



A virtual water maze revisited: Two-year changes in navigation performance and their neural correlates in healthy adults

Ana M. Daugherty^{a,b,*}, Naftali Raz^{a,c}

^a Institute of Gerontology, Wayne State University, Detroit, MI, USA

^b Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, Champaign, IL, USA

^c Department of Psychology, Wayne State University, Detroit, MI, USA

ARTICLE INFO

Article history:

Received 19 April 2016

Accepted 17 September 2016

Available online 19 September 2016

Keywords:

Brain

Iron

Hippocampus

Caudate

Aging

Longitudinal

ABSTRACT

Age-related declines in spatial navigation are associated with deficits in procedural and episodic memory and deterioration of their neural substrates. For the lack of longitudinal evidence, the pace and magnitude of these declines and their neural mediators remain unclear. Here we examined virtual navigation in healthy adults (N=213, age 18–77 years) tested twice, two years apart, with complementary indices of navigation performance (path length and complexity) measured over six learning trials at each occasion. Slopes of skill acquisition curves and longitudinal change therein were estimated in structural equation modeling, together with change in regional brain volumes and iron content (R2* relaxometry). Although performance on the first trial did not differ between occasions separated by two years, the slope of path length improvement over trials was shallower and end-of-session performance worse at follow-up. Advanced age, higher pulse pressure, smaller cerebellar and caudate volumes, and greater caudate iron content were associated with longer search paths, i.e. poorer navigation performance. In contrast, path complexity diminished faster over trials at follow-up, albeit less so in older adults. Improvement in path complexity after two years was predicted by lower baseline hippocampal iron content and larger parahippocampal volume. Thus, navigation path length behaves as an index of perceptual-motor skill that is vulnerable to age-related decline, whereas path complexity may reflect cognitive mapping in episodic memory that improves with repeated testing, although not enough to overcome age-related deficits.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Spatial navigation is a complex skill that requires maintaining cognitive representations in memory and updating them in real time, evaluating and implementing navigation strategies and deploying perceptual-motor skills—all of which demand substantial computational effort (Wolbers and Hegarty, 2010) and show significant age-related declines (Moffat, 2009). The study of age-related differences in navigation skill and neural correlates has been aided, in particular, by use of the Morris water maze (MWM; Morris, 1981) as an experimental method for investigating animal

Abbreviations: BS 95% CI, bias-corrected bootstrapped 95% confidence intervals; Cb, cerebellum; Cd, caudate nucleus; CFI, comparative fit index; FD, fractal dimension (path complexity); FIML, full information maximum likelihood; Hc, hippocampus; ICV, intracranial volume; LPFC, lateral prefrontal cortex; PHG, parahippocampal gyrus; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; SWI, susceptibility weighted imaging; vMWM, virtual Morris water maze; WRMR, weighted root mean square residual

* Correspondence to: Beckman Institute, University of Illinois at Urbana-Champaign, 405 N Mathews Ave., Urbana, IL 61801, USA.

E-mail address: adaugher@illinois.edu (A.M. Daugherty).

<http://dx.doi.org/10.1016/j.neuroimage.2016.09.044>
1053-8119/© 2016 Elsevier Inc. All rights reserved.

navigation, and its virtual analog (vMWM, Astur et al., 1998) designed to study human behavior in a comparable framework. The task requires use of cues to navigate towards a hidden platform, often repeatedly to test learning a hidden goal location and stability of this knowledge. It is assumed that, to successfully reach the goal, the individual performing the task engages several cognitive processes, including procedural and episodic memory.

Given the differential vulnerability of these memory functions to decline in normal aging (Horn and Donaldson, 1980; Lindenberger et al., 1993; Park, 2000), understanding their contributions to navigation may elucidate the mechanisms of its decline. The magnitude of age-related differences in perceptual-motor skill acquisition depends on the task (Seidler, 2006), yet several studies have replicated poorer performance in older adults compared to their younger counterparts (Wright and Payne, 1985; Raz et al., 2000; Smith et al., 2005), with the differences appearing most robust when performance approaches asymptote (Ghisletta et al., 2010). Notably, whereas perceptual-motor skills decline with time (e.g., Rodrigue et al., 2005), episodic memory improves with repeated testing, even after long delays, albeit without overcoming the initial deficit observed among older adults (Bender et al., 2015;

Daugherty et al., 2015b; Persson et al., 2016). Based upon extant studies of canonical procedural and episodic memory tasks, we can expect age deficits in navigation in general, but perhaps differential declines related to memory function and perceptual-motor skill acquisition. Indeed, compared to younger counterparts, older adults take longer time to reach the vMWM platform, cover greater virtual distance (Moffat and Resnick, 2002; Driscoll et al., 2003, 2005), travel more complex paths (Daugherty et al., 2015a, 2015c), commit more heading errors (Moffat and Resnick, 2002), and evidence greater asymmetry in turning behavior (Yuan et al., 2013). Declines in general cognitive ability, perceptual-motor skill, memory (see Moffat (2009) for a review) and cognitive mapping function (O'Keefe, 1990; Daugherty et al., 2015a, 2015c) are hypothesized to underlie this age-related decline in navigation, but the relative contributions of various cognitive processes and skills and their vulnerability to decline in a navigation context are unclear. Thus, the use of varied, complementary indices of behavior is essential for understanding the nature of navigation skill and age-related changes therein.

Although indices of navigation performance are not entirely independent, their differential neural correlates and distinct patterns of age differences suggest a degree of specificity. Thus, multiple indices of performance complement each other and yield a richer understanding of human navigation skill than is obtained from examining single measures. In the context of vMWM, such complementary assessment is exemplified by using measures of traveled distance and time from the start to goal in conjunction with quantifying the degree of meandering along the search path via its fractal dimensionality (Daugherty et al., 2015a). Although all these measures characterize efficiency of the vMWM navigation and are correlated to some extent, they show different patterns of skill acquisition and age deficits therein (Daugherty et al., 2015a, 2015c). These differences highlight a critical feature of human navigation skill: optimal navigation is fast, short and does not meander, but successful navigation paths with low complexity can be either short or long. Consistently high correlations between traveled time and distance are noted across repeated vMWM trials, whereas path complexity becomes associated with these only when performance approaches asymptote (Daugherty et al., 2015a, 2015c).

Different indices of navigation appear to vary with respect to their structural brain substrates. Persistently complex paths during acquisition are associated with smaller hippocampal and parahippocampal gyrus volumes (Daugherty et al., 2015a, 2015c), more specifically the subiculum and entorhinal cortex (a part of the parahippocampal gyrus). These medial temporal lobe structures are putative neural substrates of cognitive mapping in episodic memory (O'Keefe and Nadel, 1978; O'Keefe, 1990, 1991; see Maguire et al. (1999) for a review) and more complex paths in the vMWM suggest an impoverished cognitive map (Kirasic et al., 1992; Moffat and Resnick, 2002; Daugherty et al., 2015a). Hippocampal and parahippocampal cortex volumes also correlate with navigation time and distance, but unlike path complexity, the neural substrates of those navigation indices include in addition striatum, prefrontal cortex, and cerebellum (Rondi-Reig and Burguiere, 2005; Moffat et al., 2007; Moffat, 2009; Daugherty et al., 2015a; Korthauer et al., 2015). Spatial navigation skill relies on more than cognitive mapping, and based upon the known functional specialization of these regions, it appears that distance and time of search in part reflect decision and perceptual-motor skill components.

In late adulthood, the relationship between regional brain volumes and navigation performance can be affected by the burden of age-related vascular and metabolic risk factors (see Raz and Rodrigue (2006) and Jagust (2013) for reviews) that in turn may be related to dysfunctional cellular processes. Declining cardiovascular

health, even in otherwise healthy adults, is associated with impaired cognitive spatial learning (Barak et al., 2015) and cardiovascular risk factors, such as increased pulse pressure (Waldstein et al., 2005; Bender and Raz, 2012), have been linked to age-related deficits in episodic memory and executive functions. The impact of vascular risk is considered to reflect multifarious molecular changes brought by oxidative stress (Dröge and Schipper, 2007) and chronic neuroinflammation (Finch et al., 1969; Finch and Crimmins, 2004; Grammas, 2011). These processes are difficult to assess directly in human brains, but their proxies, such as age-related increase in brain iron content, can be measured via MRI techniques—e.g., R2* relaxometry (Daugherty and Raz, 2015). Although the literature on longitudinal changes in brain iron is sparse, at least two studies showed that increase in striatal iron content predicts its shrinkage in healthy adults after two (Daugherty et al., 2015b) and up to seven years (Daugherty and Raz, 2016). Moreover, caudate iron accumulation, not shrinkage, over two years explained poorer cognitive performance in older adults (Daugherty et al., 2015b). Of note, hippocampal and caudate iron content is greater in older adults who display elevated vascular (Rodrigue et al., 2011) and metabolic risk factors (Daugherty et al., 2015b). Yet, the combined effects of iron accumulation and cardiovascular health on navigation skill are unknown. To date, iron accumulation as a correlate of declining navigation ability has been explored only in animal models (Maaroufi et al., 2009), although age-related deficits in other cognitive and perceptual motor abilities have been linked to greater iron content in the hippocampus (Rodrigue et al., 2012) and striatum (Adamo et al., 2014) in healthy humans. Thus, indices of regional brain iron content and pulse pressure may contribute to elucidating mechanisms of age-related deficit in navigation performance.

To the best of our knowledge, there are no longitudinal studies of navigation skill acquisition, trial-by-trial and therefore the pattern and magnitude of change during aging in that domain is unknown. However, two recent studies reported change in aggregate mean performance. One examined average time of navigation in vMWM at two occasions eight years apart and found no evidence of change while it replicated commonly reported cross-sectional age differences in travel time and its correlations with hippocampal and prefrontal cortical volumes (Korthauer et al., 2015). Another longitudinal study that included virtual navigation training accompanied by treadmill walking reported greater improvement in finding target locations by older adults compared to their younger counterparts, in spite of better performance by the latter (Lövdén et al., 2012). However, without analysis of acquisition curves in these small-sample studies, the possible longitudinal stability of navigation behavior and the source of age-related differences are difficult to assess. The above-mentioned evidence notwithstanding, the current reliance on cross-sectional designs precludes assessing age-related change in navigation skill and, in particular, evaluating its potential neural mediators (Maxwell and Cole, 2007; Lindenberger et al., 2011; Raz and Lindenberger, 2011).

To address the outlined shortcomings, we examined changes in two aspects of navigation skill acquisition (path length and complexity) across repeated trials within a baseline session and at a two-years follow-up in a sizeable lifespan sample of healthy adults. Based upon the existing literature, we hypothesized that volume and iron content of several brain regions that were previously identified as substrates of spatial navigation will be associated with change in navigation skill. The selected regions were caudate nucleus (Cd), cerebellum (Cb), hippocampus (Hc), parahippocampal gyrus (PHG), and the lateral prefrontal cortex (LPFC), all showing robust shrinkage in normal aging (Raz et al., 2005; Fjell et al., 2009; Tamnes et al., 2013; Persson et al., 2014; see Kennedy and Raz (2015) for a review) and partially accounting for change in perceptual-motor skill and memory functions (Rodrigue

et al., 2005; Daugherty et al., 2015b; Persson et al., 2016). A multi-level structural equation model tested longitudinal change in non-linear acquisition curves, brain-cognition relationships, and the impact of age, sex, and vascular risk factors by introducing appropriate covariates. We hypothesized that advanced age and elevated vascular risk factors account for greater declines in the brain and navigation.

2. Materials and methods

2.1. Participants

A sample of 213 (68% female; 66% Caucasian) healthy, community-dwelling adults were recruited from the Metro Detroit area as part of a long-term longitudinal study. The study was approved by the Institutional Review Board and all participants provided informed consent. Participants (age 18–77 years at baseline) were assessed at baseline and at follow-up after a 2 year delay (average delay = 26.88 months, $SD = 3.67$). See Table 1 for a demographic profile of the sample. The exclusion criteria were neurological and cardiovascular pathology, thyroid disorder, endocrine disease, psychiatric disease, drug and alcohol abuse, and head injury with loss of consciousness for 5 min or longer. Participants reported right-hand dominance (Edinburgh Handedness Questionnaire; Oldfield, 1971) and were screened for vision and hearing problems at each assessment. To screen possible depressive symptomatology, we used a cut-off of 16 on the Center for Epidemiologic Study depression scale (CES-D; Radloff, 1977) and for dementia screening we applied a cut-off of 26 on the mini-mental state examination (MMSE; Folstein et al., 1975) at enrollment. In addition to the sample included in these analyses, 48 individuals participated in the study but did not complete the vMWM task, and 19 individuals were enrolled in the study but were dropped retrospectively for violations to qualifying criteria identified after the fact or incidental MRI findings.

Of 213 individuals who completed the vMWM task at baseline, 131 returned after approximately 2 years for repeated testing. We have previously reported on the navigation performance in the cross-sectional baseline sample (Daugherty et al. 2015a). The participants who did not return for follow-up were younger ($M = 47.40$, $SD = 16.93$; $t(211) = 2.93$, $p < 0.01$) and had a lower MMSE score ($M = 28.61$, $SD = 1.17$; $t(211) = 2.27$, $p = 0.02$) at baseline, but did not differ from persons who returned for the follow-

up in education level ($t(211) = 0.04$, $p = 0.97$), CES-D ($t(211) = -1.56$, $p = 0.12$), and average systolic ($t(211) = -0.73$, $p = 0.47$) and diastolic ($t(211) = -0.09$, $p = 0.93$) blood pressure. In line with the recommended practice (McArdle, 2009), the total sample was included in analyses and missing data were handled via full information maximum likelihood (FIML), a method that utilizes all available data to optimize estimation without imputation (Muthén et al., 1987; Larsen, 2011), and thereby introduces negligible bias in estimates of longitudinal change with the assumption of data missing at random. Because data were not missing completely at random, baseline age and MMSE were included as covariates in latent models to further reduce attrition bias. A subset of the larger sample had MRI scans for volumetry ($n = 141$ at baseline; $n = 82$ returned at follow-up) and $R2^*$ relaxometry ($n = 146$ at baseline; $n = 99$ at follow-up). We have reported longitudinal change in regional volumes (Persson et al., 2014) and $R2^*$ (Daugherty et al., 2015b) in a large part of this sample before. In this report, uniquely in application to the study of navigation skill, the neuroimaging data were added to the behavioral analyses, estimating missing cases via FIML. Thus the larger sample size of 213 individuals was maintained in all analyses.

