ELSEVIER

Neurobiology of Aging 33 (2012) 620.e9–620.e22

NEUROBIOLOGY OF AGING

www.elsevier.com/locate/neuaging

Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood

Martin Lövdén[a,b,c,*], Sabine Schaefer[a], Hannes Noack[a], Nils Christian Bodammer[a], Simone Kühn[d,e], Hans-Jochen Heinze[f,g], Emrah Düzel[f,g,h], Lars Bäckman[a,c], Ulman Lindenberger[a]

[a] Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany
[b] Department of Psychology, Lund University, Lund, Sweden
[c] Aging Research Center, Karolinska Institutet & Stockholm University, Stockholm, Sweden
[d] Department of Psychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[e] Department of Experimental Psychology, Ghent University, Ghent, Belgium
[f] Department of Neurology, Otto-von-Guericke University of Magdeburg, Magdeburg, Germany
[g] DZNE (German Centre for Neurodegenerative Disorders), Magdeburg, Germany
[h] Institute of Cognitive Neuroscience, University College London, London, UK

Received 9 April 2010; received in revised form 3 February 2011; accepted 16 February 2011

Abstract

It is unknown whether lifestyle, including mental stimulation, and appropriate training interventions, may directly improve spatial navigation performance and its underlying neural substrates. Here we report that healthy younger and older men performing a cognitively demanding spatial navigation task every other day over 4 months display navigation-related gains in performance and stable hippocampal volumes that were maintained 4 months after termination of training. In contrast, control groups displayed volume decrements consistent with longitudinal estimates of age-related decline. Hippocampal barrier density, as indicated by mean diffusivity estimated from diffusion tensor imaging, showed a quadratic shape of increased density after training followed by a return to baseline in the right hippocampus, but declined in the control groups and in the left hippocampus. We conclude that sustained experiential demands on spatial ability protect hippocampal integrity against age-related decline. These results provide the first longitudinal evidence indicating that spatial navigation experience modifies hippocampal volumes in humans, and confirm epidemiological results suggesting that mental stimulation may have direct effects on neural integrity.

Keywords: Cognitive aging; Cognitive training; Hippocampus; Plasticity; Spatial navigation

1. Introduction

Spatial navigation ability is dependent on the integration of distance, direction, and place information (Wolbers and Hegarty, 2010), and shows palpable decline across the adult life span (Lövdén et al., 2005a; Moffat et al., 2007). Older adults commonly avoid unfamiliar routes and places due to self-perceived impairments in navigation, thereby restricting their mobility and exposure to new experience (Burns, 1999). Conversely, epidemiological work indicates that individuals with a lifestyle rich in mental, physical, and social stimulation experience less cognitive decline in old age (Hertzog et al., 2009; Lövdén et al., 2005b). However, it remains unknown whether extensive training may directly improve spatial navigation ability and alleviate age-related changes in this ability and its underlying neural substrates (Wolbers and Hegarty, 2010). Here we address this issue,

* Corresponding author at: Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Tel.: +49 3082406216.
E-mail address: loevden@mpib-berlin.mpg.de (M. Lövdén).

0197-4580/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neurobiolaging.2011.02.013

predicting that extensive spatial navigation training improves navigation performance and protects against negative age-related changes in hippocampal integrity during early and late adulthood.

Spatial navigation is highly reliant on the hippocampus (Burgess et al., 2002; O'Keefe and Nadal, 1978). Hippocampal volumes decrease in adulthood and old age. Whereas cross sectional studies of healthy aging show relatively mild age-related decrements in volume (see Raz, 2000, 2004, for reviews), longitudinal studies, which reduce and control for age-differential sample selectivity, reveal median annual shrinkage rates of about 1.7% from the seventh decade of life and onwards (see Raz, 2004, for review). Evidence on earlier adulthood is scarce, but consistently reveals smaller volume decrements (e.g., Raz et al., 2005; Scahill et al., 2003). Nevertheless, volume decrements are also noticeable in early adulthood. For example, Scahill et al. (2003) reported annual decrements of about 0.75% for 30–40-year-old adults, and DeLise et al. (1997) observed shrinkage of 0.37% per year in a sample with a mean age of 28 years. In an adult life span sample, the linear component of annual decrements of hippocampal volumes amounted to 0.79%, and longitudinal decrements were visible for most 20–40 year-olds (Raz et al., 2005).

The hippocampus is a plastic brain region. For example, the nonhuman hippocampus responds to experience in younger and older adulthood with several types of changes at cellular, subcellular, molecular, and biochemical levels (Churchill et al., 2002; Jessberger and Gage, 2008; Kempermann et al., 2002; Kronenberg et al., 2006; Rosenzweig and Bennett, 1996; van Praag et al., 2000). In humans, Maguire et al. (2000, 2006) reported larger posterior and smaller anterior hippocampi of London taxi drivers relative to controls. A few recent longitudinal studies suggest that variations in human hippocampal morphology may be due to experience-dependent changes during early and late adulthood. Draganski et al. (2006) reported gray matter increases in the posterior hippocampi of medical students preparing for a major examination. In addition, transient increases of gray matter in the hippocampus were observed in older adults learning to juggle (Boyke et al., 2008). The exact microscopic nature of these morphological changes is unknown, and the results of Draganski et al. (2006) are controversial, as the gray matter expansion continued during the semester break after the examination. In addition, experience-dependent morphological changes in the adult human hippocampus have, to our knowledge, only been demonstrated with voxel-based morphometry (VBM), which has been repeatedly criticized (e.g., Bookstein, 2001; Davatzikos, 2004; but see Ashburner and Friston, 2001, for a rebuttal) and is sensitive to relatively small changes in the processing pipeline (Thomas et al., 2009). Thus, it is vital to complement these findings of experience-dependent morphological changes with manual segmentation methods, which still constitute the volumetric gold standard. In addition, longitudinal evidence supporting that spatial navigation experience may modify hippocampal volumes in humans is lacking.

Recent studies show that training of working memory and executive functions may result in performance gains that generalize (i.e., transfer) to untrained tasks in both younger (Dahlin et al., 2008; Jaeggi et al., 2008; Olesen et al., 2004; Persson and Reuter-Lorenz, 2008; Schmiedek et al., 2010) and older (Karbach and Kray, 2009; Schmiedek et al., 2010) adults, though the magnitude of gains is often reduced in older adults (e.g., Dahlin et al., 2008; for review, see Lövdén et al., 2010; Noack et al., 2009). In contrast, training of functions that critically depend on the medial temporal lobes, such as episodic memory, reveals limited transfer (Baltes and Lindenberger, 1988; Noack et al., 2009; Verhaeghen et al., 1992; but see Schmiedek et al., 2010). Considering the extent of hippocampal plasticity, this lack of transfer is unsettling and either points to suboptimal design of previous interventions or to the possibility that hippocampal plasticity may not be reflected in improved cognitive processing efficiency (see also Lövdén et al., 2010).

Here we report results from an age-comparative spatial-navigation intervention that maximized the association between task requirements and functions known to depend on the hippocampus by demanding: (1) spatial learning and memory based on allocentric processes (e.g., navigation supported by triangulation of distal cues); (2) associative memory (e.g., encoding and retrieval of associations between landmarks and navigational decisions); (3) encoding of novel information (e.g., new navigation areas during training); and (4) consolidation of information (e.g., retrieval of information encoded at previous training sessions is needed for good performance). Specifically, training involved performing a navigation task in a virtual environment (VE), while walking on an exercise treadmill (see Fig. 1). The walking component was included to increase the immersion into the VE and to leverage the interacting role that self-motion may play in forming hippocampal spatial representations (e.g., McNaughton et al., 1996; Stackman et al., 2002) and in learning-related plasticity (Kempermann, 2008). However, walking on the treadmill was not physically demanding.

To examine the effects of training on navigation performance, broad cognitive performance, and hippocampal integrity, younger and older healthy men underwent structural magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), assessment of navigation performance, and assessment of several cognitive functions before, immediately after, and 4 months after termination of a 4-month training phase. Participants in a yoked control group walked on the treadmill without the VE for an identical amount of time.

