Cerebral Cortex, 2017; 1–13

doi: 10.1093/cercor/bhw418
Original Article

ORIGINAL ARTICLE

Longitudinal Evidence for Increased Functional Response in Frontal Cortex for Older Adults with Hippocampal Atrophy and Memory Decline

Sara Pudas[1,2], Maria Josefsson[3], Anna Rieckmann[2,4] and Lars Nyberg[1,2,4]

[1]Department of Integrative Medical Biology, Umeå University, SE-901 87 Umeå, Sweden, [2]Umeå center for Functional Brain Imaging, Umeå University, SE-901 87 Umeå, Sweden, [3]Centre for Demographic and Ageing Research at Umeå University (CEDAR), Umeå University, SE-901 87 Umeå, Sweden and [4]Department of Radiation Sciences, Umeå University, SE-901 87 Umeå, Sweden

Address correspondence to Sara Pudas, Department of Integrative Medical Biology, Umeå University, SE-901 87 Umeå, Sweden.
Email: sara.pudas@umu.se

Abstract

The functional organization of the frontal cortex is dynamic. Age-related increases in frontal functional responses have been shown during various cognitive tasks, but the cross-sectional nature of most past studies makes it unclear whether these increases reflect reorganization or stable individual differences. Here, we followed 130 older individuals' cognitive trajectories over 20–25 years with repeated neuropsychological assessments every 5th year, and identified individuals with stable or declining episodic memory. Both groups displayed significant gray matter atrophy over 2 successive magnetic resonance imaging sessions 4 years apart, but the decline group also had a smaller volume of the right hippocampus. Only individuals with declining memory demonstrated increased prefrontal functional responses during memory encoding and retrieval over the 4-year interval. Regions with increased functional recruitment were located outside, or on the borders of core task-related networks, indicating an expansion of these over time. These longitudinal findings offer novel insight into the mechanisms behind age-associated memory loss, and are consistent with a theoretical model in which hippocampus atrophy, past a critical threshold, induces episodic-memory decline and altered prefrontal functional organization.

Key words: aging, fMRI, hippocampus, longitudinal study, memory decline, prefrontal cortex

Introduction

The functional organization of the brain is dynamic. Experience-dependent cortical plasticity is well established (Buonomano and Merzenich 1998; McEwen and Morrison 2013; Kolb and Gibb 2015). Plasticity of the prefrontal cortex (PFC) is particularly important in the context of cognition (Kuboshima-Amemori and Sawaguchi 2007), and may serve as a mechanism of compensation for cognitive impairment due to damage or disease (Reuter-Lorenz and Park 2014). Neuroimaging studies in humans have shown increased PFC functional responses during cognitive task performance in patients with Alzheimer's disease, which may counteract cognitive losses induced by neurodegeneration (Grady et al. 2003). Even in clinically normal older individuals, episodic (explicit) long-term-memory functioning declines with aging (Schaie 1994; Rönnlund et al. 2005). Such age-related memory decline has been linked to hippocampal atrophy and dysfunction (Petersen et al. 2000; Kramer et al. 2007; Persson et al. 2012; Hedden et al. 2016), as well as elevated PFC responses in cross-sectional functional neuroimaging studies of episodic memory encoding (Persson et al. 2006) and retrieval (Daselaar et al. 2003). In contrast to the view by which elevated PFC responses are an indicator of a dysfunctional neural system (Logan et al. 2002), it has been proposed that increased

frontal engagement is a compensatory response to hippocampal dysfunction or other neurodegenerative changes (Grady et al. 2005; Park and Reuter-Lorenz 2009). These hypotheses have typically been tested in cross-sectional studies by separating individuals into performance subgroups (Cabeza et al. 2002; Rosen et al. 2002), but in the absence of longitudinal data it cannot be ruled out that observed group differences in such studies reflect individual characteristics established in younger age (Karama et al. 2014; Pudas et al. 2014) rather than functional reorganization (Nyberg et al. 2010; Raz and Lindenberger 2013).

Cognitive aging trajectories show marked heterogeneity in large-scale population-based studies (Habib et al. 2007; Josefsson et al. 2012; Lindenberger 2014). Here, we studied 130 well-characterized older adults who had participated in a longitudinal study for 20–25 years (Nilsson et al. 1997), and were scanned with functional magnetic resonance imaging (fMRI) twice over a 4-year interval, while performing an episodic memory task. Among the participants, some showed accelerated episodic memory decline between baseline and follow-up MRI scans, relative to their prior performance trajectory over 15-20 years, whereas others remained on a stable trajectory (Fig. 1B). We hypothesized that the declining group would be characterized by hippocampus atrophy and elevated PFC functional responses over time. We investigated both the encoding and retrieval phases of the memory task. Previous cross-sectional observations have indicated process-general effects (Grady et al. 2003; Salami et al. 2012; Reuter-Lorenz and Park 2014). Therefore, we hypothesized that longitudinal elevation of PFC responses would be seen in overlapping regions for encoding and retrieval.

Materials and Methods

Participants

This research was approved by the local ethics board at Umeå University, and all participants provided written informed consent and were compensated monetarily for their participation. The sample comprised 130 participants (mean age 69 years at follow-up) from the longitudinal Betula study (Nilsson et al. 1997). The participants had taken part in cognitive testing with 5-year intervals for 20–25 years (64 of the participants were recruited at the later time point) and undergone 2 MRI sessions, approximately 4 years apart (M = 1458 days, range = 1267–1702). The participants included in this study were part of a larger imaging subsample of 376 Betula participants, which was scanned in 2009–2010, and from which 231 participants returned for the follow-up scan. Of the 231 returning participants, 176 had prior longitudinal behavioral data, which enabled inclusion in the current study. Of these, 5 individuals were excluded for having been identified as displaying cognitive decline prior to the baseline scan (according to procedures described in Josefsson et al. 2012), 22 were excluded due to diseases or brain pathology (e.g., strokes, epilepsy, multiple sclerosis, dementia), 2 had missing data (1 missing structural image, 1 missing cognitive assessment at follow-up), and 17 failed to perform the scanner task at a level that was considered to elicit reliable memory-related brain activation (performance criterion: >10 hits, <12 missing responses, out of a maximum of 24). All exclusions were made prior to imaging analyses. Demographics and characteristics of the included participants can be found in Table 1 (organized by outcome of the cognitive classification procedures described later).

Cognitive Measures and Classification of Cognitive Change

A composite score of 5 episodic memory measures was used to quantify the participants' memory performance over the duration of the study, and has been described in our previous reports (Nilsson et al. 1997; Persson et al. 2012; Pudas et al. 2013). The composite included 2 tests of immediate free recall of sentences (16 items each). The sentences were enacted by the participants during encoding in one of the tests, and studied without enactment in the other. Later, participants were asked to recall nouns from the enacted/studied sentences, with noun categories (e.g., fruits, animals) as cues. The number of correctly recalled nouns from each of these 2 delayed cued recall tests (maximum 16 per test) was added as 2 separate measures to the composite. The 5th measure in the composite was immediate free recall of a list of 12 unrelated nouns. The maximum score amounted to 76 points. Test procedures remained constant across measurement occasions, but 2 different item-lists were alternated between occasions to reduce practice effects. The composite score had a good level of

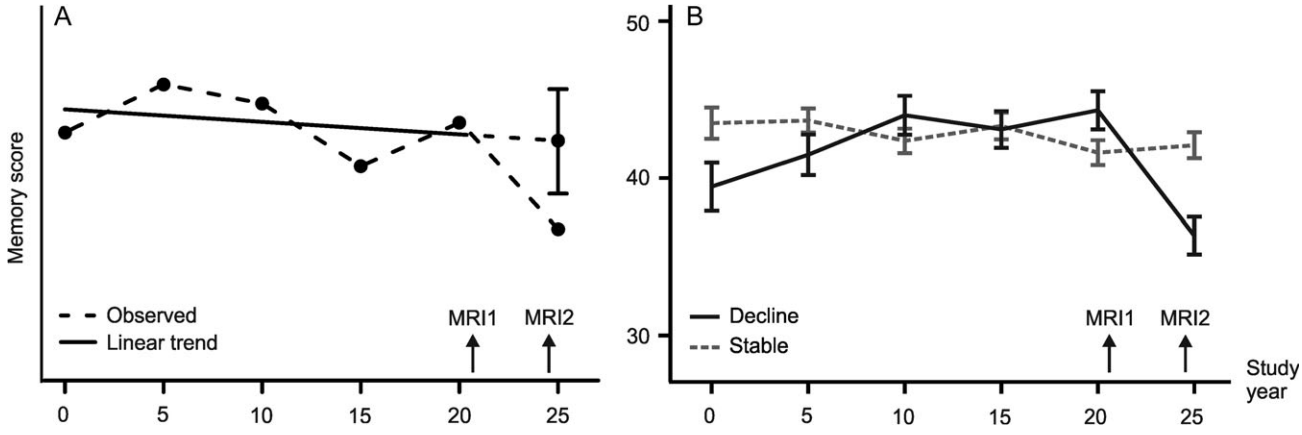

Figure 1. Classification of memory change and performance trajectories for resulting groups. Panel A: Illustration of the classification procedure (artificial data points) that was used to subdivide the sample into groups with stable and declining memory performance between the first and follow-up MRI sessions. Briefly, each participant's measured memory scores across years 0–20 were used to create a linear prediction for their performance level at the final test occasion (year 25). Those individuals whose actual performance fell below an 80% prediction interval were classified as having memory decline. Panel B: Actual group-averaged memory performance for the groups resulting from the classification procedure. Decline group $n = 41$, stable group $n = 89$. Error bars in Panel B represent ±1 SEM.

