Newly trained navigation and verbal memory skills elicit changes in task-related networks but not brain structure

Li Zheng[1,2], Zachary Boogaart[3], Andrew McAvan[1], Joshua Garren[1], Stephanie Doner[1], Bradley J. Wilkes[5], Will Groves[4], Ece Yuksel[4], Lucia Cherep[1], & Arne D. Ekstrom[1,2*], & Steven M. Weisberg[4,6*]

[1]Department of Psychology, University of Arizona, 1503 E. University Blvd., Tucson, AZ 85719
[2]Evelyn McKnight Brain Institute, University of Arizona, 1503 E. University Blvd., Tucson, AZ 85719
[3]The University of Miami Miller School of Medicine, University of Miami, 1600 NW 10th Ave #1140, Miami, FL 33136
[4]Department of Psychology, University of Florida, 945 Center Dr., Gainesville, FL 32611
[5]Department of Applied Physiology & Kinesiology, University of Florida, 945 Center Dr., Gainesville, FL 32611
[6]Evelyn McKnight Brain Institute, University of Florida, Gainesville, FL 32611

*These two authors shared equally in the work

CORRESPONDING AUTHORS:
Dr. Steven M. Weisberg, email: stevenweisberg@ufl.edu
Dr. Arne Ekstrom, email: adekstrom@arizona.edu

Funding Sources: NIH/NINDS R21NS120237-01 to SMW and ADE; NIH/NIA K01AG070333-01 to SMW, and a Pilot Project Grant from the Arizona Alzheimer's Consortium (AAC) and the Arizona Alzheimer's Disease Research Center, funded by the State of Arizona Department of Health Services (ADHS), awarded to L.Z.

CRediT statement: Li Zheng: Data curation, Methodology, Formal Analysis, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; Zachary Boogaart: Investigation, Data curation, Methodology, Project administration, Software, Validation; Andrew McAvan: Data curation, Methodology, Software, Validation; Joshua Garren: Investigation, Project administration, Validation; Stephanie Doner: Data curation, Formal Analysis, Investigation, Methodology, Project administration, Validation, Visualization; Will Groves: Investigation, Project administration, Visualization; Ece Yuksel: Data curation, Formal Analysis, Software, Validation, Visualization; Lucia Cherep: Investigation, Methodology, Project administration, Supervision; Bradley Wilkes: Formal Analysis, Investigation, Software, Visualization; Arne Ekstrom: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing; Steven Weisberg: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources,

Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing

Abstract

Training cognitive skills, such as remembering a list of words or navigating a new city, has important implications for everyday life. Yet, understanding what brain changes underlie the acquisition of complex cognitive skills remains unresolved. Here, we developed and validated intensive multiweek interventions in which participants were randomly assigned training in either navigation or verbal memory. Healthy young participants (N=75) underwent structural and functional imaging prior to and following the training. Based on pre-registered and exploratory analyses, we did not find any evidence for changes to gross hippocampal or hippocampal subfield volume, cortical brain volume, or white matter connectivity due to the training. In contrast, network-based analyses suggested changes in task-related informational connectivity, which occurred primarily between cortical areas and mostly involved putative cognitive control networks. These results suggest that cognitive interventions target more transient configurations in network connectivity rather than more durable structural changes.

Introduction

Improving cognitive skills can contribute positively to multiple aspects of daily life, such as education and professional endeavors[1,2]. Yet, the neural changes underlying cognitive skill acquisition remain an unresolved question in cognitive neuroscience, limiting our ability to target interventions designed to improve these skills. Some past studies have suggested a relationship between cognitive skills and focal gray matter volume[3-9], although many of these studies have involved comparison of experts with non-experts or specialized populations (e.g., older adults with cognitive decline). Therefore, it remains unclear whether gray matter volume differences underlie the behavioral changes that occur as part of new cognitive skill acquisition and whether the same mechanisms apply to healthy younger adults. Other studies have suggested the importance of changes in either the recruitment of specific brain areas (i.e., functional activation) or interactions between brain regions (i.e., functional connectivity) that may not result in macroscopic structural changes[1,10,11]. Resolving the neural mechanism(s) underlying the acquisition of new cognitive skills in healthy populations, with the potential for maximal behavioral improvement, is important so we can determine how to target interventions in future studies.

One possibility is that training different cognitive skills leads to changes in similar underlying neural mechanisms by training a cognitive process or enhancing function in a brain region common to both skills. Past studies suggest such overlap for a brain region thought to be broadly important to memory, the hippocampus[12,13]. People with better navigation and/or verbal memory skills typically have larger hippocampi [3,14,15], although this finding has not been consistently replicated[16,17]. Alternatively, it might be that learning different cognitive skills targets only partially overlapping or even independent brain networks, which in turn might hint at why transfer of newly trained cognitive skills to other cognitive domains is often modest[18]. For example, some studies suggest only partial overlap between navigation and verbal memory[19-21], and it is possible that some of the observed differences between experts and non-experts in past studies arose due to partial overlap between verbal memory and navigation. For example, when comparing expert navigators, street and landmark names would be intertwined with wayfinding skills, and thus it is important to determine whether superior verbal memory or navigation skills underlie greater hippocampal volume.

One issue with comparing experts and non-experts is that they involve between-participant designs, which may be underpowered compared to within-participant approaches[22] and may introduce other uncontrolled variables. One potential way to address this issue is to employ a within-participant manipulation to measure how a brain region such as the hippocampus (and connected brain networks) might change in the same participant as a result of training. Successful cognitive interventions suggest that targeted within-participant cognitive training, even for as little as 1-2 weeks, can result in improvements to specific cognitive functions, including changes in focal gray matter[4,23-27]; but see[28]. In some instances, interventions may even generalize to areas not explicitly trained but closely related to the training (termed “near transfer”)[29-33]. We developed and pre-registered two cognitive training regimens in which we attempted to separately target navigation and verbal memory skills. This allowed us to address the following novel theoretical questions: 1) what neural changes, if any, result from an intensive within-participant intervention that improves memory or navigation skills in

healthy young adults 2) if such changes occur, what is the degree of neural overlap between the acquisition of these cognitive skills.

Here, we tested three different cohorts of healthy young adults using two different cognitive interventions, each consisting of 10 training sessions targeting either navigation or verbal memory. A third active control group watched educational videos containing information about navigation and verbal memory but did not receive explicit training[34,35]. Before and after training, participants underwent high-resolution structural imaging targeting the hippocampus (T2), diffusion weighted imaging (DWI), whole brain structural imaging (T1), and task-related functional imaging during a spatial and temporal associative memory task to understand how the hippocampus and related brain networks changed. We pre-registered three novel and specific hypotheses, which are described in more detail here (https://osf.io/etxvj): (1) targeted training over two weeks will result in improvements in navigation or verbal memory, (2) improvements will be specific to the trained task and will transfer to a separate behavioral measure of that skill ("near transfer"), (3a) if the hippocampus is central to navigation, hippocampal volume should increase from pre-test to post-test following navigation training but not verbal memory training or the video control condition, (3b) if the hippocampus is central to verbal memory, hippocampal volume should increase from pre-test to post-test in the Verbal Memory group but not in the Navigation or Video Control groups, or (3c) alternatively, hippocampal volume does not reflect acquisition of new verbal memory nor navigation skills. As part of planned exploratory analyses, we also examined changes in structural connectivity (DWI), gray matter volume outside of the medial temporal lobes, functional changes to individual brain regions (i.e., univariate functional activation), and changes to brain networks (i.e., task-related informational connectivity).

Results

Behavioral Results

Training-related effects

Participants were randomly assigned to one of three conditions, each of which consisted of 10 training sessions, separated by at least one day over the course of approximately 4–8 weeks (4.12 ± 1.87 weeks). Each training session lasted two hours. In the navigation training condition, we recruited a total of 29 participants, with 27 successfully completing 10 sessions of navigation training in a large city-scale virtual environment ("Virtual Arida", Fig. 1b left). Briefly, participants were placed at one building and given the name of another building to find. Participants traveled to and from that pair of buildings until they were able to take a path not exceeding 120% of the shortest possible path. Upon reaching the criterion, they navigated between a new pair of landmarks. After one subsection of Virtual Arida was learned completely, participants were trained on integrating that subsection with previously learned subsections, with no more than one subsection learned in a day (see Methods). To quantify the training effect, we analyzed the average distance error travelled and normalized it by 120% of the length of the shortest possible path. While all 27 participants learned subsections A (although data was lost due to technical reasons for one participant), B, and A-B integration, 26 participants progressed to C, while 23 made it to A-B-C integration. 19 participants made it to subsection D, and 14 participants made it to A-B-C-D integration. Thus, to determine the training effect, we performed a repeated-measures ANOVA with

subsection as the only factor and threshold-normalized distance error as the dependent measure.

We observed a significant main effect of subsection $F(2, 42) = 13.46$, $p < .001$, $\eta^2 = .39$; removing sex and site from the model did not change the effect (Supplemental Note 2). Participants improved over the course of the training, showing higher error initially for subsection A ($M = 79.09$, $SD = 33.22$) compared to subsection B ($M = 65.52$, $SD = 26.89$, Fig. 1b right), which was significantly reduced for subsection C ($M = 47.44$, $SD = 31.22$). This was also the case for subsections involving integration $F(1,22) = 5.16$, $p = .03$, $\eta^2 = .19$, although this finding did not survive inclusion of sex and site in the model ($p = .06$). The distance error was reduced from subsections A-B ($M = 22.03$, $SD = 18.47$) to A-B-C ($M = 15.17$, $SD = 14.47$, Fig. 1b right). In sum, participants improved both in learning new subsections of our large virtual training environment and in integrating multiple areas of the training environment together, suggesting that the navigation training resulted in significant reductions in path error over the course of the training.

In the verbal memory condition, we recruited a total of 32 participants, with 27 successfully completing 10 sessions of training. Briefly, participants viewed words for one second (one at a time), completed math problems as a distractor after all words were presented, and then recalled as many words as they could by typing them in any order they wished. They were trained on a time-based method of loci ("temporal method of loci") to improve their verbal memory[36]. During the first five sessions, participants learned the same list of words, with the list growing in length after each successful trial (Same Word List, Fig. 1a left). In the last 5 sessions, a new list of words was used for each trial (New Word List, Fig. 1a left; see Methods). To quantify the training effect, we conducted a linear regression analysis on the maximum number of correctly recalled words for each day. The slopes of the regression lines for the first five days and the last five days were calculated separately (Fig. 1a right), as the training paradigm differed between these two phases (see Methods). Both slopes were significantly greater than zero, indicating a significant improvement in memory performance over the training period: first five days, $t(26) = 16.971$, $p < .001$; last five days, $t(26) = 23.579$, $p < .001$. This effect remained significant after controlling for sex and site (Supplementary Note 2), suggesting robust improvements in verbal memory performance as a result of the training.

In the active control condition, we recruited a total of 22 participants, with 21 successfully completing 10 sessions of watching videos. These videos included interviews, lectures, documentaries, and TV series on topics related to memory and navigation. After watching the videos for that session, participants responded to multiple choice questions. We did not expect to find a training effect (i.e., improved performance on the questions over sessions), but did find that the percentage of correct responses each day was significantly higher than a 50% guess rate (Fig. 1c right, $ts(20) > 8.28$, $ps < 0.001$, mean = 88.58%, SD = 5.82%), indicating that participants were engaged with the video content.

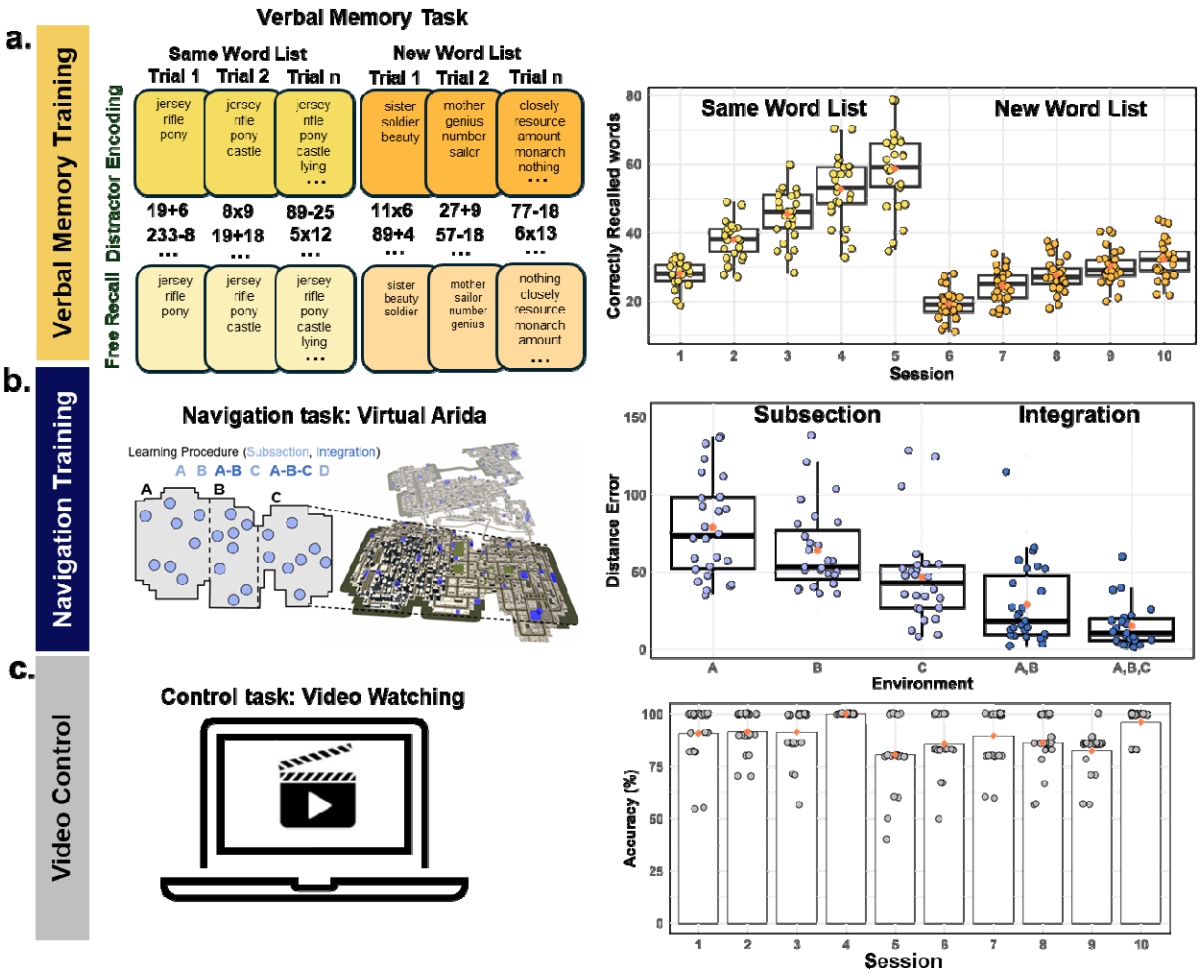

Fig. 1 Experimental design. Participants were randomly assigned to one of three training conditions as follows. (a) In the verbal memory training condition (n = 27), participants underwent 10 sessions of verbal memory training involving word free recall after a distractor task, with list complexity increasing over sessions. (b) In the navigation training condition (n = 27), participants trained in a large virtual environment, navigating between buildings until optimal paths were learned. Subsections of the environment were integrated progressively across sessions. (c) In the active control condition (n = 21), participants watched videos related to memory and navigation, answering multiple-choice questions afterward, with accuracy consistently above 50%, indicating engagement. Training schedules spanned 4–8 weeks, with each session lasting two hours.