In the course of the longitudinal study, 16 individuals developed conditions that would have been exclusionary at baseline: 1 was diagnosed with diabetes, 4 with cardiovascular and cerebrovascular conditions (e.g., heart attack, mitral valve prolapse, transient ischemic attack), 2 with cancer, 2 with an endocrine disorder, 4 participants scored above the 16 point on CES-D, and 3 had MMSE scores of 25. In order to maintain the sample size, these 16 cases were kept in the primary analyses, and later removed to examine if development of health conditions alone could account for the effects.

The inclusion criteria allowed for having a diagnosis of hypertension and taking antihypertensive and cholesterol lowering medications. By the last assessment 45 persons were taking at least one such medication: 12 were on an ACE inhibitor, 11 on a beta-blocker, 10 on an angiotensin II receptor blockers, 21 on diuretics, 9 on calcium channel blockers, and 16 on statins. At each occasion, systolic and diastolic blood pressure (in mm Hg) was measured at the beginning of three different testing sessions via a mercury sphygmomanometer (BMS 12-S25) with a standard blood pressure cuff (Omron Professional) on the left arm with the participant seated in a quiet room. Although the sample was, on average, normotensive, in the course of study, as many as 40 participants had clinically diagnosed or observed hypertension (blood pressure measurement at least 140 mm Hg systolic or 90 mmHg diastolic; see Table 1). Pulse pressure was calculated from measured blood pressure and, on average, did not change in the course of study ($t = -0.89$, $p = 0.37$). Thus, the three baseline measures were used to identify a latent pulse pressure construct to include as a covariate in analyses.

2.2. Virtual Morris water maze testing procedure and indices

We used a computerized vMWM developed by Moffat and Resnick (2002). A detailed description of the virtual environment and testing procedures can be found in previous publications (Moffat et al., 2007; Daugherty et al., 2015a; see Fig. 1). The same task performed at baseline was repeated, unaltered, at follow-up. Briefly, prior to testing, participants were introduced to two unique virtual environments and trained to control the joystick for movement. For practice, each participant completed six trials with a stationary platform that alternated between hidden and visible in a circular pool with objects around the periphery. Participants then completed six learning trials in a unique environment with a similar format as the practice, except that the stationary platform was always hidden. There was no time limit and a trial terminated

Table 1
Sample demographic profile at baseline and follow-up 2 years later.

	Baseline	Follow-up
N	213	131
% Female	68%	74%
% Caucasian	66%	73%
Age (years)	51.27 ± 15.52	55.74 ± 14.28
Education (years)	15.42 ± 2.30	15.74 ± 2.33
MMSE	28.88 ± 1.09	28.77 ± 1.18
CES-D	4.41 ± 3.91	5.04 ± 5.89
N of Hypertensive Participants	40	28
Average systolic (mmHg)	121.83 ± 12.80	121.99 ± 14.11
Average diastolic (mmHg)	75.32 ± 7.67	75.23 ± 6.83

Note: Sample means and standard deviations are reported. The total sample was included in latent modeling and missing data at the longitudinal follow-up were handled via estimation with full-information maximum likelihood (see Section 2 for details). Average systolic and diastolic blood pressure values were calculated from three measurements taken on separate occasions prior to cognitive testing. Hypertension was classified from either current diagnosis, current anti-hypertensive medication use, or observed average systolic pressure ≥ 140 mmHg, or average diastolic pressure ≥ 90 mm Hg. MMSE—mini-mental state exam, criterion > 25 ; CES-D—Center for Epidemiological Study-Depression scale, criterion < 16 .

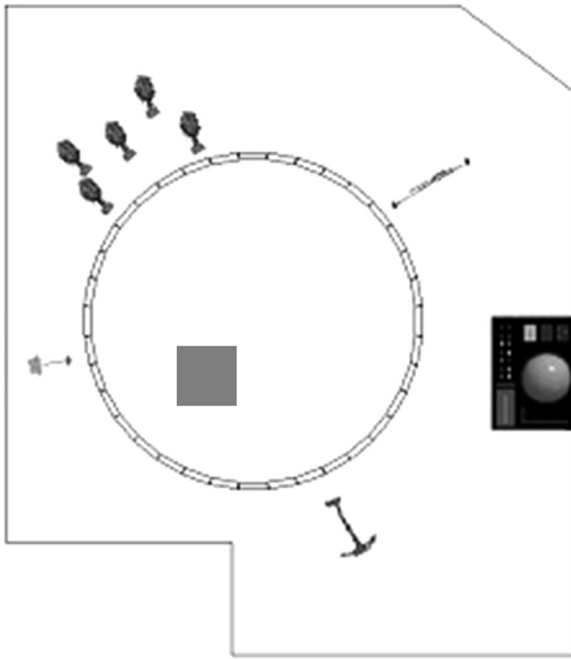


Fig. 1. An over-head diagram of the virtual Morris water maze. The location of the platform is illustrated by a square. The platform was stationary and hidden for all learning trials. The task was adopted from [Moffat and Resnick \(2002\)](#).

when the participant crossed the platform. Upon crossing the platform, it became visible and rose above the water while a sound played. There were three unique starting positions in the quadrants without the platform (trials 1–3), which were repeated in the same order on the last three trials (trials 4–6). Distance traveled from the start to the first intersection of the platform was recorded as path length. The starting locations were not equidistant to the platform, and to allow for direct comparison across trials with different starting locations, the minimum linear distance between each starting location and the platform was subtracted from navigation distances. Navigation latency, i.e. time required for finding the platform, is another popular metric of navigation efficiency. However, in this version of the vMWM, time-to-platform can be confounded by delay in initiating movement at the start of trial, potentially due to individual variability in joystick control and attention to the start of a new trial. Thus, we only report distance measures that are more robust to these sources of error. In addition to path length, path complexity was measured via fractal dimension using the x–y coordinate output (sampled at 100 Hz) calculated with the Fractal Mean function freeing estimation parameters in Fractal (v 5.20) software developed by Nams (Dalhousie University, <http://www.dal.ca/faculty/agriculture/environmental-sciences/faculty-staff/our-faculty/vilis-nams/fractal.html>; last accessed 09/08/2016). Fractal dimension for the path line in a two-dimensional plane was calculated by 1-slope of the function defined by change in $\log(\text{distance})$ and $\log(\text{spatial scale})$ (Nams, 2006) and falls on a continuum between 1 (low complexity) and 2 (highest complexity). At the end of testing, participants completed a visible platform trial for which travel time was used as a covariate to control for individual differences in joystick use (Moffat and Resnick, 2002). In addition, participants completed a questionnaire about prior experience with three-dimensional video games (7-point rating: 1—Never; 7—Daily).

Immediately before the visual platform trial, participants completed a 1-minute probe trial without the platform. Performance on the probe trial is not included in reported analyses because path length and complexity cannot be interpreted in the same manner as when the platform is actually present during

learning trials. Alternate indices are commonly used in probe trials but a single measure is insufficient for latent longitudinal change models, which is the gold standard for such analyses (McArdle, 2009) and the approach taken here.

2.3. Volumetry

All imaging was collected on the same 4T Varian Scanner (Bruker Biospin, Ettlingen, Germany) equipped with Siemens interface. A T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence was acquired for volumetry with the following parameters: $.67 \times .67 \times 1.34 \text{ mm}^3$ voxel, echo time (TE)=4.38 ms, repetition time (TR)=1600 ms, inversion time=800 ms, GRAPPA acceleration factor=2; field of view (FOV)= $256 \times 256 \text{ mm}^2$, matrix size=384 × 384, and flip angle (FA)=8°. With Analyze software (v10 and later; Biomedical Imaging Resource, Mayo Clinic College of Medicine), seven expert raters manually traced the Cd, Hc, PHG, Cb and LPFC using a stylus on a 21-inch digitizing tablet (Wacom Cintiq). All operators attained among them high inter-rater reliability measured via an ICC (2) formula that assumes random raters (Shrout and Fleiss, 1979). Across rater pairs and regions, ICC(2) of at least 0.90 was attained. In addition to regional measures, intracranial volume (ICV) was estimated through manual morphometry and used in adjustment of the regional volumes for head size and possible effects of changes in scanner-related variables over the two-year hiatus. See our previous publications (Raz et al. 2004; Persson et al. 2014) for a description of tracing procedures. To guard against rater drift and to ensure raters were blind to time in the study, both occasions of MRI data were traced randomly assigned across measurement occasions. Regional measures were corrected for head size via analysis of covariance (Jack et al., 1989; Raz et al., 2004).

2.4. $R2^*$ Relaxometry

Iron content was estimated via $R2^*$ relaxometry, a method that is based upon the strong paramagnetic properties of iron-containing particles (see Wood et al. (2004), Gurge et al. (2005) and Haacke et al. (2005)). $R2^*$ is calculated as the inverse of $T2^*$ ($=1/R2^*$), which we estimated here with a multi-echo susceptibility weighted image (SWI) sequence. The multi-echo SWI was acquired in the axial plane with the following parameters: 1 mm^3 isotropic voxel; four TE=7.53–30 ms with an inter-echo interval of 7.5 ms; TR=35 ms; FA=25°; bandwidth=200 Hz; FOV= $256 \times 208 \text{ mm}^2$. Noise was excluded from all measurements by applying a threshold that selectively isolated intensity values corresponding to image inhomogeneity. All images were pre-processed in Signal Processing in NMR (SPIN) in-house written software available at <http://mrinnovations.com/index.php?site=spin> (last accessed 09/08/2016). Boundaries of the Cd and Hc were manually traced with a stylus on a 21-inch digitizing tablet (Wacom Cintiq) by a single rater (A.M.D); see Daugherty et al. (2015b) for a description of processing and tracing procedures. $T2^*$ is biased by the presence of myelinated fibers (Fukunaga et al., 2010; Langkammer et al., 2012; Lodygensky et al., 2012) and interpretation as an index of iron content is confounded in regions including myelination, thus $T2^*$ was not measured in PHG, LPFC or Cb. The $T2^*$ images from both assessments were processed simultaneously, randomly assigned across time points. Rater reliability was confirmed for both regions with an intraclass correlation coefficient (ICC(3); Shrout and Fleiss, 1979) exceeding 0.90 in a sub-sample of 10 cases measured twice with a delay of 2 weeks. Average $T2^*$ (ms) was estimated across all slices traced of a given region and converted to $R2^*$ [s^{-1}]= $(1/T2^*) \times 1000$, for which higher values correspond to greater iron content.

2.5. Multi-level modeling: longitudinal change of latent learning curve components

We analyzed the data in a multi-level structural equation modeling framework in order to take advantage of the repeated trials within a session to estimate acquisition curves, as well as the longitudinal repeated testing after a delay for latent change scores. To determine skill acquisition, unspecified non-linear latent growth curves were fit to the data from 6 repeated trials within each session (McArdle and Anderson, 1990; Stoolmiller, 1995). This approach partially specifies the loadings of trials to identify the acquisition curve while optimizing estimation of the true non-linear function (see Supplemental Fig. 1 for an example). As fractal dimension and distance measures showed different patterns of non-linearity, the trials were fixed differently and different weights were assigned for each index of performance. These were chosen based upon inspection of the learning curves and were confirmed by models' fit indices, which were in the excellent range and comparably better than those for an alternative linear slope model. The freely estimated loadings reflect the proportional change relative to the fixed loadings—e.g., scaled to trials 2 and 3 for fractal dimension, and trials 4 and 6 for distance measures. Although the shape of the slope function cannot be interpreted, we adopted the described approach in line with the current recommendations (see Grimm et al. (2011)) as our purpose was to evaluate longitudinal change in non-linear acquisition trajectories and not to parameterize and interpret the exact type of non-linearity in the acquisition curves. In the latent growth curve specification, the first trial identified the intercept for each distance and fractal dimension acquisition curves. To enable direct comparison of the unspecified non-linear acquisition curves between occasions, the estimated factor loadings were constrained to be equal across time. To calculate longitudinal change, the acquisition curves and intercepts at baseline and follow-up were nested within latent change score models (McArdle and Hamagami, 2001; see Supplemental Fig. 1). A latent change score is conceptually similar to a difference score, yet it provides error-free estimates of change because measurement error is estimated independent of the latent construct.

A similar approach was taken to calculate longitudinal change in end-of-session performance, defined by the last three trials (see Supplemental Fig. 2). Although 6 learning trials are insufficient for reaching asymptotic levels of performance (see Daugherty et al. (2015c)), we still gain useful insights into the stability of navigation learning by examining longitudinal change in performance at the end of the task. In a latent modeling framework, several indicators optimize measurement error estimation apart from the latent construct and thus we chose to use the last three trials of performance, which correspond to the same ordered starting locations as the first three trials. The three trials identified the construct all fixed at 1 (freely estimated residuals, constrained across time), in other words, similar to an error-free average of performance at the end of the task.

Latent components of the acquisition curves that demonstrated significant variability were further assessed for possible covariates. The latter included baseline age (centered at the sample mean), a latent composite of baseline pulse pressure (measures were normed), and sex, as well as several control measures: normed baseline MMSE, performance on the visible platform trial, prior experience with three-dimensional video games, and actual delay between occasions. In addition to these, parallel latent change score models of regional volumetry and iron content ($R2^*$) were included (e.g., Hertzog et al., 2008; as implemented in Daugherty et al. (2015b)).

Some model constraints were imposed to maintain measurement invariance: the variances of behavior on each trial and repeated measures of brain volumes and iron content were constrained to be equal across time, measurement intercepts were

constrained to zero, and measures were allowed to correlate across time but were constrained to be equal across repeated measure sets (see Supplemental Fig. 1 for an example). There were a few exceptions: the variances of distance measures were free, distance intercepts on trial 1 at each occasion and trial 6 at follow-up were free, and only trial 4 was correlated across time; fractal dimension measure intercepts on trials 5 and 6 at each occasion were free; and the follow-up measure intercept of Cb volume was free. These exceptions were necessary to achieve acceptable model fit. Prior to modeling, distance measures were log transformed to address severe skew but fractal dimension measures did not require correction. Behavioral measures were normed to baseline trial 1 means and standard deviations; all coefficients can be interpreted as proportional growth from mean performance on trial 1. Univariate outliers in brain volumes and iron were winsorized and data were normed to baseline.

