We found a main effect of age group ($F(1,88) = 4.3, p<.05$), a main effect of d' type ($F(1,88) = 94.2, p<.001$), and an interaction ($F(1,88) = 10.6, p<.001$) (Figure 3B). Consistent with our predictions, post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed an increase in d' (L,F) in for aging adults compared to young ($t(88) = 3.8, 95\%CI [-1.2, -.28]$), but no difference in d' (T,L) ($t(88) = .83, 95\%CI [-.27, .59]$)².

Because d' assumes equal variances in each distribution, we chose to also use a variant of d' , d'_a . This measure makes no assumption about the standard deviation of each variable and is calculated as $d'_a = (\mu_{\text{target}} - \mu_{\text{lure}}) / \sqrt{(\sigma_{2\text{target}}^2 - \sigma_{2\text{lure}}^2) / 2}$ (Mickes, Wixted, & Wais, 2007). Using this measure, we again computed a Repeated-Measures 2×2 ANOVA, using age

²When we included the d' (T,F) in a comparable 3×2 ANOVA, we again observed a main effect of age, a main effect of d' type, an interaction, with post-hoc t-tests again revealing an effect of age for d' (L,F), but not for d' (T,F) or d' (T,L).

group (Young and Aging) and d'_a type (d'_a (T,L), and d'_a (L,F)) as factors. We found a main effect of age group ($F(1,84) = 4.3, p < .05$), a main effect of d'_a type ($F(1,84) = 101.8, p < .001$), and an interaction ($F(1,84) = 16.8, p < .001$) (Figure 3B). Consistent with our predictions, post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed an increase in d'_a (L,F) in for aging adults compared to young ($t(84) = 4.4, 95\%CI [-1.2, -.37]$), but no difference in d'_a (T,L) ($t(84) = 1.4, 95\%CI [-.15, .67]$). Thus, even with this more rigorous measure, the pattern of results is similar.

Next, we examined the confidence ratings for lures. We had hypothesized that young adults might express greater confidence in their responses to the lures compared to older adults since their lure discrimination is markedly better. Responses to lures were entered into a Chi-square analysis that revealed a difference in the distribution of response types as a function of age group ($\chi(3)=123, p < .001$; Figure 3D). While, both groups exhibited more “sure old” responses to lures than any other response type, the aging group responded with more “sure old” responses than the young group ($t(176) = 4.9, p < .05$). In some ways, this result is not surprising because we already knew that older adults made more “old” responses than young adults for the lures. However, the older adults appear to be quite confident in these lure responses, contrary to the notion that younger adults may be more sure of their lure discrimination.

Discussion

By utilizing a two-response design, we were able to evaluate the underlying difference in the center of the lure distributions for older and younger adults, again reinforcing the greater overlap in the distribution between lures and repeats for older adults. While we might have predicted a concurrent shift in d' (T,L) measure, we don't observe it here. We suspect this is because the young adults are actually endorsing more new items as old than the older adults (resulting in numerically poorer memory performance for younger adults). Nevertheless we do observe an increase in the d' (L,F) for the aging group, consistent with our hypothesis. In addition, consistent with Experiment 2, which utilized a continuous recognition test format, the extra time allowed for responses did not ameliorate the age-related deficit in LDI performance in this study-test format.

Experiment 4: Gist/Veridical, Study/Test with confidence ratings

With no specific instructions on how to respond to the similar lures, older adults again demonstrated a bias to identify highly similar lures as repeated items while not demonstrating any difficulty with traditional targets and foils. We wondered if older adults could overcome this bias if the task instructions were modified. We coined these two versions “gist”, an abstract representation of semantic content, and “veridical”, a verbatim trace of information components (Reyna & Brainerd, 1995). We compared two versions of the instructions: one that emphasized the gist, requiring the same response to both repeats and lures, and the other that emphasized the veridical distinction between the repeats and lures.

Methods

The task design for Experiment 4 was identical to Experiment 3, with the exception of the task instructions and how we collected the confidence ratings. In the study session, participants were simply told to perform the indoor/outdoor task. Upon completion of the study task, they were shown example images and were given the instructions for either the gist or veridical test. In the gist instructions, participants were instructed to respond “old” to the similar lure items, while in the veridical instructions, participants were instructed to respond “new” to those similar items. Participants received both versions: study/gist instructions and study/veridical instructions, counterbalanced for order and with 2 different sets of stimuli to minimize interference. They were instructed that the two study/test sets were independent of each other and no items would be repeated from one test to the next. We collected data from 20 Young (mean age: 21.1, range: 19–24; 13F/7M) and 20 Aging participants (mean age: 79.5, range: 66–87; 15F/5M) (5 young and 4 older additional participants were removed for using the wrong response keys or not complying with task instructions).