Pre-test and post-test behavioral transfer

We administered a set of cognitive tests related to navigation and verbal memory before and after training to track transfer from the training. This involved learning a large virtual environment not related to the training (i.e., Navigation Transfer task: Virtual Silcton[37]) and learning an untrained list of 12 words (i.e., Verbal Memory Transfer task) until all could be recalled perfectly (or after five trials). When we considered our pre-registered dependent measures specific to verbal memory (average number of words recalled, number of trials to criteria, and slope) and navigation (path error of Navigation Transfer task, pointing error of Navigation Pointing task, and map accuracy of

Navigation Model Building task), we found some evidence of improvements in slope specific to the verbal memory training group and path error specific to the navigation training group. The other dependent measures, however, did not show any clear evidence of a training-specific benefit (see Supplemental Note 3). We derived an exploratory measure that allowed us to compare between navigation and verbal memory, termed "*learning rate*." The learning rate attempts to capture how quickly a participant might learn new information at post-test given that they were trained in improving their acquisition of this information. For example, participants who underwent navigation training should show a smaller difference in path distance on the first ten trials compared to the last ten at post-test compared to pre-test, reflective of faster learning due to the training (see Methods).

To quantify the learning rate for the Navigation Transfer task, a mixed-design ANOVA was conducted with condition (Navigation/Verbal Memory/Video Control) as a between-subjects factor and session (pre/post) as a within-subjects factor. This analysis revealed a significant main effect of session ($F(1,70) = 37.52$, $p < 0.001$, $\eta^2 = 0.202$) and a marginally significant main effect of condition ($F(2, 70) = 2.66$, $p = 0.08$, $\eta^2 = 0.028$), with no significant interaction between condition and session ($F(2, 70) = 1.82$, $p = 0.17$, $\eta^2 = 0.020$). Paired-sample t-tests indicated that all three groups demonstrated an increase in learning rate from pre to post, with the Navigation group showing the largest effect (Fig. 2a; Navigation: $t(26) = 4.43$, $p < 0.001$, Cohen's $d = 1.32$; Verbal Memory: $t(26) = 2.31$, $p = 0.03$, Cohen's $d = 0.596$; Video Control: $t(18) = 3.97$, $p < 0.001$, Cohen's $d = 1.20$). Results held when controlling for sex and site (see Supplementary Note 3). The larger effects size for navigation suggests a greater benefit to the learning rate for the navigation condition, although this conclusion is tempered by the lack of a significant interaction effect.

To quantify the learning rate for the Verbal Memory Transfer task, we employed a mixed-design ANOVA, with condition (Navigation/Verbal Memory/Video Control) as a between-subjects factor and session (pre/post) as a within-subjects factor. We found a main effect of session ($F(1, 71) = 3.21$, $p = 0.078$, $\eta^2 = 0.014$) and a significant interaction between condition and session ($F(2, 71) = 2.9$, $p = 0.025$, $\eta^2 = 0.034$). Paired-sample t-tests indicated that only the Verbal Memory group showed a significant increase in learning rate from pre-test to post-test (Fig. 2b; Verbal Memory: $t(25) = 3.32$, $p = 0.003$, Cohen's $d = 0.608$; Navigation: $t(26) = 0.488$, $p = 0.63$, Cohen's $d = 0.113$; Video Control: $t(20) = -0.55$, $p = 0.588$, Cohen's $d = 0.164$). This effect remained significant after controlling for sex and site (see Supplemental Note 3). These findings suggest that verbal memory training resulted in significant improvements from pre-test to post-test in the Verbal Memory Transfer task, specific to the Verbal Memory group.

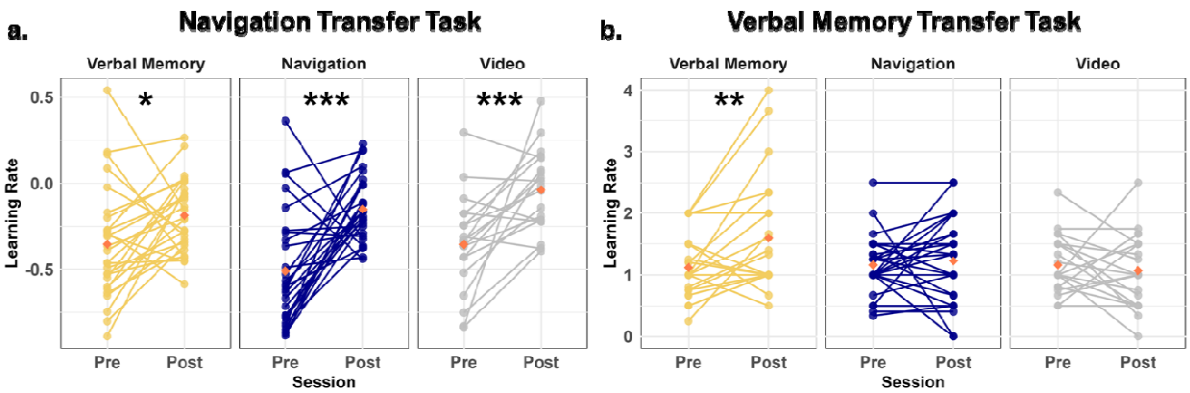

Fig. 2 Changes in learning rates across training conditions from pre-test to post-test. (a) Learning rate for spatial navigation improvement on the Navigation Transfer task. All three groups improved from pre-test to post-test, with the Navigation group demonstrating the largest effect. (b) Learning rate for verbal memory improvement in the Verbal Memory transfer task. Only the Verbal Memory group significantly improved from pre-test to post-test. Red diamonds represent the mean value. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Brain Structural Results

Medial temporal lobe volume did not change due to the training

As our task provided a separation of navigation and verbal memory skills, with some evidence of their independence in the transfer tests, we next tested the hypothesis that navigation training or verbal memory training (or both) would result in increases in hippocampal volume (as measured with T2 high-resolution imaging, see Methods). We conducted a mixed-design ANOVA with condition (Navigation/Verbal Memory/Video Control) as a between-subjects factor and session (pre/post) as a within-subjects factor. The analysis revealed no significant main effect of session ($F(1, 71) = 0.199$, $p = .657$, $\eta^2 = 0.0002$, $BF_{10} = 0.151$, moderate evidence against the inclusion of the main effect of session), no significant main effect of condition ($F(2, 71) = 0.113$, $p = .893$, $\eta^2 = 0.003$, $BF_{10} = 0.241$, moderate evidence against the inclusion of the main effect of condition), and no significant interaction between session and condition ($F(2, 71) = 1.224$, $p = .300$, $\eta^2 = 0.003$, $BF_{10} = 0.054$, strong evidence against the inclusion of the interaction effect). Paired t-tests further confirmed that none of the three training groups exhibited significant changes in total hippocampal volume from pre to post-test (Fig. 3b, ps > .265). This nonsignificant result persisted even after controlling for sex and site as covariates (ps > .256).

We further investigated changes in hippocampal volume by conducting a mixed-design ANOVA, with a particular focus on lateralization. Specifically, we included regions of interest (ROI: left/right hippocampus) and session (pre/post) as a within-subjects factors, alongside condition (Navigation/Verbal Memory/Video Control) as a between-subjects factor. Consistent with the previous analysis, no significant main effects or interactions were observed (session: $F(1, 71) = 0.199$, $p = .657$, $\eta^2 = .0002$, $BF_{10} = 0.131$, moderate evidence against the inclusion of the main effect of session; condition: $F(2, 71) = 0.113$, $p = .893$, $\eta^2 = 0.003$, $BF_{10} = 0.189$, moderate evidence against the inclusion of the main effect of condition; ROI: $F(1, 71) = 0.208$, $p = .650$, $\eta^2 = 0.0001$, $BF_{10} = 0.115$, moderate evidence against the inclusion of the main effect of ROI; and session × condition × ROI, $F(2, 71) = 0.072$, $p = .930$, $\eta^2 < 0.001$, $BF_{10} =$

0.0002, strong evidence against the inclusion of the interaction effect). Paired-sample t-tests confirmed that none of the three training groups exhibited significant changes in hippocampal volume from pre-test to post-test in either the left hippocampus (Fig. 3b, ps > .316) or the right hippocampus (ps > .265). These findings remained consistent after controlling for sex and site as covariates (left hippocampus: ps > .306; right hippocampus: ps > .256).

Because some studies have shown long-axis specialization in correlations between hippocampal volume and behavior[3,38,39], we considered whether anterior and posterior hippocampus might change differentially due to the training, which we analyzed based on the T1 whole brain structural image (see Methods). The previously described mixed-design ANOVA model was expanded to incorporate ROI (anterior and posterior hippocampus) as a within-subjects factor, thereby allowing for the examination of volume differences along the hippocampal anterior-posterior axis. Consistent with the above result, although we found a significant main effect of ROI ($F(1,71) = 22.311$, $p < 0.001$, $\eta^2 = 0.052$, $BF_{10} > 100$, strong evidence for the inclusion of the main effect of ROI), we found no significant main effect of session ($F(1, 71) = 2.914$, $p = 0.092$, $\eta^2 < 0.001$, $BF_{10} = 0.439$, weak evidence against the inclusion of the main effect of session), no significant main effect of condition ($F(2, 71) = 1.045$, $p = 0.357$, $\eta^2 < 0.022$, $BF_{10} = 0.277$, moderate evidence against the inclusion of the main effect of session), and no significant interaction effect between session × condition × ROI ($F(2, 71) = 0.032$, $p = .968$, $\eta^2 < 0.001$, $BF_{10} = 0.005$, strong evidence against the inclusion of the main effect of condition). Simple main effects confirmed that none of the three training groups exhibited significant changes in hippocampal volume from pre-test to post-test neither in the anterior hippocampus (ps > .272) nor the posterior hippocampus (ps > .115). These findings remained consistent after controlling for sex and site as covariates (anterior hippocampus: ps > .260; posterior hippocampus: ps > .106).

Next, we divided the medial temporal lobe into six subregions: three hippocampal subfields (CA1, CA23DG, and SUB) and three adjacent subregions (PRC, ERC, and PHC). We then conducted a mixed-design ANOVA to examine changes in these subfields and subregions. Although results revealed a significant main effect of ROI ($F(1.766, 125.390) = 602.143$, $p < 0.001$, $\eta^2 = 0.895$, Greenhouse-Geisser corrected; $BF_{10} > 100$, strong evidence for the inclusion of the main effect of ROI), we did not find any significant main effect of session ($F(1,71) = 0.766$, $p = 0.384$, $\eta^2 < 0.001$, $BF_{10} = 0.047$, strong evidence against the inclusion of the main effect of session), condition ($F(2,71) = 0.131$, $p = 0.878$, $\eta^2 < 0.001$, $BF_{10} = 0.002$, strong evidence against the inclusion of the main effect of session) and no significant interaction effect between session × condition × ROI ($F(3.21,110.806) = 0.888$, $p = .453$, $\eta^2 < 0.001$, Greenhouse-Geisser corrected; $BF_{10} < 0.001$, extremely strong evidence against the inclusion of the main effect of session). Paired-sample t-tests confirmed that none of the three training groups exhibited significant changes from pre to post-test in any of the six MTL subregions (Fig. 3a, ps > .588, FDR corrected). These findings remained nonsignificant after controlling for sex and site as covariates (ps > .559, FDR corrected).

We also analyzed gray matter volume changes outside of the medial temporal lobe using FreeSurfer (see Methods) to determine if any cortical or other relevant brain areas might have been affected by the training. We applied a vertex-wise analysis of cortical volume, again finding no significant differences across the entire cortex (see

Methods). This finding was further validated using the Destrieux atlas[40], which included 74 cortical parcellations per hemisphere (148 ROIs in total). Paired-sample t-tests revealed that none of the ROIs exhibited significant volume changes from pre- to post-test in any of the three groups (ps > 0.697, FDR-corrected). These findings suggest that training did not result in any measurable cortical volumetric changes.