We manually segmented the hippocampus on high-resolution structural images. Considering evidence of differential experience-dependent changes in the anterior and pos-

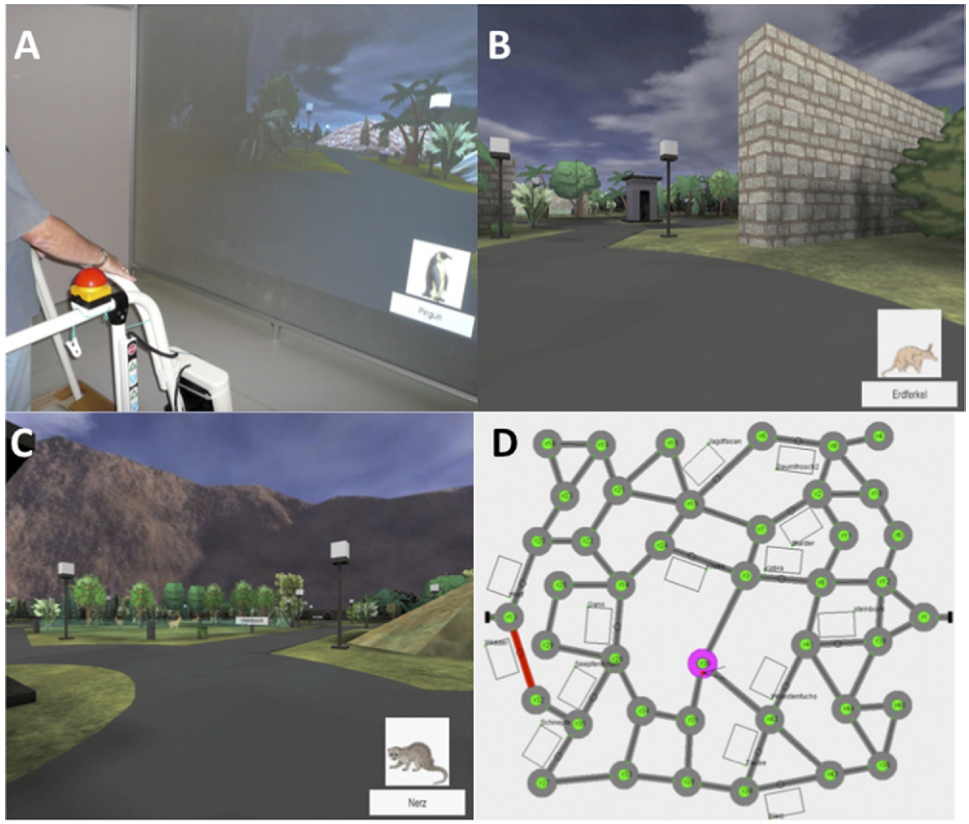

Fig. 1. (A) Participants in the navigation training group navigated in a virtual environment (VE), while walking on a normal exercise treadmill at a relatively modest and not physically demanding pace. Participants in a walk-time-yoked control group walked on a treadmill without the VE. A total of forty-two 50-minute training sessions were administered. (B, C) The task was to search for the animal currently displayed at the lower right corner of the screen. When the currently displayed animal had been found (by passing the sign displaying the animal's name; see Fig. 1C), a new target animal was displayed. After all animals had been found, the participant completed the trial by finding the exit, and started a new trial in the same zoo. After completing 4 trials in a zoo (which was not possible within one 50-minute session), the participant received a new zoo (with different animals, topography, and layout) and proceeded with the task. (D) Survey map of an exemplar zoo with a similar path structure to those used during training, pretest, and posttest (never shown to participants). Gray thick lines symbolize paths that could be followed, gray circles denotes decision points (roundabouts), rectangles depict animal cages, purple color symbolizes the current position of the participants, and red color indicates the current target animal.

terior portions of the hippocampus (Draganski et al., 2006; Maguire et al., 2000), we further segmented the hippocampus into head, body, and tail volumes. The DTI data were used to estimate hippocampal mean diffusivity (MD), which is sensitive to microstructural changes in hippocampal integrity (e.g., Kantarci et al., 2005). We predicted training-related increases in navigation performance, transfer of gains especially for cognitive tasks requiring allocentric spatial processing, and positive effects on the structural integrity of the hippocampi in both younger and older adults.

2. Methods

2.1. Participants

Fifty-six younger (aged 20–30 years) and 62 older (aged 60–70 years) men were recruited through newspaper advertisements, word-of-mouth recommendation, and fliers circulated in Berlin, Germany. Participants were all right-handed, had normal or corrected-to-normal vision, and reported no history of cardiovascular disease (except treated hypertension), neurological or psychiatric conditions, problems hindering gait or balance, or drug/alcohol abuse. They reported no use of antiseizure or antidepressant drugs. A clinical neurologist evaluated T1- and T2-weighted magnetic resonance (MR) images collected at pretest and excluded 12 participants (1 younger and 11 older) due to various brain abnormalities (e.g., infarcts, large white matter lesions), and 6 participants (3 younger and 3 older) due to imaging artifacts (e.g., movement). Four younger participants dropped out during or immediately after pretest due to lack of motivation. Five participants (4 younger and 1 older) dropped out at different times after pretest due to health issues, starting a new job, or unspecified personal problems. No other participant developed any conditions during the course of the study that violated the inclusion and exclusion criteria listed above.

The effective sample consisted of 44 younger (age range:

Table 1
General participant characteristics at pretest and time between pretest and posttests as a function of age and training group

Measure	Younger (20–30 years)				Older (60–70 years)			
	Navigators		Walkers		Navigators		Walkers	
	M	SD	M	SD	M	SD	M	SD
Age (years)	25.1	2.8	27.01	2.5	65.3	2.8	64.6	2.9
DS	57.2	11.1	57.5	12.1	42.0	6.8	37.8	9.8
Vocabulary	30.0	2.7	30.5	3.2	32.2	1.8	31.6	2.1
Raven's matrices	9.5	4.2	10.3	3.2	4.5	2.3	4.4	3.0
Navigation performance	20.9	6.1	21.9	4.6	7.6	2.9	8.8	2.7
Days between pretest and posttest 1	118	4.6	117	4.6	115	5.0	118	5.4
Days between posttest 1 and posttest 2	121	6.2	123	6.4	122	2.8	120	7.1

The Digit-Symbol Substitution Test (DS) is a perceptual speed measure from Wechsler (1981). The vocabulary test asked participants to identify real words among 4 nonwords (Lehrl et al., 1991). The 18 odd-numbered items from set II of the Raven's advanced progressive matrices were administered in computerized form. Navigation performance denotes the number of animals found within 50 minutes in the virtual zoo at pretest (section 2.4.2 for a detailed description).
Key: M, mean.

20–30 years; mean [M] = 26.0; SD = 2.8) and 47 older (age range: 60–70 years; M = 65.0; SD = 2.8) men. The sample displayed the typical age-related trends of increases in vocabulary ($t(89) = 3.34$, $p = 0.003$; younger: M = 30.3, SD = 2.7; older: M = 31.9, SD = 1.8; Lehrl et al., 1991) and decreases in perceptual speed ($t(89) = 8.25$, $p < 0.001$; younger: M = 57.3, SD = 11.5; older: M = 39.8, SD = 8.6; Digit-symbol Substitution [DS]; Wechsler, 1981). Relative to mean DS performance reported in a meta-analysis (Hoyer et al., 2004), the sample was less positively selected (about 1 SD for both young and old) than typical samples in cognitive aging research.

After completion of pretest, we matched participants on DS performance and age within age groups and randomly assigned participants to either the navigation ($n_{young} = 23$; $n_{old} = 23$) or the walk-time-yoked control group ($n_{young} = 21$; $n_{old} = 24$). The unequal n's in the effective samples reflect unequal dropout after pretest. Background characteristics of the groups at pretest are summarized in Table 1. Pretest values of the imaging measures are summarized as a function of age and experimental group in Tables 2 and 3. The groups did not differ significantly on any of these pretest measures (all F's < 1.42, p's > 0.23). One younger navigator, 4 older navigators, and 5 older walkers reported treated hypertension.

Participants were paid 1150 Euro for completion of the whole study. The ethical review board of the Otto-von-Guericke University of Magdeburg approved the imaging part of the study, and the ethical review board of the Max Planck Institute for Human Development, Berlin approved the behavioral part of the study. Written informed consent was obtained prior to the investigation.

