Aging Differentially Affects Episodic Memory and Skill Acquisition

Adam W. Broitman

University of Pennsylvania

M. Karl Healey

Michigan State University

Michael J. Kahana

University of Pennsylvania

Author Note

The authors gratefully acknowledge support from National Institutes of Health Grants AG048233 and MH55687. We thank Patrick Crutchley, Jonathan Miller, and Isaac Pedisich for assistance with designing and programming the experiment and Ada Aka, Elizabeth Crutchley, Kylie Hower, Joel Kuhn, and Logan O'Sullivan for help with data collection. All data collected as part of this study can be accessed by visiting memory.psych.upenn.edu. Correspondence concerning this article should be addressed to either Michael Kahana (kahana@sas.upenn.edu) or Karl Healey (khealey@msu.edu).

Version of February 9, 2017

Abstract

Healthy aging can differentially impact various cognitive processes. Here we report findings from a 35-session longitudinal study of episodic memory carried out over five years. Using a model-based analysis of the behavioral data, we extracted measures of age-related changes in both episodic memory and skill acquisition (Anderson, Fincham, & Douglass, 1999). This analysis revealed minimal age-related decline in episodic memory and significant age-related decline in skill acquisition and retention in older adults.

Keywords: episodic memory; aging; free recall; recognition; memory models

Aging Differentially Affects Episodic Memory and Skill Acquisition

Introduction

For a participant to perform well when faced with an unfamiliar laboratory memory task, they must be able to efficiently apply pre-existing task-general memory mechanisms to the new task. But they must also devise new task-specific strategies or skills to meet the idiosyncratic demands of the new task. Memorizing a list of words in the free recall task, for example, draws on general mechanisms for forming new episodic associations which operate in many different tasks (Healey & Kahana, 2017). But it also relies on developing new task-specific strategies (Delaney & Knowles, 2005). Cognitive aging impacts both task-general (e.g. Wahlheim, McDaniel, & Little, 2016) and task-specific (e.g., Dunlosky & Hertzog, 1998) components of performance. But it is unclear whether aging impacts these components differentially.

Here we attempt to separate the effects of aging on episodic memory mechanisms on the one hand versus task-specific skill acquisition on the other hand. To do so we introduce a five-year longitudinal data set in which participants completed a seven-session battery of free recall and recognition testing each year. This design allowed us to measure the development of task-specific skills (i.e., practice effects) as well as the loss of these skills (i.e., forgetting effects) during the separation between annual waves. To disentangle the effects of age on task-general and task-specific factors we use a model-based approach (e.g., Anderson et al., 1999; Sliwinski, Hoffman, & Hofer, 2010).

This approach allows us to simultaneously model an age-related decline in performance due to episodic memory impairments and a practice-related increase in performance due to task-specific skill accumulation. That is, does episodic memory per se get worse with increasing age after accounting for the influence of task-specific skill? Similarly, do older individuals reap less of a benefit from extensive practice with a task than do younger individuals?

Method

Participants

The data reported here are from the Penn Aging and Longitudinal Memory Study (PALMS). Participants were recruited from the cross-sectional sample collected for Healey and Kahana (2016) to take part in annual follow-ups over a five-year period. Participants were recruited from the municipal Philadelphia area, using advertisements in community centers and local newspapers. Participants ranged from 62-74 years of age at the start of the experiment and had a mean age of 66.87 years. All participants had completed a high school education or a high school equivalency, and total years of education ranged from 14-21 years. Potential participants were excluded if they suffered from any medical conditions, or regularly took medications that might affect their cognitive performance. Twelve of the 39 older adult participants who completed the original cross-sectional experiment were recruited in to the longitudinal sample.

Once recruited, participants ran the full seven-session experiment each year, with different word lists generated for each series of follow-up sessions. Since the beginning of the study, three participants have chosen to discontinue follow-ups and one participant has passed away. These participants have been excluded from the current analyses due to a lack of sufficient longitudinal data. Of the eight participants included in the present analyses, two have completed four annual waves of testing and six have completed five waves. At the beginning of each wave, The Recent Life Changes Questionnaire (Miller & Rahe, 1997) was administered to collect information about any potential changes in subjects' health or personal lives. No participants included in the current analyses developed a medical condition that would have excluded them from initial participation.