Diffusion weighted imaging (DWI) metrics did not change due to the training

For DWI, we separately evaluated free water (FW) and free water-corrected fractional anisotropy (fwcFA), as these measures capture different aspects of tissue microstructure. We also chose to analyze ROIs in gray matter (GM) and white matter (WM) separately, as these tissue types have dramatically different values for both FW and fwcFA, which involved eight GM ROIs and seven WM ROIs (see Methods). We conducted a mixed-design ANOVA with condition (Navigation/Verbal Memory/Video Control) as a between-subjects factor, with ROI and session (pre/post) as within-subjects factors. For the WM-fwcFA analysis, we observed a significant main effect of ROI ($F(4.093, 233.279) = 2340.878$, $p < .001$, Greenhouse-Geisser corrected, $\eta^2 = 0.628$, $BF_{10} > 100$, extremely strong evidence for the inclusion of the main effect of ROI). Yet, we found no main effect of session ($F(1,57) < 0.001$, $p = .981$, $\eta^2 < 0.001$, $BF_{10} = 0.090$, strong evidence against the inclusion of the main effect of session), no significant main effect of condition ($F(2,57) = 0.025$, $p = .975$, $\eta^2 < 0.001$, $BF_{10} = 0.011$, strong evidence against including the main effect of condition), and no significant interaction between session × condition × ROI ($F(6.924, 197.328) = 1.954$, $p = .064$, $\eta^2 < 0.001$, $BF_{10} < 0.001$, strong evidence against including this interaction). Paired-sample t-tests confirmed that none of the three training groups exhibited significant changes from pre- to post-test in any of the seven ROIs (ps > .083, FDR corrected). These findings remained nonsignificant after controlling for sex and site as covariates (ps > 0.063, FDR corrected). Analyses involving WM-FW, and GM regions (GM-fwcFA and GM-FW) were also not significant (see Supplementary note 4). These findings suggest that no DWI measures changed as a result of the training. Together, these findings suggest that neither navigation nor verbal memory training for 20 hours across 10 sessions was sufficient to induce changes in macroscopic hippocampal volume or microstructure of relevant gray matter regions and white matter tracts.

Improvements in the learning rate of the Verbal Memory Transfer task, but not the Navigation Transfer task, correlated with both baseline total hippocampal volume and the baseline volume of the CA23DG subfield

We then performed an exploratory analysis to determine whether individual differences in hippocampal volume independent of the training might predict how well they learned from pre-test to post-test based on the different trainings. We conducted analyses employing more stringent corrections for multiple comparisons to examine the association between hippocampal volume and the change in learning rate from pre- to post-test based on the Verbal Memory and Navigation Transfer tasks. This analysis addresses the idea of whether those with larger hippocampi might improve more from pre- to post-test based on the training. Because there were no pre-post differences in hippocampal volume, we averaged the pre- and post-test hippocampal volume. We

consider other behavioral correlations with baseline hippocampal volume in the Supplemental material; none of these were significant (see Supplementary Note 5).

For the Navigation Transfer task, the learning rate did not correlate with hippocampal volume, hemispheric volume, anterior and posterior hippocampus, or any hippocampal subfields, which was also not significant when controlling for sex and site in all three groups (Supplementary Note 5 and Supplementary table 5). For the Verbal Memory Transfer task, the changes in learning rate did not correlate with hippocampal volume ($r(23) = 0.367$, $p = .071$). When accounting for sex and site as covariates, however, we found a significant positive correlation ($r(23) = 0.605$, $p = .006$, FDR corrected). This effect was specific to the Verbal Memory group; no significant correlation was identified for either the Navigation group ($r(25) = 0.028$, $p = .889$) or the Video Control group ($r(19) = -0.206$, $p = .371$); these findings remained consistent when controlling for sex and site as covariates (Navigation: $r(25) = 0.038$, $p = .858$; Video Control: $r(19) = -0.123$, $p = .858$, FDR corrected).

We further examined the correlation between learning rate and hippocampal subfield volumes after correcting for multiple comparisons (see Methods), specifically focusing on CA1 and CA23DG. Only the CA23DG subfield showed a positive correlation with the change in learning rate from pre- to post-test in the Verbal Memory Group ($r(23) = 0.532$, $p = .036$, FDR corrected, Fig. 3c), and this correlation persisted after controlling for sex and site as covariates ($r(23) = 0.493$, $p = .017$). No significant correlations were observed for the CA23DG subfield in the Navigation group ($r(25) = -0.083$, $p = .953$, FDR corrected) or the Video Control group ($r(19) = -0.12$, $p = .953$, FDR corrected), regardless of whether sex and site were included as covariates. For correlations by hemisphere, anterior/posterior hippocampus and other subfields, please see Supplementary Note 5 and Supplementary Table 4. These findings suggested that individuals in the Verbal Memory group with larger CA23DG volumes exhibited greater improvement in verbal memory performance from pre- to post-test.

Fisher's z-tests revealed that the positive correlation in the Verbal Memory group was significantly stronger than that in the Navigation and Video Control group even after controlling for sex and site (Navigation: total hippocampus: $Z = 2.48$, $p = 0.018$; CA23DG: $Z = 1.972$, $p = .036$; Video Control group: total hippocampus: $Z = 2.596$, $p = 0.015$; CA23DG: $Z = 2.091$, $p = .036$; FDR corrected, one-side); there was no significant difference between the Navigation and Video Control group (total hippocampus: $Z = -0.518$, $p = 0.698$; CA23DG: $Z = -0.264$, $p = .604$; FDR corrected, one-side). Together, these findings suggest that baseline hippocampal volume, and CA23DG baseline subfield volume, correlated with the change in learning rate on the Verbal Memory Transfer task following verbal memory training, but not after navigation training or video control. We highlight, however, that these analyses are exploratory and somewhat underpowered for between-subject comparisons.

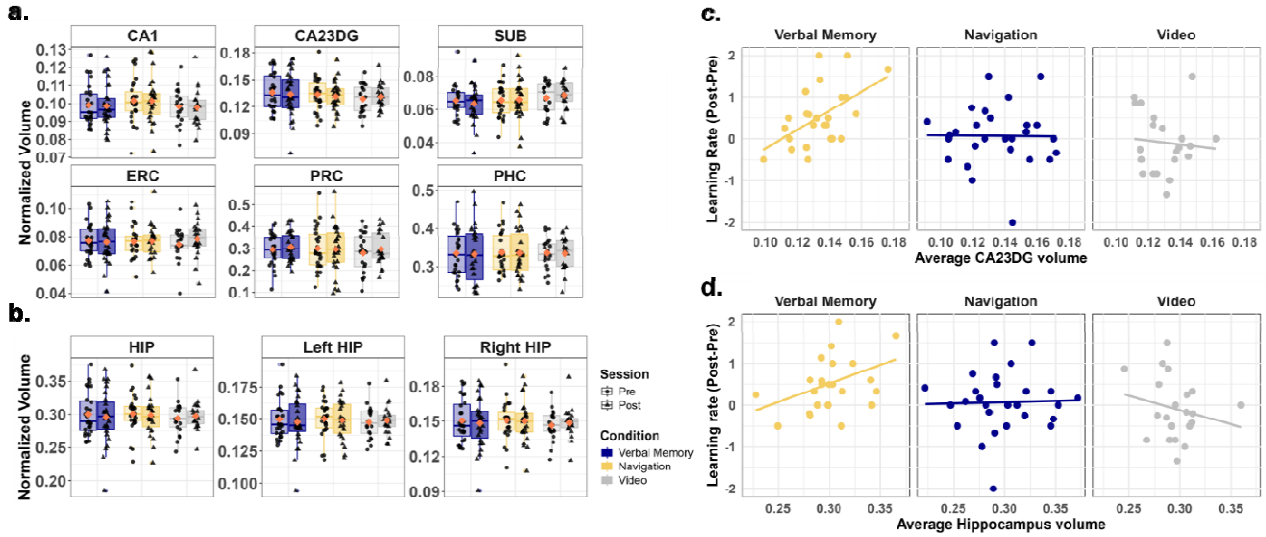

Fig. 3 Relationship Between Medial Temporal Lobe Volumes and Training-Induced Learning Rate Changes Across Conditions. (a) Normalized MTL subregion volumes across conditions (Verbal Memory/Navigation/Video Control) and sessions (pre-test/post-test). No significant changes were detected in CA1, CA23DG, SUB, ERC, PRC, or PHC volumes between pre-test and post-test. (b) Normalized hippocampal volumes (HIP), including left and right hippocampus, across conditions and sessions. No significant changes in total, left, or right hippocampal volume from pre-test to post-test. (c) Correlations between changes in learning rate (post-test minus pre-test) and average CA23DG volume across groups. A significant positive correlation was observed in the Verbal Memory group, suggesting that participants with larger CA23DG volumes exhibited greater improvements in verbal memory performance. No such correlation was observed in the Navigation group or Video Control group. (d) Correlations between the changes in learning rate (post-test minus pre-test) and average total hippocampal volume across groups. A significant positive correlation was observed only in the Verbal Memory group.

Task-related fMRI Results

Task-related informational connectivity analyses suggested network changes specific to the Navigation and Verbal Memory training groups

Although we did not observe changes in hippocampal volume as a result of training, it is likely that other brain-related changes might be mediating the improvements specific to the navigation and verbal memory training. We investigated this issue by collecting fMRI data both pre- and post-training during a source memory encoding (Fig. 4c) and retrieval task (Fig. 5a). During the encoding phase, participants learned objects by placing them within either an autobiographical timeline (temporal) or a familiar spatial layout (spatial). During the retrieval phase, they saw the objects again and indicated which of the “sources” (spatial or temporal) they had placed the object in. Because we trained the Verbal Memory group to use the temporal method of loci but trained the Navigation group to learn spatial locations, we reasoned that training-specific strategies might lead to context-specific changes (i.e., depending on whether the object was encoded in time or space) that would be reflected in network changes during the source memory tasks. We performed a univariate analysis across the whole

brain and an informational connectivity analysis using 242 pre-defined brain regions involved in spatial processing and memory (see Methods, Fig. 4b). Informational connectivity involved correlating multivariate patterns between trials within each ROI and then correlating again between all 242 ROIs (Fig. 4a). Informational connectivity could reflect changes in network configurations during source memory encoding and retrieval based on skills gained from training.

Encoding phase: Decreases in informational connectivity specific to the Navigation group and increases specific to the Verbal Memory group

We first considered both spatial and temporal source memory during the encoding phase, which could reveal how training affected the encoding of novel object-source pairings more broadly. For the univariate analyses, we found some changes from pre-test to post-test, but no interactions between session and training condition (see Supplementary Note 7, Supplementary Table 16 and Supplementary Fig. 6). When considering spatial or temporal encoding separately, we also found changes from pre-test to post-test, but no interaction between session and training group (see Supplementary Note 7, Supplementary Table 16 and Supplementary Figure 6). All statistical analyses were corrected for multiple comparisons (see Methods).

Informational connectivity changes between pre-test and post-test were then compared across the three training groups. For the Verbal Memory group, we observed significantly increased informational connectivity at post-test compared to pre-test (i.e., post > pre), relative to the combined Navigation and Video Control groups (Fig.4). Specifically, we found enhanced connectivity between the left precuneus and left angular gyrus ($t = 3.94$, $p < 0.05$), the right lateral occipital cortex (LOC) and left intracalcarine (visual) cortex ($t = 3.63$, $p < 0.05$), and the right LOC and left superior parietal lobule (SPL) ($t = 4.16$, $p < 0.05$). In contrast, no significant decreases in informational connectivity (i.e., post < pre) were observed. Considering spatial and temporal encoding separately, for the Verbal Memory group during spatial encoding, we found enhanced informational connectivity from pre-test to post-test (i.e., post > pre) between the frontal cortex and visual cortex (see Supplementary Note 6 and Fig. 4d). For temporal encoding, increased connectivity from pre- to post-test (i.e., post > pre) was observed between the frontal cortex and temporal gyrus (Fig. 4d). No significant decreases in informational connectivity from pre-test to post-test (i.e., post < pre) were observed in the Verbal Memory group relative to the combined Navigation and Video Control groups during either spatial or temporal encoding (Fig. 4d).

Conversely, for the Navigation group compared to the combined Verbal Memory and Video Control groups, we observed significantly decreased informational connectivity at post-test compared to pre-test (i.e., post < pre). Specifically, reduced connectivity was observed between the left paracingulate gyrus and right middle frontal gyrus (MFG) ($t = 3.80$, $p < 0.05$), left superior frontal gyrus (SFG) and right supramarginal gyrus (SMG) ($t = 5.14$, $p < 0.05$), left frontal pole and right MFG ($t = 3.85$, $p < 0.05$), left dorsal LOC and right ventral LOC ($t = 3.7$, $p < 0.05$) and right MFG and left MFG ($t = 3.79$, $p < 0.05$). In contrast, no significant increases in connectivity were observed from pre- to post-test (i.e., post > pre) for the Navigation group relative to the combined Verbal Memory and Video Control groups (Fig. 4d, bottom-left). When considering spatial and temporal encoding separately, during spatial encoding, reduced

connectivity was observed from pre-test to post-test (i.e., post < pre) between the primary visual cortex and higher-order visual areas, while increased connectivity was identified from pre-test to post-test (i.e., post > pre) between the parietal and temporal cortices. For temporal encoding, the Navigation group exhibited significant decreases in connectivity from pre-test to post-test (i.e., post < pre) across multiple regions, including the frontal and parietal cortices (see Supplementary Note 6), with no significant increases in connectivity from pre-test to post-test (i.e., post > pre). No significant changes in task-related informational connectivity to the hippocampus survived after multiple comparisons correction for either navigation or verbal memory training. Broadly, these findings suggest changes in informational connectivity specific to each training group (i.e., decreases in informational connectivity from pre-test to post-test in the Navigation group and increases in the Verbal Memory group), with no changes specific to the hippocampus.