Table 1 Group characteristics and cognitive performance

	Time point	Stable Mean (SD)	Decline Mean (SD)	Between-group statistics
Demographics				
N		89	41	
% Female		54%	44%	$\chi^2 = 1.13$, $P = 0.29$
Age	Baseline MRI	64.9 (6.5)	65.7 (7.4)	$t = -0.64$, $P = 0.52$
	Follow-up MRI	69.0 (6.5)	69.7 (7.4)	$t = -0.60$, $P = 0.55$
Education, years	Baseline MRI	13.5 (4.0)	13.8 (4.2)	$t = -0.38$, $P = 0.70$
Offline cognitive scores				
MMSE (Folstein et al. 1975)	Follow-up MRI	28.2 (1.4)	28.2 (1.3)	$t = 0.07$, $P = 0.95$
Episodic memory	Study entry	42.5 (6.5)	39.4 (7.8)	$t = 2.36$, $P = 0.02^*$
	Baseline MRI	41.6 (7.4)	44.3 (7.8)	$t = -1.90$, $P = 0.06$
	Follow-up MRI	42.1 (7.8)	36.3 (7.7)	$t = 3.93$, $P < 0.001^*$
Vocabulary	Study entry	24.7 (3.2)	24.5 (3.1)	$t = 0.29$, $P = 0.78$
	Follow-up MRI	24.8 (3.2)	24.3 (2.7)	$t = 0.85$, $P = 0.40$
Block design	Study entry	33.6 (8.6)	33.5 (7.9)	$t = 0.06$, $P = 0.95$
	Follow-up MRI	28.8 (8.0)	27.4 (8.3)	$t = 0.91$, $P = 0.37$
Word fluency	Study entry	25.6 (8.0)	25.1 (6.7)	$t = 0.41$, $P = 0.70$
	Follow-up MRI	24.9 (7.7)	24.3 (5.5)	$t = 0.48$, $P = 0.63$
Scanner task performance				
Number of hits	Baseline MRI	15.4 (3.1)	14.6 (3.0)	$t = 1.32$, $P = 0.19$
	Follow-up MRI	16.5 (3.1)	16.7 (3.0)	$t = -0.35$, $P = 0.72$
Retrieval RT (ms)	Baseline MRI	2623.4 (301.9)	2661.7 (314.6)	$t = -0.66$, $P = 0.51$
	Follow-up MRI	2558.1 (336.7)	2571.0 (297.9)	$t = -0.21$, $P = 0.83$
Missing responses	Baseline MRI	3.5 (3.0)	3.6 (2.7)	$t = -0.11$, $P = 0.92$
	Follow-up MRI	0.66 (1.1)	0.44 (1.0)	$t = 1.08$, $P = 0.28$
Control task responses	Baseline MRI	28.2 (8.1)	28.4 (8.0)	$t = -0.10$, $P = 0.92$
	Follow-up MRI	31.7 (0.6)	31.7 (0.6)	$t = -0.06$, $P = 0.95$
Control task RT (ms)	Baseline MRI	510.6 (95.0)	523.6 (103.2)	$t = -0.70$, $P = 0.48$
	Follow-up MRI	486.1 (62.8)	506.4 (111.1)	$t = -1.10$, $P = 0.28$

internal consistency (Cronbach's α: 0.83) and test–retest reliability ($r = 0.79$; Pearson correlation; Pudas 2013).

Memory change between the baseline and follow-up MRI sessions was characterized in the following manner. Each individual's memory performance at the final measurement occasion (i.e., follow-up MRI) was predicted from his/her linear memory change across the duration of the study, up until the first MRI session, see Figure 1A. The memory change was estimated using ordinary least squares regression across 4–5 measurement points (i.e., 15–20 years), and a prediction interval of 80% was generated for each individual's performance at the last measurement point. Those individuals whose actual, measured, performance at the last time point fell below the lower limit of the 80% prediction interval were classified as memory decliners. In this manner, the classification procedure captured accelerated (i.e., nonlinear) performance decrements. Participants who performed within their prediction intervals were considered having stable memory. The memory change slope that was used to predict performance at the final measurement point was negatively correlated with age ($r = -0.235$, $P < 0.001$ across 290 participants that were part of the imaging sample at baseline), so the classification procedure indirectly accounted for age-effects across the sample age-range. In other words, older participants with a more negative prior memory trajectory had to decline to a lower level to be considered memory decliners. However, a potential limitation with the classification model was that some individuals had a positive slope across the first few measurement occasions due to practice effects, and/or regression to the mean effects for initially low-performing individuals. Such individuals required a lesser degree of memory decline to be classified into the decline group. We therefore performed control analyses to verify our results in subgroups defined with an additional, more conservative, criterion for decline, namely a negative slope for the entire duration of the study.

To more comprehensively characterize the cognitive profiles of the stable and declining memory groups, they were also compared on 3 other cognitive measures. The first was the Block Design test from WAIS-R (Wechsler 1981), in which participants were required to reproduce spatial patterns shown to them on cards, by rearranging red and white blocks. This test captures visuospatial ability, but is also assumed to reflect fluid intelligence. The second measure, assumed to reflect semantic memory, was a 30-item vocabulary test in which participants were required to choose synonyms for target words among 5 alternatives (Nilsson et al. 1997). The third measure was a composite of 3 word fluency tests, in which participants orally generated as many words as possible satisfying the following conditions (1) starting with the letter A, (2) 5-letter words with the initial letter M; and (3) names of professions beginning with the letter B (Nilsson et al. 1997). Word fluency measures are thought to reflect a combination of vocabulary knowledge and executive functioning (Shao et al. 2014).

fMRI Task and In-Scanner Setup

The scanner task at both baseline and follow-up MRI was a face–name paired-associates task (described in detail in our previous work, e.g., Persson et al. 2011; Salami et al. 2012; Pudas et al. 2013), implemented in E-prime software (version 2.0.8.22; Psychology Software Tools, Inc.). The 10-min task comprised 6 blocks of face–name encoding, 6 blocks of cued name retrieval and 8 blocks of an active control task involving a simple perceptual discrimination

(pressing a button each time a fixation star [*] changed to a circle). Mean duration between encoding and retrieval of a given face was 85.1 s (standard deviation [SD] = 26.1 s). Block order was pseudorandomized and constant across participants. Each block comprised 4 items, which were color photographs of unfamiliar faces, presented for 4 s each. During encoding blocks, the faces were paired with a common first name. During retrieval blocks, the same faces were presented together with 3 letters, from which participants were instructed to choose the one corresponding to the first letter of the name previously paired with the face. Responses were given through a button press on a scanner-compatible response pad, and participants were instructed to guess if uncertain. All participants completed a short practice version of the task at least once prior to scanning. In the scanner room, the task was displayed on a computer screen seen through a tilted mirror on the head coil. At baseline scanning, due to technical reasons unrelated to this specific study, approximately half of the participants (46% of the decline group and 47% of the stable group) had the computer screen placed at the foot of the scanner table, whereas the other half had it placed behind the scanner (with the mirror tilted the other way). At follow-up scanning, the screen was placed behind the scanner for all participants. The set-up with the screen behind the scanner reduced the distance between the participants' head and the screen compared with the placement at the foot of the scanner. This procedural change accounts for the longitudinal increase in visual cortex activation apparent in both groups (see Table 2 and Supplementary Table 1), but could not influence any group differences since equal numbers of stable and declining individuals had the screen placed in front of the scanner at baseline MRI.

MRI Acquisition

The same 3 T General Electric (GE) scanner equipped with a 32 channel head coil was used to collect images at both baseline and follow-up. Functional images were acquired with a gradient echoplanar imaging sequence (37 transaxial slices; thickness: 3.4 mm, gap: 0.5 mm, time repetition [TR]: 2000 ms, time echo [TE]: 30 ms, flip angle: 80°, field of view: 25 × 25 cm, matrix: 96 × 96 voxels [zero-filled to 128 × 128]). Ten dummy scans were collected and discarded prior to experimental image acquisition to allow for progressive saturation of the signal. Structural T1-weighted images were collected with a 3D fast spoiled gradient echo sequence (FSPGR; 180 slices with a 1 mm thickness; TR: 8.2 ms, TE: 3.2 ms, flip angle: 12°, field of view: 25 × 25 cm). Subject head movement during scanning was minimized using cushions inside the head coil.