PALMS Experiment

The current analyses focus on the behavioral data from the PALMS experiment, which consisted of 7 sessions, each of which included 16 free recall lists followed by 16 lists of recognition. For each recall list, 16 words were presented one at a time on a computer screen followed by an immediate free recall test. The first session and half of the remaining sessions were randomly chosen to include a final free recall test before recognition, in which participants recalled words from any of the lists from the session. Each word was accompanied by a cue to perform one of two judgment tasks ("Will this item fit into a shoebox?" or "Does this word refer to something living or not living?") or no encoding task. The current task was indicated by the color and typeface of the presented item. There were three conditions: no-task lists (participants did not have to perform judgments with the presented items), single-task lists (all items were presented with the same task), and task-shift lists (items were presented with either task). The first two lists were task-shift lists, and each list started with a different task. The next fourteen lists contained four no-task lists, six single-task lists (three of each of the task), and four task-shift lists. List and task order were counterbalanced across sessions and participants. The present analyses do not include performance data from the encoding task.

Each stimulus was drawn from a pool of 1638 words. Lists were constructed such that varying degrees of semantic relatedness occurred at both adjacent and distant serial positions. Semantic relatedness was determined using the Word Association Space (WAS) model described by Steyvers, Shiffrin, and Nelson (2004). WAS similarity values were used to group words into four similarity bins (high similarity: $\cos\theta$ between words > 0.7; medium-high similarity, $0.4 < \cos\theta < 0.7$; medium-low similarity, $0.14 < \cos\theta < 0.4$; low similarity, $\cos\theta < 0.14$). Two pairs of items from each of the four groups were arranged such that one pair occurred at adjacent serial positions and the other pair was separated by at least two other items. For each list, there was a 1500 ms delay before the first word appeared on the screen. Each item was on the screen for 3000 ms, followed by jittered (i.e., variable) inter-stimulus interval of 800-1200 ms (uniform distribution). If the word was associated with a task, participants indicated their response via a keypress. After the last item in the list, there was a jittered delay of 1200-1400 ms, after which a tone sounded, a row of asterisks appeared, and the participant was given 75 seconds to attempt to recall aloud any of the just-presented items. If a session was selected for final free recall, following the immediate free recall test from the last list, Participants had 5 minutes to recall any item from the preceding lists. Final free recall data are not analyzed here.

A recognition test was administered following the free recall portion of the experiment. In this final recognition test, lures were selected from the remaining items not presented during the free recall phase, and target/lure ratio varied with session, where targets made up 80%, 75%, 62.5%, or 50% of the total items. In total, 320 words were presented one at a time on the computer screen. When a word was presented on the screen, participants were instructed to indicate whether the test word had been presented previously. Participants were told to respond verbally “pess” for old items and “po” for new items and to confirm their response by pressing the space bar. These responses (”pess” and “po”) were chosen so that both response types would initiate with the same stop consonant (or plosive), thus assisting in automated detection of word onset times. Following the old-new judgment, participants made a confidence rating on a scale of 1 to 5, with 5 being the most confident. Although recognition was self-paced, participants were encouraged to respond as quickly as possible without sacrificing accuracy. Participants were given feedback on accuracy and reaction time.

Results

Free recall performance was measured as overall recall probability; recognition performance was measured with d'. As a statistical test of the effect of wave, a linear regression provided a slope for each subject’s annual performance and we used a t-test to determine whether the distribution of slopes across subjects was different from zero. Contrary to our initial expectations, there was a reliable increase in recognition performance across waves (Mean = .0061, SD= .0673, $t(7) = 2.6, p = .03$). There was also a small but non-significant increase in free recall performance across waves (Mean = .0062, SD = .0119, $t(7) = 1.5, p = .18$).