2.5.1. Model fit and reverse-effects model evaluation

Model fit was determined by a set of customary indices (Raykov and Marcoulides, 2006): normal theory weighted chi-square statistic (a non-significant value indicates good fit); root mean square error of approximation ($RMSEA \leq 0.10$ indicates good fit); comparative fit index ($CFI \geq 0.90$ indicates excellent fit); standardized root mean square residual ($SRMR \leq 0.10$ supports good fit), or for models with categorical variables, weighted root mean square residual ($WRMR \leq 0.80$ supports good fit). To ensure reliable model specification, we compared the non-linear slope functions to a linear slope. This comparison was made with the reported fit indices, as well as nested model comparisons with Akaike information (AIC) and sample-size adjusted Bayesian information criteria (BIC), for which comparatively lower values indicate improved model fit.

To optimize model fit during hypothesis testing, final models constrained non-significant effects that did not serve as necessary control covariates (e.g., non-significant correlations with delay between assessments were maintained in the models). Unless otherwise stated, all reported effects are standardized. To avoid spurious results due to a relatively small sample size and undue influence of extreme observations, models were bootstrapped with bias correction (5000 iterations of the whole sample; Hayes and Scharkow, 2013) to estimate 95% confidence intervals (BS 95% CI) of unstandardized effects. Finally, the relatively small sample size required separate models for distance and fractal dimension measures and for each brain region, and a Bonferroni correction (α') was used to adjust for multiple comparisons. To test the direction of effects, paths in final models were reversed to have behavior predict neural correlates. Comparison of model fit and effects' significance provides evidence for choosing the direction of effects.

3. Results

3.1. Longitudinal measurement invariance

We assessed the consistency of measures over the two year hiatus outside of the latent model with Pearson correlations. Path length ($r=0.07-0.36$) and complexity ($r=0.05-0.27$) had low test-retest consistency over two years. Volume measures had excellent consistency across regions in both hemispheres from $r \geq 0.88$ for PHG to $r \geq 0.97$ for LPFC, and $R2^*$ measures were moderately consistent in the Cd (for both hemispheres, $r \geq 0.58$) and less in the Hc (for both hemispheres, $r \geq 0.31$). Constraints in the models to ensure metric invariance accounted for the lack of consistency and removed any bias from estimates of longitudinal change and individual differences therein.

3.2. Longitudinal change in navigation performance

3.2.1. Acquisition slope specification

Prior to constructing change score models, acquisition curves and intercepts, and composite performance on the last three trials (end-of-session performance) for path length and complexity were examined at each occasion. The latent growth curve models of distance ($\chi^2=70.31$, $p=0.13$; CFI=0.96; RMSEA=0.03; SRMR=0.07) and FD ($\chi^2=64.32$, $p=0.64$; CFI=1.00; RMSEA=0.00; SRMR=0.07) each had excellent fit. The non-linear acquisition curves for distance had factor loadings for trials 1–6, respectively (unstandardized loadings, * designates an estimated coefficient that was constrained across time): 0, 0.35* ($p=0.52$), 4.44* ($p < 0.001$), 2, 4.53* ($p < 0.001$), 6; and for fractal dimension: 0, 1, 2, 3.56* ($p < 0.001$), 1.86* ($p < 0.001$), 0.84* ($p=0.04$). The first trial value identified the intercept of the acquisition curve and the estimated loadings of subsequent trials can be interpreted as proportional, additive change scaled to the fixed loadings. To confirm that the fitted non-linear slope models reliably reproduced the observed data, we compared its fit to that of a linear slope model. The fit of linear slope models of distance ($\chi^2=110.17$, $p < 0.001$; CFI=0.82; RMSEA=0.06; SRMR=0.08) and FD measures ($\chi^2=90.45$, $p=0.07$; CFI=0.90; RMSEA=0.04; SRMR=0.08) was notably worse than of the non-linear slope models. Moreover, the nested model comparisons of distance (AIC=7849.62, BIC=7855.78 vs. AIC=7883.48, BIC=7889.06, non-linear and linear, respectively) and FD measures (AIC=6651.54, BIC=6655.58 vs. AIC=6671.67, BIC=6675.13, respectively) confirmed that the non-linear function reproduced the observed data with a high degree of reliability and better than the linear function.

At baseline and follow-up, the observed shortening of the path distance and reduction in fractal dimension were consistent with learning (Table 2; Fig. 2A and 2B). However, comparison of the acquisition curves between occasions revealed shallower acquisition curves for distance-to-goal at follow-up: unstandardized mean change=0.13, $p=0.009$, $\alpha'=0.03$; BS 95% CI: 0.03/0.23; Fig. 2A2. Thus, a reduced rate of learning, as measured by path shortening, was observed after a two-year hiatus. Further, the composite performance on the last three trials was worse at follow-up than at baseline (unstandardized mean change=0.76, $p < 0.001$, $\alpha'=0.03$; BS 95% CI: 0.49/1.03; Fig. 2A2), indicating

longitudinal decline also in the magnitude of improvement after repeated testing within a session.

Conversely, acquisition curves for fractal dimension at follow-up were more negative (unstandardized mean change=−0.14, $p=0.004$; BS 95% CI: −0.25/−0.07; Fig. 2B2) and the end-of-session performance composite was significantly better (unstandardized mean change=−0.52, $p < 0.001$; BS 95% CI: −0.74/−0.30; Fig. 2B2). Thus, learning measured by reduction of search path complexity was faster and evidenced greater gains after a two-year break. Fig. 3 presents an example of search paths taken from learning trial 5, which was included in the end-of-session performance composite. Intercept (corresponding to performance on the first trial) was unrelated to acquisition curves at baseline and follow-up for distance (−0.29 and 0.12, $p \geq 0.38$) and fractal dimension (−0.20, $p \geq 0.34$), and did not change longitudinally for either index (see Table 2). The models of longitudinal change in acquisition curves and end-of-session performance had excellent fit: $\chi^2 \leq 74.64$, $p \geq 0.13$; CFI ≥ 0.95 ; RMSEA ≤ 0.03 ; SRMR ≤ 0.08 .

Inspection of the acquisition curves (Fig. 2A1 and 2B1) revealed that heterogeneity of performance on trial 4 might have been increased. To confirm that analyses of change were not biased by inclusion of that trial, we re-evaluated the slopes after omitting it. Excluding trial 4 data did not alter the pattern of longitudinal change in distance and FD. Search distance reduction over trials was significantly less at follow-up compared to the baseline (unstandardized mean change=0.51, $p=0.04$; variance=0.04, $p=0.49$), whereas FD acquisition improved: unstandardized mean change=−0.14, $p=0.047$; variance=0.17, $p=0.11$. Both models fit well: $\chi^2 < 55.67$, $p > 0.08$; CFI > 0.90 ; RMSEA < 0.04 ; SRMR < 0.07 . Because performance at any trial in the vMWM acquisition may depend upon previous trials, we chose to retain trial 4 data in subsequent analyses, in spite of their anomalous appearance.

3.3. Individual differences in navigation performance change

Because individuals varied in several components of the learning curves assessed by search distance and fractal dimension of the path (see Fig. 4 and Table 2), we tested the influence of possible covariates: baseline age, pulse pressure and sex, as well as control measures. Latent change models that included covariates

Table 2

Two year longitudinal change in latent components of navigation acquisition curves and individual variability.

Latent component	Distance			Fractal dimension		
	Mean (95% CI)	Variance	d	Mean (95% CI)	Variance	d
Intercept						
Baseline	−0.09 (−0.40/0.24)	1.38*		0.02 (−0.08/0.13)	0.21*	
Follow-up	−0.33 (−0.61/0.02)	1.23*		0.10 (−0.01/0.24)	0.21*	
Change	−0.28 (−0.62/0.04)	0.54*	−0.24	0.12 (0.002/0.24)	0.03	0.26
Acquisition curve						
Baseline	−0.26* (−0.34/−0.17)	0.02		−0.07 (−0.15/−0.02)	0.07*	
Follow-up	−0.13* (−0.26/−0.05)	0.001		−0.16* (−0.26/−0.08)	0.04	
Change	0.13* (0.03/0.23)	0.001	0.92	−0.14* (−0.25/−0.07)	0.07*	−0.53
End performance						
Baseline	−1.46* (−1.67/−1.28)	1.15*		−0.22* (−0.39/−0.04)	0.38*	
Follow-up	−0.71* (−0.94/−0.49)	1.52*		−0.74* (−0.92/−0.55)	0.42*	
Change	0.76* (0.49/1.03)	0.94*	0.71	−0.52* (−0.74/−0.30)	0.07	−0.84

Note: Unstandardized latent means and variances are reported for data normed to baseline trial 1 values. Bootstrapped bias-corrected 95% confidence intervals (BS 95% CI) of unstandardized means are reported in parentheses. * indicates significance after correction for multiple comparisons, $p < 0.01$, $\alpha'=0.03$. d is a standardized effect size for latent change that can be compared between measures: $d = \text{Mean Change} / \sqrt{\text{Variance at Baseline}}$. Latent intercept corresponds to performance on task trial 1 at each occasion defined by the latent growth curves centered at that trial. Latent acquisition curves were estimated as unspecified non-linear slopes with weights constrained to be equal across time. Latent end-of-session performance was a composite of navigation measures on trials 4–6 at each occasion. (See Section 2 for more detail.)

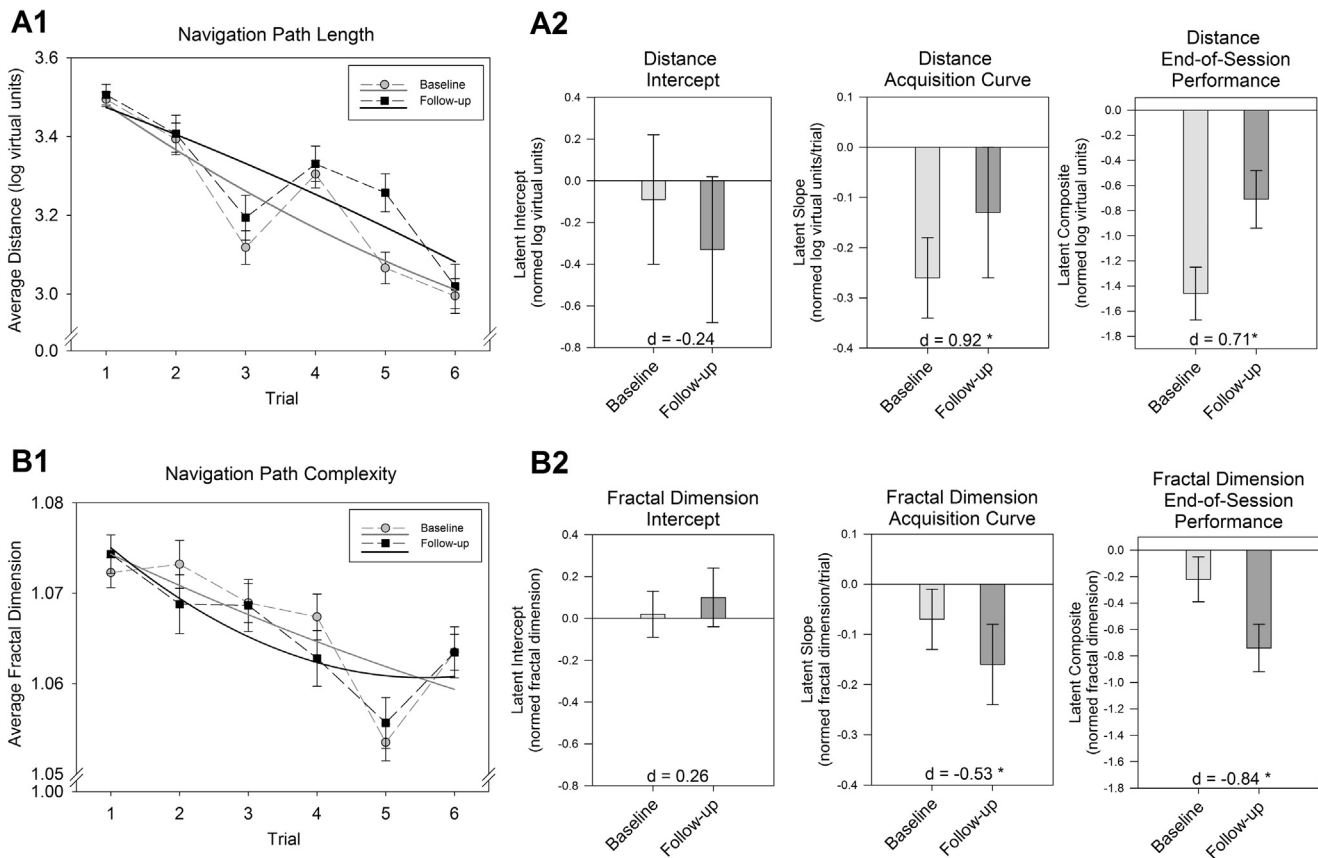


Fig. 2. Mean acquisition curves across six repeated trials within a session at baseline and follow-up after approximately 2 years. Figures **A1** (path length) and **B1** (path complexity) display the original data as sample average performance and standard error of the means for each trial, and non-linear regression lines were fit to the acquisition curves—greater distance and fractal dimension indicate worse navigation skill. The lack of complete overlap between acquisition curves (solid lines) fit to the original data at each occasion is in accord with the significant latent change scores estimated from fitted non-linear growth curves. Figures **A2** and **B2** display the latent means at each occasion and bars represent bias-corrected bootstrapped 95% confidence intervals of the three components to the acquisition curves derived from the multi-level latent change score models; for all components, greater, more positive values correspond to worse performance. d is a standardized effect size of longitudinal latent change that can be compared between measures; * indicates significant change $p < 0.01$, Bonferroni correction $\alpha = 0.03$. Intercept corresponds to performance on trial 1 as defined by the latent growth curves. Acquisition curves were defined as unspecified non-linear latent growth curves with loadings constrained to be equal between baseline and follow-up. End-of-session performance was a latent composite of performance on trials 4–6 (see Section 2 for more detail). As seen in panel **A**, path length decreased across repeated trials within a session to indicate learning, but the acquisition was slower after the longitudinal delay and end-of-session performance was worse. In panel **B**, decreasing path complexity across trials supports learning, and steeper acquisition curves and better end-of-session performance at follow-up suggest greater repeated-testing gains after exposure to the task 2 years prior.

and control measures had excellent fit: $\chi^2 \leq 184.27$, $p \geq 0.05$; CFI ≥ 0.94 ; RMSEA ≤ 0.04 ; WRMR ≤ 0.81 .