2.2. General study design

The study followed a 2 (training group: navigation training/walk-time yoked control) × 2 (age group: young/old) × 3 (time: pretest/posttest 1/posttest 2) design, with posttest 1 starting after the completion of the 4-month long training phase. Posttest 2 was administered to assess the maintenance of training-related changes 4 months after completion of posttest 1, with no training administered between the 2 posttests. For all participants, pretests and posttests included MRI assessment, assessment of navigation performance, and administration of cognitive tests. For navigators, training involved a navigation task while walking in a VE equipped with a walking interface (see Fig. 1). Participants in the walk-time-yoked control group walked on an identical treadmill for the same amount of time as the navigators, but without the navigation task and the VE. Training ses-

Table 2
Means and standard deviations of total adjusted hippocampus volumes (mm^3) as a function age group, training group, hemisphere, and time

Hemisphere	Time	Younger (20–30 years)				Older (60–70 years)			
		Navigators		Walkers		Navigators		Walkers	
		M	SD	M	SD	M	SD	M	SD
Left	Pretest	3583	339	3526	398	3477	350	3534	365
	Posttest 1	3599	337	3503	373	3499	337	3478	387
	Posttest 2	3581	336	3512	375	3501	338	3514	395
Right	Pretest	3765	381	3733	388	3658	390	3702	427
	Posttest 1	3770	363	3712	373	3648	378	3668	436
	Posttest 2	3760	356	3705	373	3664	378	3654	452

The time interval between each measurement point is approximately 4 months.
Key: M, mean.

Table 3
Means and standard deviations of hippocampal mean diffusivity (10^{-3} mm^2/second) as a function of age group, training group, hemisphere, and time

Hemisphere	Time	Younger (20–30 years)				Older (60–70 years)			
		Navigators		Walkers		Navigators		Walkers	
		M	SD	M	SD	M	SD	M	SD
Left	Pretest	0.9410	0.0302	0.9333	0.0222	0.9524	0.0297	0.9492	0.0288
	Posttest 1	0.9416	0.0223	0.9341	0.0188	0.9569	0.0380	0.9488	0.0288
	Posttest 2	0.9479	0.0241	0.9449	0.0214	0.9585	0.0346	0.9548	0.0335
Right	Pretest	0.9315	0.0255	0.9306	0.0219	0.9415	0.0288	0.9474	0.0295
	Posttest 1	0.9256	0.0172	0.9344	0.0223	0.9396	0.0265	0.9460	0.0244
	Posttest 2	0.9328	0.0172	0.9378	0.0227	0.9475	0.0298	0.9492	0.0358

The time interval between each measurement point is approximately four months.
Key: M, mean.

sions were scheduled every other day for both walkers and navigators. A total of 42 50-minute training sessions were administered. If a participant missed a scheduled session (e.g., due to illness), the training period was prolonged accordingly. The average number of days between pretest and posttests (see Table 1) did not differ significantly between age and experimental groups (all F's < 1.11, p's > 0.29).

2.3. Training protocol

All participants selected their own preferred walking speed before assessment of navigation performance at pretest (see below) and before participants were assigned to their training status (navigation or walking), but after familiarization with the navigation task. Participants were instructed to select a comfortable and nondemanding speed of walking ("as if they took a Sunday walk in the park"). This self-selected walking speed was kept constant throughout the entire study. Note that the visual flow of the virtual world was held constant for all participants, so that the selected walking speed did not influence navigation performance. Training groups did not differ significantly in their mean walking speed (t's < 0.81; $M_{young\ navigators} = 3.6$ km/hour; $SD_{young\ navigators} = 0.7$ km/hour; $M_{young\ walkers} = 3.7$ km/hour; $SD_{young\ walkers} = 0.7$ km/hour; $M_{old\ navigators} = 3.1$ km/hour; $SD_{old\ navigators} = 0.9$ km/hour; $M_{old\ walkers} = 3.3$ km/hour; $SD_{old\ walkers} = 0.7$ km/hour), but there was an expected age difference, $t(89) = 3.00$, $p = 0.004$, with older adults walking slower. These walking speeds were not physically demanding and self-selected walking speed is likely better at minimizing individual differences in physical demands than a preselected speed for all participants.

In the navigation laboratory, a VE was back-projected onto a screen situated approximately 150 cm in front of the walking area of a normal motorized treadmill that was positioned at the level of the floor (see Fig. 1A). The projection area was 176 × 236 cm, which allowed for an approximately 75 degrees wide field of view. Two buttons attached to the handrail were used to control navigation decisions (turning left or right) in the VE. Maps of the VE were rendered in first person view using the Quake 3 engine (www.idsoftware.com/games/quake/quake3-arena) and constructed in GtkRadiant (www.qeradiant.com/cgi-bin/trac.cgi). An interface for the experimenter was designed and programmed by us in Java to allow for control of the treadmill and the VE, and for access to a MySQL database storing the relevant data.

The task for participants in the navigation group was to start walking at the entrance to a virtual zoo and search for the animal currently displayed at the lower right corner of the screen (see Fig. 1B and C). The participants never had access to any survey map of the environment (such as the 1 displayed in Fig. 1D). When the currently displayed animal had been found (by passing the sign displaying the animal's name; see Fig. 1C), a new target animal was displayed in the lower right corner of the screen and participants started searching for this animal. After all animals had been found, the participant completed the trial by finding the exit of the zoo, received feedback on the number of minutes needed to complete the trial, and started a new trial in the same zoo. The order of animals to be found was randomly generated for each of the trials, but kept constant across participants. After completing 4 trials in a zoo, the participant received a new zoo (with different animals, topography, and layout) and proceeded with the task in the same way as described above.

Twenty different training zoos were constructed. All of these zoos had identical skylines and distal landmarks (e.g., mountains, towers; see Fig. 1B and C) situated at identical places surrounding the area of the zoo (which was enclosed by walls). The length of the direct paths between the animals was kept constant across trials and zoos. The direct full path was generated such that it was impossible to complete 4 trials in a zoo within one 50-minute session and so that consecutive targets were not located next to each other. The path structure of all training zoos was randomly generated under the constraint that they had 42 decision points (i.e., roundabouts with a different number of connecting paths; see Fig. 1C and D): 1 (±1) 1-decision point; 10 (±2) 2-decision points; 18 (±2) 3-decision points; 12 (±2) 4-decision points, and 1 (±1) 5-decision points that were connected by paths of identical total length across zoos. Num-

bers in parentheses denote the freedom for the random generation of the number of decision points of each type. Participants could not deviate from the paths (see Fig. 1B–D) in the VE and could not stop walking (and moving in the VE) unless they requested the experimenter to stop the task (which was discouraged and a very rare event). In all zoos, 14 animals were randomly distributed over the map (but placed next to paths). In each of the 20 zoos, an approximately equal amount of local landmarks (e.g., walls, trees, hills, cafeterias, lampposts) were manually distributed over the zoo.

Participants entered a roundabout clockwise (by clicking the left button) or counterclockwise (by clicking the right button). If no decision was made then participants entered counterclockwise. They exited a roundabout with the appropriate button press (left or right depending on direction of movement). After each button press, an arrow symbolizing the decision was displayed in the upper center half of the screen until the decision was executed or another decision overrode the previous 1.

All participants received the different training zoos in the same order, but, depending on their performance, completed different amounts of zoos during the training period (i.e., time of training was kept constant: 42 50-minute sessions).

2.4. Pretest and posttests

Pretest included, in the following order: (1) an introductory group session; (2) 2 individual 50-minute sessions familiarizing participants with the VE and the navigation task; (3) 1 individual 50-minute session assessing navigation performance in an easier 10-animal version of the navigation task; (4) 1 MRI session; (5) two 90-minute sessions of cognitive testing in small (4–6 participants) group settings; and (6) a final individual assessment of navigation in a version of the navigation task similar to that administered during training (intermediate difficulty; see 2.4.2. Assessment of navigation performance). Both posttests included: (1) one 50-minute individual session refamiliarizing walkers to the VE and the navigation task; (2) one 50-minute session assessing navigation performance in a version of the navigation task with intermediate difficulty; (3) one 50-minute session assessing navigation performance in a more difficult version of the navigation task (more objects that to a higher degree prevent participants from seeing the animals from a distance); (4) 1 MRI session; and (5) two 90-minute sessions of cognitive testing. All sessions were completed on separate days. Below we describe the sessions in more detail.

2.4.1. Familiarization

The 2 sessions at pretest included familiarization with walking on the treadmill and navigating in a smaller version (3 animals, 5 decision points) of the zoos at various walking speeds. These sessions were administered to familiarize participants to walking on a treadmill, and to enable response mappings of navigating in the VE. Participants selected their preferred speed at the end of the second session (see 2.3. Training protocol, for details). At posttest, walkers were refamiliarized with the navigation task, by navigating in the same smaller versions of the zoos as used at pretest.

2.4.2. Assessment of navigation performance

We collected performance data for zoos of varying difficulty (easy, intermediate, and difficult). Difficulty was manipulated by a combination of the number of animals included and the density of objects in the VE. Higher density of objects increases difficulty by preventing participants from seeing the animals from a distance. The different difficulties were implemented to ensure that we covered the whole measurement space, including younger and older adults before and after training. As the version of the navigation task with intermediate difficulty was most similar to the trained task, it was administered at all measurement occasions, and showed acceptable psychometric properties (no ceiling or floor effects for any of the groups), we base the results on data from this version and describe these sessions in more detail here.