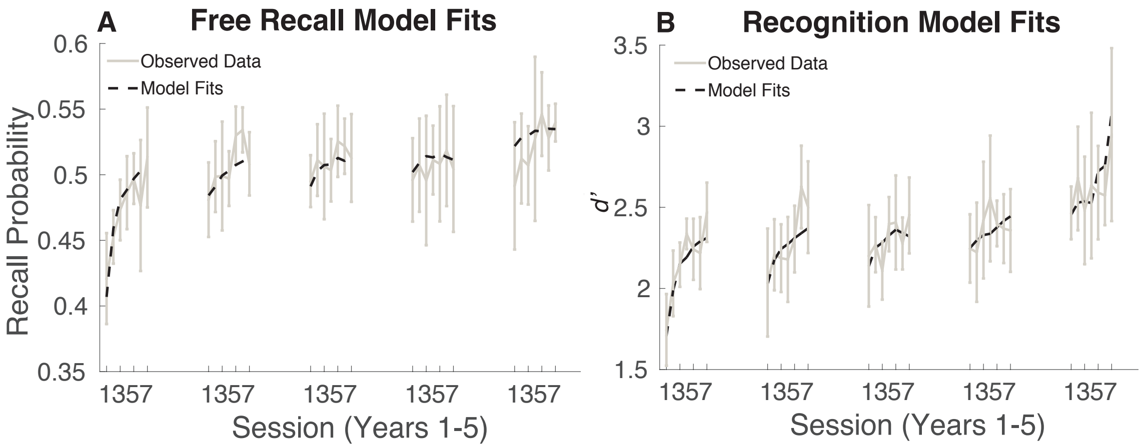

Figure 1. Mean observed performance by session (gray) is plotted with mean model-estimated performance (black). Error bars are 95% within-subject confidence intervals (Loftus & Masson, 1994). N=8 for years 1-4. N=6 for year 5.

Across the seven sessions within each wave, we observed substantial practice effects (Figure 1). We hypothesized that in our linear regression, these practice effects were being confounded with age-related episodic memory changes, and we determined that adding a model of practice to our linear model could help disentangle these two aspects of performance. Whereas there are several existing models that can be applied to predict practice effects in a multi-session study, we selected John Anderson's strength accumulation model (Anderson et al., 1999) due to its simple, integrated term for the estimation of practice and forgetting effects. In our adaptation of this model, memory performance is a function of both the linear effects of age-related episodic memory change and the power-law effects of practice:

$$\text{Perf.} = \beta_0 + \beta_{age} Age + \left(\beta_{prac} - \frac{\beta_{prac}}{\sum_{i=1}^{n} t_i^{-d}}\right) \quad (1)$$

where at session i, memory performance is a function of four free parameters: β_0, an intercept which represents the participant's performance at the time of their birth; β_{age}, the amount by which performance changes daily as a result of aging; β_{prac}, the maximum performance benefit that can be gained as a result of practice; and d, which modulates the rate at which practice and forgetting affect memory performance. As the d parameter grows larger, the rate of practice accumulation slows, while the rate of practice dissipation in between trials increases.