At baseline, older participants traveled longer paths on trial 1 (intercept, 0.67, $p < 0.001$; BS 95% CI: 0.50/0.83) and end-of-session performance (0.70, $p < 0.001$; BS 95% CI: 0.49/0.88), but age did not independently explain differences in longitudinal change in end-of-session performance distance measures: -0.08 , $p = 0.72$ (constrained). Greater pulse pressure was associated with longer paths at baseline trial 1 (intercept, 0.34, $p < 0.001$; BS 95% CI: 0.12/0.36), but was unrelated to differences in distance end-of-session performance (both ≤ 0.37 , $p \geq 0.13$, constrained). Regarding path complexity, older age and greater pulse pressure were associated with higher fractal dimensionality at baseline trial 1 (age = 0.68, $p < 0.001$, BS 95% CI: 0.17/0.37; pulse pressure = 0.34, $p = 0.01$, BS 95% CI: 0.02/0.19), slower acquisition curves (age = 0.25, $p = 0.047$, BS 95% CI: 0.01/0.13; pulse pressure = 0.30, $p = 0.03$, BS 95% CI: $-0.01/0.10$), and worse end-of-session performance (age = 0.58, $p < 0.001$, BS 95% CI: 0.23/0.49; pulse pressure = 0.28, $p = 0.02$, BS 95% CI: 0.06/0.43), which in turn accounted for lesser longitudinal improvement at the end of the task measured at follow-up (indirect effects age = 0.53, $p < 0.001$, BS 95% CI: 0.32/0.74; pulse pressure = 0.25, $p = 0.03$, BS 95% CI: 0.04/0.43). Increased pulse pressure partially mediated the effects of age on higher path complexity on the last three trials at follow-up (indirect

effect = 0.13, $p = 0.03$, BS 95% CI: 0.01/0.22). However, neither factor explained individual differences in longitudinal change of fractal dimensionality during the acquisition trials (-0.08 and -0.04 , $p \geq 0.58$), or in change of distance intercept (-0.09 and 0.25, $p \geq 0.14$).

Sex differences in navigation performance were limited. At baseline, men traveled shorter paths to the goal as indicated by lower intercept (-0.33 , $p < 0.001$, BS 95% CI: $-0.53/-0.15$) and showed steeper improvement (reduction) of fractal dimensionality of the path (-0.29 , $p = 0.03$, BS 95% CI: $-0.14/-0.02$). However they did not differ from women in navigation performance change and learning as assessed by either index (-0.01 and -0.08 , $p \geq 0.70$, constrained). Examining end-of-session performance revealed that, compared to women, men traveled shorter (-0.68 , $p < 0.001$, BS 95% CI: $-0.65/-0.25$) and less complex paths (-0.42 , $p = 0.04$, BS 95% CI: $-0.28/-0.04$) on the last three trials at baseline. Change rate over the two-year period did not differ between the sexes in either distance (-0.03 , $p = 0.91$, constrained) or in fractal dimensionality (-0.35 , $p = 0.42$, constrained).

Finally, previous experience with three-dimensional computer games was introduced as a nuisance covariate as younger adults reported more game experience at both occasions (correlations with age $r = -0.49$ and -0.47 , $p < 0.001$), and men reported more

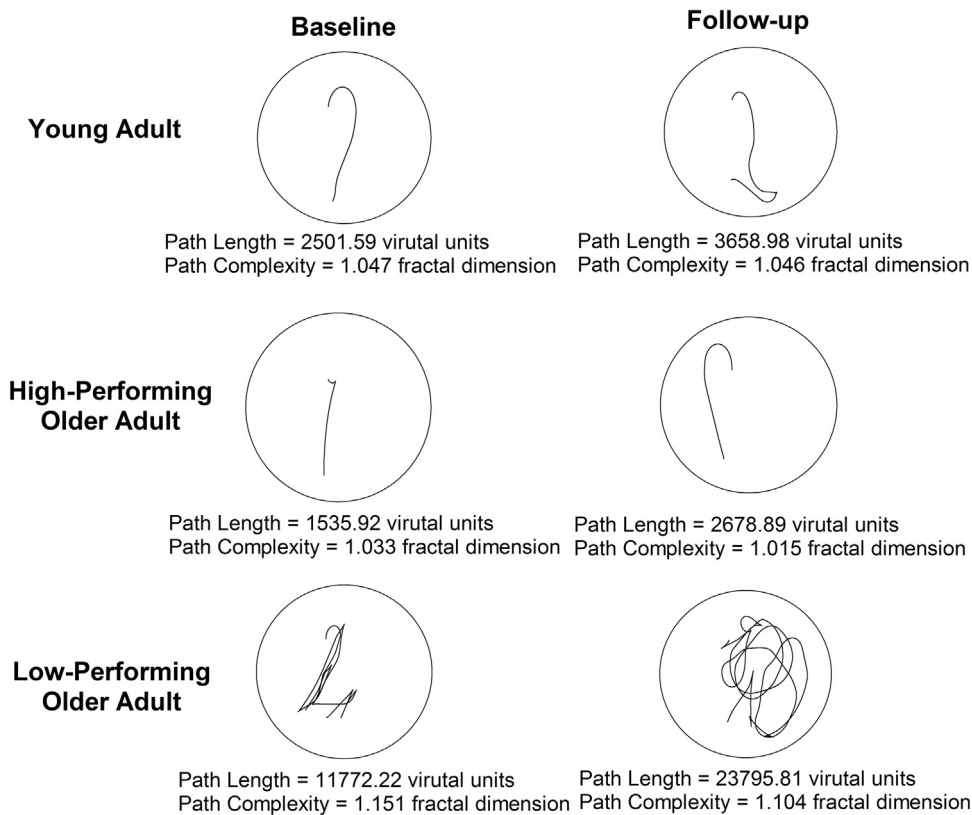


Fig. 3. Example navigation paths from task trial 5 at baseline and follow-up. In each example, path length was greater at follow-up as compared to baseline, but path complexity was either similar or improved.

experience at baseline ($t(209) = -3.51, p = 0.001$) though not at follow-up ($t(168) = -1.25, p = 0.21$). On average, reported experience did not change after a two-year delay ($t(167) = -1.27, p = 0.21$) and baseline experience was unrelated to baseline vMWM performance or change in navigation paths as measured by either distance (all $p \geq 0.28$) or fractal dimension (all $p \geq 0.09$). Baseline game experience was, nonetheless, negatively correlated with performance on the visible platform trial ($-0.22, p = 0.001$), which is a more specific control measure for the vMWM task. Thus, game experience was removed from further testing, but the visible platform trial was kept as a necessary control variable.

3.4. Neural correlates of navigation performance

We examined several putative neural correlates of navigation skill. Briefly, the Cd (unstandardized mean change = $-0.07, p < 0.001$; variance = $0.02, p < 0.001$), Hc (unstandardized mean change = $-0.10, p < 0.001$; variance = $0.05, p < 0.001$), PHG ($-0.08, p = 0.02$; variance = $0.06, p < 0.001$), and Cb (unstandardized mean change = $-0.20, p < 0.001$; variance = $0.17, p < 0.001$) significantly shrank after two years, but LPFC did not (unstandardized mean change = $0.01, p = 0.65$; variance = $0.02, p = 0.001$). Notably, all regions exhibited significant individual variability in change. In addition, estimated iron content increased in the Cd (unstandardized mean change = $0.20, p = 0.01$; variance = $0.47, p < 0.001$) but not in the Hc (unstandardized mean change = $-0.09, p = 0.45$; variance = $1.12, p < 0.001$) and individuals varied in longitudinal change in both regions. These factors were added to multi-level models of navigation performance indices to predict change in acquisition curve components and individual differences therein. In these models, sex as a categorical variable was unrelated to volume and iron measures (all $p \geq 0.07$) and its inclusion compromised model fit. Therefore, it

was removed from these analyses, but the other covariates and control measures were retained.

3.4.1. Search path length (distance traveled)

Latent change score models of distance acquisition curves and end-of-session performance that included Cd volume and iron, as well as Cb and LPFC volumes had good-to-excellent fit: $\chi^2 = 171.75-444.08, p = 0.001-0.03$; CFI = $0.94-0.98$; RMSEA = $0.03-0.046$; SRMR = $0.07-0.09$. Larger Cd volume (intercept, $-0.46, p = 0.001, \alpha' = 0.01$; BS 95% CI: $-0.63/-0.29$; Fig. 5) and lesser estimated iron content (intercept, $0.23, p = 0.02, \alpha' = 0.01$; BS 95% CI: $0.03/0.44$) were associated with shorter distance of search at baseline trial 1, although the latter effect did not survive correction for multiple comparisons. Better baseline end-of-session performance was nominally associated with larger Cd volume ($-0.24, p = 0.03, \alpha' = 0.01$; BS 95% CI: $-0.48/-0.07$; Fig. 5) and when taking this into account, worse performance at baseline explained greater longitudinal decline ($-0.36, p < 0.001, \alpha' = 0.01$; BS 95% CI: $-0.49/-0.18$)—those who demonstrated a deficit at baseline were more vulnerable to further decline longitudinally. A trend for greater longitudinal iron accumulation predicting greater longitudinal decline in end-of-session performance ($0.25, p = 0.06, \alpha' = 0.01$; BS 95% CI: $0.11/0.80$) was supported by the bootstrapped confidence intervals. Further, smaller Cd volume ($-0.59, p < 0.001$) and greater iron content ($0.54, p < 0.001$) in older age together explained greater age-related decline in end-of-session performance (cumulative indirect effect = $-0.11, p = 0.01$; BS 95% CI: $-0.22/-0.06$).

Smaller Cb volume directly predicted greater longitudinal decline in the end-of-session performance ($-0.43, p = 0.01, \alpha' = 0.01$; BS 95% CI: $-0.51/-0.13$), whereas smaller LPFC volume was associated with longer distance traveled at baseline trial 1 (intercept, $-0.64, p = 0.001, \alpha' = 0.01$; BS 95% CI: $-0.80/-0.46$) and end-of-session performance ($-0.32, p = 0.007, \alpha' = 0.01$; BS 95% CI: $-0.61/$

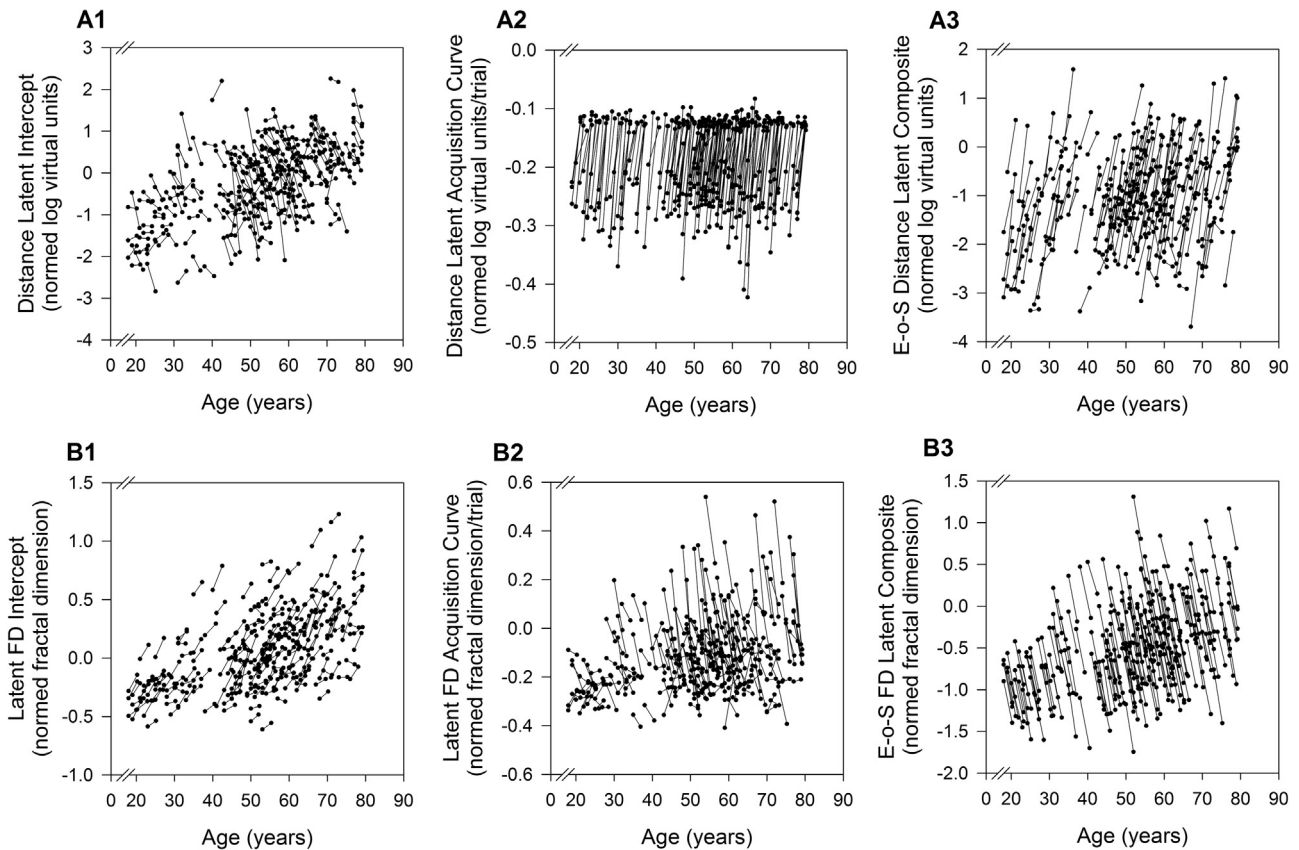


Fig. 4. Individual longitudinal change trajectories between baseline and follow-up for path length (distance traveled) and complexity (fractal dimension, FD) in each of the three acquisition curve components: intercept (corresponding to task trial 1), acquisition curve (non-linear change across repeated trials within a session), and end-of-session (E-o-S) performance (composite of trials 4–6). Values are latent estimates derived from the multi-level latent change score models. For all components, greater, more positive values correspond to worse performance. Individuals significantly varied ($p < 0.01$; $\alpha = 0.03$) in longitudinal change in all components, except for distance acquisition curve and fractal dimension end-of-session performance.