Participants started walking at the entrance to a virtual zoo and searched for the animal currently displayed at the lower right corner of the screen (see Fig. 1B and C). When the currently displayed animal had been found, a new target animal was displayed and participants started searching for this animal. After all 14 animals had been found, the participant completed the trial by finding the exit of the zoo, received feedback on the number of minutes needed to complete the trial, and started a new trial in the same zoo. The dependent variable was the number of targets (animals and exits) found within 50 minutes of navigation (minimum = 0; maximum = 59).

Three different zoos were constructed, 1 for each assessment (pretest, posttest 1, and posttest 2). The 3 zoos had an identical structure of the paths and decision points and the animals were placed at identical locations (see Fig. 1D). In addition, the correct path between the animals was identical across zoos. However, the zoos contained different animals, skylines, distal landmarks (e.g., mountains, towers) placed at identical locations, and different layout of proximal landmarks (e.g., trees, hills, buildings). In this way, difficulty could be held constant across the zoos, although the visual appearance of the zoos was different. At each measurement occasion, the same zoo was administered to all participants. No participant reported being aware of that the structure of the paths was identical across zoos. Piloting confirmed that the 3 different zoos were of equal difficulty.

2.4.3. Assessment of cognitive performance (transfer)

Identical versions of the cognitive tests were administered at pretest and posttests using identical procedures. The tests were: Raven's advanced progressive matrices (Raven et al., 1998), mental rotations (Vandenberg and Kuse, 1978), vocabulary (Lehrl et al., 1991), Digit-Symbol Substitution (Wechsler, 1981), route memory (Jäger et al.,

1997), location memory (Jäger et al., 1997), Guilford-Zimmerman spatial orientation (Guilford and Zimmerman, 1948), and a battery of tests from the Cogito study (Schmiedek et al., 2010): object-position memory, numerical memory updating, numerical and figural comparison, spatial 2-back, word-list recall, and number-noun pairs. Though several of these tests may pick up performance drawing on abilities involved in our navigation task, the Guilford-Zimmerman task is arguably the 1 that best measures the spatial allocentric (perspective-taking) ability critical to performance on our navigation task. Thus, we primarily predicted navigation-related changes in performance for this task. A detailed description of all tests is available in Supplementary data.

2.4.4. Magnetic resonance imaging

High-resolution T_1-weighted images and diffusion-weighted images were acquired on a 3 Tesla Magnetom Trio tomograph (Siemens, Erlangen, Germany), with an 8-channel phased-array head coil. For T_1-weighted imaging, a magnetization prepared rapid gradient echo (MPRAGE) sequence was used (TE = 5.12 ms, TR = 2600 ms, TI = 1100 ms, flip angle = 7°, bandwidth = 140 Hz/pixel, matrix = 320 × 320 × 240, isometric voxel size = 0.8 mm^3).

The diffusion tensor imaging protocol consisted of 2 blocks of measurements with and without diffusion-weighting accumulating a total of 8 averages — 4 averages per block — for each diffusion gradient setting. One non-diffusion-weighted acquisition preceded the diffusion-weighted scans in each block and was later used for motion correction. Diffusion weighting was applied along 12 non-collinear directions, chosen according to the standard Siemens DTI acquisition scheme, with a b value of 1000 seconds/mm^2. To compensate for eddy currents diffusion weighting was conducted by the standard twice-refocused spin-echo sequence. Parallel imaging (GRAPPA) with an acceleration factor of 3 and 25% phase oversampling was applied to obtain the images (TE = 89 ms, TR = 5000 ms, matrix = 128 × 128, isometric voxel size = 2.0 mm^3). The slices were aligned to the connection line between genu and splenium of the corpus callosum.

In order to improve the coregistration of T_1-weighted and diffusion-weighted data, we also acquired an axial T_2-weighted TSE sequence (TE = 78 ms, TR = 3300 ms, 72 = slices, 2 mm slice thickness, matrix = 256 × 192, field of view (FOV) = 256 mm × 192 mm, with the same slice orientation as for the DTI protocol).

2.5. Data analysis

2.5.1. Behavioral data

We used a mixed 2 (training group: navigators vs. walkers) × 2 (age group: young vs. old) × 3 (time: pretest, posttest 1, posttest 2) analysis of variance (ANOVA) to detect possible training-related changes in navigation performance and performance on the transfer tests. Polynomial contrasts were used to detect linear and quadratic effects involving time. With this design and analysis, the critical intervention effect is revealed by an interaction between training group and time; that is, changes between pretest and posttests for navigation groups that are significantly different from those for the walking groups. Similarly, age-related differences in intervention effects are revealed by an interaction among age, training group, and time. In the presence of significant effects involving training × time, we followed up with ANOVAs and paired t tests to trace the source of the effect. The alpha level was set to $p = 0.05$ for all analyses.

2.5.2. Imaging data

All manual tracing was performed based on the T_1-weighted MPRAGE images using a stylus on a Wacom DTU-710 pen tablet (Wacom Technology Corp., Vancouver, WA, USA) and the Analyze 8.1 (AnalyzeDirect Inc., Overland Park, KS, USA) software package. The raters were always blind to group and measurement occasion.

2.5.2.1. Intracranial area (ICA). ICA has proven a valid surrogate measure for intracranial volume (ICV; Ferguson et al., 2005; Nandigam et al., 2007). In line with these findings, results revealed a high correlation between the geometric mean of sagittal and axial ICA and intracranial volume ($r = 0.94$) estimated from a subsample of images from this study (see Supplementary data, for details). Taking advantage of the lower labor costs, we therefore determined ICA for the entire data set.

Analyses of ICA revealed a significant linear effect, $F(1,87) = 5.68$, $p = 0.019$, and a marginally significant quadratic trend of time, $F(1,87) = 3.53$, $p = 0.064$. ICA increased with an average of 0.3% from pretest to posttest 1, $F(1,87) = 7.59$, $p = 0.007$, but was stable (change < 0.1%) between posttest 1 and posttest 2, $F < 1$. Importantly, this effect of time did not interact with age or training group, F's < 1. Thus, the observed changes do not threaten the interpretation of training-related effects, which rely on observing training group by time interactions. Nevertheless, we statistically controlled for ICA in the analyses of hippocampal volumes. Note that statistically controlling for total brain volume (as estimated with SPM5; www.fil.ion.ucl.ac.uk/spm; Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK) revealed substantively identical results to those reported here with ICA correction.

2.5.2.2. Hippocampus (HC) segmentation. Two experienced raters performed segmentation. Images were displayed in native space at a magnification factor of 3. For convenience, we first manually aligned all images (see Supplementary data). Next, we determined the anterior and posterior limits of the hippocampus to define the set of relevant slices. To minimize rater-related bias in single data sets, both raters always worked on just a subset of randomly chosen slices (Raz et al., 2004). Each of these subsets was of the same size. To enhance intraperson reliability across

the 3 measurement time points, we used a 2-step segmentation procedure. In the first step, 1 data set from each participant was chosen randomly and segmented according to the protocol described below. The resulting regions of interest (ROIs) were then coregistered to the remaining 2 data sets of that person using automatic rigid-body coregistration routines provided in Analyze 8.1. Next, all data sets were anonymized with respect to the measuring time point again, such that the raters were unaware of the origin of the template ROI. The final segmentation was done on the basis of both the segmentation protocol and the template that overlaid the images from each time point.

The entire hippocampus was segmented anterior-to-posterior using the coronal view as default (see Supplementary Fig. 1, for an example). Sagittal and axial planes were also consulted, however, to aid and validate the decision process. In general, segmentation was based on intrinsic anatomic properties of the hippocampus (Duvernoy, 2005). However, because there were several aspects of the hippocampus where such decisions could not be made with high reliability, we formulated a number of guiding rules. With a few exceptions, our approach followed the protocol provided by Pruessner and colleagues (Pruessner et al., 2000). Full details of our protocol are reported in the Supplementary data.

Before starting to trace these data, raters were trained using data sets from other studies. Training was continued until the 2 operators achieved an interrater agreement (ICC2; Shrout and Fleiss, 1979) of no less than $r = 0.90$ for each ROI volume. To investigate the reliability of the ROIs that entered our data analysis, we randomly chose a subsample of 20 data sets after the segmentation of the whole sample was completed. We then reverted to the ROIs that already existed and added a new parallel version such that each rater now processed those slices that the other rater had processed before. Because half of the samples originated from a random time point of the segmenting process, the average reliability over time should be better captured by our measure of interrater reliability than by a pure a posteriori evaluation. The interrater agreements for all slice-based and volume-based comparisons exceeded 0.93.