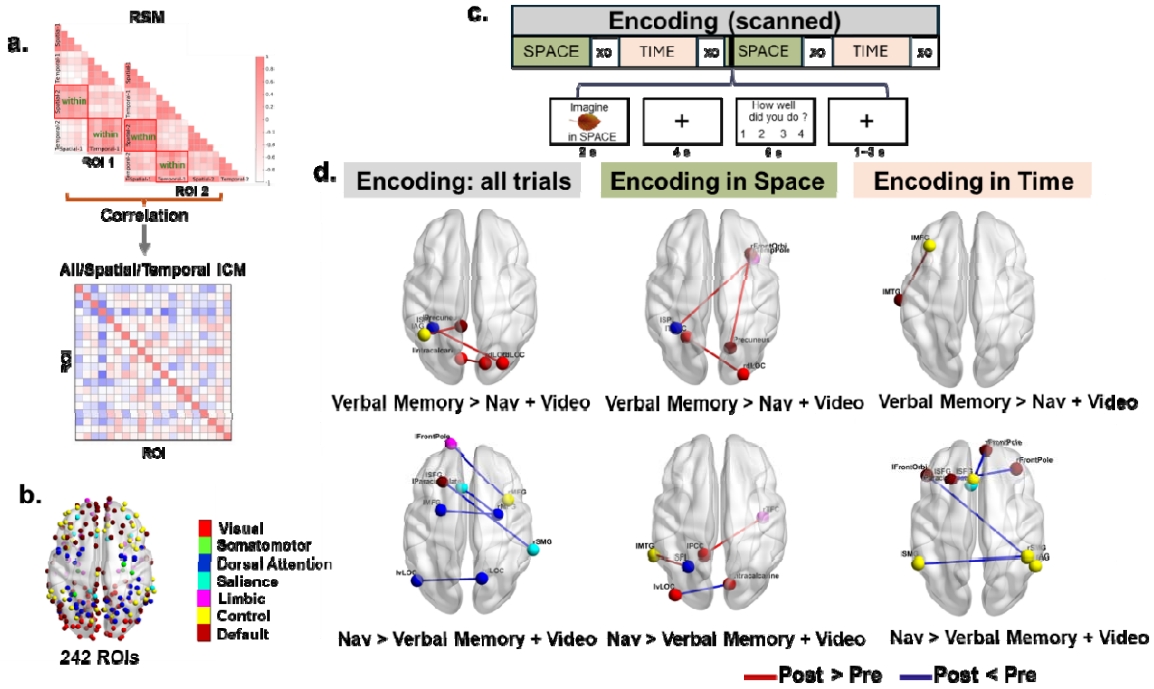

Fig. 4 Task-related informational connectivity changes during encoding as a result of verbal memory and navigation interventions. (a) Representational similarity matrices (RSMs) that illustrate within-context correlations (spatial and temporal) for each of the 242 brain ROIs. Informational connectivity between 242 ROIs was derived by correlating the RSMs across regions within different contexts, resulting in three types of informational connectivity matrices (ICMs): the spatial ICM, the temporal ICM, and the combined ICM for all within-context trials.(b) Visualization of the 242 predefined ROIs, color-coded by functional networks. (c) Schematic of the experimental design for the source memory task during the encoding stage, highlighting tasks related to spatial and temporal contexts. (d) Differences in task-related informational connectivity during encoding (Verbal Memory > Navigation + Video; Navigation > Verbal Memory + Video) across training conditions. Results are shown for all trials, as well as separately for spatial and temporal encoding contexts. Red lines indicate regions with significantly increased connectivity (post > pre), while blue lines indicate regions with

significantly decreased connectivity (post < pre). The top panels display results for the Verbal Memory group, while the bottom panels display results for the Navigation group. Nav: Navigation

Retrieval phase: Increases in information connectivity specific to the Navigation group and decreases specific to the Verbal Memory group

For the univariate analyses for both spatial and temporal source memory retrieval, no significant changes were observed from pre-test to post-test, nor were there any interactions between session and training condition (see Supplementary Table 16). When examining spatial and temporal retrieval separately, we similarly found no significant changes from pre-test to post-test and no interactions between session and training group (see Supplementary Table 16).

We next examined informational connectivity changes from pre-test to post-test during the source memory retrieval stage (Fig. 5a), finding a pattern somewhat opposite to that seen during encoding. In the Verbal Memory group, we observed significantly decreased informational connectivity in post-test compared to pre-test (i.e., post < pre) relative to the combined Navigation and Video Control groups. These reductions were identified between the right inferior frontal gyrus (IFG) and right frontal pole ($t = 4.01$, $p < 0.05$), left frontal pole and right SFG ($t = 4.48$, $p < 0.05$), right medial frontal gyrus and right posterior cingulate cortex (PCC) ($t = 4.42$, $p < 0.05$), right ventral LOC and left temporo-occipital fusiform cortex (TOFC) ($t = 4.30$, $p < 0.05$), and right lingual gyrus and left SPL ($t = 3.82$, $p < 0.05$). No significant increases in connectivity from pre-test to post-test (i.e., post > pre) were observed in the Verbal Memory group relative to the combined Navigation and Video Control groups (Fig. 5b, top-left). Considering spatial and temporal source retrieval separately, spatial retrieval showed decreased connectivity from pre- to post-test (i.e., post < pre) between frontal regions (e.g., left and right superior frontal gyri) and between the right frontal pole and PCC, and increased connectivity from pre to post (i.e., post > pre) between temporal and parietal regions (e.g., left middle temporal gyrus [LMTG] and right SMG) and between the visual cortex and right SMG. For temporal retrieval, decreased connectivity from pre-test to post-test (i.e., post < pre) was prominent across multiple regions, including the frontal pole, angular gyrus, precuneus, and lateral occipital cortices. One increase specific to the Verbal Memory group was observed from pre-test to post-test (i.e., post > pre), however, in the left ventral lateral occipital cortex and left MFG (Fig. 5b, see details in Supplementary Note 6).

In the Navigation group during retrieval, we found significantly increased informational connectivity in the post-test compared to the pre-test (i.e., post > pre) relative to the combined Verbal Memory and Video Control groups. These increases were found between the right occipital pole and dorsal LOC ($t = 4.27$, $p < 0.05$), and the right SPL and right angular gyrus (AG) ($t = 4.09$, $p < 0.05$). Considering spatial and temporal retrieval separately, spatial retrieval showed enhanced connectivity from pre-test to post-test (i.e., post > pre) between regions such as the left frontal pole and the occipital and parahippocampal cortices. Temporal retrieval revealed increases from pre- to post-test (i.e., post > pre) between the precuneus and SFG, lingual gyrus and occipital gyrus, and SFG and MTG. No significant decreases in connectivity were observed from pre-test to post-test (i.e., post < pre) in the Navigation group relative to the combined Verbal Memory and Video Control groups (Fig. 5b, bottom).

Again, no significant changes in task-related informational connectivity during source retrieval were present to the hippocampus when controlling for multiple comparisons. To summarize, verbal memory training resulted in decreased connectivity during retrieval across frontal, parietal, and occipital regions, especially during temporal retrieval, with limited increases. In contrast, navigation training consistently increased connectivity for both spatial and temporal retrieval, particularly between frontal, occipital, parietal, and temporal regions, with no significant decreases observed. These findings suggest that verbal memory primarily alters network connectivity during temporal contextual retrieval, whereas navigation training enhances connectivity more broadly across both contexts.

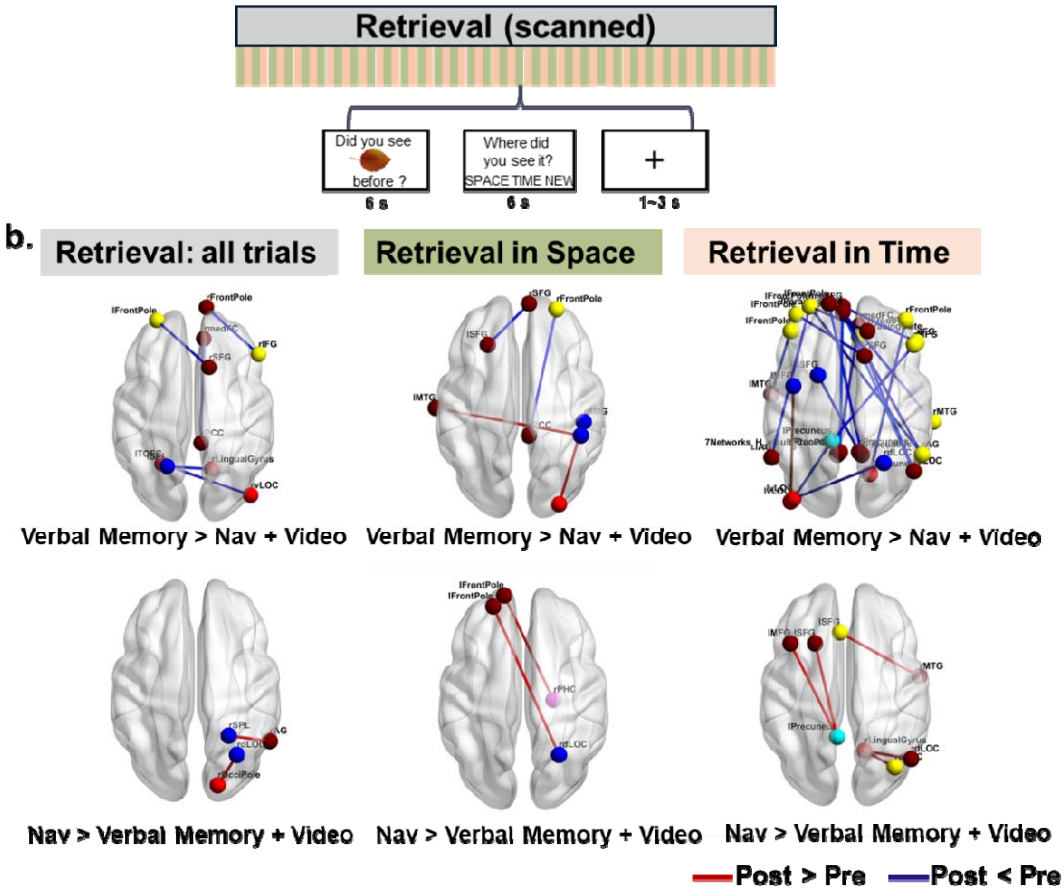

Fig. 5 Task related informational connectivity changes during source retrieval as a result of verbal memory and navigation interventions. (a) Schematic of the source memory task structure during retrieval. Participants were scanned during source retrieval, in which they identified whether an item was previously seen and determined its spatial or temporal context. (b) Differences in task-related informational connectivity during source retrieval (Verbal Memory > Navigation + Video; Navigation > Verbal Memory + Video) across training conditions. Results are shown for all trials, as well as separately for spatial and temporal encoding contexts. Red lines indicate regions with significantly increased connectivity (post > pre), while blue lines indicate regions with significantly decreased connectivity (post < pre). The top panels display results for the Verbal Memory group, while the bottom panels display results for the Navigation group. Nav: Navigation

Specific changes in network-wide informational connectivity pattern distance during retrieval: Enhanced due to verbal memory but not navigation training

The above analysis focused on changes in informational connectivity between pre-test and post-test for individual connectivity edges between pairs of ROIs. Next, we examined changes in multivariate informational connectivity patterns, which we defined as the connectivity between a single ROI, and the remaining 241 ROIs. This analysis helped address network-wide changes in connectivity rather than those restricted to two different ROI pairs. Multivariate informational connectivity patterns were extracted separately for spatial and temporal informational connectivity matrices (ICMs), and the distance between the two patterns was computed using 1 - r (Pearson's, see Methods).

While no significant changes in pattern distance were observed during the encoding stage for any of the three conditions, a greater distance between spatial and temporal patterns in the left SFG network connectivity during source retrieval was found in the Verbal Memory group at post-test compared to pre-test ($t(23) = -4.275$, $p = 0.043$, FDR corrected; Fig. 6b). This effect was not observed in the Navigation and Video Control groups (Fig. 6c). A similar result was obtained using the Spearman correlation to compute the distance, with a greater distance in the left SFG for the Verbal Memory group at post-test relative to pre-test ($t(23) = -4.400$, $p = 0.050$, FDR corrected). No significant differences were found in the Navigation or Video Control groups. These findings also suggest some network-wide changes that are specific to verbal memory training.

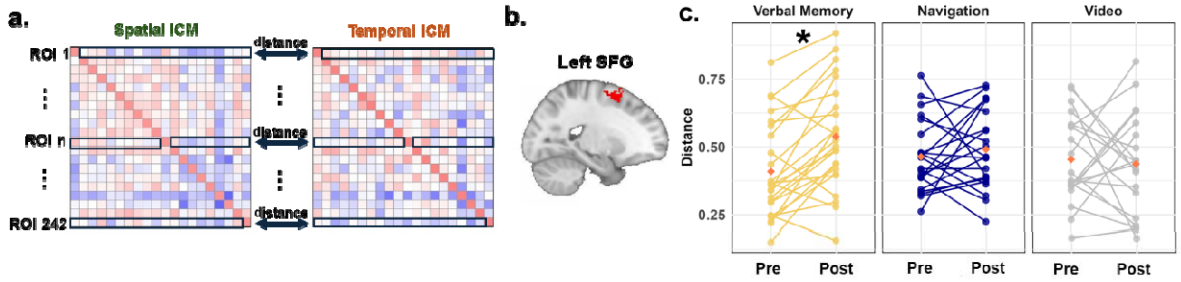

Fig. 6 Multivariate informational connectivity pattern distance between pre-test and post-test. (a) Schematic of multivariate distance analysis. For each ROI, informational connectivity between the current ROI and the remaining 214 ROIs was extracted from both spatial and temporal ICMs, and distances were calculated as 1 minus the Pearson or Spearman correlation coefficient. (b) and (c) Among all 242 ROIs, only the SFG showed a greater distance between spatial and temporal ICMs in post-test compared to pre-test in the Verbal Memory group, but not in the Navigation or Video Control groups. ICM: informational connectivity matrix. *$p < 0.05$, FDR corrected.