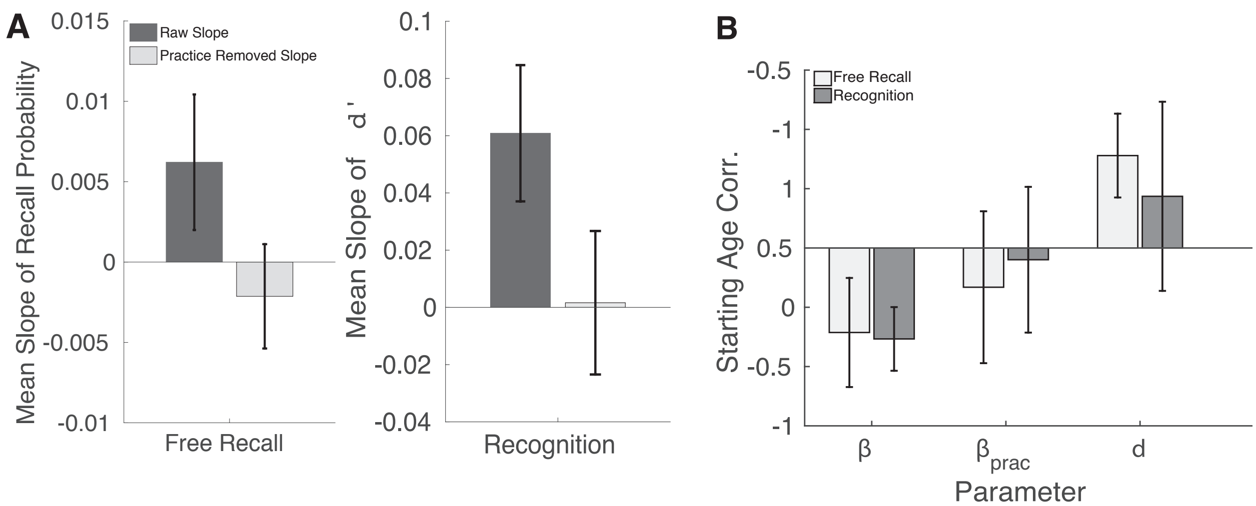

Figure 2. A) Slope of annual performance in observed data & data with estimated practice effects removed. Slopes were calculated by averaging annual performance across all participants and applying a linear regression model. B) Parameter correlations with starting age. β_{age} estimates age-related cognitive change, β_{prac} estimates the maximum possible performance gain an individual can acquire due to practice effects, and d modulates the rate of practice accumulation and forgetting in between test trials. Error bars in subplot A and B are 95% confidence intervals obtained from the standard error of the mean. Error bar in subplot C are 95% bootstrapped confidence intervals.

Model-Based Analysis

We fit the model separately to the free recall and recognition performance of each individual participant by minimizing the χ^2 difference value between the model predictions and observed data using the equation $\chi^2 = \Sigma_{i=1}^{n} \left(\frac{s_i - \hat{s_i}}{SE_{\bar{s}}}\right)$ where s is performance data for session i. To search for the best-fitting parameters, we used MATLAB's GlobalSearch function, which conducts a nonlinear interior-point optimization search from multiple starting points to find the parameter values that yield the global minimum χ^2 difference value. β_{age} was constrained so that it could not change annual performance by more than 10% annually. β_{prac} was constrained from 0 to 1 for free recall and from 0 to 3 for d' estimates. d was constrained from 0 to 1 for all estimates. Figure 1 shows the model's fits with observed data.

Optimized values for β_{prac} were significantly positive for fits to free recall (Mean=.137, SD=.0521, $t(7) = 7.45, p < .01$) as well as recognition data (Mean=.914, SD=.433, $t(7) = 6.0$, $p < .01$). Optimized values for d were also significantly positive in both free recall (Mean = .331, SD=.240, $t(7) = 3.9$, $p < .01$) and recognition data fits (Mean =.396, SD = .173 $t(7) = 6.5, p < .01$). However, the best-fitting β_{age} values were not significantly different from zero (FR: Mean = 1.15×10^{-6}, SD=1.25×10^{-5}, $t(7) = -.26, p = .81$; Recog: Mean=9.65×10^{-6}, SD=5.73×10^{-5} $t(7) = .48, p = .65$), indicating that the behavioral data do not reflect significant age-related episodic memory decline.

To provide a residual estimate of age-related change, we removed estimated practice effects from observed data using the fitted values of the $\beta_{prac} - \frac{\beta_{prac}}{\sum_{i=1}^{n} t_i^{-d}}$ term for each participant. Figure 2a shows that with practice effects removed, we did not observe annual performance increases (as we had with the linear model). However, as shown by the error bars on Figure 2a, we also did not observe significant age-related decreases in free recall or recognition performance. Although small and non-significant annual performance decreases were observed in the practice-removed free recall data ($t(7) = -1.1$, $p = .30$), the practice-removed recognition performance remained highly stable ($t(7) = -.25$, $p = .81$). This result was consistent with our expectations, as it is rare to observe drastic annual changes in episodic memory performance in a population of healthy adults (Salthouse, 2009), and the previous literature reports minimal recognition memory impairments in healthy older adults (Craik, 1971; Jacoby, 1999; Ratcliff, Thapar, & McKoon, 2004; Schonfield & Robertson, 1966)