Scanner Stability

The scanner underwent standard maintenance and upgrades of hardware and software in the interval between the baseline and follow-up scans of this study. A quality assurance routine was run weekly since November 2010 to assess signal stability. For functional scans, a procedure described by Friedman and Glover (2006) was used to assess scanner performance. The same fMRI protocol as in the study was used, and 300 time-points were collected for the GE brain phantom. Four quality assurance measures were used to describe potential changes in the scanner over the time period between baseline and follow-up scans; static spatial signal-to-noise ratio (SNR), signal-to-fluctuation-noise ratio (SFNR), percent drift, and percent fluctuations. For the time period from November 2010 to January 2011, the following averaged measurements were obtained—SNR: 152 (SD = 3.78); SFNR: 149 (SD = 1.72); percent drift: 0.45 (SD = 0.17); and percent fluctuations: 0.072 (SD = 0.006). The corresponding measurements for the follow-up scan period (October 2013 to August 2014) were SNR: 136 (SD = 8.33); SFNR: 136 (SD = 5.13); percent drift: 0.24 (SD = 0.14); and percent fluctuations: 0.099 (SD = 0.070). There was no significant variation of the drift and fluctuations in the time signal of the phantom, while a decline in the 2 signal-to-noise measures was observed, resulting in a reduction of SNR of about 8–10% at follow-up, relative to baseline. The fact that the values of SNR and SFNR are very similar indicates that the within-scan scanner instability is very small so that the SFNR is determined mainly by the spatial SNR. The spatial noise is assumed to be uncorrelated and will therefore be much reduced by the smoothing being performed in the processing of the functional data so that the impact of the reduced SNR on the results will be minor (Greve et al. 2011).

To assess the stability of volumetric measurements, the same T1-weighted FSPGR protocol as in the study was used to obtain volume data for the GE phantom. Data were thresholded well above the noise level and the selected voxels were used to calculate the volume of the phantom. The relative volume change between baseline and follow-up was 0.45%. This change is relatively small compared with the average change in total gray matter volume observed in this study (−1.83% across all participants). Furthermore, the positive change in the measured volume of the phantom suggests that the volumetric decline observed in the study might be slightly underestimated.

Preprocessing of Functional MRI Data

Functional data from both baseline and follow-up were preprocessed using SPM12 (Wellcome Trust Centre for Neuroimaging, Functional Imaging Laboratory, http://www.fil.ion.ucl.ac.uk/spm), implemented in MATLAB R2014b (MathWorks). First, all images were corrected for differences in acquisition time (slice timing). Second, head movement correction was performed by the realign and unwarp function, by which each volume was rigidly aligned to the first volume of the series. Thereafter, spatial normalization was performed in a multistep procedure employing DARTEL (Ashburner 2007). This involved segmenting each individual's structural T1-image into gray-matter, white-matter, and cerebrospinal fluid components, and coregistering the individual's functional images to these. Separate coregistrations were performed on data from baseline and follow-up MRI sessions. Thereafter, DARTEL was used to create a template image of baseline and follow-up data for each participant, and these individual template images were subsequently merged into a group-level DARTEL template comprising both the baseline and follow-up scans. The subject-specific flow fields from these transformations were applied to the functional images to transfer them into template space. The images were affinely aligned to 2 × 2 × 2 mm Montreal Neurological Institute (MNI) space, and smoothed with an isotropic 8 mm full width at half maximum Gaussian kernel.

Statistical Analyses of Functional MRI Data

Analyses were performed with the statistical parametric mapping approach implemented in SPM12. The preprocessed functional data were high-pass filtered (128 s), and voxel-wise general linear models were set up, with the experimental conditions from the scanner task (encoding, retrieval, and control) as regressors. Each regressor was modeled as a boxcar, convolved with the standard hemodynamic response function. Separate models were set up for baseline and follow-up fMRI data. The models

included the 6 realignment parameters from the motion correction step of the preprocessing as covariates of no interest. Thereafter, subject-level contrast images were generated, comparing the experimental conditions of the scanner task, encoding versus control and retrieval versus control, separately for baseline and follow-up data. These contrast images were then carried on to random-effects group analyses, which proceeded in several steps. First, in order to assess group differences at baseline fMRI, between-group two-sample *t*-tests were used to compare the groups with stable and declining memory. Thereafter, we sought to identify brain regions in which the activation of the decliners differed from the stable group at follow-up, and/or changed in relation to their own activation levels at baseline. This was tested by between-group *t*-contrasts for the follow-up data, and within-group (paired) *t*-contrast for each group separately. Thereafter, the parameter estimates (beta values) from the first-order contrasts (encoding-control task and retrieval-control task) were extracted from all significant PFC clusters from the whole-brain analyses and tested in a 3-way ANOVA (group × time × task condition) in IBM SPSS Statistics software (version 22, SPSS, Inc.) to verify whether the group × time interaction was significant. This analysis approach was chosen since whole-brain 3-way interactions cannot currently be modeled in SPM.

Between- and within-group whole-brain analyses were considered significant at an uncorrected threshold of $P = 0.0005$, with a minimum cluster size of 20 voxels. We also performed Monte Carlo-simulations using 3dClustSim implemented in AFNI (version 16.2.09, https://afni.nimh.nih.gov/afni/) to estimate the probability of false positives. 3dClustSim determines the cluster extent threshold that, at a given voxel-wise threshold, produces a corrected alpha level of 0.05. Spatial correlation across voxels in our data was assessed using AFNI's 3dFWHMx, and the long-tailed (non-Gaussian) model for spatial autocorrelation of fMRI noise was used when running 3dClustSim (Cox et al. 2016). The simulations were restricted to the frontal cortex by using the WFU Pickatlas frontal lobe mask (Maldjian et al. 2003), given our a priori interest in this region. This approach produced a recommended cluster extent threshold of 61 voxels for the uncorrected voxel-wise threshold of $P = 0.0005$. To foreshadow our results, 2 of our reported clusters survived this threshold, but since 3 additional smaller clusters showed a highly similar pattern of results (selective increases over time in the memory decline group), we are confident that these too reflect true effects rather than false positives. We indicate which clusters survive the more stringent threshold with a footnote in Table 2. Region-of-interest and follow-up analyses of extracted parameter estimates from the frontal peaks identified in the whole-brain analyses were considered significant at $P = 0.05$, given our a priori interest in frontal cortex functional dynamics.

Extraction of individual-level parameter estimates (beta values) was performed using an in-house developed software, DataZ. Beta values were averaged across a 5 mm radius sphere around peak coordinates from the whole-brain group analyses (exact coordinates indicated by footnotes in Table 2). The extracted values were then entered into IBM SPPS Statistics software for further testing.

Gray Matter Volumetry

To obtain estimates of total gray matter volume, hippocampal volume, and estimated total intracranial volume (ICV), the structural T1-weighted images were first processed using the standard processing stream in Freesurfer v.5.3 (http://surfer.nmr.mgh.harvard.edu/), separately for baseline and follow-up MRI images. Technical details of this procedure have been documented online and in previous publications (e.g., Fischl et al. 2002). Thereafter, the images from baseline and follow-up were processed through the Freesurfer longitudinal processing stream, which creates an unbiased within-subject template image to increase reliability of the segmentation and parcellation of brain regions over time (Reuter et al. 2012). The raw values for left and right hippocampal volumes obtained from Freesurfer were then corrected for differences in ICV, via the analysis of covariance method (Jack et al. 1989). This adjusts each individual's volume by an amount that is proportional to the difference between that individual's ICV and the mean ICV for the sample (adjusted volume = raw volume − b × [ICV − mean ICV]; where b is the slope of the regression line between the raw volume and ICV). The adjusted volumes were screened for outliers, and one individual (belonging to the cognitively stable group) was removed from further analyses due to a segmentation failure of left hippocampal volume at baseline MRI.

Results

Cognitive Classification and Characterization of Groups

Of the 130 included participants, 41 met the criteria for memory decline according to our classification procedure, in which each participant's prior memory performance change across 15–20 years was used to predict an 80% confidence interval for their performance at the final assessment (see Fig. 1 A). Participants performing below the lower confidence limit at the final follow-up were classified as showing memory decline. By the same logic, the remaining 89 individuals were considered having intact (stable) memory. The groups did not differ significantly with respect to age, gender, or education (Table 1). The average memory scores of the groups across 25 years can be seen in Figure 1B, which illustrates the average drop between time-points 5 and 6 (year 20 and 25) in the decline group, coinciding with the baseline and follow-up MRI sessions. The groups did not significantly differ on memory scores at baseline MRI, but they did differ at follow-up (Table 1, and significant group by time interaction: $F = 92.2$, $P < 0.001$). Figure 1B also shows that there was a group difference in memory performance already at study entry, with the decline group performing significantly lower than the stable group ($t = 2.36$, $P = 0.02$). However, the decline group still performed significantly lower at the final follow-up compared with their own performance at study entry (paired *t*-test: $t = 3.10$, $P = 0.004$; the corresponding difference for the stable group was nonsignificant: $t = 0.54$, $P = 0.59$).