In addition to the change in task instructions, we also changed the design of the confidence rating responses. We were concerned that the four-choice response design in Experiment 3 might be overwhelming, particularly for older adults who must make both judgments (old/new and confidence) in a single, speeded response. Therefore, for Experiment 4 we changed the design to a two-step response, first the old/new response, followed by a confidence response (very sure, somewhat sure, somewhat unsure, very unsure). Using this new design, we also hoped to get a more even distribution of confidence ratings in order to analyze these data with receiver operator characteristics (ROCs).

Results

The data were entered in to a within-subjects Repeated-Measures 2×2 ANOVA, using age group (Young and Aging) and instruction type (Gist and Veridical) (see Supplemental Table 3 for response rates in each condition). We focused our analysis on the LDI_ON (lures identified as old – lures identified as new) to evaluate the effect of task instructions on this discrimination. We found a main effect of age group ($F(1, 38) = 13.7, p < .001$), a main effect of instructions ($F(1, 38) = 34.2, p < .001$), and an interaction ($F(1,38) = 4.4, p < .05$) (Figure 4A). Post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a difference in aged adults compared to young on their LDI_ON scores for the veridical instructions ($t(76) = 4.1, 95\% \text{ CI } [-.78, -.22]$), but not for the gist instructions ($t(76) = 1.2, 95\% \text{ CI } [-.42, -.50]$). Importantly, the shift in task instructions resulted in a closure of the difference between young and aged adults for the gist version but aged group was still impaired when instructed to respond based on the veridical instructions.

As in Experiment 3, we evaluated this data based on the d' distributions for each of the three distributions. For each version of the instructions, we entered these d' scores into a Repeated-Measures 2×2 ANOVA, using age (Young and Aging) and d' type (d' (T,L), and d' (L,F)) as factors. For the gist instructions, we found a main effect of d' ($F(1,76) = 58.5, p > .001$), but no main effect of age ($F(1,76) = .02, p = .90$) or interaction ($F(1,76) = 1.0, p = .32$; Figure 4B). In contrast, for the veridical instructions, we again found no main effect of age

($F(1,76) = .31, p=.58$) or d' ($F(1,76) = 2.5, p=.11$), but there was an interaction ($F(1,76) = 33.1, p<.001$) (Figure 4C). Post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a decrease in d' (T,L) ($t(76) = 3.7, 95\%CI [.23, .99]$) and a corresponding increase in d' (L,F) in for aging adults compared to young adults ($t(76) = 4.4, 95\%CI [-1.12, -.36]$). While the change in task instructions did alter performance in both the young and aging groups, older adults were unable to overcome the bias to provide an “old” response to lure items.

Also as in Experiment 3, we calculated d'_a to account for unequal distributions of variance for each of the conditions. Again, for each version of the instructions, we entered these d'_a scores into a Repeated-Measures 2×2 ANOVA, using age (Young and Aging) and d'_a type (d' (T,L), and d' (L,F)) as factors. For the gist instructions, we found a main effect of d'_a for instruction type ($F(1,76) = 45.5, p>.001$), but no main effect of age ($F(1,76) = .08, p=.78$) or interaction ($F(1,76) = .45, p=.51$; Figure 4B). In contrast, for the veridical instructions, we again found no main effect of age ($F(1,76) = .02, p=.87$) or d'_a instruction type ($F(1,76) = .02, p=.88$), but there was an interaction ($F(1,76) = 43.6, p<.001$) (Figure 4C). Post-hoc t-tests (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a decrease in d'_a (T,L) ($t(76) = 4.6, 95\%CI [.34, 1.0]$) and a corresponding increase in d'_a (L,F) in for aging adults compared to young adults ($t(76) = 4.7, 95\%CI [-1.05, -.37]$).

Due to the change in the task design for confidence ratings, there was enough data to evaluate the receiver operator characteristics (ROCs) underlying these two experiments. Like d' , ROCs relate the proportion of hits and false alarms across variations in response criteria (here, the confidence rating related to each response). This framework has been used in the study of memory to dissociate the recollection and familiarity components of recognition memory (Wixted, 2007; Wixted, Mickes, & Squire, 2010; Yonelinas & Parks, 2007). ROC analyses provide a far richer representation of the data than simple d' analyses (Yonelinas et al., 1996). Here, we are interested in any shifts in response bias that may interact with confidence ratings instead of making any claims regarding the underlying contributions of recollection or familiarity to these judgments.