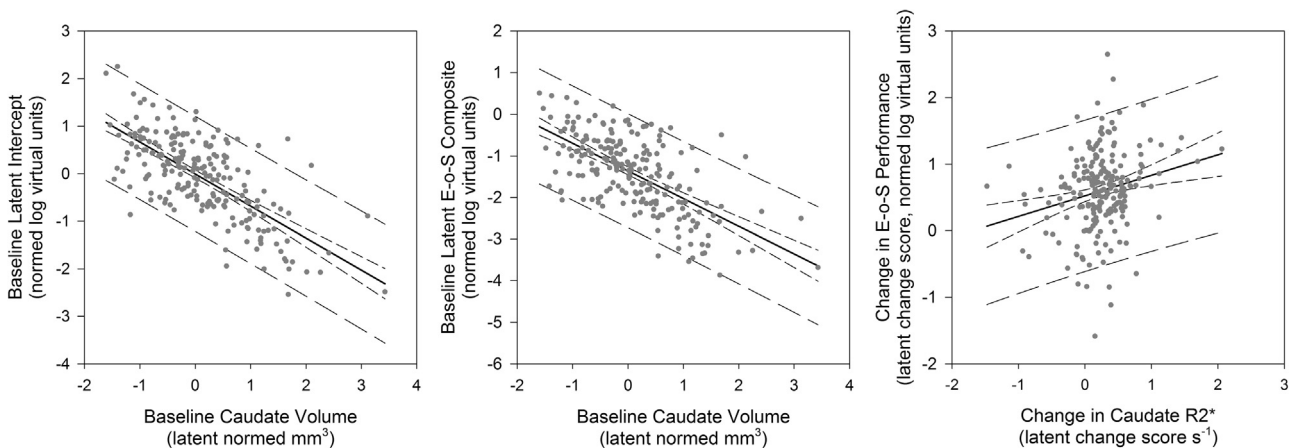


Fig. 5. Small caudate volume and greater iron content predicted longer paths at baseline and greater longitudinal decline in distance traveled. Values are latent estimates derived from the final multi-level latent change score models that included baseline and longitudinal change in caudate volume and iron, as well as baseline age, pulse pressure, and control variables. A regression line was fit to the scatter plots, including prediction intervals (long dash line) and 95% confidence intervals (short dash line). Greater, more positive values correspond to larger volumes, greater increase in iron content (change in $R2^*$), and longer paths. Individuals with smaller caudate volume at baseline traveled longer paths on trial 1 (i.e., intercept; $p = 0.001$) and at end-of-session (E-o-S; $p = 0.03$). Individuals who demonstrated greater increase in iron content across 2 years showed a trend for greater longitudinal increase in end-of-session performance path length (i.e., greater decline in navigation skill; $p = 0.06$).

–0.11). Further, smaller LPFC volume mediated the effects of age on worse end-of-session performance at baseline (indirect effect = 0.21, $p = 0.01$, BS 95% CI: 0.07/0.44) and follow-up (0.16, $p = 0.01$, BS 95% CI: 0.07/0.44). Hc volume, Hc iron content and PHG volume were unrelated to distance measures (see Table 3, effects constrained). Final models of medial temporal lobe

structures had acceptable fit: $\chi^2 \leq 391.58$, $p \geq 0.01$; CFI ≥ 0.96 ; RMSEA ≤ 0.03 ; SRMR ≤ 0.08 .

3.4.2. Path complexity (fractal dimension)

Whereas path length was unrelated to measures of the medial temporal lobe structures, volume and iron content of these regions

Table 3
Neural correlates of baseline and change in latent components of navigation acquisition curves.

Region	Baseline index	Distance				Fractal dimension				
		Intercept		End performance		Intercept	Acquisition curve		End performance	
		Baseline	Change	Baseline	Change		Baseline	Baseline	Change	Baseline
Cb	Volume	-0.04 (0.71)	-0.18 (0.16)	0.03 (0.81)	-0.43 (0.01)	0.08 (0.56)	-0.05 (0.68)	-0.14 (0.33)	-0.04 (0.70)	<i>-0.04 (0.70)</i>
Cd	Iron	0.23 (0.02)	-0.21 (0.24)	-0.01 (0.96)	<i>-0.12 (0.09)</i>	-0.15 (0.24)	0.12 (0.35)	-0.04 (0.77)	-0.11 (0.33)	<i>-0.11 (0.34)</i>
	Volume	-0.46 (0.001)	-0.15 (0.43)	-0.24 (0.03)	<i>0.09 (0.07)</i>	-0.004 (0.98)	-0.10 (0.46)	-0.12 (0.41)	-0.10 (0.36)	<i>-0.10 (0.36)</i>
Hc	Iron	-0.04 (0.69)	0.01 (0.95)	0.10 (0.43)	-0.14 (0.43)	-0.14 (0.41)	0.27 (0.03)	-0.20 (0.04)	0.29 (0.01)	0.27 (0.01)
	Volume	0.08 (0.44)	-0.13 (0.51)	0.09 (0.43)	-0.004 (0.99)	-0.06 (0.73)	0.06 (0.73)	-0.06 (0.71)	0.08 (0.45)	<i>0.08 (0.45)</i>
PHG	Volume	-0.01 (0.90)	-0.10 (0.68)	0.04 (0.78)	-0.14 (0.54)	-0.001 (0.97)	0.05 (0.09)	-0.27 (0.01)	0.08 (0.52)	<i>0.08 (0.52)</i>
LPFC	Volume	-0.64 (0.001)	-0.10 (0.59)	-0.32 (0.01)	-0.18 (0.48)	-0.39 (0.002)	-0.18 (0.22)	-0.09 (0.54)	-0.29 (0.03)	-0.20 (0.03)

Note: Effects are standardized coefficients (p-value) of the latent component for navigation learning regressed on the respective baseline neural correlate in models that included baseline age, pulse pressure and control variables. Only components that demonstrated significant variability were tested for neural correlates. Boldface effects are nominally significant, supported by bias-corrected bootstrapped 95% confidence intervals (see text) estimated in final models, with Bonferroni-corrected $\alpha=0.01$. Effects shown in italics indicate estimated indirect paths. Effects that were not nominally significant ($p > 0.05$) were constrained in the final models to improve fit. All models had good-to-excellent fit (see text). Volumes are the latent composite of the hemispheres, adjusted for intracranial volume; iron is the latent composite of R2* relaxometry from each hemisphere (see Section 2). Cb—cerebellum; Cd—caudate nucleus; Hc—hippocampus; PHG—parahippocampal gyrus; LPFC—lateral prefrontal cortex.

selectively predicted variability in path complexity. The models of effects including these regions had good-to-excellent fit: $\chi^2=154.78-384.79$, $p=0.01-0.06$; CFI=0.96–0.97; RMSEA=0.03–0.04; SRMR=0.08–0.09. Greater Hc iron content (0.27, $p=0.03$, $\alpha'=0.01$; BS 95% CI: 0.01/0.15; Fig. 6), but not volume (0.06, $p=0.73$, constrained), was a nominally significant predictor of slower learning (shallower acquisition curves) at baseline. In general, faster learning at baseline was associated with lesser improvement between the two measurement occasions (-0.74 , $p < 0.001$, $\alpha'=0.01$; BS 95% CI: $-1.37/-0.42$), possibly due to a floor effect in the decrease in path complexity across trials. Yet, despite possible greater longitudinal gains, individuals with greater Hc iron content who had shallower acquisition curves at baseline still demonstrated slower learning at follow-up (indirect effect = -0.20 , $p=0.04$; BS 95% CI: $-0.16/-0.02$). Greater Hc iron content was associated with worse end-of-session performance at baseline (0.29, $p=0.01$, $\alpha'=0.01$; BS 95% CI: 0.05/0.42) and follow-up (indirect effect = 0.27, $p=0.01$, $\alpha'=0.01$; BS 95% CI: 0.07/0.40; Fig. 6), whereas Hc volume was not (both 0.08, $p=0.52$, constrained). Older adults had greater Hc iron content (0.30, $p < 0.001$; BS 95% CI: 0.08/0.31) and higher pulse pressure (0.42,

$p < 0.001$; BS 95% CI: 0.22/0.41) at baseline, and these factors conjointly mediated the effects of age to explain worse end-of-session performance after a 2-year delay (cumulative indirect effect = 0.21, $p=0.001$, $\alpha'=0.01$; BS 95% CI: 0.08/0.23).

A mixture of effects related to PHG and LPFC volumes further distinguished variability in path complexity. Larger PHG volume was associated with greater longitudinal improvement in acquisition rate (-0.27 , $p=0.014$, $\alpha'=0.01$; BS 95% CI: $-0.21/-0.04$), but its relation to all other measures were not significant (see Table 3, effects constrained). Smaller LPFC volume was associated with more complex paths at baseline trial 1 (intercept, -0.39 , $p=0.002$, $\alpha'=0.01$; BS 95% CI: $-0.31/-0.09$), as well as worse end-of-session performance, both at baseline (-0.29 , $p=0.03$, $\alpha'=0.01$; BS 95% CI: $-0.31/-0.06$) and follow-up (indirect effect = -0.20 , $p=0.03$, $\alpha'=0.01$; BS 95% CI: $-0.34/-0.07$). Smaller LPFC volumes in older age (-0.64 , $p < 0.001$; BS 95% CI: $-0.67/-0.28$), together with elevated pulse pressure, mediated age-related decline in end-of-session performance at both occasions: cumulative indirect effect = 0.32, $p=0.001$, BS 95% CI: 0.10/0.31; and 0.31, $p=0.001$, BS 95% CI: 0.11/0.34, $\alpha'=0.01$ (baseline and follow-up, respectively). Variability in fractal dimensionality

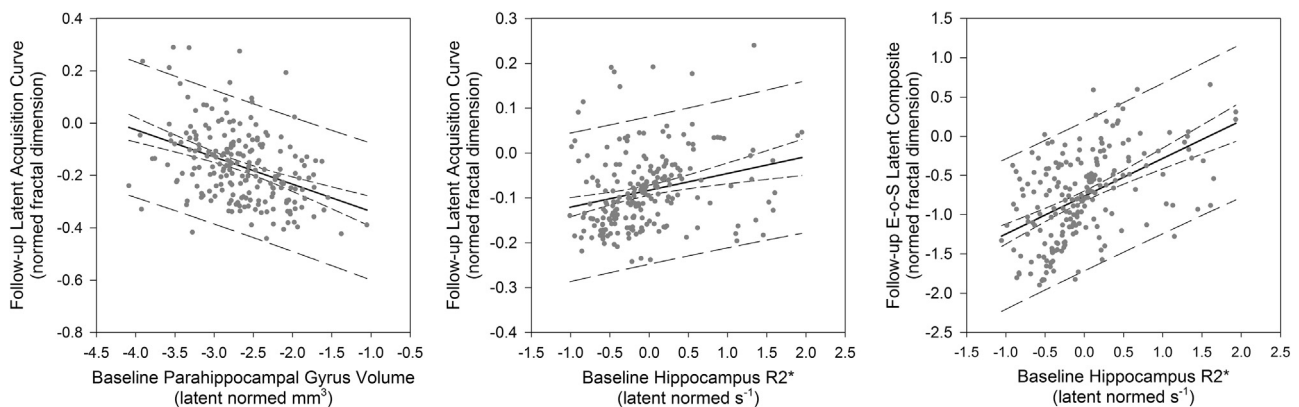


Fig. 6. Greater hippocampal iron content and smaller parahippocampal gyrus volume predicted more complex paths (fractal dimension) at baseline and lesser longitudinal improvement. Values are latent estimates derived from the final multi-level latent change score models that included baseline and longitudinal change in hippocampal volume and iron content, or parahippocampal gyrus volume, as well as baseline age, pulse pressure, and control variables. A regression line was fit to the scatter plots, including prediction intervals (long dash line) and 95% confidence intervals (short dash line). Greater, more positive values correspond to larger volumes, greater iron content (R2*), and more complex paths. Individuals with smaller parahippocampal gyrus volume ($p=0.01$) and greater hippocampal iron ($p=0.04$) at baseline showed slower (less negative) acquisition curves at follow-up. Greater hippocampal iron at baseline also predicted lesser longitudinal improvement in end-of-session performance (E-o-S; $p=0.01$).

was unrelated to Cd volume, its iron content, or Cb volume (see Table 3, effects constrained). Final models of these effects had acceptable fit: $\chi^2 \leq 426.29$, $p \geq 0.001$; CFI ≥ 0.95 ; RMSEA ≤ 0.04 ; SRMR ≤ 0.08 .

3.5. Reverse models

To test the direction of effects, paths in the final models were reversed and subsequent decline in model fit or non-significant effects provided further evidence for the direction of effects reported. Reverse models had incrementally worse fit: $\chi^2 = 175.33$ – 417.75 , $p = 0.001$ – 0.03 ; CFI = 0.95 – 0.97 ; RMSEA = 0.03 – 0.05 ; SRMR = 0.07 – 0.10 . All reversed effects were not significant (all = -0.19 – -0.27 , $p \geq 0.08$), except for LPFC volume regressed on fractal dimension measured from trial 1 (intercept, -0.36 , $p = 0.01$) and end-of-session performance (-0.27 , $p = 0.02$) at baseline, and the direction of these effects should be interpreted with caution. Finally, all models were estimated again after excluding the 16 individuals who had developed exclusionary health conditions in the course of study. All effects remained nominally significant ($p \leq 0.04$), except for LPFC predicting distance intercept (-0.13 , $p = 0.27$, BS 95% CI: $-0.38/0.09$) and change in Cd iron content predicting change in distance end-of-session performance (0.20 , $p = 0.19$; BS 95% CI: $-0.009/0.73$) but the bootstrapped confidence intervals of the latter only slightly overlap with zero and thus the test may be underpowered.

4. Discussion

In the first longitudinal study of its kind, we provide evidence of differential cognitive and neural changes that underlie deficits in navigation skill acquisition in older age. Although learning was reliably observed in both performance indices (search path length and complexity) at baseline and follow-up, the end-of-session gains from baseline were not maintained over a two-year hiatus. Notably, performance on the first trial at follow-up was the same as at baseline, as if participants had never performed the task; yet, individuals varied in longitudinal change in trial 1 path length, but not path complexity. Examination of longitudinal change in acquisition curve components revealed a clear dissociation between the indices of performance. Whereas the degree of meandering in search of the goal was significantly better at follow-up than at baseline, the search path lengthened after the two year hiatus. Thus, lesser path complexity did not translate into shorter distance to the goal. Greater caudate iron content and smaller volume, as well as smaller Cb volume predicted worse navigation performance indexed by distance, whereas greater Hc iron and smaller PHG volume were linked to greater meandering. Dissociation between navigation path length and complexity in their putative neural correlates suggests two complementary cognitive processes underlying navigation ability: perceptual-motor skill and cognitive mapping in episodic memory.