Because skill acquisition depends on cognitive processes that might themselves be vulnerable to age-related decline, we can predict that the oldest individuals in our sample will benefit less from practice than the youngest individuals. To test this prediction, we investigated whether age at the start of the experiment correlated with the model's parameter estimates. Figure 2b shows the correlations between age at the start of the experiment and the model's parameter estimates. Starting age correlated negatively with β_{age} in both free recall ($r(6) = -.71$, $p = .04$) and recognition estimates ($r(6) = -.77$, $p = .02$), indicating that the oldest participants experienced more negative age-related episodic memory changes. Estimates for the d parameter had a positive, significant correlation with starting age in free recall fits ($r(6) = .78$, $p = .02$) and a modest, non-significant positive correlation in recognition fits ($r(6) = .45$, $p = .28$). A positive correlation between the d parameter and starting age indicates that the oldest participants have a slower rate of practice accumulation, as well as a more rapid rate of forgetting in between testing sessions. Finally, the β_{prac} parameter had a small, negative correlation with starting age that was non-significant (FR: $r(7) = -.33$, $p = .42$; Recog: $r(7) = -.10$, $p = .81$). These estimates suggest that the rate of decline in episodic memory increases with age, while the ability to accumulate and retain performance gains from practice decreases.

Discussion

Our aim was to separately evaluate the effects of cognitive aging on task-general episodic memory mechanisms versus task-specific skill acquisition. We designed a longitudinal study in which, once a year, participants received intensive practice with the free recall and recognition tasks. We then separately modeled age-related declines in episodic memory on the one hand and task-specific practice effects on the other hand. Our analysis revealed that although overall performance increased across annual waves, this was largely due to task-specific skills and not general episodic memory ability. Indeed, after the effect of task-specific skill was removed, episodic memory ability was highly stable across the five years of the study, which is consistent with observations from previous work (Salthouse, 2009). There was, however, evidence that the oldest participants in our sample were beginning to show significant episodic memory decline.

Our modeling also revealed important effects of age on task-specific skill acquisition. There was no correlation between a participant's initial age and the model's estimate of the maximum benefit they would receive from practice. This suggests little age-related decline in the ability to acquire practice-related performance gains. Although, the maximum practice-related performance gain did not seem to differ as a function of age, there was evidence that older participants approached this maximum gain more slowly and that their task-specific skills deteriorated more rapidly when not practicing the tasks.

In summary, we have shown that while episodic memory performance remained relatively stable in our study population over a four-year period, the rate of age-related episodic memory decline, as well as the rate of skill dissipation (forgetting) in between practice trials, appear to increase as a function of age.

References

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1120–1136.

Craik, F. I. M. (1971). Age differences in recognition memory. *The Quarterly Journal of Experimental Psychology*, *23*(3), 316–323.

Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing effects in the free recall of unmixed lists. *Journal of Memory and Language*, *52*(1), 120-130.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging*, *13*, 597–607.

Healey, M. K., & Kahana, M. J. (2016). Age-related changes in the dynamics of memory encoding processes provide a biomarker of successful aging. *Submitted*.

Healey, M. K., & Kahana, M. J. (2017). Contiguity in episodic memory. *submitted*.

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 3-22.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Miller, M. A., & Rahe, R. H. (1997). Life changes scaling for the 1990s. *Journal of Psychosomatic Research*, *43*(3), 279–292.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*(4), 408–424.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514.

Schonfield, D., & Robertson, B. A. (1966). Memory storage and aging. *Canadian Journal of Psychology*, *20*, 228-236.

Sliwinski, M., Hoffman, L., & Hofer, S. (2010). Modeling retest and aging effects in a measurement burst design. In P. C. Molenaar & K. M. Newell (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (p. 37-50). Washington D.C.: American Psychological Association.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer.* Washington, DC: American Psychological Association.

Wahlheim, C. N., McDaniel, M. A., & Little, J. L. (2016). Category learning strategies in younger and older adults: Rule abstraction and memorization. *Psychology and Aging*, *31*(4), 346.