Discussion

Acquiring complex cognitive skills is a fundamental human capability, but we have little knowledge about the neural changes that accompany such new learning. As suggested in past work, we are capable of remarkable improvements in performance across a wide range of domains, yet the neural basis for these changes remains

unclear. Here, we build on expert training regimens in verbal memory (e.g., the method of loci) and spatial navigation (e.g., London taxi driver training) by employing a within-subject intervention approach to determine the neural correlates of training two related but distinct complex skills: navigation and verbal memory. Our results show that non-expert young adults substantially improved in the task they were trained on and exhibited some limited near transfer (to new lists of words or new environments). Unlike the neuroimaging literature on experts or impaired populations (like older adults or individuals with dementia who cannot perform these tasks as well), we found no evidence that training modified brain structure, as measured by hippocampal volume and DWI microstructure in relevant GM regions and WM tracts. Rather, analyses of task-based functional imaging data showed training-specific differences in informational connectivity, a neural metric that indexes how similar representations are across brain regions[41]. Our findings have implications for both basic research and clinical translation. From a basic research perspective, our results suggest that structural changes in gray and white matter are relatively insensitive to cognitive intervention but are instead reflected in more transient changes to network-based connectivity patterns. From a clinical-translational perspective, our results suggest promise in training new complex cognitive skills such as navigation and verbal memory, although the efficacy of our specific training regimen in clinical populations remains to be tested.

Our study involved two novel and extensive in-lab training regimens, which shared key features but allow for dissociable cognitive training. While previous training studies have examined network configurations before and after participants learned the method of loci or evaluated navigation training, these experiments often involved expert groups (memory athletes or London taxi drivers)[3,42], or older adults[9]. In our experiment, we evaluated healthy young adults while carefully controlling the duration and nature of the verbal memory and navigation trainings to match the intensity and the degree to which the skill was scaffolded. In this way, one training serves as a control for the other, while also allowing us to compare with an active control group who did not gain a complex skill (video control) but was passively exposed to similar material.

Both training regimens were effective: participants in the Verbal Memory group were able to learn longer word lists, while participants in the Navigation group were able to reduce the length of their paths and required fewer trials to learn landmarks in new subsections of the environment. Yet, it was difficult to ascertain whether either training generalized beyond immediate transfer. Our near transfer measures exhibited strong individual differences at baseline, which were difficult to overcome with 10 sessions of training. Both near transfer measures also likely had practice and ceiling effects, which dampened our ability to measure the effects of training (e.g., Supplementary Figure 2). Exploratory analyses, however, revealed training-specific differences in the learning rate on the transfer tasks from pre- to post-test. In addition to good face validity as a match for the skills we were training (that is, for example, in the navigation training, participants learned how to quickly acquire spatial information about the whole environment after only 10 trials compared to those who did not undergo that training), the change in learning rate from pre- to post-test was also the only dependent measure we found to correlate with baseline hippocampal volume. These findings, though exploratory, suggest that baseline hippocampal volume may enable more receptiveness to training, aligning with concepts like neural reserve[43]. In addition to replicating these findings,

future research using these training regimens should develop a more comprehensive near-transfer measure and especially evaluate far transfer to completely untrained tasks.

Our navigation training is particularly notable in that it partially mimics the training undertaken by London taxi drivers, who first learn individual subsections of an environment before connecting these subsections to one another. While other virtual navigation trainings exist[44,45], their environments are often small, simplified, or non-hierarchical. Thus, a key contribution of this work is the development of a navigation training environment that is suitable for training wayfinding skills. We also developed a novel verbal memory training paradigm that has not been reported previously in the literature. Using this training, we were able to dramatically increase the number of freely recalled words of the average participant, with the typical participant showing an improvement from about 25-30 words freely recalled (repeated) words on day 1 to 60-65 words after only five days of training.

Despite the efficacy of the training, neither the Navigation nor the Verbal Memory group showed macroscopic neural structural differences in the hippocampus, which has been linked to verbal memory and/or navigation function in past studies[3,38,46]. These results supply causal evidence supporting the notion that hippocampal structural changes may not impart superior navigation or verbal memory skills. Although previous work has shown effects of training, to our knowledge, this work has occurred in specialized populations or involved between-subject comparisons. For example, London taxi drivers who undergo extensive training to learn the city of London show increased hippocampal volume from pre- to post-training compared to taxi drivers who fail to complete the training[3]. Older adults who undergo an exercise-only training or a navigation + exercise training show reduced hippocampal atrophy (i.e., less volume loss) compared to a no-contact control[47]. Here, in healthy younger adults, who nevertheless showed robust individual differences in navigation and verbal memory and showed strong evidence of improvement across both types of training, there were no observed structural changes to hippocampal gray or white matter tracts, nor were there any significant structural changes to other brain regions.

In exploratory analyses, informational connectivity analyses on fMRI data from a spatial and temporal source memory task revealed unique network reconfiguration signatures as a result of navigation and verbal memory training. These changes were based on a strict correction for multiple comparisons and suggest that verbal memory and navigation training results in robust changes in task-related functional connectivity patterns. Interestingly, the vast majority of these changes were to areas outside of the medial temporal lobe, an area often considered central to both navigation and memory. During encoding, we observed increases in connectivity due to the verbal memory training primarily in the frontal and occipital lobes. Conversely, we observed decreases in informational connectivity in some of these same areas as a result of navigation training. We observed somewhat the opposite pattern during retrieval: increases in informational connectivity during navigation and decreases during verbal memory training. Again, many of these connectivity changes were areas within the frontal cortex, with some longer-range changes to parietal, parahippocampal, and occipital cortex. Finally, a multivariate approach centered around superior frontal gyrus as a hub revealed differences in network connectivity patterns for spatial compared to temporal

retrieval, suggesting task-related network-level changes due to the training as well. Similarly, we note that two studies, which trained participants on an n-back task, also observed changes in dynamic functional connectivity in frontal and parietal cortices[48,49].

Thus, our results suggest that the networks most malleable during acquisition of complex cognitive skills largely involve areas related to cognitive control, with one possibility being that such changes allow modulation of existing networks that include the medial temporal lobe[50]. At the same time, navigation and verbal memory training resulted in differential changes within these frontal-parietal networks, suggesting that such cognitive skills involve at least partially dissociable changes. While it is reasonable to conjecture that microstructural changes in brain tissue (as measured with DWI) or even macrostructural gray matter changes may occur after considerable expertise is developed, such structural reorganization would likely be preceded by network changes in functional connectivity. In this scenario, functional changes that reflect training effects could potentially become solidified through long-term maintenance or the development of domain-specific expertise.

Methods

Preregistration and Power Analysis

This study was preregistered on the Open Science Framework (OSF; https://osf.io/etxvj). The preregistration was completed on October 30, 2023, after approximately 80% of data collection had been completed, but prior to any analysis of the primary outcome variables. The preregistration outlines the study hypotheses, design, target sample size, and planned behavioral and neuroimaging analyses, including longitudinal ROI comparisons and statistical correction procedures.

A priori power analysis was conducted using G*Power 3.1 to estimate the required sample size for detecting a Group × Time interaction in a mixed-design ANOVA. Assuming a small-to-medium effect size ($f = 0.35$), we determined that 24 participants per group would provide 80% power to detect a significant effect at $\alpha = 0.05$. To allow for potential attrition and data exclusion (e.g., due to excessive motion or incomplete datasets), we targeted recruitment of 30 participants per group across two study sites.

All primary hypotheses, analytic plans, and inference criteria are documented in the preregistration. Exploratory analyses are clearly delineated in both the preregistration and the present manuscript.

Participants

83 participants between the ages of 18-45 were recruited for enrollment in the study across two sites (University of Florida and University of Arizona). Participants were recruited through advertisements and flyers posted on campus. After recruitment, participants were screened for: 1) being right-handed; 2) having no metal in their bodies or other MRI contraindications; 3) having no COVID-19 symptoms. Participants also completed and passed an MRI prescreen before fMRI scans. This study included 75 participants (Supplementary Table 1) who fully completed the entire study: 27 in the verbal memory condition, 27 in the navigation condition, and 21 in the video control condition. One verbal memory participant, despite completing training, was excluded

from MRI scanning due to ear piercings, resulting in 26 participants for brain structure and connectivity analyses. Some participants were excluded from certain tests due to outlier performance, excessive head movement during scanning, or missing data. For detailed sample size information for each test, please refer to Supplementary Table 2.

Overview

Participants were recruited via flyers or word of mouth. During prescreening, each participant was randomly assigned to one of three training conditions (Navigation, Verbal Memory, or Video Control). Participants then completed a prescreening phone call for MRI eligibility, COVID safety, and availability to determine participation in the study. They then received instructions for their first session verbally and via email. The study took place over 12 sessions, each of which occurred on a separate day. These 12 sessions took place over the course of three to four weeks (average duration = 4.116 weeks). The first and last sessions were 3-hour sessions, referred to as pre-test and post-test, and consisted of behavioral and neuroimaging measures. These sessions were identical for all participants, regardless of which condition they were assigned to. Following the pre-test session, participants completed ten 2-hour training sessions, during which they completed training based on their assigned condition. Following training, participants completed the post-test session, were debriefed and released. Participants were paid for their participation. See Supplementary Table 3 for more details about conditions and scheduling.

Training Conditions*: Navigation Training*

Participants assigned to the Navigation group navigated a virtual environment in Unity 3D called “Virtual Arida” using an Xbox game controller on a desktop computer. Virtual Arida has six individual subsections (A, B, C, D, E, F), with each containing eight different target locations. Participants began training on day 1 in subsection A and were spawned in front of a target location. They were then asked to navigate using the shortest possible route to a second target location. If the participants did not perform a sufficiently direct route, they repeated the route; their route was determined to be sufficiently direct if it did not exceed 120% of the shortest possible path between two buildings. Once they successfully employed a sufficiently direct route, they moved on to the next route. Participants learned all 28 possible routes in the subsection. When they learned all routes successfully, they completed random target location pairings for the remainder of the 2-hour session. After completing a subsection, participants moved on to a new subsection. In some cases, the new “subsection” involved navigating between two previously learned subsections. For example, once individual subsections A and B were completed, participants navigated between targets in subsection A and subsection B. This involved “integrating” their previously learned knowledge from prior subsections (A+B). Participants then learned the next individual subsection (subsection C), and after that, the next integrated subsection A+B+C. This continued in this manner through the final integrated subsection A+B+C+D+E+F (for those participants that reached this far).

Training Conditions*: Verbal Memory Training*

Participants placed in the verbal memory condition completed free recall training on a desktop computer using a task built with Unity 3D. Similar to the verbal memory

transfer task during pre/post-test, in this task participants were shown a list of words from the Toronto word pool one at a time, which they memorized using a trained strategy (the "temporal" method of loci)[51]. After being shown the list of words, they completed a distractor task by typing the numeric answer for five simple addition/subtraction questions. Then, they recalled the words by typing them into the provided space. The strategy that participants were given to memorize the words was a modified method of loci termed the "temporal method of loci"[51]. At the beginning of Day 1 of training, participants wrote 30 of their most memorable, vivid, and non-traumatic memories from their life thus far. They were asked to order these memories into a mental timeline to help them with the word recall task. Participants were asked to associate each word they were given with a memory from their mental timeline. Then, when they reached the recall portion of the task, participants were to use their mental timeline memories to help them recall the associated words in the current list. The number of words increased with each successful trial. If there were multiple unsuccessful trials, the number of words decreased. Once they were successful with this shorter list, the previous list was repeated. For example, if the list was "red," "orange," "yellow," and the participant did not recall all 3 words, the next trial's list would be "red," "orange." On successful completion, the next trial would be "red", "orange," "yellow," "blue."

Session 1-5 of verbal memory training (same words list)

For the first five sessions of training, participants recalled a list that increased in length on each trial when they successfully recalled all the words. This was done using an algorithm based on a power law learning function. For example, they could start with the words: "Jersey," "rifle," "pony." Upon successful completion of this list, the list would grow to "Jersey," "rifle," "pony," "castle." (Fig. 1a).

Session 6-10 *of verbal memory training (new word list)*

For the final 5 sessions of training, once participants successfully recalled a list, they learned a completely new list of words that was increased in length in the same manner as described above. For example, they could start with the words: "sister," "soldier," "beauty." Upon successful recall of this list, the next trial, the next list could be "mother," "genius," "number," "sailor." (Fig. 1b).

Active Control Condition- Video Watching

Participants in this condition watched a wide range of videos on a desktop computer. The types of videos varied, but included interviews, lectures, documentaries, shows, and YouTube videos covering different aspects of memory and navigation. Participants were given a set of questions for each video to ensure that they paid attention. Videos and quiz questions were administered using Qualtrics.

Pre-test overview

In the Pre-test session, participants provided informed consent, completed a demographics questionnaire, and confirmed their MRI prescreen information. Participants then completed pre-test behavioral measures, which included a navigation transfer task (Virtual Silcton), a verbal memory transfer task (a new set of words from

the Toronto word pool), and an attention task (Posner cueing task). Next, the experimenter described the source memory task that would take place during the fMRI scan. The participants created the temporal and spatial contexts for this task, then after a short break to prepare for the scanner, they completed the source memory task in the scanner (see details in Supplementary Note 1 and Supplementary Table 3).