The hippocampi were further subsegmented into head, body, and tail based on the head-body and body-tail boundaries determined by visual inspection (see Supplementary data). To provide high comparability of the HC subregions over time, the MPRAGE datasets — containing the horizontally aligned hippocampi — were coregistered again for each subject individually using the pretest dataset as a template. This was performed with SPM5, and the same transformation matrices were then applied to the head, body, and tail binary masks that were generated from the manual segmentation. The division into subsegments was performed by calculating for each subject the average of the coregistered anterior-posterior boundary coordinates over time and then applying coronal cuts at these locations to the horizontally aligned hippocampi. Test-retest correlations (pretest-posttest1-posttest2) for the volumes of these subsegments were all high ($r > 0.97$), again indicating high stability and reliability of these measures.

The volumes of the head, body, and tail were adjusted for ICA with an analysis of covariance approach (e.g., Raz et al., 2005): adjusted volume = raw volume $- b \times$ (ICA-mean ICA), where b is the slope of regression of the raw volume on ICA. The adjusted volumes were analyzed in a similar way as the behavioral data, but the main ANOVA additionally included hemisphere (left/right) and segment (head/body/tail) as factors.

2.5.2.2. DTI data. The DTI data were used for obtaining mean MD values for the left and the right hippocampus. Here, we focused on mean values across the whole hippocampus in order to maximize reliability. In a first step, the DTI data were motion-corrected by coregistering the second acquisition block to the first block by means of the non-diffusion-weighted dataset for each block. Then, maps of the diffusion tensor elements were calculated using singular value decomposition, and voxel-wise diagonalization of the diffusion tensors the eigenvalues was used to calculate the MD for each voxel. Voxels with negative eigenvalues were excluded from statistical analysis. To obtain mean hippocampal MD values, the manual segmentations were coregistered to the non-diffusion-weighted images, using SPM5. For achieving optimal coregistration, the T_2-weighted TSE images were used as an intermediate step. Specifically, first the MPRAGE images and hippocampus ROIs were coregistered to the T_2-weighted images by rigid body transforms and by using normalized mutual information as cost function (because of the different contrasts to be coregistered). Second, the hippocampus ROIs were coregistered to the distorted DTI space of the same subject by means of the T_2-weighted image volumes, using a mixture of linear and nonlinear transforms with sum of squared differences used as cost function (because of the similar contrast in the T_2-weighted images and the non-diffusion-weighted images). To enable a voxel-by-voxel matching of the hippocampus masks with the MD maps, the latter were regridded without interpolating the data.

Four participants had to be excluded from the analyses of MD due to coregistration problems. Test-retest correlations (pretest-posttest1-posttest2) for the remaining MD estimates of left and right hippocampus ranged from 0.63 to 0.72, indicating acceptable stability and reliability of these measures. The estimates of MD were analyzed in a similar way as the behavioral data, but the main ANOVA additionally included hemisphere as a factor (left/right).

2.5.3. Correlation between performance and imaging measures

To examine brain-behavior correlations, we computed partial correlations, controlling for age, between changes (i.e., difference scores; posttest 1-pretest) over training in the cognitive performance measures and the adjusted hip-

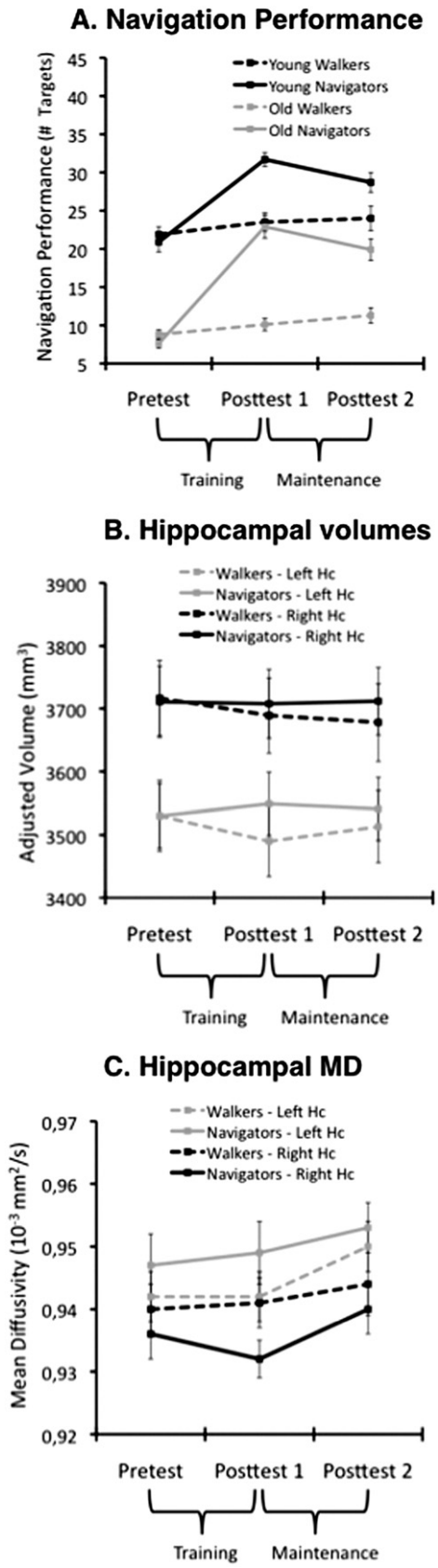

Fig. 2. (A) Mean (±SE) navigation performance (number of targets found within 50 minutes of navigation) as a function of training group (navigation/walk-time-yoked control), age group (young/old), and time of assessment (pretest/posttest 1/posttest 2). (B) Mean (±SE) hippocampal volumes as a function of hemisphere (left/right), training group (navigation/walk-time yoked control), and time of assessment (pretest/posttest 1/posttest 2). The training by time interaction reached significance, indicating that navigation training protected against the longitudinally observed decrements in hippocampal volumes. Neither hemisphere nor age qualified this interaction. (C) Mean (±SE) hippocampal mean diffusivity (MD) as a function of hemisphere (left/right), training group (navigation/walk-time yoked control), and time of assessment (pretest/posttest 1/posttest 2). Statistical analyses revealed significant training-related quadratic change over time for the right hippocampus (HC), but not for the left HC.

pocampal volumes and DTI estimates within the navigators. As cognitive performance measures, we used navigation performance and the transfer measures showing significant training × time interactions. In addition, we computed these correlations separately for each age group. The alpha level was adjusted for multiple comparisons (see 3. Results).

3. Results

3.1. Behavioral data

3.1.1. Navigation performance

Analyses addressing training-related changes in navigation performance (Fig. 2A) revealed significant linear, $F(1,87) = 40.18$, $p < 0.001$ and quadratic, $F(1,87) = 55.41$, $p < 0.001$, training × time interactions. Navigation training resulted in larger improvements than walking, which in turn resulted in marginal increases that likely reflect retest effects (see Fig. 2A). Older adults showed lower performance than younger adults, $F(1,87) = 173.07$, $p < 0.001$, but, as indicated by a non-significant trend for a linear age × training × time interaction, rather tended to show larger linear training-related improvements than younger adults, $F(1,87) = 2.80$, $p = 0.098$. This trend was confirmed by a significant interaction among age, training, and time in a follow-up ANOVA contrasting pretest with posttest 1 only, $F(1,87) = 4.41$, $p = 0.039$. Notably, older adults' performance after navigation training (i.e., at posttest 1) reached the performance of younger controls that had been walking (see Fig. 2A; $t(42) = 0.29$, $p > 0.77$). The net effect of navigation training (mean $\text{gain}_{\text{posttest 1 - pretest}}/\text{SD}_{\text{pretest}}$ for walkers subtracted from mean $\text{gains}_{\text{posttest 1 - pretest}}/\text{SD}_{\text{pretest}}$ for navigators) was 1.42 SD for younger adults and 3.34 SD for older adults. Training-related improvements were partly maintained 4 months after termination of training. However, as indicated by the significant quadratic training × time interaction, and a significant interaction between training and time in a follow-up ANOVA comparing the 2 posttests, $F(1,87) = 9.01$, $p = 0.004$, the training-related effect declined in magnitude during the maintenance period.