First, we calculated the “old” responses to the targets, lures, and foils and ranked them on a 7-point scale based on their confidence ratings (Yonelinas & Parks, 2007). Then we plotted the average value for each group and calculated the ROC curves (Figure 5a) and also z-transformed the average endorsement rate and calculated the slopes (Figure 5b). For the group comparisons, we z-transformed each participant’s endorsement rate and compared the group averages of the unbiased slopes and the unbiased intercepts. We ran a Repeated-Measures 2×2 ANOVA with age (Young and Aging) and instruction type (Gist and Veridical) on the unbiased, z-transformed slopes for each of the trial type comparisons: targets vs. foils (TvF), targets vs. lures (TvL), and lures vs foils (LvF). There were no significant differences among the slopes, for either group, either instruction type, or any interactions.

Next, we analyzed the unbiased z-transformed intercepts using the same parameters. There were no differences for the TvF intercepts for either instruction type, but there was a main effect of age group for the LvF intercepts, $F(1,75) = 11.9, p<.001$. Post-hoc comparisons

(Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a difference between the young and aging group in the veridical condition ($t(75) = 2.9$, 95%CI [-1.6, -.07]), but no difference in the gist condition ($t(75) = 1.9$, 95%CI [-1.3, .22]). For the TvL ANOVA, there was a main effect of age group ($F(1,74) = 12.5$, $p < .001$), instruction type ($F(1,74) = 15.6$, $p < .001$), and an interaction ($F(1,75) = 4.5$, $p < .05$). Post-hoc comparisons (Sidak correction for multiple comparisons with a p-value threshold of .05) revealed a higher intercept for the young group in the veridical condition than the young group in the gist condition ($t(74) = 4.2$, 95%CI [.26, 1.2]) and than the aging group in the veridical condition ($t(74) = 3.9$, 95%CI [.21, 1.1]).

Discussion

The results of Experiment 4 emphasize that young adults are able to alter their performance in this task based on task instructions. They can readily group the lure items with either the old items (gist instructions) or with the foil items (veridical instructions). Older adults, however, cannot do this and treat the lure items much like the target items. With the gist-based instructions they were unimpaired (and numerically doing “better” than the younger individuals). With the veridical instructions, they were again, markedly impaired. Whereas the younger adults were able to modify their responses based on task instructions, the aging adults were unable to overcome their impairment in lure discrimination to improve their performance in the veridical condition. These data again support the conclusion that underlying neurobiological changes are responsible for this decrease in lure discrimination that cannot be overcome by task instructions.

General Discussion

In a series of experiments, we sought to address possible confounds or strategies that might affect the aging impairment in lure discrimination (Holden & Gilbert, 2012; Holden et al., 2013; Stark et al., 2013). We first addressed whether attentional or intentional strategies applied at encoding could rescue the aging bias towards identifying lures as old items instead of similar. However, providing older and younger adults with overt instructions about the three stimulus types and subsequent three-choice recognition test was not sufficient to overcome this bias in older adults. Since prior knowledge of the recognition task did not have a detrimental impact on recognition or mnemonic similarity performance, we concluded the MST is a sensitive task that can be utilized repeatedly in the same person to ascertain change over time, making it an ideal tool for use in clinical trials and other intervention studies with multiple time points.

We also wondered about the minimum number of trials per condition that were necessary to get a reliable measure of this lure discrimination bias. Performance on both the lure discrimination and recognition of repeated items was consistent with as few as 16 items per condition. However, the between-experiment correlation for the 16-item condition was quite low (likely from quantization error), leading us to conclude that a minimum of 20-items per condition can produce a reliable measure. It is worth noting that with the full 64-item per condition set, the lures can be distributed into lure bins based on their similarity (Lacy et al., 2011), which has proven to be an important measure for correlating with functional activity (Yassa et al., 2011) and perforant path integrity (Yassa, Muftuler, & Stark, 2010b). With the

reduced sets, there are not a sufficient number of lures to produce this lure bin analysis or at least to do so with the same resolution. Nevertheless, the reduced sets might be optimal for clinical trials, where the time to take the test is a critical factor. The results from Experiment 1 and Supplement 1 emphasize the utility of this task for repeated administration, which remains robust across testing and does not rely upon the incidental encoding and surprise recognition test. While back-to-back administration would likely not be utilized in clinical or intervention trials, these data suggest that repeated administration of this task does not appear to drastically alter the results.