Post-test overview

The post-test session was similar to the pre-test session. The only changes in the post-test session were different words for the verbal memory transfer task, a different section of Virtual Silcton, and different images for the source memory task. After reviewing the MRI prescreening form for any changes, participants completed the post-test measures (navigation, verbal memory, and attention tasks). Participants then reviewed the spatial and temporal contexts created in the pre-test session, took a short break, and completed the source memory task in the scanner. Finally, participants completed a debriefing survey and received payment (see details in Supplementary Note 1 and Supplementary Table 2).

Pre-post tests

Navigation Transfer task: Virtual Silcton

The pre-test and post-test navigation transfer task was completed on a laptop computer with an Xbox game controller using the "Virtual Silcton" environment recreated in Unity 3D[37] (http://www.virtualsilcton.com/). Participants were instructed to freely navigate from their starting location to one of the eight possible destination buildings within the environment. If the participant walked for more than 60 seconds during a trial, a compass appeared at the bottom of the screen to point them to the target building. Over the trials, participants learned the spatial environment from the routes taken between the buildings. Participants completed trials until all routes were completed, or until 30 minutes elapsed. After the learning phase, participants completed an on-site pointing task, in which they were placed next to one of the eight buildings and had to point to each of the other seven in a random order. Next, they completed a model-building task, in which they dragged and dropped images of the eight buildings around an on-screen box. These tasks demonstrate configural knowledge of the environment.

Verbal Memory Transfer task

Participants memorized and recalled a list of 12 words from the Toronto Word pool. The structure of the task was the same as that for the verbal memory training task although it was programmed in PsychoPy. All participants had a maximum of five attempts to recall all 12 of the words correctly.

Source Memory task

Prior to entering the scanner, participants were asked to generate a *temporal* and a *spatial* context to use during the source memory task. For their temporal context, they were asked to create a list of 10-15 salient memories of events from their life. We then asked them to write these down on a piece of paper and arranged them into a mental timeline of life events. For their spatial context, participants drew a blueprint of their

living space (e.g., dorm room or apartment) on a second piece of paper. Participants were given time to review their contexts to ensure they knew them well enough prior to entering the scanner. They were then instructed on how they would use these contexts to recall objects. They were provided with an example of how to use each context and practiced verbally by using these contexts verbally.

Participants underwent functional MRI scanning while performing the source memory task. This task was comprised of two phases, each with four runs. During the encoding phase (runs 1-4), participants were presented with images and instructed to encode them within either a temporal or spatial context, consistent with prior out-of-scanner practice. During the retrieval phase (runs 5-8), participants were presented with images and asked to make two judgments: first, whether the image was new, and second, if recognized, the context in which it had been previously encoded. All participants viewed the same images, but in counterbalanced orders.

Source Memory Encoding

This source memory task included 4 encoding and 4 retrieval runs. In the encoding task, participants viewed an image and were asked to imagine it in either "time" or "space", referring to their previously generated temporal or spatial contexts. Participants were to either visualize the image as a part of their mental timeline if it was to be imagined in "time" or visualize this object existing in their living space if it was indicated that it was to be imagined in "space". Following this visualization period, participants were asked to indicate how vividly they visualized this item on a Likert scale by pressing 1-4 on the button box (see image below). All trials were jittered. At the end of the block, participants completed a short odd/even task. In this 60-second task, an "X" or an "O" would be displayed. Participants pressed "1" if it was an "X" and "2" if it was an "O". This served as a baseline task for hippocampal activation[52].

Source Memory Retrieval

In the retrieval task, participants were shown an image of an object and asked whether they recognized it from the encoding phase or if it was a new image by pressing "1" for recognition and "2" for new. They were then asked if this image was previously viewed in "time", or "space", or if it was new by pressing "1", "2" or "3", respectively. Four retrieval runs were completed. All trials were jittered. Detailed descriptions of other pre-post tests are provided in Supplementary Note 1.

Behavioral Data Analysis

Navigation task (navigation training in Virtual Arida)

To quantify the training effect, we analyzed the average distance error travelled, normalized by the threshold (120% of the length of the shortest possible path). We removed all trials that occurred after all unique pairs had been learned. Because we controlled the amount of time per task (2 hours per session), not all participants progressed to the same extent. Repeated-measures ANOVAs were conducted with subsection or integration as the within-subjects factors and threshold-normalized distance error as the dependent measure.

Verbal Memory task (verbal memory training with the modified method of loci)

To quantify the training effect, we performed linear regression analysis on the maximum number of correctly recalled words per day. Regression slopes were calculated separately for the first five and last five days (Fig. 1a right), reflecting the distinct training paradigms implemented during these two phases (see Methods).

Learning rate in the Navigation Transfer task (Virtual Silcton) during the pre-/post-test

To assess navigation performance in the Navigation Transfer task, for each trial, we calculated distance error by subtracting the optimal distance from the participant's actual navigated distance between their starting location and the target building. Since distance errors are inherently influenced by the distance between the starting location and the target building, we normalized these errors by dividing each distance error by the corresponding optimal distance. The task consisted of 20 trials, during which participants progressively learned the environment. As expected, normalized distance errors decreased over time, indicating improved navigation performance in later trials. To quantify this learning effect, we calculated a learning rate, defined as the negative of the difference between the normalized distance errors of the first 10 trials and the last 10 trials, divided by their sum (see Eq. 1). A higher/positive learning rate suggests relatively stable and consistent performance across early and late trials whereas a lower/negative learning rate reflects a greater difference in performance between early and late trials.

$$Learning\ Rate = -\left(\frac{Distance\ errors\ of\ first\ 10\ trials - Distance\ errors\ of\ last\ 10\ trials}{Distance\ errors\ of\ first\ 10\ trials + Distance\ errors\ of\ last\ 10\ trials}\right) \quad \text{[Eq. 1]}$$

Learning rate in the Verbal Memory Transfer task during the pre-/post-test

For the Verbal Memory Transfer task, we calculated a learning rate to quantify memory performance. To determine the learning rate, we first calculated the difference in the number of words correctly recalled between the first and last trials. This difference was then divided by the total number of trials to yield the learning rate (see Eq. 2).

$$Learning\ Rate = \frac{Last\ trial - First\ trial}{\text{number of trials}} \quad \text{[Eq. 2]}$$

Correlation analysis

To examine the relationship between learning rate of either task and brain structure (volume/DWI, see below), we collapsed the session factor by averaging values from pre- and post-training sessions for each ROI, given the absence of significant session effects or interactions between session and training condition. We conducted Pearson correlation analyses (parametric) for data that met the normality assumption based on the Shapiro-Wilk test. If normality was violated, Spearman correlation analysis (nonparametric) was used instead. Sex and site were included as covariates in the corresponding partial correlation analyses.

fMRI Data Analysis

MRI data acquisition

Scanning was performed using a 32-channel 3T Siemens “Skyra” scanner at the University of Arizona and 32-channel 3T Siemens “Prisma” scanner at the University of Florida. Visual stimuli were presented on a screen positioned behind the scanner and viewed by participants through a mirror attached to the head coil. Stimuli and responses were presented and collected using PsychoPy (https://www.psychopy.org) running on a Windows 10 laptop. High-resolution anatomical images of the hippocampus and surrounding cortex were acquired with a T2-weighted turbo-spin echo (TSE) anatomical sequence (FOV□=□200□mm□×□200□mm, matrix□=□448□× 448, TR□=□4200.0□ms, TE□=□93.0□ms, flip angle□=□139 degree, slice thickness□=□1.9□mm, 28 slices, bandwidth□=□199□Hz/pixel). High-resolution structural images of the whole brain were obtained using a 3D, T1-weighted MPRAGE (1□mm^3 isotropic) sequence (FOV□=□256□mm, matrix□=□256□×□256, slice thickness□=□1□mm, TR□=□2300□ms, TE□=□2.41□ms, flip angle□=□8 degree, bandwidth□=□330□Hz/pixel). Functional images were acquired using a whole-brain echo planar imaging (EPI) sequence (TR□=□1560□ms, TE□=□30□ms, flip angle□=□70 degree, field of view (FOV)□=□220□mm, matrix□=□88□×□88, slice thickness□=□2.5□mm, slices□=□48, bandwidth = 2030 Hz/pixel), involving a voxel resolution of 2.5 × 2.5 × 2.5 mm. Diffusion Weighted Imaging (DWI) data were acquired using two sequences with opposite phase-encoding directions to correct for distortion artifacts without signal loss. Acquisition parameters included: TR = 9200 ms, TE = 86 ms, FOV = 256 mm, 30 diffusion directions, 60 slices with a thickness of 2.0 mm, and a voxel size of 2.0 × 2.0 × 2.0 mm. The diffusion-weighted images were obtained using an echo plane sequence with the following parameters: number of b0 images = 1, b-value = 1000 s/mm^2, number of directions = 30, TR□=9200□ms, TE□=□86□ms, and voxel size: 2 × 2 × 2 mm. High-resolution resting state images were acquired using another whole-brain EPI sequence(TR□=□3000□ms, TE□=□36□ms, flip angle□=□90 degree, field of view (FOV)□=□240□mm, matrix□=□160□×□160, slice thickness□=□2.5□mm, slices□=□48, bandwidth = 802 Hz/pixel), involving a voxel resolution of 1.5 × 1.5 × 2.5 mm. Resting state data were not considered in this manuscript.

DWI data preprocessing

We performed pre-processing of DWI data using a customized pipeline that combines tools from the fMRIB Software Library 6.0[53], Advanced Normalization Tools[54], and MRtrix3[55]. First, we removed Gaussian noise present in the DWI data by fitting a Marchenko-Pastor distribution to the signal matrices to generate a threshold for PCA denoising[56-58]. We then removed Gibbs-ringing artifacts that can occur at tissue borders such as the outer surface of the brain and near ventricles[55]. Next, we generated brain masks for the DWI images using the dwi2mask function from MRtrix3, which uses information from both diffusion-weighted and non-diffusion weighted (b=0) volumes to generate an accurate brain mask. We then corrected for Eddy current and movement related distortions using FSL Eddy[59-61]. Using the root mean square motion output provided by Eddy, we then applied a motion threshold, excluding participants with > 2 mm absolute displacement or > 0.5 mm relative displacement between diffusion weighted volumes. Application of this motion threshold at both timepoints resulted in 20

participants in the Verbal Memory group, 22 participants in the Navigation group and 18 participants in the Video control group in final DWI analysis.

Consistent with prior work[62-66], we reconstructed the DWI data using a bi-tensor model, in which freely diffusing water is modeled by one tensor and anisotropic water diffusion is modeled with a separate tensor after removing the contribution from isotropic free water. In this model, the free water (FW) compartment of each voxel is interpreted primarily as originating from extracellular water diffusion and the second tensor represents the tissue compartment after removing the contribution from FW. We performed whole brain estimation of FW and calculated free water corrected tensor metrics using custom MATLAB scripts (R2023a, The Mathworks, Natick, MA, USA). Briefly, we calculated FW from single shell diffusion data based on minimization of a variational regularization framework outlined in Pasternak et al. (2009). Initialization of the free water estimate in this pipeline used mean diffusivity (MD) maps calculated from a single tensor fit. Thus, prior to FW estimation we performed conventional single tensor reconstruction using FSL's DTIFIT. Next, conventional MD maps were for initialization of the bi-tensor reconstruction in MATLAB. Voxels with MD values greater than 0.8 x d (*i.e.,* d is constant diffusivity) were assumed to be comprised of CSF and omitted from the fitting process. Fitting a bi-tensor model with DWI data from a single non-zero diffusion weight (i.e. b-value) has been described as an ill-posed problem with nearly infinite solutions, thus we employed sensible biological constraints to the minimization process, as performed in Pasternak et al. (2009). We set the reference MD_t at 0.6 $\mu m^2/ms$. and diffusivities were limited to λ_{max} = 2.5 $\mu m^2/ms$ and λ_{min} = 0.1 $\mu m^2/ms$. We also assumed isotropic water diffusion at 37° C, corresponding to human body temperature, as a constant (d = 3.0 x 10^{-3} $\mu m^2/ms$). After initialization, 100 iterations were used to refine the FW estimates corresponding free water corrected metrics. We used an automated quality assurance procedure and visually inspected output images (e.g., FW and fwcFA) to confirm appropriate data quality.

We used ANTs to perform nonlinear registration to MNI standard space by warping each participant's FA image to the HCP 1065 template[67]. We used participant's uncorrected FA for the registration process because the HCP 1065 template was created with FA images from single tensor reconstruction, not fwcFA images. We then applied the same transformation matrix to align FW and fwcFA images from the same participant into MNI space. Following registration, we extracted FW and fwcFA values from ROIs in MNI space. We separately evaluated ROIs from gray matter (GM) regions and white matter (WM) tracts, selected based on prior literature and relevance to navigation. Eight bilateral GM ROIs from the Mayo Clinic Adult Lifespan Template[68] were used: hippocampus, parahippocampal gyrus, entorhinal cortex, inferior temporal cortex, cuneus, superior parietal cortex, retrosplenial cortex, and caudate nucleus[69]. Seven bilateral WM ROIs from the Johns Hopkins University (JHU) white-matter tractography atlas were used: three subregions of the corpus callosum (body, genu, and splenium), fornix, posterior thalamic radiation, cingulum (hippocampus), and the fornix cres / stria terminalis (these two small tracts are not able to be differentiated at 2mm voxel resolution).