3.1.2. Transfer tasks

We observed significant main effects of age for all transfer tests, all F's $(1,87) > 9.46$, p's < 0.003, with older

adults performing less well than younger adults on all tests except the vocabulary test, which showed an age-related increase in performance. A nonsignificant trend for a quadratic training group by time interaction was observed for the Guilford-Zimmerman test, $F(1,87) = 3.33$, $p = 0.072$. Navigators tended to improve more than walkers on this measure of spatial orientation. The age by training group by time interaction did not reach significance, $F < 1$. For the Guilford-Zimmerman test, the overall net increase, subtracting control group gains, in test scores (based on posttest 1-pretest gains) was 0.29 SD (0.31 for younger adults and 0.22 for older adults). The training group × time interaction did not approach significance for any of the other transfer tasks, all F's $(1,87) < 2.27$, p's > 0.13.

3.2. Imaging data

3.2.1. Hippocampal volumes

Analyses of the manually segmented hippocampal head, body, and tail volumes, adjusted for ICA, revealed linear main effects of segment, $F(1,87) = 1718.71$, $p < 0.001$ (tail < body < head), and hemisphere, $F(1,87) = 117.97$, $p < 0.001$ (right > left). No other main effects, including the effect of age group, reached significance, all F's $(1,87) < 1.55$, p's > 0.21.

The hemisphere × time interaction, $F(1,87) = 4.58$, $p = 0.035$, as well as the linear, $F(1,87) = 5.81$, $p = 0.018$, and quadratic, $F(1,87) = 4.41$, $p = 0.039$, training group × time interactions were significant. These interactions were qualified by a quadratic training × hemisphere × time interaction, $F(1,87) = 7.74$, $p = 0.007$. The walkers decreased in both total left, $t(44) = 3.22$, $p = 0.002$, and total right, $t(44) = 2.38$, $p = 0.022$, hippocampal volumes over the training phase. For the right hippocampus, this decrease tended to continue during the maintenance phase (see Fig. 2B), but this change did not reach significance, $t(44) = 0.98$, $p = 0.334$. The left hippocampus volume showed mild increases, $t(44) = 2.11$, $p = 0.041$, which is the main source of the quadratic 3-way interaction. The volumes of the navigators were relatively stable across all time points, t's $(45) < 1.34$. Importantly, the training × time interactions were neither qualified by age group nor by segment (F's < 1), indicating that the training-related effects were not statistically different for younger and older adults or for the head, body, and tail.

Overall, the decreases from pretest to posttest 2 for the walkers correspond to a 0.75% annual decrease of the left (0.59% for younger and 0.85% for older adults) and a 1.59% decrease of the right (1.12% for younger and 1.94% for older adults) total hippocampal volume (see also Table 2). For the left hippocampus, these estimates are in the range of the estimates reported in previous longitudinal studies (e.g., DeLise et al., 1997; Scahill et al., 2003; Raz et al., 2005; see Raz, 2004, for review). For the right hippocampus, the decreases were somewhat larger than expected from the few available longitudinal studies, but not more extreme than what can be expected by sampling error. The estimates are also typical in the sense that they show a numerical trend for less longitudinal decline at a younger adult age (e.g., Raz et al., 2005; Scahill et al., 2003), but this age effect did not reach significance in an age × hemisphere × time ANOVA on hippocampus volumes for the walkers (i.e., the age × time interactions were not significant; F's < 1.08). Notably, this effect was also not significant in the corresponding analysis of the navigators' data (F's < 1), indicating similar stability of volumes in both younger and older age groups.

In summary, the main message from these analyses (see Fig. 2B) is that both younger and older adults in navigation training displayed stable hippocampal volumes both across the 4 month training phase and 4 months after termination of training, whereas participants in the walk-time-yoked control groups displayed declining volumes consistent with normal age-related decline.

3.2.2. Hippocampal mean diffusivity

Analyses of the hippocampal MD estimates (see Table 3) revealed a main effect of age, $F(1,83) = 7.95$, $p = 0.006$, with older adults displaying higher MD than younger adults. In addition, the linear main effects of hemisphere, $F(1,83) = 16.12$, $p < 0.001$, and time, $F(1,83) = 5.63$, $p = 0.020$, reached significance. These effects were qualified by a reliable training group × hemisphere interaction, $F(1,83) = 5.84$, $p = 0.018$, which was further qualified by a quadratic training group × hemisphere × time interaction, $F(1,83) = 4.04$, $p = 0.048$. Figure 2C displays the means relevant for interpreting this 3-way interaction. To further help tracing the source of this interaction, we ran mixed age by time ANOVAs separately for the training groups and for hemisphere. For right hippocampal MD, the navigators displayed a significant quadratic effect over time, $F(1,43) = 4.61$, $p = 0.037$, with a trend for decreases after training and return to baseline across the maintenance phase, whereas the walkers did not display any significant changes for the right hippocampal MD estimates, F's $(1,40) < 1.28$, p's > 0.265. For left hippocampal MD, both training groups displayed nonsignificant trends for linear increases, $F(1,40) > 3.42$, $p = 0.072$ for walkers and $F(1,43) = 3.65$, $p = 0.063$ for navigators. No other effects reached significance. Again, age did not interact significantly with any other factor in the model (F's < 1).

To summarize the analyses of the MD estimates, older adults showed higher MD than younger adults and there was a training-related quadratic change over time for the right hippocampus, but not for the left hippocampus.

3.3. Brain-behavior correlations

No significant ($p = 0.05/8 = 0.006$) partial correlations in the group of navigators, controlling for age, were observed between navigation performance and the Guilford-Zimmerman test, on the 1 hand, and hippocampal volumes and MD estimates, on the other, all p's > 0.072. Computing

these correlations separately for the age groups did not show any significant associations either, all p's > 0.065.

4. Discussion

This study shows that spatial navigation training improves performance on the criterion task and protects the hippocampus against age-related volume changes, as observed longitudinally. Younger and older adults in navigation training exhibited stable hippocampal volumes during training that were further maintained 4 months after termination of training, whereas individuals in the walk-time yoked control groups displayed volume decrements in the range of previously reported estimates of longitudinally observed age-related decline (e.g., DeLise et al., 1997; Scahill et al., 2003; Raz et al., 2005; see Raz, 2004, for review). This finding extends cross sectional (Maguire et al., 2000, 2006) and longitudinal (Boyke et al., 2008; Draganski et al., 2006) studies reporting experience-dependent morphological changes in the adult human hippocampus. Critically, the findings reported in this article are, to our knowledge, the first longitudinal evidence on hippocampal alterations in humans that has been established with manual segmentation of hippocampal volumes. Our study is also the first to report longitudinal evidence indicating that spatial navigation experience may modify hippocampal volumes in humans. This proof-of-principle suggests that previously observed morphological differences between individuals with different histories of demands on spatial navigation (e.g., Maguire et al., 2000, 2006) could indeed reflect experiential effects operating in adulthood. In contrast to the cross sectional studies on London taxi drivers by Maguire and colleagues (Maguire et al., 2000, 2006), our findings generalized across 3 subsegments (head, body, and tail) of the hippocampus.

The biological nature of experience-dependent moderation of changes in gray matter morphology is unknown. The list of potential candidate mechanisms is long (see Roth et al., 2010, for a discussion), including experience-dependent changes in number and size of blood vessels, glia, and neurons (either in the cell body, dendrites, or axons, or a combination of these). Though neurogenesis alone (without assuming also cascading effect on, for example, blood vessels) is unlikely to explain the volume effects, the similarity of our paradigm and the resulting findings to several animal studies of neurogenesis (e.g., Kempermann et al., 2002; Kronenberg et al., 2006) is striking. In relation to the present data of stable hippocampal volumes for navigators as compared with declining volumes for controls, such experience-dependent changes may offset negative hippocampal changes. Alternatively, use of its structure may protect the hippocampus from negative changes in adulthood, such as decrease in cell size and synaptic density.

Our analyses of hippocampal mean diffusivity partially inform the nature of the observed effects on volume. Because elevated diffusion of water molecules in tissue indicates a reduced density of membranes (Beaulieu, 2002), mean diffusivity might reflect barrier sparseness of the hippocampi. Thus, across the 8 months of the study, volume declined for the walkers and barrier sparseness within the volume increased marginally (at least in the left hippocampus; see Fig. 2C). Conceivably, both of these findings indicate negative changes. For the navigators, MD estimates of the left hippocampus displayed marginal increases. In other words, training status (navigation or walking) did not moderate the left hippocampal MD estimates. It is, therefore, unlikely that different dynamics of changes in cell membranes and tissue fluid underlies the differential volume changes in the left hippocampus for the 2 training groups. Contrasting with these increases, the MD estimates of the right hippocampus of navigators indicated quadratic change, with a decrease of barrier sparseness (i.e., increased density) after training followed by a return to baseline. Thus, the maintained right hippocampal volumes for navigators may not emanate from a volume equilibrium of decreased number of cell membranes and increased tissue fluid. Rather, for the right hippocampus, the results indicate increases in barrier density accompanied by maintained volume. Conceivably, both of these changes denote positive effects as compared with patterns of the walkers.