We were further interested in the effects of the task design for Experiment 2, where we used a continuous recognition design, which we had utilized in earlier functional imaging studies (Kirwan & Stark, 2007; Yassa et al., 2010c; Yassa et al., 2010a). Functional imaging has revealed activity in the hippocampus, specifically in combined DG/CA3 hippocampal subfields, during both incidental encoding (Bakker et al., 2008; Lacy et al., 2011) and the three-choice continuous recognition paradigm (Kirwan & Stark, 2007; Yassa et al., 2010c; Yassa et al., 2010a), with no difference between the two tasks in the hippocampus (Motley & Kirwan, 2012). We were curious as to whether the lure discrimination bias in older adults would be affected by this task design. While lure discrimination performance is improved in the continuous recognition design compared to the study/test design, the age-related impairment in lure discrimination is present in both designs, while recognition performance is spared.

Next, we were interested in the underlying distribution of responses for the lures and how task instructions might influence those responses. In Experiment 3, we shifted to an old/new response paradigm and collected confidence ratings. Again, we observed an age-related bias towards old responses for lure items. Compared to young adults, older adults exhibit a shift in their lure distribution towards the distribution of the repeats. We then manipulated task instructions in Experiment 4 to encourage responses based on gist (respond “old” to lures) or veridical (respond “new” to lures) instructions. Young adults were able to shift their lure responses based on the instructions, but older adults continued to endorse lures as old items, even when instructed to endorse them as new. A detailed ROC analysis, taking into account confidence ratings of these responses, revealed the same overall effect. The underlying neurobiological changes that result in reduced lure discrimination performance cannot be overcome by changes to task instructions.

Taken together, these task manipulations demonstrate the robustness of the age-related deficit in lure discrimination. In any cognitive paradigm, we make an effort to isolate the cognitive construct that we are testing. However, it is important to address how other strategies, and thus, other cognitive processes, may impact the outcome of task performance, particularly when drawing conclusions between groups. In this set of experiments, we set out to address possible strategies or confounds that could account for the age-related differences in lure discrimination that have been repeatedly reported. Despite differences in timing, response type, or task instructions, the age-related impairment in lure discrimination remained. These data support the notion that underlying changes in the brain circuitry, specifically within the hippocampus, cannot be overcome to improve performance on this task.

It is worth noting that in many of the experiments here, females were disproportionately represented over males. This bias was unintentional and merely a byproduct of the population available and interested in participating in the research. We have not observed gender differences here or in the MST when our male sample was larger (Stark et al., 2013). While there are many domains of memory that have reported sex differences in performance (Andreano & Cahill, 2009), many of these effects are mediated by interactions between stress and sex hormones, which we have not manipulated here. For example, an emotional modulation of the MST has revealed a selective remembering of gist, with a loss of details, which correlated with depressive symptoms (Leal et al., 2014).

Models of hippocampal aging have identified several changes occurring within the hippocampal circuit that could contribute to the differences in pattern separation³ performance of older adults observed here (Wilson, Gallagher, Eichenbaum, & Tanila, 2006). First, there is reduced connectivity from the entorhinal cortex to the hippocampus via the perforant path, possibly due to synapse loss (Geinisman, deToledo-Morrell, Morrell, Persina, & Rossi, 1992). Consistent with this rodent data, high-resolution diffusion tensor imaging (DTI) data have shown reduced perforant path integrity in older adults compared to younger adults (Yassa et al., 2010b) with the amount of reduction correlating with lure discrimination performance for aging individuals (Yassa et al., 2011). In addition, reduced input to the DG is coupled with a decrease in modulation by inhibitory neurons in this region (Vela, Gutierrez, Vitorica, & Ruano, 2003) resulting in a hypoactive DG. In aging, the reduction in input from the EC via the perforant path and decreased interneuron activity in the DG may result in a failure of the DG to reduce the similarity of input patterns and project this on to the CA3, leading to a decrease in pattern separation performance.

Further, reduced inhibitory drive in the recurrent collaterals in the CA3 region may lead to rigidity in place cell firing patterns, resulting in an increased propensity for pattern completion in the network (Hasselmo, Schnell, & Barkai, 1995). Specifically, reduced cholinergic input releases the CA3 auto-associative fibers from inhibition, thereby promoting completion of the familiar pattern with an existing representation. Consistent with this model, there is evidence for similar representational rigidity in the combined DG/CA3 subfields in humans using high-resolution BOLD fMRI (Yassa et al., 2010a). Although the mechanism for this representational rigidity cannot be determined from these data, the degree of rigidity correlated with lure discrimination performance on this task. Thus, the MST has demonstrated strong links to two different neurobiological features of the aging hippocampal network: (1) representational rigidity in the DG/CA3, and (2) perforant path integrity.