DWI statistics

Similar to previous work[64], we performed separate statistical analyses of ROIs from GM and WM regions, as these tissue types have dramatically different values for DWI metrics. We also separately evaluated FW and fwcFA, as these metrics provide separate but complementary information about extracellular and intracellular contributions to tissue microstructure. For each of these four combinations (GM-FW, GM-fwcFA, WM-FW, WM-fwcFA), we performed mix-designed ANOVAs with condition (Navigation/Verbal Memory/Video Control) as a between-subjects factor and ROI and session (pre/post) as within-subjects factors with corresponding post-hoc comparisons. We also performed correlations (Pearson or Spearman, as appropriate) to evaluate the relationship between DWI metrics and performance on behavioral measures. Correlations were corrected for multiple comparisons using the false discovery rate (FDR) method[70], which was applied across p-values for each ROI, performed separately for FW and fwcFA analyses.

MTL subfield demarcation

Automatic hippocampal subfield segmentation software (ASHS)[71,72] was used to segment the subfields of the MTL based on each participant's high-resolution T2-weighted MRI image. We used ASHS with the ASHS-Princeton-1.0.0-Young-Adult[72]. The MTL was segmented into CA1, CA2/3, DG, and subiculum (SUB), as well as the perirhinal cortex (PRC), entorhinal cortex (ERC), and parahippocampal cortex (PHC). We combined the CA2/3 and DG subfields as finer distinctions could not be made at the acquired resolution[73]. Each participant's subfield segmentations were manually inspected to ensure accuracy of the segmentation protocol. The hippocampus was further subdivided into anterior (head) and posterior (body + tail) regions along its longitudinal axis using the T1-weighted MPRAGE sequence and FreeSurfer 7.4.1 software (https://surfer.nmr.mgh.harvard.edu/). To ensure reliable volume estimates, images were processed using FreeSurfer's longitudinal pipeline [74]. This method creates an unbiased within-subject template using a robust, inverse-consistent registration approach[75]. Visual quality control was conducted by three trained raters who inspected the skull stripping, surface reconstruction, and segmentation accuracy at both the within-subject template and individual timepoints. Manual edits were primarily performed on the within-subject template to correct segmentation errors—particularly in challenging regions such as the hippocampus—since edits to the template automatically propagate to all associated timepoints. Raters followed standardized FreeSurfer longitudinal editing guidelines to ensure consistent correction across subjects and timepoints. Discrepancies between raters were resolved via consensus discussions. This approach enhanced segmentation accuracy and consistency across longitudinal scans, thereby improving the reliability of volumetric and morphometric analyses. These quality control procedures ensured that anatomical registration was accurate and that atlas-based ROI extractions were not confounded by segmentation or alignment errors. To account for individual differences in brain size, we normalized the volume of each MTL subregion by dividing each volume by intracranial volume (ICV) and scaling by a factor of 100.

Whole brain volumetric analysis

We conducted whole-brain structural analyses using FreeSurfer (version 7.4.1; https://surfer.nmr.mgh.harvard.edu). T1-weighted anatomical images were processed with the longitudinal processing pipeline [74].

Vertex-wise analyses of cortical volume were conducted using FreeSurfer's general linear modeling tool, mri_glmfit. Group-level statistical comparisons were corrected for multiple comparisons using mri_glmfit-sim, which implements cluster-wise correction based on Monte Carlo simulations. A vertex-wise threshold of $Z > 3.0$ (corresponding to $p < 0.001$, two-sided) was used to identify both positive and negative effects. Clusters were considered significant if they survived a cluster-wise corrected p-value of less than 0.05.

In addition to vertex-wise analyses, cortical parcellation was performed using the Destrieux atlas[40] (aparc.a2009s), which delineates 74 cortical regions per hemisphere, resulting in a total of 148 cortical regions of interest (ROIs). To account for inter-individual variability in brain size, each ROI volume was normalized by the estimated intracranial volume (ICV) and scaled by a factor of 100. Longitudinal statistical comparisons were conducted using paired-sample t-tests. To control for multiple comparisons across ROIs, FDR correction was applied at $q < 0.05$.

Regions of Interest

To identify brain regions associated with memory and spatial navigation, we conducted meta-analyses using Neurosynth (https://neurosynth.org/). We first defined 11 maps based on Neurosynth association tests, using the following 11 key terms: "memory," "episodic memory," "autobiographical memory," "recognition memory," "subsequent memory," "memory encoding," "memory retrieval," "memory performance," "memory processes," "memory tasks," and "memory test." These maps were then combined to create a single, comprehensive memory-related brain mask. Then, following the same steps, we created a single, comprehensive spatial navigation-related mask using six key terms: "navigation", "spatial", "spatial information", "spatial temporal", "time task", and "visual spatial". The memory-spatial map was then created by summing up the memory-related and navigation-related maps. Finally, this combined map was overlapped with the 400 ROIs from the "Schaefer 2018 parcellation" to yield 242 ROIs, each containing a minimum of 30 voxels. Prior to ROI extraction, all registration steps—from individual subject space to MNI space—were visually inspected for each participant to ensure accurate alignment between the functional data and the Schaefer atlas.

fMRI data preprocessing

We performed fMRI data preprocessing using FEAT (FMRI Expert Analysis Tool), version 6.00, implemented in FSL (http://www.fmrib.ox.ac.uk/fsl). The EPI images underwent motion correction, slice-timing correction, and temporal filtering with a nonlinear high-pass filter (100 s cutoff). Six motion parameters were included as confounding regressors in the model. Additionally, outlier timepoints which identified using the FSL's motion outlier detection tool (framewise displacement [FD] > 0.9 mm) were incorporated as additional confounds in the first-level general linear model (GLM)

analysis. If more than 20% of the volumes in a run had an FD exceeding 0.9 mm, or if the absolute head movement exceeded half the voxel size (i.e., 1.25 mm), the entire run was excluded from further analyses. For single-trial estimation, no spatial smoothing was applied. All functional images were linearly registered to the middle image of the first run, and all analyses were conducted in MNI standard space.

Univariate Activation Analysis

We examined encoding and retrieval-related neural activity using the general linear model (GLM) within the FILM module of FSL. During the encoding stage, the GLM included 5 regressors: (1) the encoding stage of all pictures (i.e., the 6 s of the trial in which the participants were presented with an image and asked to encode that image using a temporal or spatial context); (2) rating period (i.e., reaction time, RT) when the participant made a rating judgment of all trials; (3) the remaining time period after the participant made a rating judgment of all trials (i.e., 6s minus RT) while the stimulus stayed on the screen;(4) The duration of inter-trial interval (ITI) fixations; (5) all the instruction period within the same run. The “X” or “O” judgement baseline period was not coded and thus was treated as an implicit baseline. Events were modeled at the time of the stimulus onset and convolved with canonical hemodynamic response function (double gamma function). The encoding phase consisted of four runs, equally divided between spatial and temporal encoding runs. For the retrieval phase, remembered and forgotten pictures were separately modeled for both item memory retrieval and source memory retrieval. Source memory retrieval was further categorized into spatial and temporal retrieval trials. The source memory retrieval effect was defined as the difference in activity between correctly retrieved contextual pictures and the implicit baseline.

A second-level analysis was conducted to compute cross-run averages for spatial encoding effects, temporal encoding effects, combined encoding effects (spatial and temporal concatenated), spatial source memory retrieval effects, temporal source, and combined source memory retrieval effects (spatial and temporal concatenated) using a fixed-effects model. These contrasts were then used in a group analysis with a random-effects model using full FMRIB’s Local Analysis of Mixed Effect 1 with automatic outlier detection[76,77]. Unless otherwise noted, group images were thresholded using cluster detection statistics, with a height threshold of $z > 3.1$ and a cluster probability of $P < 0.05$, corrected for whole-brain multiple comparisons using Gaussian Random Field Theory. To specifically examine univariate activation within the hippocampus, a key region of interest in our experiment, we conducted small volume correction using the hippocampal mask derived from the Harvard-Oxford subcortical atlas in MNI152 space.

Single-trial response estimates

General linear models (GLMs) were conducted separately to estimate the activation patterns for each of 80 encoding trials and 100 retrieval trials. In each single-trial model, a Least Square–Separate (LS-S) approach was used, in which the trial of interest was modeled as one regressor, with all other trials modeled as a separate regressor[78]. Specifically, during the encoding stage, each single-trial GLM included six regressors: (1) the trial of interest (i.e., the 6 s of the trial where the participants were

presented with an image and asked to encode that image using a temporal or spatial context); (2) all other remaining trials within the same run; (3) rating period (i.e. reaction time, RT) when the participant made a rating judgment of all trials; (4) the remaining time period after participant made a rating judgment of all trials (i.e., 6s minus RT) while the stimulus stayed on the screen; (5) The duration of inter-trial interval (ITI) fixations; (6) the instruction period within the same run. The “X” or “O” judgement period was not coded and thus was treated as an implicit baseline.

Similarly, for the retrieval stage, each single-trial GLM included six regressors: (1) the trial of interest (i.e., the RT of the trial in which participants were asked to retrieve source information for the image); (2) all other remaining source retrieval trials within the same run; (3) the remaining time period after the participant made a source judgment (i.e., 6 seconds minus RT) while the stimulus remained on the screen; (4) all item retrieval periods in which participants judged whether they could recognize the item from the encoding phase by pressing “Yes” or “No” within the same run; (5) the remaining time period after the participant made a Yes/No judgment (i.e., 6 seconds minus RT) while the stimulus stayed on the screen; and (6) all instruction periods within the same run. The ITI fixation period was not modeled and was treated as an implicit baseline. Each event was modeled at the time of stimulus onset and convolved with a canonical hemodynamic response function (double gamma). To control the effects of head motion, six motion parameters were included in the GLM model as a covariate, as well as a regressor for each TR that was flagged as having greater framewise displacement (FD) than 0.9 during preprocessing. The t-map for each trial was used for multivariate pattern similarity analysis to increase the reliability by normalizing for noise[79].

Multivariate pattern similarity analysis (MPS)

Multi-voxel pattern similarity (MPS) was used to measure the similarity of activation patterns by calculating the correlation between trials that were either encoded or correctly retrieved, within each ROI[80]. Following the approach of Power[81], volumes with a framewise displacement greater than 0.9 mm were censored, and trials containing any censored frames within the duration of the modeled GLM response were excluded. Specifically, within-context pattern similarity (PS) during encoding was calculated using pairwise Pearson correlation coefficients between trials within the same context (i.e., spatial or temporal). Within-spatial context was calculated using only spatial trials, while within temporal context was calculated using only temporal trials. For source memory retrieval, we performed analogous calculations for within-context PS, separately measuring within-spatial context PS and within-temporal context PS, using correct retrieval trials. All MPS analyses were performed across trials from different runs to avoid temporal autocorrelations that could otherwise inflate or bias results. The resulting correlation coefficients were then transformed into Fisher’s z-scores.

Informational connectivity analysis

We performed informational connectivity among the 242 predefined ROIs by calculating pairwise correlations between representational similarity matrices (RSMs)[82-86]. This was done by correlating the within-context PS for each ROI with another ROI. In contrast to traditional functional connectivity (FC) analyses, which quantify temporal

correlations of BOLD signal fluctuations between brain regions, the informational connectivity approach examines the similarity of information coding patterns across regions[83,84,86]. While FC analyses rely on averaged BOLD signals within regions, representational similarity analysis (RSA) captures the fine-grained spatial patterns of neural responses to different experimental conditions[85]. Specifically, RSA quantifies how similarly or differently a brain region responds to different stimuli or conditions by comparing multivariate activity patterns, creating a RSM that characterizes each region's representational structure[85]. By correlating these RSMs between regions, we can assess whether different brain areas encode information in similar ways, providing insights into information sharing and processing that complement traditional FC measures.

Three types of within-context RSMs were calculated: one based on all trials, one based on spatial trials, and one based on temporal trials (Fig. 4a top). All Pearson correlation coefficients (r values) were Fisher-Z transformed prior to statistical analysis. Significance thresholds for informational connectivity were determined using 10,000 permutation simulations. The reported results have been corrected for multiple comparisons using the FDR method, with a q-value threshold of less than 0.05.

Multivariate informational connectivity pattern distance analysis

To assess changes in multivariate informational connectivity patterns between pre and post, distances were computed for each ROI by comparing spatial and temporal informational connectivity patterns before and after tranining. Specifically, for each ROI, the informational connectivity with the remaining 241 ROIs was extracted separately from the spatial and temporal informational connectivity matrices (ICM). The distance between the two connectivity patterns was calculated using 1 minus the Pearson or Spearman correlation coefficient (Fig. 6a). Paired t-tests were conducted to evaluate statistical differences between pre and post, with FDR correction applied across the 242 ROIs at a q-value threshold of < 0.05.

References

1 Kraft, J. N. *et al.* Task-based functional connectivity of the useful field of view (UFOV) FMRI task. *GeroScience* **45**, 293–309 (2023).

2 Edwards, J. D., Fausto, B. A., Tetlow, A. M., Corona, R. T. & Valdés, E. G. Systematic review and meta-analyses of useful field of view cognitive training. *Neuroscience & Biobehavioral Reviews* **84**, 72–91 (2018).