Path length from the start to the hidden goal may be construed as an index of perceptual-motor skill that, while readily acquired within session, worsened after the two-year delay. This pattern of intact learning combined with poor retention of gains after two years is similar to a pattern observed on classic perceptual-motor skill learning tasks in samples selected according to similar criteria (Kennedy and Raz, 2005; Rodrigue et al., 2005), and is in accord with commonly reported age deficits in acquisition (Wright and Payne, 1985; Smith et al., 2005; Seidler, 2006; Janacsek et al., 2012; Coats et al., 2013) and retention of the task-relevant skills after long delays (Fisk et al., 1994). Deficits indexed by navigation distance were conjointly explained by smaller volume and greater iron content of the caudate nuclei, as well as smaller volume of

cerebellar hemispheres—regions commonly implicated in perceptual-motor skill acquisition and maintenance (e.g., Raz et al., 2000; Ungerleider et al., 2002; Cavaco et al., 2011) as well as spatial navigation (Rondi-Reig and Burguiere, 2005; Moffat, 2009). This is partly consistent with our previous cross-sectional report of correlation between navigation time and Cb volume (Daugherty et al., 2015a), although in that study, Cd volume was unrelated to navigation efficiency. Thus, path length appears to capture the perceptual-motor skill learning component of navigation that is vulnerable to decline, especially in older age.

Whereas distance may be an index of perceptual-motor skill, path complexity operationalized as fractal dimensionality appears to be more closely related to cognitive mapping in episodic memory function. Path fractal dimensionality improved across repeated trials at baseline and follow-up, and unlike the longitudinal declines in path length, repeating the same task after a two-year delay produced faster improvement over trials and greater end-of-session gains. Such outcome is consistent with developing a cognitive map during baseline acquisition that remains relatively intact after hiatus (O'Keefe and Nadel, 1978; O'Keefe, 1990). In that sense, vMWM navigation behaves similar to episodic and working memory tasks that improve longitudinally (Thorvaldsson et al., 2005; Bender et al., 2015; Daugherty et al., 2015b; Persson et al., 2016), with lesser improvement shown by older adults. Variability in path complexity was associated with hippocampal iron content and parahippocampal gyrus volume—characteristics of medial temporal lobe structures that are acknowledged as brain regions critical for episodic memory in general and memory for places in particular (Squire and Zola-Morgan, 1991; Epstein and Kanwisher, 1998; Eichenbaum et al., 1999; Burgess et al., 2002; Smith and Mizumori, 2006; Chrastil et al., 2015). These regions are also viewed as neural substrates of cognitive mapping (see Maguire et al. (1999) for a review). Indeed, a recent case study reported a patient with medial temporal lobe damage who improved in vMWM navigation, despite demonstrated amnesia, but showed a specific deficit in navigation precision (Kolarik et al., 2016), and patients with parahippocampal lesions evidence specifically impaired performance on a “dry” version of the MWM (Bohbot et al., 2000). The longitudinal results obtained here are in line with cross-sectional findings linking path complexity with PHG volume (Daugherty et al., 2015a, 2015c). We also show the association between path complexity and age-related iron accumulation in the Hc that is consistent with an association between Hc iron and episodic memory observed in an independent sample (Rodrigue et al., 2013). But we found no evidence for Hc volume and path complexity, unlike our previous study (Daugherty et al., 2015c). Thus, search path complexity that is likely related to cognitive mapping is relatively stable and potentially improves longitudinally, although older adults experience lesser repeated-testing gains.

Age-related deficits in navigation reported here are in accord with extant cross-sectional studies (Moffat and Resnick, 2002; Driscoll et al., 2003, 2005; Moffat, 2009). In replicating and extending our previous cross-sectional reports (Moffat et al., 2007; Yuan et al., 2014; Daugherty et al., 2015a, 2015c), we observe here that older adults traveled longer and more complex paths than younger counterparts, despite showing improvement in both indices across repeated trials within a session. After two years, older adults showed greater declines in navigation ability—greater prolongation in distance-to-goal measures and lesser reduction in meandering on the last three trials. The observed sex differences were also in line with other reports (Sandstrom et al., 1998; Moffat et al., 2001; Driscoll et al., 2003; Nowak and Moffat, 2011)—men traveled shorter and less complex paths at baseline, but did not differ from women in longitudinal change.

We detect these sources of variability in change even when the magnitude of individual differences in acquisition curves was notably smaller at follow-up as compared to baseline, particularly

in search path length. The reduction in variability may in part be due to the constraints in the model to ensure measurement invariance—a necessary assumption for valid estimates of change. Mastering a skill via practice and gradual increase in automaticity results, as a rule, in compression of individual variability in performance (Kanfer and Ackerman, 1989). Lesser individual variability in acquisition curves at follow-up may reflect such automatization of navigation skill, whereas variability in the magnitude of change between measurement occasions is an indication of individual differences in aging. Although we speculate that navigation skill becomes automated with practice, six vMWM trials were unlikely to be sufficient to reach asymptotic levels of skill. Thus, it is possible with additional trials, participants might have recovered performance to the levels reached at the baseline testing. Between-trial heterogeneity in the average acquisition curves suggests that the difficulty of the search might have depended on the starting position and thus initial location in the virtual arena might have affected differences in performance through interaction with age-related cognitive decline. Due to the limited number of trials, we cannot test this source of variability apart from normal fluctuations in performance during the acquisition process, but should consider such a test in future study. Yet, even when navigation distance reaches asymptote in a vMWM, older adults still travel longer paths than their younger counterparts (Daugherty et al. 2015c).

To understand the differential pattern of longitudinal change in each index of navigation and its vulnerability in aging, path length and complexity should be considered as complementary (not redundant) metrics of navigation efficiency. Optimal navigation will be short and not meander, but in this task that required successful location of the platform, an individual may favor less complex paths over optimizing path length, as suggested in the longitudinal trend here. For example, a path traveled around the circumference of the pool, presumably in reference to the landmarks on the perimeter, will be longer than a straight path from the start but will still have relatively low complexity (see Daugherty et al. (2015a)). Although decrease in path complexity partially explains improvement in navigation distance (Daugherty et al. 2015a, 2015c), the trade-off as described above is a realistic possibility. Regrettably, we were unable to examine the relationship between the two indices here due to the constraints on the number of model parameters imposed by the sample size. In the future, such trade-off analysis may provide further insights into a shift in behavior to emphasize less meandering over shorter distance that occurs in aging.

This shift in behavior may mirror the alteration of navigation strategies that is hypothesized to underlie age-related decline in performance (Wiener et al., 2013; Bohbot et al., 2012; Harris et al., 2012). When navigating environments like the vMWM, older adults appear to favor a beacon strategy—directed movement towards a landmark near a goal without required encoding of directional information (Wiener et al., 2013; Bohbot et al., 2012; Harris et al., 2012). Compared to an allocentric strategy, which is believed to draw on a cognitive map, the beacon approach is less efficient in accomplishing the navigation task at hand (Wiener et al., 2013). Moreover, these different navigation strategies dissociate by neural correlates: beacon-directed movement engages the striatum whereas allocentric navigation is dependent upon medial temporal lobe structures (Bohbot et al., 2007; Packard and McGaugh, 1992; Doeller et al., 2008; Wiener et al., 2013). With practice, striatum-dependent beacon navigation proceduralizes and becomes less amenable to change (Packard and McGaugh, 1996), a characteristic that is consistent with the longitudinal worsening of distance acquisition curves seen here. Conversely, practice of allocentric navigation dependent upon the medial temporal lobe is expected to produce gains and retention over

time, as noted in regard to longitudinal change in meandering. In the course of aging, navigation strategies shift from allocentric to beacon-based, which may reflect, in part, declines in the relevant neural substrates (Bohbot et al., 2012; Konishi et al., 2013; Wiener et al., 2013; Schuck et al., 2015). Nonetheless, proceduralization of navigation skill does not completely negate the contributions from cognitive mapping in episodic memory (Packard and McGaugh, 1996) and the putative shift in strategy use is expected to be gradual and continuous during the aging process.

Under conditions of uncertainty and potentially conflicting navigation strategies, the prefrontal cortex may act as arbiter (Chersi and Burgess, 2015). As the prefrontal cortex is viewed as a neural substrate of cognitive control in skill acquisition and maintenance as well as strategic aspects of procedural and episodic memory (Miller and Cohen, 2001; Poldrack et al., 2005; Seidler et al., 2012; Ofen and Shing, 2013), individuals with larger LPFC volume may be better able to adapt strategies that optimize path length and complexity (O'Keefe, 1991). Hence, the common correlations of these measures with LPFC observed here. Smaller LPFC volumes in older adults partially mediated declines after the two-year hiatus and could be related to a shift towards less-effective strategies (Moffat and Resnick, 2002; Rodgers et al., 2012; Wiener et al., 2013; Harris and Wolbers, 2014). Further, individual differences in neural substrates and navigation strategies are likely to shape navigation paths through complex interactions, either as compensation following age-related decline (McDonald and White, 1993; Harris et al., 2012; Schuck et al., 2015) or as competing influences (Chersi and Burgess, 2015; Poldrack and Packard, 2003). A systems-level analysis of neural and cognitive changes would require multiple measurement occasions and a substantially larger sample than presented here, and remains an important goal for future study.

Age-related deficits in navigation were only partially explained by regional shrinkage; iron accumulation and decline in vascular health were additional factors that mediated age effects. Increased pulse pressure—a proxy for endothelial dysfunction and elevated vascular risk—was associated with longer and more complex paths and, together with shrinkage and iron accumulation, mediated the effects of age to partly explain greater decline after two years in older adults.

Age-related brain iron accumulation proved to be an essential predictor of longitudinal decline in navigation ability. Greater baseline Hc iron, and not smaller volumes, accounted for more complex navigation paths, lesser repeated-testing gains after two years, and partially mediated age effects at baseline and in longitudinal change. Our previous cross-sectional reports identified an association between path complexity and age-related differences in Hc volumes (Daugherty et al., 2015a, 2015c), and others have reported correlations with distance (see Moffat (2009) for a review) although evidence of this in older adults is inconsistent (e.g., Moffat et al., 2007). Here we find that longitudinal increase in Hc iron content, but not change in volume, predicted declines in navigation skill indexed by path complexity. Accumulation of iron causes oxidative stress that in turn disrupts function via impaired metabolism and neurotransmission (Mills et al., 2010; Ward et al., 2014), as well as by contributing to structural declines (Daugherty et al., 2015b). Thus, brain iron accumulation in aging may be a sensitive marker of impending cognitive decline and may contribute to declines that are in part shared with shrinkage. Indeed, greater Cd baseline iron content together with smaller volume predicted larger losses in end-of-session navigation distance with advanced age, and a trend for greater increase in Cd iron content directly predicted greater declines in performance at the end of the task. Cross-sectional studies report that greater brain iron content in older adults is associated with deficits in episodic memory (Bartzokis et al., 2011; Rodrigue et al., 2013) and

cognitive-motor control (Sullivan et al., 2009; Adamo et al., 2014). In a previous report on this sample, greater Cd iron content predicted lesser repeated-testing gains in verbal working memory after 2 years (Daugherty et al., 2015b), and similarly found that iron accumulation predicted functional decline when volume did not. However, in that study, Hc iron content was unrelated to change in a general episodic memory composite (Daugherty et al., 2015b). The lack of association between Hc iron and change in episodic memory may be in part due to the nature of the latent composite used in that study. Here, we partially replicate our previous report with latent navigation acquisition curves that were indexed by six repeated trials of the same task, compared to cumulative performance on two different memory tasks. Recognizing the important differences between this study and our previous one, we provide further evidence for a role of iron accumulation in neural and cognitive aging (see Daugherty and Raz (2015) for a review).

Iron accumulation in PHG, LPFC, and Cb is also expected to promote cognitive declines, and may be a stronger predictor thereof than volume alone. However, we were unable to assess iron content in these regions via R2* due to possible confounds in these regions (Fukunaga et al., 2010; Langkammer et al., 2012; Lodygensky et al., 2012). R2* relaxometry is not a specific index of brain iron and can be biased by presence of calcium deposits (Naderi et al., 1993), myelinated fibers (Fukunaga et al., 2010; Langkammer et al., 2012; Lodygensky et al., 2012) and amyloid (Quintana et al., 2006). The risk of these confounds in the Cd and Hc are nontrivial and although in this healthy sample they are likely to be low, they cannot be discounted, especially with respect to the Hc. Nor is R2* a specific proxy of oxidative stress-driven damage, although it is related to increases in ferritin-bound non-heme iron and thus signals a general shift in iron homeostasis (Morita et al., 1981; Salgado et al., 2010). Such shift and ensuing production of free-radicals is a plausible explanation of the effects observed here (see Daugherty and Raz (2015) for a review).

The results reported here should be interpreted in the context of several other limitations. Although we report differential patterns of longitudinal change in navigation indices and their neural correlates, we did not test competing hypotheses with all measures in a single model and thus cannot show exclusive correlates. The limitations in modeling stem from a relatively small sample size that was further compromised by non-random attrition: 39% of the sample did not return for repeated assessment and the drop-outs were younger and had a lower MMSE at baseline. We, nonetheless, handled missing data via full information maximum likelihood, which leverages all available data during estimation without imputation, and inclusion of baseline age and MMSE corrected for non-random attrition. Thus, we capitalize on the larger sample to improve validity of the estimates, although we cannot completely discount the possibility of some bias in the analyses. Our correction for measurement inconsistency mitigated bias due to poor trial-performance reliability from estimates of change, yet the necessary constraints may have minimized the magnitude of effects detected. Further, the constraints imposed on the model in order to meet the assumption of measurement invariance and the unspecified non-linear slope estimates were sample-specific. This is true of any longitudinal analysis and a strong point of the SEM methods used here is explicit testing of the assumptions. Such assumption testing leads to the most valid estimates of change that are feasible for the available data (McArdle, 2009). Therefore, the results reported here are generalizable only to the extent the current sample is representative of an aging population. In other words, generalization of the reported findings is restricted to older adults selected under rather stringent health criteria, and such restriction could have minimized the magnitude of change in navigation performance and

individual differences therein. An additional limitation of the modeling implemented here is having only two measurement occasions, which allows estimating only linear change (McArdle and Hamagami, 2001) and precludes testing lead-lag relationships. Finally, it is important to emphasize that we cannot make any claims of causality from these data (Winship and Morgan, 1999; Rubin, 2005).