Age-related changes to this system lead to predictions of impairments on those forms of memory which all place strong demands on pattern separation and to a greater reliance on gist rather than specific details (Yassa & Stark, 2011), such as we observed here. However, it is worth noting that the computations of pattern completion and pattern separation can certainly act in parallel and there are clear opportunities for them to interact or compete as to

³It is worth noting that while we refer to pattern separation behavior, it is impossible to directly measure the computational process of pattern separation.

which one sends the strongest signal forward to the next stage. For example, in the CA3 subfield, the auto-associative recurrent collaterals are typically hypothesized to provide a mechanism for pattern completion. If we assume a pattern separation signal arriving from the Mossy Fibers from the dentate, we have a clear example of the computations being performed separately, but interacting or competing to set activity in CA3.

The behavioral data presented here, coupled with the neuroimaging and integrity measures, emphasize the utility of this task for measuring pattern separation behavior, possibly as a proxy for the integrity of the DG and may be a useful task for assessing changes in other disorders associated with disruptions of the dentate gyrus, such as Schizophrenia, depression, and sleep disorders. The MST has already been used as an outcome measure in a recent clinical trial in amnesic mild cognitive impairment (Bakker et al., 2012) and documented improvement in pattern separation performance with two-week exposure to a low-dose antiepileptic medication (Levetiracetam). These findings highlight the usefulness of using the MST as a sensitive measure for detecting behavioral changes for investigational drugs aimed at improving memory.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by a grant from the National Institutes on Aging R01 AG034613, and an Alzheimer's Disease Research Center Project Award AG016573. We thank Emily Thai, Jessica German, Neil Patel, Ryan Abrigo, Max Chen, and Farid Ghamsari for assistance in data collection. Finally, we would like to thank Dr. Peter Wais and Dr. John Wixted for help with calculating the ROC curves.

References

- Alley BA, Hussey EP, Ko PC, Molitor RJ. Pattern separation and pattern completion in Alzheimer's disease: evidence of rapid forgetting in amnesic mild cognitive impairment. *Hippocampus*. 2013; 23:1246–1258. [PubMed: 23804525]
- Andreano JM, Cahill L. Sex influenced on the neurobiology of learning and memory. *Learning and Memory*. 2009; 16:248–266. [PubMed: 19318467]
- Bakker A, Kirwan CB, Miller M, Stark CE. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*. 2008; 319(5870):1640–1642.10.1126/science.1152882 [PubMed: 18356518]
- Bakker A, Krauss GL, Albert MS, Speck CL, Jones LR, Stark CE, Gallagher M. Reduction of hippocampal hyperactivity improves cognition in amnesic mild cognitive impairment. *Neuron*. 2012; 74(3):467–474.10.1016/j.neuron.2012.03.023 [PubMed: 22578498]
- Brandt J, Spencer M, Folstein MF. The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*. 1988; 1:111–117.
- Chawla MK, Guzowski JF, Ramirez-Amaya V, Lipa P, Hoffman KL, Marriott LK, Barnes CA. Sparse, environmentally selective expression of Arc RNA in the upper blade of the rodent fascia dentata by brief spatial experience. *Hippocampus*. 2005; 15(5):579–586. [PubMed: 15920719]
- Folstein M, Folstein SE, McHugh PR. "Mini Mental State": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry*. 1975; 12:189–198.
- Geinisman Y, deToledo-Morrell L, Morrell F, Persina IS, Rossi M. Age-related loss of axospinous synapses formed by two afferent systems in the rat dentate gyrus as revealed by the unbiased