3 Maguire, E. A. *et al.* Navigation-related structural change in the hippocampi of taxi drivers. *Proc Natl Acad Sci U S A* **97**, 4398–4403 (2000). https://doi.org/10.1073/pnas.070039597

4 Draganski, B. *et al.* Changes in grey matter induced by training. *Nature* **427**, 311–312 (2004).

5 Boyke, J., Driemeyer, J., Gaser, C., Büchel, C. & May, A. Training-Induced Brain Structure Changes in the Elderly. *The Journal of Neuroscience* **28**, 7031–7035 (2008). https://doi.org/10.1523/jneurosci.0742-08.2008

6 Wenger, E., Papadaki, E., Werner, A., Kühn, S. & Lindenberger, U. Observing Plasticity of the Auditory System: Volumetric Decreases Along with Increased

Functional Connectivity in Aspiring Professional Musicians. *Cerebral Cortex Communications* **2** (2021). https://doi.org/10.1093/texcom/tgab008

7 Hänggi, J., Koeneke, S., Bezzola, L. & Jäncke, L. Structural neuroplasticity in the sensorimotor network of professional female ballet dancers. *Human Brain Mapping* **31**, 1196–1206 (2010). https://doi.org/https://doi.org/10.1002/hbm.20928

8 Draganski, B. *et al.* Temporal and Spatial Dynamics of Brain Structure Changes during Extensive Learning. *The Journal of Neuroscience* **26**, 6314–6317 (2006). https://doi.org/10.1523/jneurosci.4628-05.2006

9 Engvig, A. *et al.* Effects of memory training on cortical thickness in the elderly. *NeuroImage* **52**, 1667–1676 (2010). https://doi.org/https://doi.org/10.1016/j.neuroimage.2010.05.041

10 Dresler, M. *et al.* Mnemonic training reshapes brain networks to support superior memory. *Neuron* **93**, 1227–1235. e1226 (2017).

11 Braun, U. *et al.* Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc Natl Acad Sci U S A* **112**, 11678–11683 (2015). https://doi.org/10.1073/pnas.1422487112

12 Spiers, H. J. *et al.* Unilateral temporal lobectomy patients show lateralized topographical and episodic memory deficits in a virtual town. *Brain* **124**, 2476–2489 (2001).

13 Buzsaki, G. & Moser, E. I. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat Neurosci* **16**, 130–138 (2013). https://doi.org/10.1038/nn.3304

14 Brunec, I. K. *et al.* Cognitive mapping style relates to posterior–anterior hippocampal volume ratio. *Hippocampus* **29**, 748–754 (2019).

15 Poppenk, J., Evensmoen, H. R., Moscovitch, M. & Nadel, L. Long-axis specialization of the human hippocampus. *Trends in cognitive sciences* **17**, 230–240 (2013).

16 Clark, I. A. *et al.* Does hippocampal volume explain performance differences on hippocampal-dependant tasks? *NeuroImage* **221**, 117211 (2020).

17 Weisberg, S. M., Newcombe, N. S. & Chatterjee, A. Everyday taxi drivers: Do better navigators have larger hippocampi? *Cortex* **115**, 280–293 (2019). https://doi.org/10.1016/j.cortex.2018.12.024

18 Sala, G. & Gobet, F. Does Far Transfer Exist? Negative Evidence From Chess, Music, and Working Memory Training. *Current Directions in Psychological Science* **26**, 515–520 (2017). https://doi.org/10.1177/0963721417712760

19 Ekstrom, A. D. & Hill, P. F. Spatial navigation and memory: A review of the similarities and differences relevant to brain models and age. *Neuron* **111**, 1037–1049 (2023). https://doi.org/10.1016/j.neuron.2023.03.001

20 Fan, C. L., Abdi, H. & Levine, B. On the relationship between trait autobiographical episodic memory and spatial navigation. *Memory & Cognition* **49**, 265–275 (2021).

21 Weisberg, S. M. & Newcombe, N. S. How do (some) people make a cognitive map? Routes, places, and working memory. *J Exp Psychol Learn Mem Cogn* **42**, 768–785 (2016). https://doi.org/10.1037/xlm0000200

22 Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022). https://doi.org/10.1038/s41586-022-04492-9
23 Driemeyer, J., Boyke, J., Gaser, C., Büchel, C. & May, A. Changes in Gray Matter Induced by Learning—Revisited. *PLOS ONE* **3**, e2669 (2008). https://doi.org/10.1371/journal.pone.0002669
24 Lövdén, M. *et al.* Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiol Aging* **33**, 620.e629–620.e622 (2012). https://doi.org/10.1016/j.neurobiolaging.2011.02.013
25 Colom, R. *et al.* Gray matter volumetric changes with a challenging adaptive cognitive training program based on the dual n-back task. *Personality and Individual Differences* **98**, 127–132 (2016). https://doi.org/https://doi.org/10.1016/j.paid.2016.03.087
26 Miró-Padilla, A. *et al.* Sustained and transient gray matter volume changes after n-back training: A VBM study. *Neurobiology of Learning and Memory* **178**, 107368 (2021). https://doi.org/https://doi.org/10.1016/j.nlm.2020.107368
27 Chavan, C. F., Mouthon, M., Draganski, B., van der Zwaag, W. & Spierer, L. Differential patterns of functional and structural plasticity within and between inferior frontal gyri support training-induced improvements in inhibitory control proficiency. *Hum Brain Mapp* **36**, 2527–2543 (2015). https://doi.org/10.1002/hbm.22789
28 Ganesan, K. *et al.* Cognitive control training with domain-general response inhibition does not change children's brains or behavior. *Nat Neurosci* **27**, 1364–1375 (2024). https://doi.org/10.1038/s41593-024-01672-w
29 Jaeggi, S. M., Buschkuehl, M., Shah, P. & Jonides, J. The role of individual differences in cognitive training and transfer. *Memory & cognition* **42**, 464–480 (2014).
30 Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D. & Dalgleish, T. Training the emotional brain: improving affective control through emotional working memory training. *J Neurosci* **33**, 5301–5311 (2013). https://doi.org/10.1523/jneurosci.2593-12.2013
31 Karbach, J. & Verhaeghen, P. Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychol Sci* **25**, 2027–2037 (2014). https://doi.org/10.1177/0956797614548725
32 Judd, N. & Klingberg, T. Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour* **5**, 1548–1554 (2021). https://doi.org/10.1038/s41562-021-01118-4
33 Schmiedek, F., Lövdén, M. & Lindenberger, U. Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. *Front Aging Neurosci* **2** (2010). https://doi.org/10.3389/fnagi.2010.00027
34 Simons, D. J. *et al.* Do "brain-training" programs work? *Psychological science in the public interest* **17**, 103–186 (2016).
35 Gobet, F. & Sala, G. Cognitive Training: A Field in Search of a Phenomenon. *Perspect Psychol Sci* **18**, 125–141 (2023). https://doi.org/10.1177/17456916221091830

36 Bouffard, N., Stokes, J., Kramer, H. J. & Ekstrom, A. D. Temporal encoding strategies result in boosts to final free recall performance comparable to spatial ones. *Memory & Cognition* **46**, 17–31 (2018). https://doi.org/10.3758/s13421-017-0742-z
37 Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F. & Epstein, R. A. Variations in cognitive maps: understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **40**, 669 (2014).
38 Brunec, I. K. *et al.* Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Current biology* **28**, 2129–2135. e2126 (2018).
39 Angeli, P. A., DiNicola, L. M., Saadon-Grosman, N., Eldaief, M. C. & Buckner, R. L. Specialization of the human hippocampal long axis revisited. *Proceedings of the National Academy of Sciences* **122**, e2422083122 (2025).
40 Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).
41 Mooraj, Z. *et al.* Toward a functional future for the cognitive neuroscience of human aging. *Neuron* **113**, 154–183 (2025). https://doi.org/10.1016/j.neuron.2024.12.008
42 Maguire, E. A., Valentine, E. R., Wilding, J. M. & Kapur, N. Routes to remembering: the brains behind superior memory. *Nat Neurosci* **6**, 90–95 (2003). https://doi.org/10.1038/nn988
43 Stern, Y. Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology* **11**, 1006–1012 (2012).
44 Lövdén, M. *et al.* Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiology of aging* **33**, 620. e629–620. e622 (2012).
45 McLaren-Gradinaru, M. *et al.* (2020).
46 Burgess, N., Maguire, E. A. & O'Keefe, J. The Human Hippocampus and Spatial and Episodic Memory. *Neuron* **35**, 625–641 (2002). https://doi.org/10.1016/S0896-6273(02)00830-9
47 Lövdén, M. *et al.* Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiology of Aging* **33**, 620.e629–620.e622 (2012). https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2011.02.013
48 Bassett, D. S., Yang, M., Wymbs, N. F. & Grafton, S. T. Learning-induced autonomy of sensorimotor systems. *Nat Neurosci* **18**, 744–751 (2015).
49 Finc, K. *et al.* Dynamic reconfiguration of functional brain networks during working memory training. *Nature communications* **11**, 2435 (2020).
50 Schedlbauer, A. M. & Ekstrom, A. D. Flexible network community organization during the encoding and retrieval of spatiotemporal episodic memories. *Network Neuroscience* **3**, 1070–1093 (2019).
51 Bouffard, N., Stokes, J., Kramer, H. J. & Ekstrom, A. D. Temporal encoding strategies result in boosts to final free recall performance comparable to spatial ones. *Mem Cognit* **46**, 17–31 (2018). https://doi.org/10.3758/s13421-017-0742-z

52 Squire, L. R., Stark, C. E. & Clark, R. E. The medial temporal lobe. *Annu. Rev. Neurosci.* **27**, 279–306 (2004).
53 Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. Fsl. *Neuroimage* **62**, 782–790 (2012).
54 Tustison, N. J. *et al.* The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific reports* **11**, 9068 (2021).
55 Tournier, J.-D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* **202**, 116137 (2019).
56 Cordero-Grande, L., Christiaens, D., Hutter, J., Price, A. N. & Hajnal, J. V. Complex diffusion-weighted image estimation via matrix recovery under general noise models. *Neuroimage* **200**, 391–404 (2019).
57 Veraart, J., Fieremans, E. & Novikov, D. S. Diffusion MRI noise mapping using random matrix theory. *Magnetic resonance in medicine* **76**, 1582–1593 (2016).
58 Veraart, J. *et al.* Denoising of diffusion MRI using random matrix theory. *Neuroimage* **142**, 394–406 (2016).
59 Andersson, J. L., Graham, M. S., Zsoldos, E. & Sotiropoulos, S. N. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage* **141**, 556–572 (2016).
60 Bastiani, M. *et al.* Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *Neuroimage* **184**, 801–812 (2019).
61 Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208–S219 (2004).
62 Mitchell, T. *et al.* Advanced diffusion imaging to track progression in Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy. *NeuroImage: Clinical* **34**, 103022 (2022).
63 Pasternak, O., Sochen, N., Gur, Y., Intrator, N. & Assaf, Y. Free water elimination and mapping from diffusion MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **62**, 717–730 (2009).
64 Wilkes, B. J. *et al.* Cortico-basal ganglia white matter microstructure is linked to restricted repetitive behavior in autism spectrum disorder. *Molecular Autism* **15**, 6 (2024).
65 Wilkes, B. J. *et al.* Distinct cortical and subcortical predictors of Purdue Pegboard decline in Parkinson's disease and atypical parkinsonism. *npj Parkinson's Disease* **9**, 85 (2023).
66 Shin, Y. S. *et al.* Transcallosal white matter and cortical gray matter variations in autistic adults aged 30–73 years. *Molecular Autism* **16**, 16 (2025). https://doi.org/10.1186/s13229-025-00652-6
67 Yeh, F.-C. Population-based tract-to-region connectome of the human brain and its hierarchical topology. *Nature communications* **13**, 4933 (2022).
68 Schwarz, C. *et al.* (2018).
69 Ekstrom, A. D., Huffman, D. J. & Starrett, M. Interacting networks of brain regions underlie human spatial navigation: a review and novel synthesis of the literature. *Journal of neurophysiology* **118**, 3328–3344 (2017).

70 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).
71 Yushkevich, P. A. *et al.* Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum Brain Mapp* **36**, 258–287 (2015). https://doi.org/10.1002/hbm.22627
72 Yushkevich, P. A. *et al.* Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. *Neuroimage* **111**, 526–541 (2015). https://doi.org/10.1016/j.neuroimage.2015.01.004
73 Zeineh, M. M., Engel, S. A., Thompson, P. M. & Bookheimer, S. Y. Unfolding the human hippocampus with high resolution structural and functional MRI. *The Anatomical Record: An Official Publication of the American Association of Anatomists* **265**, 111–120 (2001).
74 Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* **61**, 1402–1418 (2012).
75 Reuter, M., Rosas, H. D. & Fischl, B. Highly accurate inverse consistent registration: a robust approach. *Neuroimage* **53**, 1181–1196 (2010).
76 Beckmann, C. F., Jenkinson, M. & Smith, S. M. General multilevel linear modeling for group analysis in FMRI. *Neuroimage* **20**, 1052–1063 (2003). https://doi.org/10.1016/s1053-8119(03)00435-x
77 Woolrich, M. Robust group analysis using outlier inference. *Neuroimage* **41**, 286–301 (2008).
78 Mumford, J. A., Turner, B. O., Ashby, F. G. & Poldrack, R. A. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* **59**, 2636–2643 (2012). https://doi.org/10.1016/j.neuroimage.2011.08.076
79 Walther, A. *et al.* Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188–200 (2016).
80 Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences* **103**, 3863–3868 (2006).
81 Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014). https://doi.org/10.1016/j.neuroimage.2013.08.048
82 Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* **2**, 4 (2008). https://doi.org/10.3389/neuro.06.004.2008
83 Gao, Z. *et al.* Context free and context-dependent conceptual representation in the brain. *Cerebral Cortex* **33**, 152–166 (2022). https://doi.org/10.1093/cercor/bhac058
84 Zheng, L. *et al.* Hippocampal contributions to novel spatial learning are both age-related and age-invariant. *Proceedings of the National Academy of Sciences* **120**, e2307884120 (2023). https://doi.org/doi:10.1073/pnas.2307884120

85 Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* **2**, 249 (2008).

86 Aly, M. & Turk-Browne, N. B. Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences* **113**, E420–E429 (2016). https://doi.org/doi:10.1073/pnas.1518931113