5. Conclusions

In spatial navigation, perceptual-motor skill and cognitive mapping as well as their putative neural substrates are differentially associated with age-related declines. Navigation performance indexed by search path length behaves as a typical perceptual-motor skill as it shows improvement with repetition but little saving over a two-year hiatus. That aspect of navigation is weaker in older adults and is linked to changes in the brain structures commonly associated with perceptual-motor skill acquisition and retention: caudate nucleus and cerebellum. In contrast, ability to navigate without meandering indexed by path complexity improves within a few repeated trials with the gains retained and even augmented after a two-year break. The higher level of performance, greater degree of improvement and stability of gains are related to younger age, lesser vascular risk, larger volumes of brain substrates of episodic memory and lower level of a marker of cellular damage (iron) therein. Both aspects of navigation are related to the LPFC volumes that possibly mediate the strategic aspects of navigation. Thus, several cognitive functions comprise spatial navigation skill and differential decline in these and respective neural substrates partly explain mixed general decline, but not complete failure, in navigation during healthy aging.

Acknowledgments

This work was supported by National Institute on Aging at the National Institutes of Health grant R37 AG011230 to N.R. and Blue Cross Blue Shield of Michigan Foundation student awards program grant 1893.SAP to A.M.D. We thank Drs. Cheryl Dahle, Andrew Bender, Peng Yuan, and Yiqin Yang for their assistance in data collection, and Dr. E. Mark Haacke for advice on the imaging protocol.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.neuroimage.2016.09.044>.

References

- Adamo, D.E., Daugherty, A.M., Raz, N., 2014. Brain iron content and grasp force-matching ability in older women. *Brain Imaging Behav.* 8 (4), 579–587. <http://dx.doi.org/10.1007/s11682-013-9284-6>.
- Astur, R.S., Ortiz, M.L., Sutherland, R.J., 1998. A characterization of performance by men and women in a virtual Morris water task: a large and reliable sex difference. *Behav. Brain Res.* 93, 185–190.
- Barak, B., Feldman, N., Okun, E., 2015. Cardiovascular fitness and cognitive spatial learning in rodents and in humans. *J. Gerontol. A Biol. Sci. Med. Sci.* 70 (9), 1059–1066.
- Bartzokis, G., Lu, P., Tingus, K., Peters, D.G., Amar, C.P., Tishler, T.A., Finn, J.P., Willablanca, P., Altschuler, L.L., Mintz, J., Neely, E., Connor, J.R., 2011. Gender and iron genes may modify associations between brain iron and memory in healthy aging. *Neuropsychopharmacology* 36, 1375–1384.

- Bender, A.R., Raz, N., 2012. Age-related differences in memory and executive functions in healthy APOE $\epsilon 4$ carriers: the contribution of individual differences in prefrontal volumes and systolic blood pressure. *Neuropsychologia* 50 (5), 704–714. <http://dx.doi.org/10.1016/j.neuropsychologia.2011.12.025>.
- Bender, A.R., Prindle, J.J., Brandmaier, A.M., Raz, N., 2015. White matter and memory in healthy adults: coupled changes over two years. *Neuroimage* [Epub ahead of print].
- Bohbot, V.D., Allen, J.J., Nadel, L., 2000. Memory deficits characterized by patterns of lesions to the hippocampus and parahippocampal cortex. *Ann. N. Y. Acad. Sci.* 911, 355–368.
- Bohbot, V.D., Lerch, J., Thorndyraft, B., Iaria, G., Zijdenbos, A.P., 2007. Gray matter differences correlate with spontaneous strategies in a human virtual navigation task. *J. Neurosci.* 27 (38), 10078–10083.
- Bohbot, V.D., McKenzie, S., Konishi, K., Fouquet, C., Kurdi, V., Schachar, R., Boivin, M., Robaey, P., 2012. Virtual navigation strategies from childhood to senescence: evidence for changes across the lifespan. *Front. Aging Neurosci.* 4, 28.
- Burgess, N., Maguire, E.A., O'Keefe, J., 2002. The human hippocampus and spatial and episodic memory. *Neuron* 35 (4), 625–641. [http://dx.doi.org/10.1016/S0896-6273\(02\)00830-9](http://dx.doi.org/10.1016/S0896-6273(02)00830-9).
- Cavaco, S., Anderson, S.W., Correia, M., Magalhaes, M., Pereora, C., Tuna, A., Taipa, R., Pinto, P., Pinto, C., Cruz, R., Lima, A.B., Castro-Caldas, A., Martins da Silva, A., Damasio, H., 2011. Task-specific contribution of the human striatum to perceptual-motor skill learning. *J. Clin. Exp. Neuropsychol.* 33 (1), 51–62.
- Chersi, F., Burgess, N., 2015. The cognitive architecture of spatial navigation: hippocampal and striatal contributions. *Neuron* 88 (1), 64–77.
- Chrastil, E.R., Sherrill, K.R., Hasselmo, M.E., Stern, C.E., 2015. There and back again: hippocampus and retrosplenial cortex track homing distance during human path integration. *J. Neurosci.* 35 (46), 15442–15452. <http://dx.doi.org/10.1523/JNEUROSCI.1209-15.2015>.
- Coats, R.O., Snapp-Childs, W., Wilson, A.D., Bingham, G.P., 2013. Perceptuo-motor learning rate declines by half from 20s to 70/80s. *Exp. Brain Res.* 225, 75–84.
- Daugherty, A.M., Bender, A.R., Yuan, P., Raz, N., 2015a. Changes in search path complexity and length during learning of a virtual water maze: Age differences and differential associations with hippocampal subfield volumes. *Cereb. Cortex* . <http://dx.doi.org/10.1093/cercor/bhv061>.
- Daugherty, A.M., Haacke, E.M., Raz, N., 2015b. Striatal iron content predicts its shrinkage and changes in working memory after two years in healthy adults. *J. Neurosci.* 35 (17), 6731–6743. <http://dx.doi.org/10.1523/JNEUROSCI.4717-14.2015>.
- Daugherty, A.M., Raz, N., 2015. Appraising the role of iron in brain aging and cognition: promises and limitations of MRI. *Neuropsychol. Rev.* 25 (3), 272–287. <http://dx.doi.org/10.1007/s11065-015-9292-y>.
- Daugherty, A.M., Raz, N., 2016. Accumulation of iron in the putamen predicts its shrinkage in healthy older adults: A multi-occasion longitudinal study. *NeuroImage* (Epub ahead of print).
- Daugherty, A.M., Yuan, P., Dahle, C.L., Bender, A.R., Yang, Y., Raz, N., 2015c. Path complexity in virtual water maze navigation: Differential associations with age, sex, and regional brain volume. *Cereb. Cortex* 25 (9), 3122–3131. <http://dx.doi.org/10.1093/cercor/bhu107>.
- Driscoll, I., Hamilton, D.A., Petropoulos, H., Yeo, R.A., Brooks, W.M., Baumgartner, R. N., Sutherland, R.J., 2003. The aging hippocampus: cognitive biochemical and structural findings. *Cereb. Cortex* 13, 1344–1351.
- Driscoll, I., Hamilton, D.A., Yeo, R.A., Brooks, W.M., Sutherland, R.J., 2005. Virtual navigation in humans: the impact of age, sex, and hormones on place learning. *Horm. Behav.* 47, 326–335.
- Dröge, W., Schipper, H.M., 2007. Oxidative stress and aberrant signaling in aging and cognitive decline. *Aging Cell* 6, 361–370.
- Doeller, C.F., King, J.A., Burgess, N., 2008. Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *PNAS* 105 (15), 5915–5920.
- Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., Tanila, H., 1999. The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron* 23 (2), 209–226. [http://dx.doi.org/10.1016/S0896-6273\(00\)80773-4](http://dx.doi.org/10.1016/S0896-6273(00)80773-4).
- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature* 392, 598–601.
- Finch, C.E., Crimmins, E.M., 2004. Inflammatory exposure and historical changes in human life-spans. *Science* 305, 1736–1739.
- Finch, C.E., Foster, J.R., Mirsky, A.E., 1969. Ageing and the regulation of cell activities during exposure to cold. *J. Gen. Physiol.* 54, 690–712.
- Fisk, A.D., Hertzog, C., Lee, M.D., Rogers, W.A., Anderson-Garlach, M.M., 1994. Long-term retention of skilled visual search: Do young adults retain more than old adults? *Psychol. Aging* 9, 206–215.
- Fjell, A.M., Walhovd, K.B., Fennema-Notestine, C., McEvoy, L.K., Hagler, D.J., Holland, D., Brewer, J.B., Dale, A.M., 2009. One-year brain atrophy evident in healthy aging. *J. Neurosci.* 29 (48), 15223–15231.
- 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, 189–198.
- Fukunaga, M., Li, T.Q., van Gelderen, P., de Zwart, J.A., Shmueli, K., Yao, B., Lee, J., Maric, D., Aronova, M.A., Zhang, G., Leapman, R.D., Schenck, J.F., Merkle, H., Duyn, J.H., 2010. Layer-specific variation of iron content in cerebral cortex as a source of MRI contrast. *Proc. Natl. Acad. Sci. USA* 107, 3834–3839.
- Ghisletta, P., Kennedy, K.M., Rodrigue, K.M., Lindenberger, U., Raz, N., 2010. Adult age differences and the role of cognitive resources in perceptual-motor skill acquisition: application of a multilevel negative exponential model. *J. Gerontol.: Psychol. Sci.* 65B (2), 163–173.
- Churge, N.R., Coates, T.D., Nelson, M.D., Wood, J.C., 2005. Mechanisms of tissue-iron relaxivity: nuclear magnetic resonance studies of human liver biopsy specimens. *Magn. Reson. Med.* 54 (5), 1185–1193.
- Grammas, P., 2011. Neurovascular dysfunction, inflammation and endothelial activation: implications for the pathogenesis of Alzheimer's disease. *J. Neuroinflamm.* 8, 26. <http://dx.doi.org/10.1186/1742-2094-8-26>.
- Grimm, K.J., Ram, N., Hamagami, F., 2011. Nonlinear growth curves in developmental research. *Child Dev.* 82 (5), 1357–1371.
- Haacke, E.M., Cheng, N.Y.C., House, M.J., Liu, Q., Neelavalli, J., Ogg, R.J., Khan, A., Ayaz, M., Kirsch, W., Obenaus, A., 2005. Imaging iron stores in the brain using magnetic resonance imaging. *Magn. Reson. Imaging* 23 (1), 1–25.
- Harris, M.A., Wiener, J.M., Wolbers, T., 2012. Aging specifically impairs switching to an allocentric navigational strategy. *Front. Aging Neurosci.* 4, 29.
- Harris, M.A., Wolbers, T., 2014. How age-related strategy switching deficits affect wayfinding in complex environments. *Neurobiol. Aging* 35 (5), 1095–1102. <http://dx.doi.org/10.1016/j.neurobiolaging.2013.10.086>.
- Hayes, A.F., Scharkow, M., 2013. The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: does method really matter? *Assoc. Psychol. Sci.* 24 (10), 1918–1927.
- Hertzog, C., von Oertzen, T., Ghisletta, P., Lindenberger, U., 2008. Evaluating the power of latent growth curve models to detect individual differences in change. *Struct. Equ. Model.* 15, 541–563.
- Horn, J.L., Donaldson, G., 1980. Cognitive development in adulthood. In: Brim Jr., O. G., Kagan, J. (Eds.), *Constancy and Change in Human Development*. Harvard University Press, Cambridge, pp. 445–529.
- Jack Jr, C.R., Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., Cascino, G.D., 1989. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. *Radiology* 172, 549–554.
- Jagut, W., 2013. Vulnerable neural systems and the borderland of brain aging and neurodegeneration. *Neuron* 77 (2), 219–234.
- Janacek, K., Fiser, J., Nemeth, D., 2012. The best time to acquire new skills: age-related differences in implicit sequence learning across the human lifespan. *Dev. Sci.* 15, 496–505.
- Kanfer, R., Ackerman, P.L., 1989. Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. *J. Appl. Psychol.* 74 (4), 657–690.
- Kennedy, K.M., Raz, N., 2005. Age, sex and regional brain volumes predict perceptual-motor skill acquisition. *Cortex* 41, 560–569.
- Kennedy, K.M., Raz, N., 2015. Normal aging of the brain. In: Toga, A.W. (Ed.), *Brain Mapping: An Encyclopedic Reference* vol. 3. Academic Press: Elsevier, Cambridge, pp. 603–617.
- Kirasic, K.C., Allen, G.L., Haggerty, D., 1992. Age-related differences in adults' macrospatial cognitive processes. *Exp. Aging Res.* 18 (1), 33–39.
- Kolarik, B.S., Shallice, K., Hassan, A., Borders, A.A., Kaufman, K.C., Gurkoff, G., Yonelinas, A.P., Ekstrom, A.D., 2016. Impairments in precision, rather than spatial strategy, characterize performance on the virtual Morris water maze: a case study. *Neuropsychologia* 80 (8), 90–101. <http://dx.doi.org/10.1016/j.neuropsychologia.2015.11.013>.
- Konishi, K., Etchamendy, N., Roy, S., Marighetto, A., Rajah, N., Bohbot, V.D., 2013. Decreased functional magnetic resonance imaging activity in the hippocampus in favor of the caudate nucleus in older adults tested in a virtual navigation task. *Hippocampus* 23 (11), 1005–1014.
- Korthauer, L.E., Mowak, N.T., Moffat, S.D., An, Y., Rowland, L.M., Barker, P.B., Resnick, S.M., Driscoll, I., 2015. Correlates of virtual navigation performance in older adults. *Neurobiol. Aging* . <http://dx.doi.org/10.1016/j.neurobiolaging.2015.11.003>.
- Langhammer, C., Krebs, N., Goessler, W., Scheurer, E., Yen, K., Fazekas, F., Ropele, S., 2012. Susceptibility induced gray-white matter MRI contrast in the human brain. *NeuroImage* 59, 1413–1419.
- Larsen, R., 2011. Missing data imputation versus full information maximum likelihood with second-level dependencies. *Struct. Equ. Model.: Multidiscip. J.* 18 (4), 649–662.
- Lindenberger, U., Mayr, U., Kliegl, R., 1993. Speed and intelligence in old age. *Psychol. Aging* 8 (2), 207–220.
- Lindenberger, U., von Oertzen, T., Ghisletta, P., Hertzog, C., 2011. Cross-sectional age variance extraction: What's change got to do with it? *Psychol. Aging* 26 (1), 34–47. <http://dx.doi.org/10.1037/a0020525>.
- Lodygensky, G.A., Marques, J.P., Maddage, R., Perroud, E., Sizonenko, S.V., Hüppi, P.S., Gruetter, R., 2012. In vivo assessment of myelination by phase imaging at high magnetic field. *NeuroImage* 59, 1979–1987.
- Lövdén, M., Schaefer, S., Noack, H., Bodammer, N.C., Kühn, S., Hainze, H.-J., Düzel, E., Bäckman, L., Lindenberger, U., 2012. Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiol. Aging* 620, e9–e22.
- Maaroufi, K., Ammari, M., Jeljeli, M., Roy, V., Sakly, M., Abdelmelek, H., 2009. Impairment of emotional behavior and spatial learning in adult Wistar rats by ferrous sulfate. *Physiol. Behav.* 96, 343–349.
- Maguire, E.A., Burgess, N., O'Keefe, J.O., 1999. Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. *Curr. Opin. Neurobiol.* 9, 171–177.
- Maxwell, S.E., Cole, D.A., 2007. Bias in cross-sectional analyses of longitudinal mediation. *Psychol. Methods* 12 (1), 23–44.
- McArdle, J.J., 2009. Latent variable modeling of differences and changes with longitudinal data. *Ann. Rev. Psychol.* 60, 577–605.