Clearly, more research on the underlying nature of experience-dependent changes in gray matter morphology in humans is needed. Microstructural imaging methods, such as DTI, may be promising in this respect, especially considering their higher sensitivity to microstructural changes in hippocampal integrity that predict progression to dementia better than volume (Kantarci et al., 2005). The present finding that hippocampal MD shows significant cross sectional age differences, whereas volume, in line with previous studies reporting mild cross sectional age differences (see Raz, 2000, 2004, for reviews), do not, further point to the usefulness of these methods.

The functional relevance of recently observed experience-dependent changes in hippocampal gray matter morphology (e.g., Boyke et al., 2008; Draganski et al., 2006) is unknown (see also Wolbers and Hegarty, 2010). Though issues such as the power to detect these associations, restricted reliability of difference scores, and restricted individual differences in change come to front, we note that this study did not detect any correlations between brain and behavioral changes. Though these results are similar the negligible cross sectional correlations between memory performance and hippocampal volume (Van Petten, 2004), there is a need for investigating the functional relevance of experience-dependent change in gray matter morphology in larger samples and with methods suitable for analyzing change (e.g., structural equation modeling; see e.g., Schmiedek et al., 2010). Further, we did not detect any sizable transfer of the improvements in navigation performance to the untrained tasks. One possible exception, which showed

a nonsignificant trend for a training by time interaction, was performance on the Guilford-Zimmermann task (Guilford and Zimmerman, 1948). This task was the only 1 in our battery that measured performance on spatial perspective-taking (Hegarty and Waller, 2004), which is a key part of spatial navigation based on a cognitive map strategy and dependent on the hippocampal formation (e.g., Burgess et al., 2002; Maguire et al., 1998; O'Keefe and Nadal, 1978). Thus, it makes sense to see transfer effects to this task. However, the marginal significance of this finding, together with the absence of a brain-behavior correlation, suggests that this effect should be interpreted with caution.

Notably, the observed effects of navigation training on both brain and behavioral data were, with 1 exception, not influenced by age. The exception, that older adults tended to improve more in navigation performance from navigation training, runs counter to several studies reporting smaller effects of cognitive training in older than in younger adults (e.g., Brehmer et al., 2007; Dahlin et al., 2008; Kliegl et al., 1990; Verhaeghen and Marcoen, 1996). One explanation for this discrepancy is that the age effect on training gains observed in this study possibly stems from factors extraneous to navigation performance per se. Specifically, though we administered 2 familiarization sessions before pretest, older adults may have taken longer time to familiarize themselves with the VE and its interface. An additional possibility is that the training was more demanding and thus induced a larger mismatch between the demands on spatial functioning and the available resources in older adults, thus constituting a stronger relative treatment counteracting possible age-related reductions in plasticity (see also Lövdén et al., 2010).

The present study included men exclusively. This design feature was implemented to reduce interindividual differences in navigation strategies. On average, men tend to rely more on a cognitive map strategy (Lövdén et al., 2007) and to activate the hippocampus (Grön et al., 2000) to a greater extent during navigation tasks than women. Further points to note are that cardiovascular fitness training has been shown to affect cognitive functioning (Colcombe and Kramer, 2003), and that self motion may play an important role in learning-related plasticity (Kempermann, 2008). Based on our comparison to the walk-time yoked control group, we can rule out that the reported effects of navigation training reflect effects of the walking component itself. However, we cannot rule out that the walking component is a contributing factor in interaction with the navigation component. That said, the low demands of the slow walking speed used in this study are unlikely to cause any major effects on physical fitness.

In general, this study offers reasons for being moderately optimistic that mental stimulation through directed cognitive training may improve spatial navigation and may have important protective effects against negative age-related hippocampal changes in adulthood and old age. Spatial abilities (e.g., Lövdén et al., 2005a; Moffat et al., 2007) and hippocampal volumes (Raz, 2000, 2004) decline over the adult lifespan. Older adults commonly restrict their mobility to avoid unfamiliar environments due to self-perceived impairments in navigation (Burns, 1999), likely reducing mental, physical, and social stimulation, which in turn may lead to accelerated cognitive decline in old age (Hertzog et al., 2009; Lövdén et al., 2005b). The experience-related gains in spatial navigation and the microlongitudinal preservation of hippocampal integrity observed in this study nurture the hope that spatial cognition and its neural substrates can be preserved and improved throughout the adult life span by engaging in spatially challenging tasks. Future research should investigate whether experiential factors attenuate aging-related hippocampal deterioration over longer periods (e.g., years and decades) than the 8-month period investigated in this study.

Disclosure statement

The authors disclose no conflicts.

The ethical review board of the Otto-von-Guericke University of Magdeburg approved the imaging part of the study, and the ethical review board of the Max Planck Institute for Human Development, Berlin approved the behavioral part of the study. Written informed consent was obtained prior to the investigation.

Acknowledgements

Funded by the Max Planck Society; the Sofja Kovalevskaja Award (to ML) administered by the Alexander von Humboldt Foundation and donated by the German Federal Ministry for Education and Research (BMBF); the German Research Council; the BMBF; the Swedish Research Council; Swedish Brain Power; and by an Alexander von Humboldt Research Award (to LB).

The authors thank Colin Bauer, Daniel Bittner, Marcel Derks, Isabel Dziobek, Gabriele Faust, Martin Kanowski, Jörn Kaufmann, Kristen Kennedy, Shireen Kwiatkowska-Naqvi, Naftali Raz, Karen Rodrigue, Michael Schellenbach, Claus Tempelmann, Elisabeth Wenger, and all student assistants.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neurobiolaging.2011.02.013.

References

Ashburner, J., Friston, K.J., 2001. Why voxel-based morphometry should be used. Neuroimage 14, 1238–1243.

Baltes, P.B., Lindenberger, U., 1988. On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. Behav. Ther. 19, 283–300.

Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system - a technical review. NMR Biomed. 15, 435–455.

Bookstein, F.L., 2001. "Voxel-based morphometry" should be not be used with imperfectly registered images. Neuroimage 14, 1454–1462.

Boyke, J., Driemeyer, J., Gaser, C., Büchel, C., May, A., 2008. Training-induced brain structure changes in the elderly. J. Neurosci. 28, 7031–7035.

Brehmer, Y., Li, S.C., Müller, V., von Oertzen, T., Lindenberger, U., 2007. Memory plasticity across the lifespan: Uncovering children's latent potential. Dev. Psychol. 43, 465–478.

Burgess, N., Maguire, E.A., O'Keefe, J., 2002. The human hippocampus and spatial and episodic memory. Neuron 35, 625–641.

Burns, P.C., 1999. Navigation and the mobility of older drivers. J. Gerontol. B Psychol. Sci. Soc. Sci. 54, 49–55.

Churchill, J.D., Galvez, R., Colcombe, S., Swain, R.A., Kramer, A.F., Greenough, W.T., 2002. Exercise, experience and the aging brain. Neurobiol. Aging 23, 941–955.

Colcombe, S., Kramer, A.F., 2003. Fitness effects on the cognitive function of older adults: A meta-analytic study. Psychol. Sci. 14, 125–130.

Dahlin, E., Stigsdotter-Neely, A., Larsson, A., Bäckman, L., Nyberg, L., 2008. Transfer of learning after updating training mediated by the striatum. Science 320, 1510–1512.

Davatzikos, C., 2004. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. Neuroimage 23, 17–20.

DeLise, L.E., Sakuma, M., Tew, W., Kuschner, M., Hoff, A.L., Grimson, R., 1997. Schizophrenia as a chronic active brain process: a study of progressive brain structural change subsequent to the onset of schizophrenia. J. Psychiatr. Res. 74, 129–140.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H.G., Winkler, J., Büchel, C., May, A., 2006. Temporal and spatial dynamics of brain structure changes during extensive learning. J. Neurosci. 26, 6314–6317.

Duvernoy, H.M., 2005. The Human Hippocampus: Functional Anatomy, Vascularization and Serial Sections With MRI, third ed. Springer-Verlag, Berlin.

Ferguson, K.J., Wardlaw, J.M., Edmond, C.L., Deary, I.J., Maclullich, A.M., 2005. Intracranial Area: A validated method for estimating intracranial volume. J. Neuroimaging 15, 76–78.