In order to further characterize the groups we compared them with respect to 3 other cognitive measures: semantic memory (assessed by a vocabulary test), visuospatial ability (WAIS-R Block Design), and executive functioning/fluency as assessed by a composite measure of 3 word fluency tests. There were no group differences in performance on these tests, neither at study entry nor at the last measurement occasion at follow-up MRI (see Table 1 for statistics). This observation, together with the lack of group differences on the mini-mental state evaluation (MMSE) test (Folstein et al. 1975), and the fact that all participants scored above 24 on MMSE, underscores the fact that memory decline in the decliner group was subtle and not yet indicative of clinically relevant cognitive impairment.

In the MRI scanner, both groups showed evidence of practice effects over time for number of hits (main effect of time: $F = 25.13$, $P < 0.001$) and response times (RTs; main effect of time: $F = 9.32$, $P = 0.003$), with no significant group differences (see Table 1),

or group × time interactions (hits: $F = 2.30$, $P = 0.13$; RTs: $F = 0.25$, $P = 0.62$). Likewise, both groups showed improvement on the control condition task (perceptual discrimination), both in terms of number of responses (main effect of time: $F = 19.72$, $P < 0.001$) and RTs (main effect of time: $F = 11.65$, $P = 0.001$), but no significant group differences (Table 1) or group × time interactions were observed (all Ps > 0.70). Scanner task performance correlated with offline memory performance, $r = 0.53$, $P < 0.001$, based on 272 individuals' data at follow-up MRI.

Functional Imaging Results

Core Encoding and Retrieval Networks

Across all participants, encoding and retrieval regions were defined relative to the active control condition of the scanner task. Recruitment of the identified regions (including bilateral hippocampus) was generally stable over groups and time (Supplementary Fig. 1).

Increased Frontal BOLD-Signal Over Time for Memory Decliners

At baseline MRI, there were no group differences in BOLD-responses that survived the predefined statistical threshold, either in the encoding-control task or the retrieval-control task contrasts. The absence of baseline group differences was confirmed at a more liberal statistical threshold ($P = 0.001$ uncorrected, minimum 20 voxels).

At the follow-up MRI session, the decline group had higher task-related BOLD-signal than the stable group in the right dorsolateral and ventrolateral PFC during memory encoding and retrieval, as well as in the left dorsolateral PFC during retrieval (Fig. 2, top left panel; Table 2). A within-group contrast of BOLD-signal changes from baseline to follow-up corroborated the pattern observed in the between-group analysis. For decliners, increases in retrieval-related BOLD-signal over time were observed in 3 clusters in the left superior and middle frontal gyrus (Fig. 2, lower left panel; Table 2). One of these clusters (peak x, y, z = −36, 32, 44) overlapped 16 voxels with the left dorsolateral cluster identified in the between-group analysis (peak x, y, z = −32, 32, 42; Fig. 2, bar graph a). Thus, whole-brain analyses suggested process-general increases in PFC BOLD-magnitudes during memory encoding and retrieval for individuals with declining memory. These observations were confirmed with 3-way repeated-measures ANOVAs (group × time × task condition) on individual-level parameter estimates from the 5 identified frontal clusters from the whole-brain analyses. Group × time interactions reached significance in all 5 clusters (right dorsolateral PFC: $F = 5.56$, $P = 0.02$; left dorsolateral PFC: $F = 9.00$, $P = 0.003$; right ventrolateral PFC: $F = 5.90$, $P = 0.017$; left anterior PFC: $F = 8.90$, $P = 0.003$; left dorsal anterior PFC: $F = 7.88$, $P = 0.006$), and were driven by increases in the declining group (Fig. 2, bar graphs a–e). Interactions involving task condition were nonsignificant across all 5 clusters (Ps > 0.07; see Supplementary Fig. 2 for condition-specific parameter estimates).

We further explored whether there were any significant longitudinal PFC BOLD-signal decreases in the decline group, or PFC regions in which their BOLD-magnitude was lower than the stable group's at follow-up, but no significant effects were observed (for whole-brain results, see Table 2 and Supplementary Table 1). Furthermore, there were no significant changes (increases or decreases) in the PFC over time in the stable group (Supplementary Table 1). As a control analysis, we also tested whether the increase in PFC BOLD-responses in the decline group could be driven by the observed practice effect on the scanner task (Table 1). No significant correlations were observed between change in scanner task performance and change in PFC BOLD-responses in the decline group (encoding: $r = 0.040$, $P = 0.803$; retrieval: $r = 0.047$, $P = 0.772$; parameter estimates averaged across all 5 PFC clusters).

To aid interpretation of the group differences in time-dependent change of the PFC BOLD-signal, Figure 2 includes

Table 2 fMRI results

Region	Side	BA	Peak x, y, z	t-Value	Size
Group differences at follow-up, Decline > Stable					
Encoding > Control task					
Inferior frontal gyrus	Right	45/47	54, 34, −8	4.31	233[a]
Middle frontal gyrus	Right	9	28, 40, 42	4.13	105[a]
Occipital cortex	Right/left	18	0, −82, 32	3.73	20
Retrieval > Control task					
Middle frontal gyrus	Right	9	26, 40, 42[b]	3.95	108[a]
Inferior frontal gyrus	Right	45/47	52, 34, −10[b]	4.18	71[a]
Middle frontal gyrus	Left	9	−32, 32, 42[b]	3.88	42
Decline group, Follow-up > Baseline					
Retrieval > Control task					
Occipital cortex	Right/left	17/18/19	16, −92, 26	6.43	2872[a]
Occipital cortex	Left	18	−8, −96, −10	6.85	889[a]
Orbitofrontal/anterior cingulate cortex	Left	11	−2, 36, 0	4.90	100[a]
Occipital cortex	Right	18	14, −88, 10	4.86	95[a]
Posterior cingulate cortex/precuneus	Left	7/31	−6, −40, 54	4.22	44
Superior frontal gyrus	Left	46/10	−22, 56, 26[b]	4.12	39
Superior frontal gyrus	Left	10	−18, 62, 16[b]	3.95	34
Posterior cingulate cortex/precuneus	Right	7/31	18, -36, 50	4.13	30
Middle frontal gyrus	Left	9	−36, 32, 44	4.37	28

BA, Brodmann area; x, y, z, coordinates in MNI space; Size indicates cluster size in number of 2 × 2 × 2 mm voxels.

[a]Clusters that survive multiple comparisons at alpha level 0.05 as determined by Monte Carlo-simulations.

[b]Coordinates that were used to extract beta values for follow-up analyses. Note the overlap in coordinates for the frontal peaks in the encoding and retrieval contrasts.