- McArdle, J.J., Anderson, E., 1990. Latent variable growth models for research on aging. In: Birren, J.E., Schaie, K.W. (Eds.), *The handbook of the psychology of aging*, 3rd ed. Plenum Press, New York, pp. 21–43.
- McArdle, J.J., Hamagami, F., 2001. Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In: Collins, L.M., Sayer, A.G. (Eds.), *New Methods for the Analysis of Change*. Decade of Behavior, APA, Washington, DC, pp. 139–175.
- McDonald, R.J., White, N.M., 1993. A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behav. Neurosci.* 107 (1), 3–22.
- Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Mills, E., Dong, X.P., Wang, F., Xu, H., 2010. Mechanisms of brain iron transport: Insight into neurodegeneration and CNS disorders. *Future Med. Chem.* 2 (1).
- Moffat, S.D., 2009. Aging and spatial navigation: what do we know and where do we go? *Neuropsychol. Rev.* 19, 478–489.
- Moffat, S.D., Kennedy, K.M., Rodrigue, K.M., Raz, N., 2007. Extra-hippocampal contributions to age differences in human spatial navigation. *Cereb. Cortex* 17 (6), 1274–1282.
- Moffat, S.D., Resnick, S.M., 2002. Effects of age on virtual environment place navigation and allocentric cognitive mapping. *Behav. Neurosci.* 116 (5), 851–859.
- Moffat, S.D., Zonderman, A.B., Resnick, S.M., 2001. Age differences in spatial memory in a virtual environment navigation task. *Neurobiol. Aging* 22 (5), 787–796.
- Morita, R., Yoshii, M., Nakajima, K., Kohsaka, T., Miki, M., Torizuka, K., 1981. Clinical evaluation of serum ferritin to iron ratio in malignant diseases. *Eur. J. Nucl. Med.* 6 (7), 331–336.
- Morris, R., 1981. Spatial localization does not require the presence of local cues. *Learn. Motiv.* 12 (2), 239–260.
- Muthén, B.O., Kaplan, D., Hollis, M., 1987. On structural equation modeling with data that not missing completely at random. *Psychometrika* 52, 431–462.
- Naderi, S., Colakoglu, Z., Lüleci, G., 1993. Calcification of basal ganglia associated with pontine calcification in four cases: a radiologic and genetic study. *Clin. Neurol. Neurosurg.* 95, 155–1557.
- Nams, V.O., 2006. Improving accuracy and precision in estimating fractal dimension of animal movement paths. *Acta Biotheor.* 54, 1–11.
- Nowak, N.T., Moffat, S.D., 2011. The relationship between second to fourth digit ratio, spatial cognition, and virtual navigation. *Arch. Sex. Behav.* 40, 575–585.
- Ofen, N., Shing, Y.L., 2013. From perception to memory: Changes in memory systems across the lifespan. *Neurosci. Behav. Rev.* 37, 2258–2267. <http://dx.doi.org/10.1016/j.neubiorev.2013.04.006>.
- O'Keefe, J., 1990. A computational theory of the hippocampal cognitive map. *Prog. Brain Res.* 83, 301–312.
- O'Keefe, J., 1991. The hippocampal cognitive map and navigational strategies. In: Paillard, J. (Ed.), *Brain and Space*. Oxford University Press, New York, pp. 273–295.
- O'Keefe, J., Nadel, L., 1978. *The Hippocampus as a Cognitive Map*. Oxford University Press, New York (NY).
- Oldfield, R.C., 1971. The assessment and analysis of handedness. *Neuropsychologia* 9, 97–113.
- Packard, M.G., McGaugh, J.L., 1992. Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: further evidence for multiple memory systems. *Behav. Neurosci.* 106 (3), 439–446.
- Packard, M.G., McGaugh, J.L., 1996. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol. Learn. Mem.* 65 (1), 65–72.
- Park, D.C., 2000. The basic mechanism accounting for age-related decline in cognitive function. In: Park, D.C., Schwarz, N. (Eds.), *Cognitive Aging: A Primer*. Psychology Press, Philadelphia.
- Persson, N., Ghisletta, P., Dahle, C.L., Bender, A.R., Yang, Y., Yuan, P., Daugherty, A.M., Raz, N., 2014. Regional brain shrinkage over two years: individual differences and effects of pro-inflammatory genetic polymorphisms. *NeuroImage* 103, 334–348. <http://dx.doi.org/10.1016/j.neuroimage.2014.09.042>.
- Persson, N., Ghisletta, P., Dahle, C.L., Bender, A.R., Daugherty, A.M., Yang, Y., Yuan, P., Raz, N., 2016. Regional brain shrinkage and change in cognitive performance over two years: the bidirectional influences of the brain and cognitive reserve factors. *NeuroImage* 126, 15–26.
- Poldrack, R.A., Packard, M.G., 2003. Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* 41, 245–251.
- Poldrack, R.A., Sabb, F.W., Foerdes, K., Tom, S.M., Asarnow, R.F., Bookheimer, S.Y., Knowlton, B.J., 2005. The neural correlates of motor skill automaticity. *J. Neurosci.* 25 (22), 5356–5364.
- Quintana, C., Bellefqih, S., Laval, J.Y., Guerquin-Kern, J.L., Wu, T.D., Avila, J., Ferrer, I., Arranz, R., Patiño, C., 2006. Study of the localization of iron, ferritin, and hemosiderin in Alzheimer's disease hippocampus by analytical microscopy at the subcellular level. *J. Struct. Biol.* 153 (1), 42–54.
- Radloff, L.S., 1977. The CES-D scale: A self-report depression scale for research in the general population. *J. Struct. Biol.* 153, 42–54.
- Raykov, T., Marcoulides, G.A., 2006. *A First Course in Structural Equation Modeling*, 2nd ed. Lawrence Erlbaum, Mahwah.
- Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., Acker, J.D., 2004. Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. *Neurobiol. Aging* 25, 377–396.
- Raz, N., Lindenberger, U., Rodrigue, K.M., Kennedy, K.M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., Acker, J.D., 2005. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cereb. Cortex* 15, 1676–1689.
- Raz, N., Lindenberger, U., 2011. News of cognitive cure for age-related brain shrinkage is premature: a comment on Burgmans et al., (2009). *Neuropsychol.* 24 (2), 255–257.
- Raz, N., Rodrigue, K.M., 2006. Differential aging of the brain: patterns, cognitive correlates and modifiers. *Neurosci. Biobehav. Rev.* 30 (6), 730–748.
- Raz, N., Williamson, A., Gunning-Dixon, F., Head, D., Acker, J.D., 2000. Neuroanatomical and cognitive correlates of adult age differences in acquisition of a perceptual-motor skill. *Microsc. Res. Tech.* 51, 85–93.
- Rodgers, M.K., Sindone 3rd, J.A., Moffat, S.D., 2012. Effects of age on navigation strategy. *Neurobiol. Aging* 33 (1), 202.e15–22.
- Rodrigue, K.M., Haacke, E.M., Raz, N., 2011. Differential effects of age and history of hypertension of regional brain volumes and iron. *NeuroImage* 54, 750–759.
- Rodrigue, K.M., Daugherty, A.M., Haacke, E.M., Raz, N., 2012. The role of hippocampal iron content and hippocampal volume in age-related differences in memory. *Cereb. Cortex* 23 (7), 1533–1541. <http://dx.doi.org/10.1093/cercor/bhs139>.
- Rodrigue, K.M., Kennedy, K.M., Raz, N., 2005. Aging and longitudinal change in perceptual-motor skill acquisition in healthy adults. *J. Gerontol.: Psychol. Sci.* 60B (4), 174–181.
- Rondi-Reig, L., Burguiere, E., 2005. Is the cerebellum ready for navigation? *Prog. Brain Res.* 148, 199–212.
- Rubin, D.B., 2005. Causal inference using potential outcomes: design, modeling, decisions. *J. Am. Stat. Assoc.* 100, 322–331.
- Salgado, J.C., Olivera-Nappa, A., Gerdtsen, Z.P., Tapia, V., Theil, E.C., Conca, C., Nuñez, M.T., 2010. Mathematical modeling of the dynamic storage of iron in ferritin. *BMC Syst. Biol.* 4, 147.
- Sandstrom, N.J., Kaufman, J., Huettel, S.A., 1998. Males and females use different distal cues in a virtual environment navigation task. *Cogn. Brain Res.* 6 (4), 351–360.
- Schuck, N.W., Doeller, C.F., Polk, T.A., Lindenberger, U., Li, S.-C., 2015. Human aging alters the neural computation of and representation of space. *NeuroImage* 117, 141–150.
- Seidler, R.D., 2006. Differential effects of age on sequence learning and sensorimotor adaptation. *Brain Res. Bull.* 70 (4–6), 337–346.
- Seidler, R.D., Bo, J., Anguera, J.A., 2012. Neurocognitive contributions to motor skill learning: the role of working memory. *J. Mot. Behav.* 44 (6), 445–453.
- Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428.
- Smith, C.D., Walton, A., Loveland, A.D., Umberger, G.H., Kryscio, R.J., Gash, D.M., 2005. Memories that last in old age: motor skill learning and memory preservation. *Neurobiol. Aging* 26, 883–890.
- Smith, D.M., Mizumori, S.J.Y., 2006. Hippocampal place cells, context, and episodic memory. *Hippocampus* 16, 716–729.
- Squire, L.R., Zola-Morgan, M., 1991. The hippocampus, memory, and space. *Hippocampus* 1 (3), 269–271.
- Stoolmiller, M., 1995. Using latent growth curves to study developmental processes. In: Gottman, J.M. (Ed.), *The Analysis of Change*. Lawrence Erlbaum, Mahwah, NJ, pp. 103–138.
- Sullivan, E.V., Adalsteinsson, E., Rohlfing, T., Pfefferbaum, A., 2009. Relevance of iron deposition in deep gray matter brain structures of cognitive and motor performance in healthy elderly men and women: exploratory findings. *Brain Imaging Behav.* 3, 167–175. <http://dx.doi.org/10.1007/s11682-008-9059-7>.
- Tammes, C.K., Walhovd, K.B., Dale, A.M., Østby, Y., Grydeland, H., Richardson, G., Westlye, L.T., Roddey, J.C., Hagler Jr, D.J., Due-Tønnessen, P., Holland, D., Fjell, A.M., Alzheimer's Disease Neuroimaging Initiative, 2013. Brain development and aging: overlapping and unique patterns of change. *Neuroimage* 68, 63–74.
- Thorvaldsson, B., Hofer, S.M., Hassing, L.B., Johansson, B., 2005. Cognitive change as conditional on age heterogeneity in onset of mortality-related processes and repeated testing effects. In: Hoffer, S.M., Alwin, D.F. (Eds.), *Handbook of Cognitive Aging: Interdisciplinary Perspectives*. Sage Publications, Thousand Oaks, p. 248ff.
- Ungerleider, L.G., Doyon, J., Karni, A., 2002. Imaging brain plasticity during motor skill learning. *Neurobiol. Learn. Mem.* 78 (3), 553–564.
- Waldstein, S.R., 2005. Nonlinear relations of blood pressure to cognitive function: the Baltimore longitudinal study of aging. *Hypertension* 45, 374–379.
- Ward, R.J., Zucca, F.A., Duyn, J.H., Crichton, R.R., Zecca, L., 2014. The role of iron in brain aging and neurodegenerative disorders. *Lancet Neurol.* 13, 1045–1060.
- Wiener, J.M., de Condappa, O., Harris, M.A., Wolbers, T., 2013. Maladaptive bias for extrahippocampal navigation strategies in aging humans. *J. Neurosci.* 33 (14), 6012–6017.
- Winship, C., Morgan, S.L., 1999. The estimation of causal effects from observational data. *Ann. Rev. Sociol.* 26, 659–706.
- Wolbers, T., Hegarty, M., 2010. What determines our navigational abilities? *Trends Cogn. Sci.* 14 (3), 138–146. <http://dx.doi.org/10.1016/j.tics.2010.01.001>.
- Wood, J.C., Fassler, J.D., Meade, T., 2004. Mimicking liver iron overload using liposomal ferritin preparations. *Magn. Reson. Med.* 51 (3), 607–611.
- Wright, B.M., Payne, R.B., 1985. Effects of aging on sex differences in psychomotor reminiscence and tracking proficiency. *J. Gerontol.* 40 (2), 179–184.
- Yuan, P., Daugherty, A.M., Raz, N., 2013. Turning bias in virtual spatial navigation: age-related differences and neuroanatomical correlates. *Biol. Psychol.* 8–19.