Grön, G., Wunderlich, A.P., Spitzer, M., Tomczak, R., Riepe, M.W., 2000. Brain activation during human navigation: Gender-different neural networks as substrates of performance. Nat. Neurosci. 3, 404–408.

Guilford, J.P., Zimmerman, W.S., 1948. The Guilford-Zimmerman Aptitude Survey. J. Appl. Psychol. 32, 24–34.

Hegarty, M., Waller, D., 2004. A dissociation between mental rotation and perspective-taking spatial abilities. Intelligence 32, 175–191.

Hertzog, C., Kramer, A.F., Wilson, R.S., Lindenberger, U., 2009. Enrichment effects on adult cognitive development: Can functional capacity of older adults be preserved and enhanced. Psychol Sci. Public Interest 9, 1–65.

Hoyer, W.J., Stawski, R.S., Wasylyshyn, C., Verhaghen, P., 2004. Adult age and digit symbol substitution performance: a meta-analysis. Psychol. Aging 19, 211–214.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., Perrig, W.J., 2008. Improving fluid intelligence with training on working memory. Proc. Natl. Acad. Sci. U. S. A. 105, 6829–6833.

Jessberger, S., Gage, F.H., 2008. Stem-cell-associated structural and functional plasticity in the aging hippocampus. Psychol. Aging 23, 684–691.

Jäger, A.O., Süss, H.-M., Beauducel, A., 1997. Test für das Berliner Intelligenzstrukturmodell (BIS) [in German]. Hogrefe, Göttingen.

Kantarci, K., Petersen, R.C., Boeve, B.F., Knopman, D.S., Weigand, S.D., O'Brien, P.C., Shiung, M.M., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Jack, C.R., 2005. DWI predicts future progression to Alzheimer disease in amnestic mild cognitive impairment. Neurology 64, 902–904.

Karbach, J., Kray, J., 2009. How useful is executive control training? Age differences in near and far transfer of task-switching training. Dev. Sci. 12, 978–990.

Kempermann, G., 2008. The neurogenic reserve hypothesis: what is adult hippocampal neurogenesis good for? Trends Neurosci. 31, 163–169.

Kempermann, G., Gast, D., Gage, F.H., 2002. Neuroplasticity in old age: Sustained fivefold induction of hippocampal neurogenesis by long-term environmental enrichment. Ann. Neurol. 52, 135–143.

Kliegl, R., Smith, J., Baltes, P.B., 1990. On the locus and process of magnification of age differences during mnemonic training. Dev. Psychol. 26, 894–904.

Kronenberg, G., Bick-Sander, A., Bunk, E., Wolf, C., Ehninger, D., Kempermann, G., 2006. Physical exercise prevents age-related decline in precursor cell activity in the mouse dentate gyrus. Neurbiol. Aging 27, 1505–1513.

Lehrl, S., Merz, J., Burkard, G., Fischer, B., 1991. Manual zum MWT-a [in German]. Perimed, Erlangen, Germany.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., Schmiedek, F., 2010. A theoretical framework for the study of adult cognitive plasticity. Psychol. Bull. 136, 659–676.

Lövdén, M., Schellenbach, M., Grossman-Hutter, B., Krüger, A., Lindenberger, U., 2005a. Environmental topography and postural control demands shape aging-associated decrements in spatial navigation performance. Psychol. Aging 20, 683–694.

Lövdén, M., Ghisletta, P., Lindenberger, U., 2005b. Social participation attenuates decline in perceptual speed in old and very old age. Psychol. Aging 20, 423–434.

Lövdén, M., Herlitz, A., Schellenbach, M., Grossman-Hutter, B., Krüger, A., Lindenberger, U., 2007. Quantitative and qualitative sex differences in spatial navigation. Scand. J. Psychol. 48, 352–358.

Maguire, E.A., 1998. Knowing Where and Getting There: A Human Navigation Network. Science 280, 921–924.

Maguire, E.A., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S., Frith, C.D., 2000. Navigation-related structural change in the hippocampi of taxi drivers. Proc. Natl. Acad. Sci. U. S. A. 97, 4398–4403.

Maguire, E.A., Woollett, K., Spiers, H.J., 2006. London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. Hippocampus 16, 1091–1101.

McNaughton, B.L., Barnes, C.A., Gerrard, J.L., Gothard, K., Jung, M.W., Knierim, J.J., Kudrimoti, H., Qin, Y., Skaggs, W.E., Suster, M., Weaver, K.L., 1996. Deciphering the hippocampal polyglot: The hippocampus as a path integration system. J. Exp. Biol. 199, 173–185.

Moffat, S.D., Kennedy, K.M., Rodrigue, K.M., Raz, N., 2007. Extrahippocampal contributions to age differences in human spatial navigation. Cereb. Cortex 17, 1274–1282.

Nandigam, R.N., Chen, Y.W., Gurol, M.E., Rosand, J., Greenberg, S.M., Smith, E.E., 2007. Validation of intracranial area as a surrogate measure of intracranial volume when using clinical MRI. J. Neuroimaging 17, 74–77.

Noack, H., Lövdén, M., Schmiedek, F., Lindenberger, U., 2009. Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. Restor. Neurol. Neurosci. 27, 435–453.

O'Keefe, J., Nadal, L., 1978. The Hippocampus as a Cognitive Map. Oxford University Press, Oxford.

Olesen, P.J., Westerberg, H., Klingberg, T., 2004. Increased prefrontal and parietal activity after training of working memory. Nat. Neurosci. 7, 75–79.

Persson, J., Reuter-Lorenz, P.A., 2008. Gaining control: Training executive function and far transfer of the ability to resolve interference. Psychol. Sci. 19, 881–888.

Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. Cereb. Cortex 10, 433–442.

Raz, N., 2000. Aging of the brain and its impact on cognitive performance: integration of structural and functional findings, in: Craik, F.I.M., Salthouse, T.A. (Eds.), The Handbook of Aging and Cognition, second ed. Erlbaum, Mahwah, NJ, pp. 1–90.

Raz, N., 2004. The aging brain observed in vivo: differential changes and their modifiers, in: Cabeza, R., Nyberg, L., Park, D.C. (Eds.), Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging. Oxford University Press, New York, pp. 17–55.

Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., Acker, J.D., 2004. Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. Neurobiol. Aging 25, 377–396.

Raz, N., Lindenberger, U., Rodrigue, K.M., Kennedy, K.M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., Acker, J.D., 2005. Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. Cereb. Cortex 15, 1676–1689.

Raven, J.C., Raven, J.E., Court, J.H., 1998. Progressive Matrices. Oxford Psychologist Publishing Group, Oxford, England.

Rosenzweig, M.R., Bennett, E.L., 1996. Psychobiology of plasticity: effects of training and experience on brain and behavior. Behav. Brain Res. 78, 57–65.

Roth, T.C., Brodin, A., Smulders, T.V., LaDage, L.D., Pravosudov, V.V., 2010. Is bigger always better? A critical appraisal of the use of volumetric analysis in the study of the hippocampus. Philos. Trans. R. Soc. B 365, 915–931.

Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., Fox, N.C., 2003. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. Arch. Neurol. 60, 989–994.

Schmiedek, F., Lövdén, M., Lindenberger, U., 2010. Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. Frontiers in Aging. Neuroscience 2, 1–10.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.

Stackman, R.W., Clark, A.S., Taube, J.S., 2002. Hippocampal spatial representations require vestibular input. Hippocampus 12, 291–303.

Thomas, A.G., Marrett, S., Saad, Z.S., Ruff, D.A., Martin, A., Bandettini, P.A., 2009. Functional but not structural changes associated with learning: An exploration of longitudinal Voxel-Based Morphometry (VBM). Neuroimage 48, 117–125.

Van Petten, C., 2004. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. Neuropsychologia 42, 1394–1413.

van Praag, H., Kempermann, G., Gage, F.H., 2000. Neural consequences of environmental enrichment. Nat. Rev. Neurosci. 1, 191–198.

Verhaeghen, P., Marcoen, A., Goossens, L., 1992. Improving memory performance in the aged through mnemonic training: a meta-analytic study. Psychol. Aging 7, 242–251.

Verhaeghen, P., Marcoen, A., 1996. On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. Psychol. Aging 11, 164–178.

Vandenberg, S.G., Kuse, A.R., 1978. Mental rotation, a group test of three-dimensional spatial visualization. Percept. Mot. Skills 64, 907–914.

Wechsler, D., 1981. Manual for the Wechsler Adult Intelligence Scale-Revised. Psychological Corporation, New York.

Wolbers, T., Hegarty, M., 2010. What determines our navigational abilities. Trends Cogn. Sci. 14, 138–146.