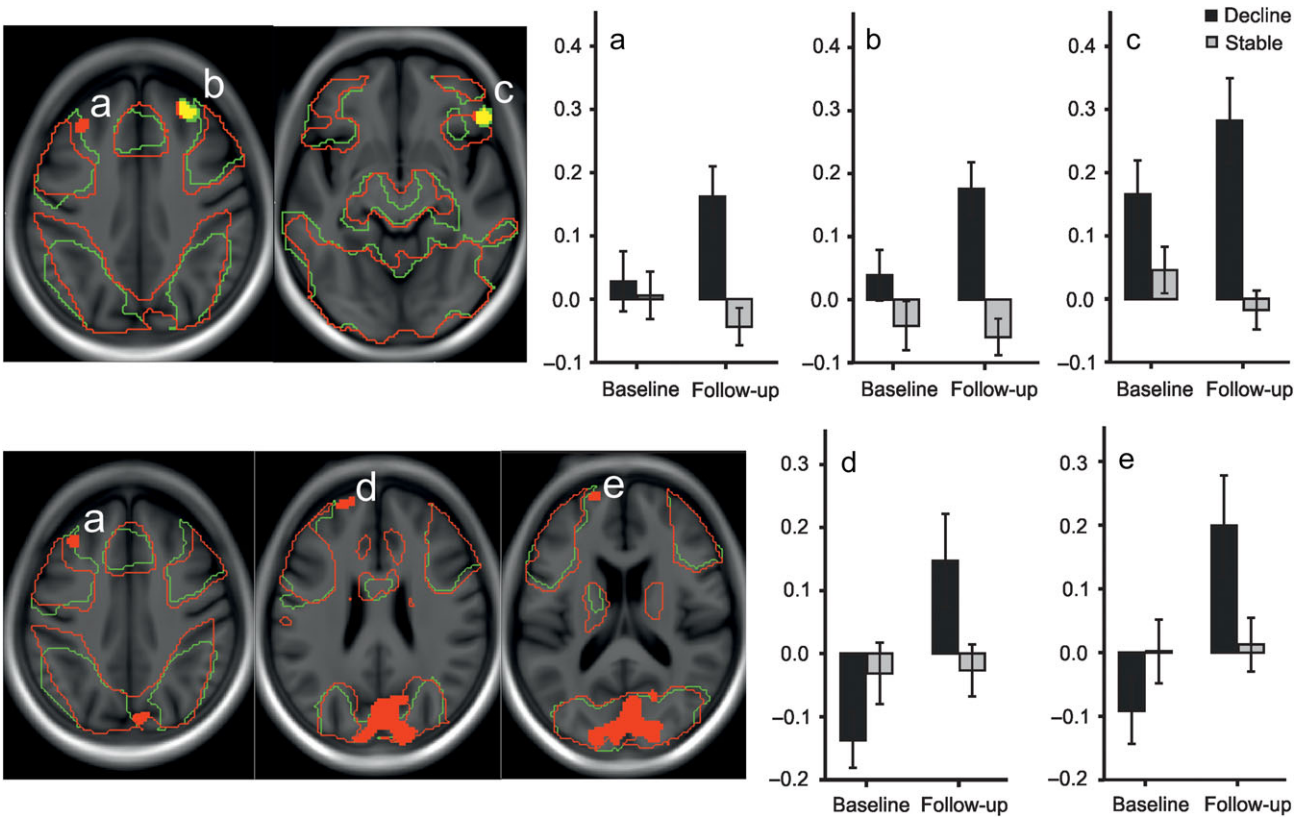

Figure 2. Functional imaging findings. The upper panel brain slices display regions in which the decline group had higher BOLD-signal than the stable group at follow-up MRI. Green color represents significant effects for the memory encoding condition, red for memory retrieval, and yellow denotes overlap between conditions. Contours show the main effects of memory encoding (green) and retrieval (red) compared with the control condition of the scanner task across all participants at baseline MRI (see Supplementary Fig. 1). Lower panel brain slices show significant BOLD-signal increases in the decline group (color coding same as above). The cluster of voxels in the medial parieto-occipital cortex is likely driven by a change in screen set-up in the MRI-room, as described in the Materials and Methods section (a similar effect is seen in the stable group, see Supplementary Table 2). All contrasts were thresholded at $P = 0.0005$ (uncorrected for multiple comparisons) with a cluster extent threshold of 20 voxels. Brain images are displayed according to neurological convention, that is, right hemisphere displayed on the right-hand side. Bar graphs show group average parameter estimates (beta values) for significant clusters with the encoding and retrieval conditions collapsed (see Supplementary Fig. 2 for graphs with the conditions separated). The beta values are difference scores, relative to the control condition of the scanner task. Error bars represent ±1 SEM.

contours that delineate the core encoding and retrieval networks (Supplementary Fig. 1). The PFC regions in which the decliners showed increases over time were consistently located outside, or on the borders of, the core mnemonic networks, which suggests that these networks expanded spatially over time in the decline group. This is further illustrated by the bar graphs that show that the stable group did not strongly recruit the regions at MRI sessions 1 or 2, and the same was true for decliners at MRI session 1 with the possible exception for right ventrolateral PFC, where a nonsignificant trend towards a group difference was seen already at baseline MRI (multivariate ANOVA across encoding and retrieval conditions: $F = 1.95$, $P = 0.15$).

PFC BOLD-Signal Magnitude at Follow-Up is Higher for the Lowest Performing Decliners

We next investigated the relationship between PFC BOLD-responses and offline memory performance. Individual-level BOLD-signal magnitudes across the 5 identified PFC clusters were significantly and positively correlated with each other ($0.31 < r < 0.72$, $Ps < 0.001$), so to reduce the number of statistical tests, we calculated an individual-level average frontal BOLD-signal magnitude for encoding and retrieval, respectively, across all clusters identified. These means were used in all subsequent analyses. Within the decline group, average PFC BOLD-signal magnitude during memory retrieval at the follow-up MRI session was negatively correlated with offline memory performance ($r = -0.31$, $P = 0.05$), indicating that individuals with the lowest memory performance engaged these PFC regions the most. The corresponding correlation for memory encoding BOLD-signal magnitude was in the same direction, but nonsignificant ($r = -0.20$, $P = 0.17$). The reason for only finding a significant association for retrieval BOLD is probably the overall stronger PFC-engagement for memory retrieval, compared with memory encoding, in our scanner task (see bar graphs in Supplementary Fig. 1). This stronger degree of task-modulation likely contributed to the fact that the parameter estimates from the retrieval condition were more likely to show significant associations to other variables. Within the stable group, BOLD-signal magnitudes were not related to memory performance at follow-up MRI (encoding: $r = -0.07$, $P = 0.51$; retrieval: $r = 0.09$, $P = 0.34$).

Verification of Results in Subgroups with a More Conservative Criterion for Memory Decline

A potential shortcoming with the current way of identifying memory decline is that individuals with a positive performance trajectory over the first few test occasions could more easily be classified as decliners. This is because the model

would generate an unreasonably high prediction for their performance level on the final follow-up due to their prior performance improvement. As can be seen in Figure 1B, an initial increase in memory performance was apparent in the decline group, which was most pronounced between the first and third measurement points. The mechanism behind this increase is likely a combination of practice effects, resulting from previous exposure to the cognitive tests, and a statistical artifact, regression towards the mean. To investigate whether the observed results may have been influenced by individuals in the decline group with such an improvement in memory performance, a subgroup analysis was performed. In this analysis, an additional criterion had to be met for being included in the decline group: a negative slope of memory change across the whole study duration (i.e., from study entry to follow-up MRI). This criterion resulted in the removal of 14 individuals from the decline group. Since these were younger than the 27 remaining decliners (65.3 vs. 72.0 years at follow-up), an age-matched stable subgroup was created for comparison ($n = 60$; mean age 71.9 years). In order to account for both age and potential misclassification of individuals with negative slopes as stable (analogous to the trimming of the decline group), the selection procedure for the new stable subgroup omitted all stable individuals under the age of 65 that had a negative slope from study entry to follow-up MRI, as well as 8 other randomly selected younger individuals.

The mean memory performance over time for the new subgroups can be seen in Figure 3A. There was no significant group difference in performance at study entry ($t = 0.55$, $P = 0.581$). For all other cognitive scores reported in Table 1, the subgroups showed the same pattern as the full sample, that is, no significant differences other than episodic memory at follow-up MRI. PFC BOLD-responses were tested with a repeated-measures ANOVA on the average BOLD-responses from the PFC-clusters identified previously. Similar to the results in the full sample, there was a significant group × time interaction ($F = 5.77$, $P = 0.018$), and a nonsignificant group × time × condition interaction ($F = 0.23$, $P = 0.629$), see Figure 3B-C. The analysis was repeated while controlling for gender since the decliner subgroup had a larger proportion of males than the stable subgroup, and the results remained unchanged (the main effect of gender and all interactions involving gender were nonsignificant; $Ps > 0.237$). Thus, the processes-general increases in PFC BOLD-responses in the full decliner group was replicated in the more conservatively defined decliner subgroup.

Hippocampus Volume and Atrophy

Gray matter volume data for the stable and decline groups are presented in Supplementary Table 2. All volumes were adjusted for estimated total ICV using the analysis of covariance approach (Jack et al. 1989), thereby removing differences related to body size (and by extension gender). In the full sample, the memory groups did not differ in ICV ($t = 1.20$, $P = 0.23$), or total gray matter volume at baseline ($t = 0.40$, $P = 0.69$) or follow-up ($t = 0.29$, $P = 0.77$). Total gray matter volume declined significantly over time across both groups (significant main effect of time: $F = 94.78$, $P < 0.001$), and there was no significant difference in rates of decline between groups (group × time interaction: $F = 0.15$, $P = 0.70$). The same pattern of results held true for the more conservatively defined subgroups of stable and declining individuals. That is, no significant group differences in ICV ($t = 1.84$, $P = 0.07$; trend for larger volumes in decline subgroup due to more men) or total gray matter volume (baseline: $t = 0.55$, $P = 0.58$; follow-up: $t = 1.13$, $P = 0.26$), a significant main effect of time for total gray matter ($F = 66.0$; $P < 0.001$), and no significant group × time interaction ($F = 1.49$, $P = 0.23$).