- stereological dissector technique. *Hippocampus*. 1992; 2(4):437–444.10.1002/hipo.450020411 [PubMed: 1308200]
- Hasselmo ME, Schnell E, Barkai E. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience*. 1995; 15(7):5249–5262. [PubMed: 7623149]
- Holden HM, Gilbert P. Less efficient pattern separation may contribute to age-related spatial memory deficits. *Frontiers in Aging Neuroscience*. 2012; 4:1–9. [PubMed: 22514534]
- Holden HM, Toner CK, Pirogovsky E, Kirwan CB, Gilbert P. Visual object pattern separation varies in nondemented older adults. *Learning & Memory*. 2013; 20:358–362. [PubMed: 23774765]
- Jennings JM, Jacoby LL. Automatic versus intentional uses of memory: aging, attention, and control. *Psychological Aging*. 1993; 8(2):283–293.
- Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychological Bulletin*. 1993; 114:3–28. [PubMed: 8346328]
- Kirwan CB, Hartshorn A, Stark SM, Goodrich-Hunsaker NJ, Hopkins RO, Stark CEL. Pattern separation deficits following damage to the hippocampus. *Neuropsychologia*. 2012; 50(10):2408–2414. [PubMed: 22732491]
- Kirwan CB, Stark CE. Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. *Learning and Memory*. 2007; 14(9):625–633. [PubMed: 17848502]
- Kensinger EA, Schacter DL. When true memories suppress false memories: effects of aging. *Cognitive Neuropsychology*. 1999; 16:399–415.
- Koutstaal W, Schacter DL, Brenner C. Dual task demands and gist-based false recognition of pictures in younger and older adults. *Journal of Memory and Language*. 2001; 44:399–426.
- Lacy JW, Yassa MA, Stark SM, Muftuler LT, Stark CE. Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning and Memory*. 2011; 18(1):15–18. [PubMed: 21164173]
- Leal SL, Tighe SK, Yassa MA. Asymmetrical effects of emotion on mnemonic interference. *Neurobiology of Learning and Memory*. 2014; 111:41–48. [PubMed: 24607286]
- Leutgeb JK, Leutgeb S, Moser MB, Moser EI. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*. 2007; 315(5814):961–966. [PubMed: 17303747]
- Lyle KB, Bloise SM, Johnson MK. Age-related binding deficits and the content of false memories. *Psychological Aging*. 2006; 21:86–95.
- McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*. 1995; 102(3):419–457. [PubMed: 7624455]
- Mickes L, Wixted JT, Wais PE. A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*. 2007; 14(5):858–865. [PubMed: 18087950]
- Motley SE, Kirwan CB. A parametric investigation of pattern separation processes in the medial temporal lobe. *Journal of Neuroscience*. 2012; 32(38):13076–13084. [PubMed: 22993425]
- Naveh-Benjamin M, Guez J, Shulman S. Older adults' associative deficits in episodic memory: assessing the role of decline in attentional resources. *Psychonomic Bulletin & Review*. 2004; 11(6):1067–1073. [PubMed: 15875977]
- Neuneubel JP, Knierim JJ. CA3 retrieves coherent representations from degraded input: direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*. 2014; 81(2):416–427. [PubMed: 24462102]
- Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review*. 2003; 110(4):611–646. [PubMed: 14599236]
- Reyna VF, Brainerd CJ. Fuzzy-trace theory: Some foundational issues. *Learning & Individual Differences*. 1995; 7:145–162.
- Schacter DL, Koutstaal W, Johnson MK, Gross MS, Angell KE. False recollection induced by photographs: comparison of older and younger adults. *Psychological Aging*. 1997; 12:203–215.

- Stanislaw H, Todorov N. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*. 1999; 31(1):137–149.
- Stark SM, Yassa MA, Lacy JW, Stark CE. A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*. 2013; 51(12):2442–2449. [PubMed: 23313292]
- Stark SM, Yassa MA, Stark CE. Individual differences in spatial pattern separation performance associated with healthy aging in humans. *Learning and Memory*. 2010; 17(6):284–288. [PubMed: 20495062]
- Toner CK, Pirogovsky E, Kirwan CB, Gilbert PE. Visual object pattern separation deficits in nondemented older adults. *Learning and Memory*. 2009; 16(5):338–342. [PubMed: 19403797]
- Treves A, Rolls ET. Computational analysis of the role of the hippocampus in memory. *Hippocampus*. 1994; 4(3):374–391. [PubMed: 7842058]
- Vela J, Gutierrez A, Vitorica J, Ruano S. Rat hippocampal GABAergic molecular markers are differentially affected by ageing. *Journal of Neurochemistry*. 2003; 85(2):368–377. [PubMed: 12675913]
- Wilson IA, Gallagher M, Eichenbaum H, Tanila H. Neurocognitive aging: prior memories hinder new hippocampal encoding. *Trends in Neuroscience*. 2006; 29(12):662–670.
- Wixted JT. Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*. 2007; 114(1):152–176. [PubMed: 17227185]
- Wixted JT, Mickes L, Squire LR. Measuring recollection and familiarity in the medial temporal lobe. *Hippocampus*. 2010; 20(11):1195–1205. [PubMed: 20848603]
- Yassa MA, Lacy JW, Stark SM, Albert MS, Gallagher M, Stark CE. Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*. 2010a; 21(9):968–979. [PubMed: 20865732]
- Yassa MA, Muftuler LT, Stark CE. Ultrahigh-resolution microstructural diffusion tensor imaging reveals perforant path degradation in aged humans in vivo. *Proceedings of the National Academy of Sciences, U S A*. 2010b; 107(28):12687–12691.
- Yassa MA, Stark SM, Bakker A, Albert MS, Gallagher M, Stark CE. High-resolution structural and functional MRI of hippocampal CA3 and dentate gyrus in patients with amnesic Mild Cognitive Impairment. *Neuroimage*. 2010c; 51(3):1242–1252. [PubMed: 20338246]
- Yassa MA, Mattfeld AT, Stark SM, Stark CE. Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences, U S A*. 2011; 108(21):8873–8878.
- Yassa MA, Stark CE. Pattern separation in the hippocampus. *Trends in Neuroscience*. 2011; 34(10):515–525.
- Yonelinas AP, Dobbins I, Szymanski MD, Dhaliwal HS, King L. Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition*. 1996; 5:418–441. [PubMed: 9063609]
- Yonelinas AP, Parks CM. Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychological Bulletin*. 2007; 133(5):800–832. [PubMed: 17723031]