Hippocampal volumes for the full sample can be seen in Figure 4. Both the full sample and the conservatively defined subgroups showed significant hippocampal atrophy over time in the left (full sample: $F = 29.68$, $P < 0.001$; subgroups: $F = 29.04$, $P < 0.001$) and right hemispheres (full sample: $F = 47.92$, $P < 0.001$; subgroups: $F = 67.86$; $P < 0.001$). No significant group × time interactions were observed ($Fs < 1.08$, $Ps > 0.30$), indicating similar atrophy rates in individuals with stable and declining memory. However, for the right hippocampus, there was a significant main effect of group in the full sample ($F = 4.05$, $P = 0.046$), such that the decliners had smaller volumes overall than the stable group. This effect could not be verified in the

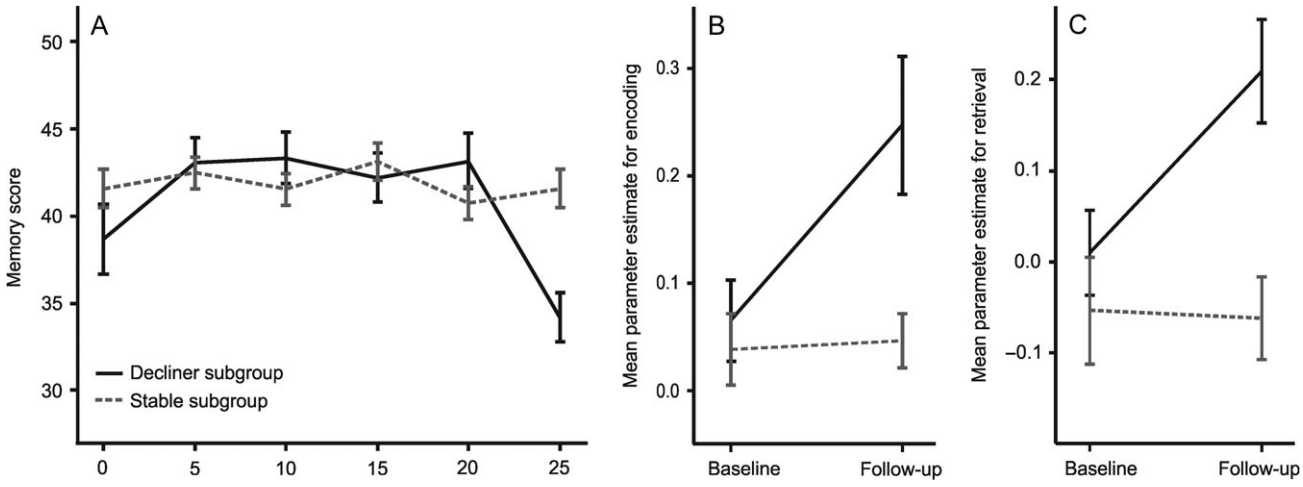

Figure 3. Control analysis in more conservatively defined subgroups. Panel A shows average memory performance for the subgroup of individuals ($n = 27$) from the decliner group that passed the more conservative criteria for being considered a memory decliner, as well as an age-matched subgroup of individuals with stable memory ($n = 60$). Panel B displays average PFC parameter estimates for the encoding condition of the scanner task for these subgroups. The average is based on the 5 PFC clusters identified in the whole-brain analyses in the full sample. Panel C shows the corresponding average parameter estimates for the retrieval condition.

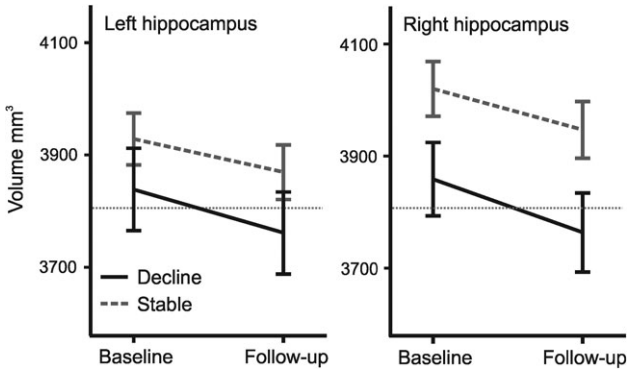

Figure 4. Hippocampus volumes. Group-averaged left and right hippocampus volumes at baseline and follow-up MRI sessions. The horizontal dashed line shows the midpoint between the decline group's baseline and follow-up volume, used to approximate a hypothetical threshold for when memory function and prefrontal engagement becomes affected (see main text). Error bars represent ±1 SEM.

smaller subgroups ($F = 1.37$, $P = 0.245$), which were both older and fulfilled a stricter criterion for memory decline.

Exploratory Threshold-Based Hippocampal Volume Analyses in Relation to Increased Frontal BOLD-Responses

The pattern of observed results invites the question of whether hippocampus volume may be directly related to elevated PFC BOLD-responses. Based on our observations, we wanted to test the possibility that hippocampal atrophy causes memory decline and elevated PFC-responses only once it has reached a critical threshold (see also Petersen et al. 2000; Mormino et al. 2014). The stable memory group, whose hippocampus volume at follow-up was still larger than that of the decline group at baseline, despite significant atrophy (see Fig. 4), maintained their memory ability as well as their PFC BOLD-responses over time. The decline group, on the other hand, displayed memory decline and elevated PFC-responses only at follow-up MRI, when their hippocampus volumes had been significantly reduced. This suggests that the memory decline group may have reached a critical level of hippocampal atrophy at some point between baseline and follow-up MRI. If correct, elevated PFC functional responses should be possible to identify in relation to hippocampal volume at follow-up. To test this threshold-based hypothesis in the full sample, we used the midpoint between the decline group's average (over left and right hemisphere) hippocampal volume at baseline and follow-up (3805.39 mm^3; dashed line in Fig. 4) as a threshold for dividing our sample into 2 new subgroups. This point should approximate a hypothetical threshold at which memory decline and, potentially, PFC BOLD increases were induced. At the follow-up MRI session, 77 individuals (19 from decline group, 58 from stable group, $M_{age} = 66.5$ years) had hippocampal volumes that were larger than this threshold, while the volumes of the remaining 52 individuals were below the threshold (22 decliners, 30 stable, $M_{age} = 73.0$). The subgroups were then matched for age, leaving 40 individuals in each group. The below-threshold group (i.e., those with smaller volumes) consisted of 16 memory decliners and 24 stable participants, whereas the above-threshold (i.e., larger volume) group comprised 6 decliners and 34 stable individuals. We then tested whether the new hippocampal volume subgroups also differed on memory decline and frontal BOLD-signal magnitude in the previously identified clusters. The below-threshold hippocampus group was found to have significantly more memory decline between baseline and follow-up MRI (–4.68 vs. –0.48 points, $t = 3.12$, $P = 0.002$). Furthermore, at follow-up but not baseline MRI, the below-threshold group showed significantly stronger magnitude of average frontal BOLD-signal during memory retrieval ($t = 2.19$, $P = 0.032$), with a nonsignificant group difference in the same direction for memory encoding ($t = 1.55$, $P = 0.126$). These findings are consistent with a theoretical model in which hippocampal atrophy, past a critical threshold, causes memory decline and upregulation of PFC responses.

Performing this threshold-based analysis with a cut-off instead based on the midpoint between baseline and follow-up hippocampus volumes from the more conservatively defined decliner subgroup only resulted in a different threshold classification for 4 individuals. None of the reported results were affected if these individuals were removed.

Discussion

This study identified individuals with declining episodic-memory functioning based on patterns of memory change across the preceding 20–25 years. Individuals who performed lower than predicted at the latest test wave of the study were considered having memory decline. This long-term, within-person, approach is a novel and unique way of capturing emerging memory dysfunction in older individuals. Both the declining and stable groups displayed hippocampal atrophy over time, with a smaller right hippocampal volume for the decliners in the full sample. No group differences in functional brain responses during memory encoding or retrieval were seen at the first MRI session, when the groups had comparable memory performance. At follow-up, the memory decliners showed elevated functional responses in the PFC in relation to the stable group, as well as their own activation levels 4 years earlier. No significant changes in PFC BOLD responses were observed in the stable memory group.

Functional Reorganization of the PFC is Related to Cognitive Decline

The current data represent the first instance of large-scale longitudinal functional imaging evidence linking PFC reorganization to emerging memory dysfunction in aging. Few functional neuroimaging studies have investigated PFC dynamics longitudinally, but 2 prior studies have found increasing frontal responses over time in healthy older individuals (Goh et al. 2013; Hakun et al. 2015). In these instances, the increased PFC responses were associated with declining executive functions. Critically, our study extends these findings to the memory domain. Collectively, these longitudinal observations of increased frontal responses in older individuals with cognitive decline contrast with the conceptualization from some cross-sectional studies that elevated PFC engagement is a characteristic of successful cognitive aging (Cabeza et al. 2002; Eyler et al. 2011). This is further underscored by our observation of a negative correlation between memory retrieval-related PFC BOLD-signal magnitude at follow-up, and offline memory performance within the group of decliners. That is, individuals with the most affected memory performance displayed the strongest pattern of frontal upregulation at follow-up. These novel longitudinal observation make important theoretical contributions since cross-sectional designs leave open the possibility that observed group differences reflect individual characteristics established in younger age (Karama et al. 2014; Pudas et al. 2014) rather than

age-related functional reorganization (Nyberg et al. 2010; Raz and Lindenberger 2013).