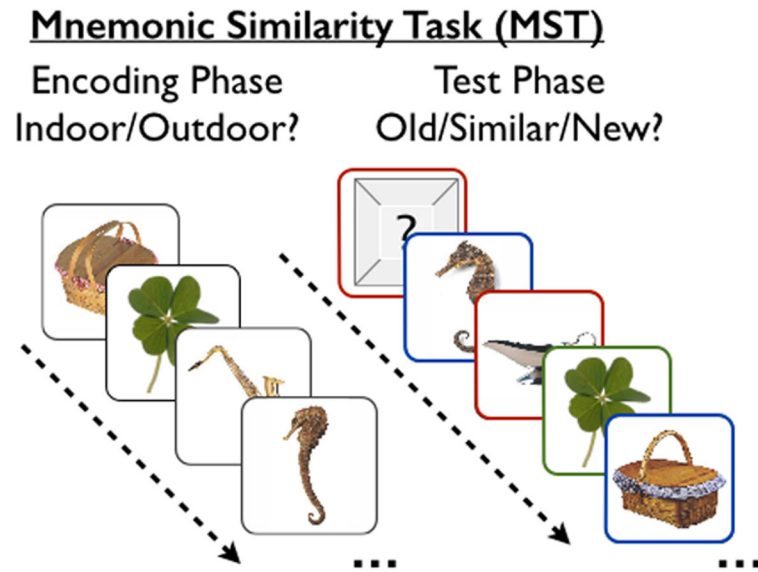


Figure 1.

The Mnemonic Similarity Task (MST) typically consists of two phases: an incidental-encoding phase, followed by a surprise recognition test with a three-choice response. Importantly, the test consists of three stimulus types: exact repetitions of earlier items (outlined in green), lures that are similar but not identical to earlier items (outlined in blue), and novel foils (outlined in red). The colored outlines are for illustrative purposes only and were not present in the actual experiment.

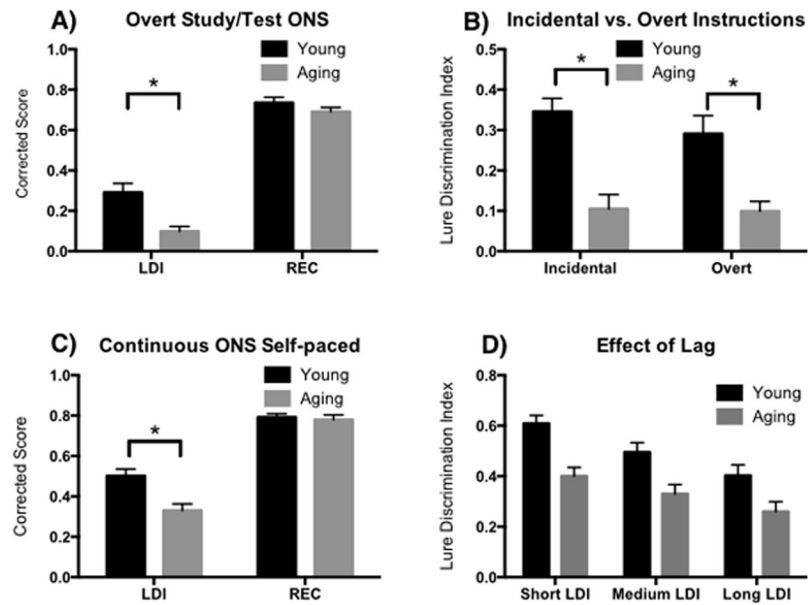


Figure 2.