Functional Reorganization of the PFC is Process-General and Spatially Widespread

The elevated PFC responses in memory decliners were common across memory encoding and retrieval, indicating a process-general function (Grady et al. 2003; Salami et al. 2012; Reuter-Lorenz and Park 2014). In a previous multivariate analysis of the baseline MRI data, a process-general 'control' network, involving multiple frontal regions, was identified and found to be more engaged by low-performing older adults for whom the specific encoding and retrieval networks showed signs of dysfunction (Salami et al. 2012). The process-generality of the effects is consistent with prior cross-sectional studies that reported elevated functional responses in older compared with younger individuals across diverse cognitive domains (Dennis and Cabeza 2008; Spreng et al. 2010). This fact, together with the location of the current effects, notably those in the anterior and dorsolateral PFC (Koechlin et al. 1999; Nyberg et al. 2003; Marklund et al. 2007), strengthens the interpretation that the effects reflect wide-ranging processes such as cognitive control (Miller and Cohen 2001; Koechlin et al. 2003). An increased reliance on cognitive control mechanisms, even when found in a group with memory decline, can be interpreted as an adaptive response in a dysfunctional neural system for episodic memory, in other words, a compensatory response. This is in line with observations in Alzheimer's disease (Grady et al. 2003). The currently observed pattern of results is also compatible with the notion of reduced neural efficiency, that is, that more neural resources have to be allocated in order to achieve similar levels of task performance (Mattay et al. 2006; Nyberg et al. 2014). Given that our current observations are correlational we cannot conclusively decide between these interpretations, neither can we rule out that the increased responses are reflections of some other detrimental age-related changes underlying episodic memory decline (Logan et al. 2002), or merely coevolving with it.

An interesting characteristic of the observed BOLD-signal increases in the decliner group is their widespread distribution within the PFC, encompassing both dorsal and ventral regions in both hemispheres. However, a common denominator is that they were consistently located outside, or bordering, the core task networks associated with memory encoding and retrieval (relative to the control condition) at baseline MRI, see Figure 2. Thus, the task networks appeared to expand spatially over time in the memory decline group. This is in line with prior cross-sectional findings of higher engagement of brain regions outside canonical task-related networks among lower performing older adults (Steffener et al. 2009; Düzel et al. 2011).

Hippocampal Atrophy and Potential Structure-Function Associations

In the current data set, both the declining and stable memory groups showed significant hippocampal and total gray matter atrophy over 4 years, possibly in combination with an overall smaller right hippocampus for memory decliners. In the light of prior longitudinal studies linking differential hippocampal atrophy to individual differences in age-related memory decline (Kramer et al. 2007; Murphy et al. 2010; Persson et al. 2012), it is notable that we found comparable atrophy rates in the stable and decline groups. One possible explanation could be the nature of the sample. Being able to participate in a longitudinal study for 20–25 years perhaps indicates that even the memory decline group is somewhat healthier and more high-functioning than the average population, which could also be associated with less pronounced atrophy rates. The estimated annual hippocampal atrophy rate in the decline group was −0.57% (averaged across left and right hemisphere, see Supplementary Table 2) and comparable to atrophy estimates in a meta-analysis of longitudinal imaging studies in healthy aging (Fraser et al. 2015). It would be expected that individuals with a more progressed and clinically relevant memory dysfunction, and those with brain markers for Alzheimer's disease-related neuropathology, show accelerated atrophy rates (Risacher et al. 2010; Rieckmann et al. 2016).

The observation that the stable memory group maintained their memory performance and PFC responses over time despite the presence of significant hippocampus atrophy is noteworthy. This shows that hippocampus atrophy is not in itself sufficient to induce memory decline, in line with previous observations of significant age-related atrophy even for very healthy older adults and those with a low risk for developing Alzheimer's disease (Resnick et al. 2000; Fjell et al. 2013). One explanation could be that the stable individuals' atrophy had not yet progressed to a level at which neurocognitive function was affected (Petersen et al. 2000). As evident in Figure 4, the average hippocampal volume of the stable group at follow-up was still larger than that of the decliner group at baseline. Although the group difference in right hippocampus volume in the full sample needs to be interpreted with caution due to its failure to replicate in the more conservatively defined memory subgroups, the larger average volume of the stable group could entail that they can sustain more age-related hippocampal atrophy before memory decline becomes evident. This would be in line with the brain reserve hypothesis, that is, the more general idea that individuals with larger neural structures can sustain more insult before a cognitive deficit becomes apparent (Katzman et al. 1988; Satz 1993). Conversely, a smaller hippocampus for some individuals (Lupien et al. 2007), like those in our decliner group, could result in earlier memory decline once age-related hippocampal atrophy sets in and reaches a certain threshold. The smaller hippocampus volume of the decline group could alternatively be explained by earlier onset of age-related atrophy, but this possibility would require additional longitudinal measurements points to be tested, preferably beginning in midlife.

Although provisional, we also observed a link between hippocampal volume and PFC BOLD-responses. Specifically, when differentiating the sample based on hippocampus volume at follow-up, rather than longitudinal memory change, those with a sub-threshold hippocampus volume showed more memory decline, along with elevated functional responses in the PFC during memory retrieval. The reason for not observing a significant effect for encoding may be related to the overall weaker task-modulation for this task condition, compared with retrieval, as mentioned previously (see also Supplementary Fig. 1). Nevertheless, our current observations are consistent with a theoretical account in which age-related hippocampal atrophy past a certain threshold induces functional reorganization in the PFC, as has previously been hypothesized based on cross-sectional observations (Park and Reuter-Lorenz 2009). There is substantial prior evidence that these structures are intimately connected and work together in a dynamic, task-dependent, manner during tasks involving learning and memory (Laroche et al. 2000; Simons and Spiers 2003), which could explain why atrophy in one structure might influence neural processing in the other. Animal work has also

demonstrated that prefrontal neural circuits that are involved in certain forms of contextual fear learning can support this ability in the presence of experimental lesions to the dorsal hippocampal regions that are mainly responsible for it under normal circumstances (Zelikowsky et al. 2013). This lends support to the general idea that alternate neural circuitry than those normally underlying a hippocampus-mediated memory function may, at least in some instances, provide compensatory plasticity in order to accomplish that function when the hippocampus is compromised. However, our threshold-based analysis was exploratory and evoked to explain the lack of PFC upregulation and memory decline in the stable memory group despite significant hippocampal atrophy. It does not rule out that other factors than hippocampal atrophy may be driving the observed effects in the decline group. Hence, a causal link cannot be established based on our findings. Our way of identifying the hypothetical threshold for when hippocampus atrophy induces PFC upregulation must also be considered a crude approximation, since large interindividual differences in hippocampal volume (Lupien et al. 2007) would render such a threshold highly individual and not well approximated from group-level averages. Nevertheless, our current findings are at least compatible with hippocampal atrophy-induced PFC upregulation, which remains an interesting hypothesis to be further investigated in both experimental animal studies and longitudinal imaging studies in healthy and clinical human populations.

Study Limitations

A potential limitation of the current findings is that the observed group differences in frontal BOLD-changes over time could be driven, in part, by changes in neural processing during the control condition of the scanner task (perceptual discrimination), to which the memory encoding and retrieval conditions were compared. This is a common issue in fMRI research, since the BOLD-signal is always relative. Due to the lack of an implicit baseline in our scanner task, we could not directly test whether neural processing during the control condition changed differently in our memory groups. However, some observations speak against this possibility. First, there were no behavioral group differences in number of responses or RTs during the control condition, neither did the groups' performance change differentially over time (see Results section and Table 1). This makes it less likely that the groups' neural processing during the control condition changed differentially over time, although it still cannot be ruled out. Secondly, brain regions that were more active during the control condition than during memory processing were found to largely overlap with those belonging to the default mode network (DMN; Raichle et al. 2001). When visually inspected, our main PFC effects in the decline group did not overlap with DMN regions, with the possible exception of the 2 most anterior PFC clusters (clusters d and e in Fig. 2) which partially overlapped with a medial frontal DMN-region. Thus, although these 2 clusters may have been contaminated by changes in neural processing during the control condition, rather than during memory processing, the remaining 3 clusters were not. Therefore, our main conclusions regarding elevated PFC BOLD-responses during memory processing in individuals with memory decline should not have been predominantly driven by neural processing during the scanner task control condition.

Conclusion

In conclusion, the present findings provide novel longitudinal insights into the neural mechanisms behind age-associated memory loss. Older individuals with a small hippocampus volume, in combination with significant atrophy over 4 years, displayed memory decline as well as functional reorganization of PFC responses during memory encoding and retrieval. Although preliminary, this is the first instance of longitudinal human functional imaging data consistent with animal studies (Zelikowsky et al. 2013) and a theoretical model (Park and Reuter-Lorenz 2009) in which hippocampus atrophy, past a critical threshold, induces memory decline and PFC upregulation. It remains an important task for future research to investigate potential causal biological mechanisms behind these observations.

Supplementary Material

Supplementary material are available at *Cerebral Cortex* online.

Funding

A grant from the Knut and Alice Wallenberg Foundation, and a grant from the Söderberg's Foundation to L.N.

Notes

We thank Micael Andersson at UFBI for assistance with preprocessing, Greger Orädd for scanner stability assessment, as well as the staff and participants of the Betula project. The Freesurfer analyses were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at HPC2N in Umeå. *Conflict of Interest*: None declared.

References

Ashburner J. 2007. A fast diffeomorphic image registration algorithm. Neuroimage. 38:95–113.

Buonomano DV, Merzenich MM. 1998. Cortical plasticity: from synapses to maps. Annu Rev Neurosci. 21:149–186.