A) Experiment 1. LDI (Lure Discrimination Index) is impaired in aging, while REC (Recognition) is intact following overt instructions prior to encoding regarding the design of the recognition test. B) Experiment 1. Overt instructions resulted in comparable LDI performance when compared to performance following incidental instructions. C) Experiment 2. In a continuous recognition design with self-paced responses, older adults are still impaired compared to young in LDI, but their REC scores remain matched. D) Experiment 2. There is a lag effect (number of intervening trials between the first presentation and lure) with a main effect of age, but no interaction.

* denotes $p < .05$

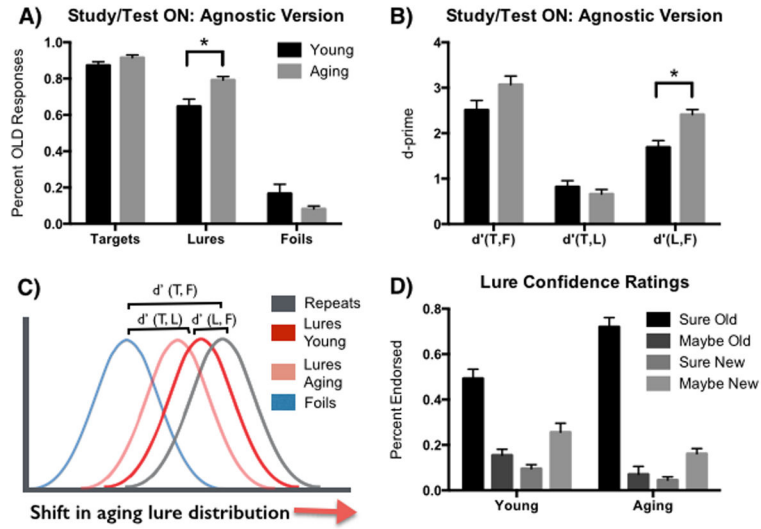


Figure 3. Experiment 4. A) When restricted to old/new responses during the recognition task and no instructions on how to respond to the similar lures, older adults demonstrated a greater bias to respond “old” to those lures. B) When we calculated d' scores for each of the three test item distributions, older adults showed a greater disparity between the lure and foil distributions than younger adults. C) These data support the theory that there is an age-related shift in the lure distribution towards the repeated items (T=target, L=Lure, F=Foils). D) An evaluation of confidence ratings for lure items reflected high confidence in responding “old” for the aging group. * denotes $p < .05$

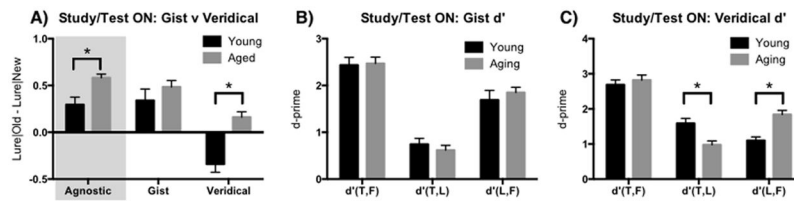


Figure 4.

Experiment 4. A) Providing instructions to bias participants to respond based on gist (respond old to the lures) versus veridical (respond new to lures) representations attenuated the age-related difference. Data from the agnostic instructions from Experiment 3 are presented in the shaded box for comparison. The shift in the underlying distributions for the three test types can be observed in the veridical condition (C), but not the gist condition (B). * denotes $p < .05$

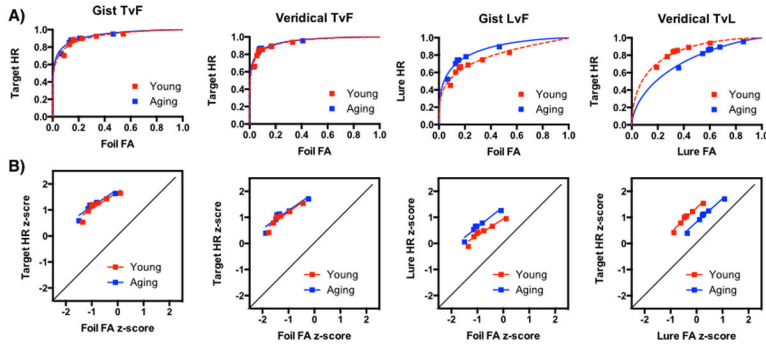


Figure 5. Experiment 4. A) ROC curves and B) slopes on the z-scores for “old” responses to targets and foils are the same for each set of instructions. When given the veridical instructions, young adults are better at discriminating between the targets and lures than older adults. In contrast, with the gist instructions, the aging group is better at discriminating the lures from the foils (i.e. they are more likely to identify the lures as “old”).