Cabeza R, Anderson ND, Locantore JK, Mcintosh AR. 2002. Aging gracefully: compensatory brain activity in high-performing older adults. Neuroimage. 17:1394–1402.

Cox RW, Reynolds RC, Taylor PA. 2016. AFNI and clustering: false positive rates redux. bioRxiv. 1–15. doidx.doi.org/101101/065862.

Daselaar SM, Veltman DJ, Rombouts ARB, Raaijmakers JGW, Jonker C. 2003. Neuroanatomical correlates of episodic encoding and retrieval in young and elderly subjects. Brain. 126:43–56.

Dennis NA, Cabeza R. 2008. Neuroimaging of healthy cognitive aging. In: Craik FIM, Salthouse TA, editors. The Handbook of aging and cognition. 3rd ed. New York (NY): Psychology Press. p. 1–54.

Düzel E, Schütze H, Yonelinas AP, Heinze H-J. 2011. Functional phenotyping of successful aging in long-term memory: preserved performance in the absence of neural compensation. Hippocampus. 21:803–814.

Eyler LT, Sherzai A, Kaup AR, Jeste DV. 2011. A review of functional brain imaging correlates of successful cognitive aging. Biol Psychiatry. 70:115–122.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, et al. 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron. 33:341–355.

Fjell AM, McEvoy L, Holland D, Dale AM, Walhovd KB. 2013. Brain changes in older adults at very low risk for Alzheimer's disease. J Neurosci. 33:8237–8242.

Folstein MF, Folstein SE, McHugh PR. 1975. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 12:189–198.

Fraser MA, Shaw ME, Cherbuin N. 2015. A systematic review and meta-analysis of longitudinal hippocampal atrophy in healthy human ageing. Neuroimage. 112:364–374.

Friedman L, Glover GH. 2006. Report on a multicenter fMRI quality assurance protocol. J Magn Reson Imaging. 23: 827–839.

Goh JO, Beason-Held LL, An Y, Kraut MA, Resnick SM. 2013. Frontal function and executive processing in older adults: process and region specific age-related longitudinal functional changes. Neuroimage. 69:43–50.

Grady CL, McIntosh AR, Beig S, Keightley ML, Burian H, Black SE. 2003. Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. J Neurosci. 23:986–993.

Grady CL, McIntosh AR, Craik FIM. 2005. Task-related activity in prefrontal cortex and its relation to recognition memory performance in young and old adults. Neuropsychologia. 43: 1466–1481.

Greve DN, Mueller BA, Liu T, Turner JA, Voyvodic J, Yetter E, Diaz M, McCarthy G, Wallace S, Roach BJ, et al. 2011. A novel method for quantifying scanner instability in fMRI. Magn Reson Med. 65:1053–1061.

Habib R, Nyberg L, Nilsson L-G. 2007. Cognitive and noncognitive factors contributing to the longitudinal identification of successful older adults in the Betula study. Aging Neuropsychol Cogn. 14:257–273.

Hakun JG, Zhu Z, Brown CA, Johnson NF, Gold BT. 2015. Longitudinal alterations to brain function, structure, and cognitive performance in healthy older adults: a fMRI-DTI study. Neuropsychologia. 71:225–235.

Hedden T, Schultz AP, Rieckmann A, Mormino EC, Johnson KA, Sperling RA, Buckner RL. 2016. Multiple brain markers are linked to age-related variation in cognition. Cereb Cortex. 26:1388–1400.

Jack C, Twomey C, Zinsmeister A, Sharbrough FW, Petersen R, Cascino G. 1989. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. Radiology. 172:549–554.

Josefsson M, de Luna X, Pudas S, Nilsson L-G, Nyberg L. 2012. Genetic and lifestyle predictors of 15-year longitudinal change in episodic memory. J Am Geriatr Soc. 60: 2308–2312.

Karama S, Bastin ME, Murray C, Royle NA, Penke L, Muñoz Maniega S, Gow AJ, Corley J, Valdés Hernández M, Lewis JD, et al. 2014. Childhood cognitive ability accounts for associations between cognitive ability and brain cortical thickness in old age. Mol Psychiatry. 19:555–559.

Katzman R, Terry R, DeTeresa R, Brown T, Davies P, Fuld P, Renbing X, Peck A. 1988. Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques. Ann Neurol. 23:138–144.

Koechlin E, Basso G, Pietrini P, Panzer S, Grafman JH. 1999. The role of the anterior prefrontal cortex in human cognition. Nature. 399:148–151.

Koechlin E, Ody C, Kouneiher F. 2003. The architecture of cognitive control in the human prefrontal cortex. Science. 302: 1181–1185. (80-).

Kolb B, Gibb R. 2015. Plasticity in the prefrontal cortex of adult rats. Front Cell Neurosci. 9:1–11.

Kramer JH, Mungas D, Reed BR, Wetzel ME, Burnett MM, Miller BL, Weiner MW, Chui HC. 2007. Longitudinal MRI and cognitive change in healthy elderly. Neuropsychology. 21:412–418.

Kuboshima-Amemori S, Sawaguchi T. 2007. Plasticity of the primate prefrontal cortex. Neurosci. 13:229–240.

Laroche S, Davis S, Jay TM. 2000. Plasticity at hippocampal to prefrontal cortex synapses: dual roles in working memory and consolidation. Hippocampus. 10:438–446.

Lindenberger U. 2014. Human cognitive aging: Corriger la fortune? Science. 346:572–578. (80-).

Logan JM, Sanders AL, Snyder AZ, Morris JC, Buckner RL. 2002. Under-recruitment and nonselective recruitment: dissociable neural mechanisms associated with aging. Neuron. 33: 827–840.

Lupien SJ, Evans A, Lord C, Miles J, Pruessner M, Pike B, Pruessner JC. 2007. Hippocampal volume is as variable in young as in older adults: implications for the notion of hippocampal atrophy in humans. Neuroimage. 34:479–485.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. 19:1233–1239.

Marklund P, Fransson P, Cabeza R, Petersson KM, Ingvar M, Nyberg L. 2007. Sustained and transient neural modulations in prefrontal cortex related to declarative long-term memory, working memory, and attention. Cortex. 43:22–37.

Mattay VS, Fera F, Tessitore A, Hariri AR, Berman KF, Das S, Meyer-Lindenberg A, Goldberg TE, Callicott JH, Weinberger DR. 2006. Neurophysiological correlates of age-related changes in working memory capacity. Neurosci Lett. 392:32–37.

McEwen B, Morrison J. 2013. The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. Neuron. 79:16–29.

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 24:167–202.

Mormino EC, Betensky RA, Hedden T, Schultz AP, Amariglio RE, Rentz DM, Johnson KA, Sperling RA. 2014. Synergistic effect of β-amyloid and neurodegeneration on cognitive decline in clinically normal individuals. JAMA Neurol. 71:1379–1385.

Murphy EA, Holland D, Donohue M, McEvoy LK, Hagler DJ, Dale AM, Brewer JB. 2010. Six-month atrophy in MTL structures is associated with subsequent memory decline in elderly controls. Neuroimage. 53:1310–1317.

Nilsson L-G, Bäckman L, Erngrund K, Nyberg L, Adolfsson R, Bucht G, Karlsson S, Widing M, Winblad B. 1997. The Betula prospective cohort study: memory, health, and aging. Aging Neuropsychol Cogn. 4:1–32.

Nyberg L, Andersson M, Kauppi K, Lundquist A, Persson J, Pudas S, Nilsson L-G. 2014. Age-related and genetic modulation of frontal cortex efficiency. J Cogn Neurosci. 26: 746–754.

Nyberg L, Marklund P, Persson J, Cabeza R, Forkstam C, Petersson KM, Ingvar M. 2003. Common prefrontal activations during working memory, episodic memory, and semantic memory. Neuropsychologia. 41:371–377.

Nyberg L, Salami A, Andersson M, Eriksson J, Kalpouzos G, Kauppi K, Lind J, Pudas S, Persson J, Nilsson L-G. 2010. Longitudinal evidence for diminished frontal cortex function in aging. Proc Natl Acad Sci USA. 107:22682–22686.

Park DC, Reuter-Lorenz P. 2009. The adaptive brain: aging and neurocognitive scaffolding. Annu Rev Psychol. 60:173–196.

Persson J, Kalpouzos G, Nilsson L-G, Ryberg M, Nyberg L. 2011. Preserved hippocampus activation in normal aging as revealed by fMRI. Hippocampus. 21:753–766.

Persson J, Nyberg L, Lind J, Larsson A, Nilsson L-G, Ingvar M, Buckner RL. 2006. Structure-function correlates of cognitive decline in aging. Cereb Cortex. 16:907–915.

Persson J, Pudas S, Lind J, Kauppi K, Nilsson L-G, Nyberg L. 2012. Longitudinal structure-function correlates in elderly reveal MTL dysfunction with cognitive decline. Cereb Cortex. 22: 2297–2304.

Petersen RC, Jack CR Jr., Xu YC, Waring SC, O'Brien PC, Smith GE, Ivnik RJ, Tangalos EG, Boeve BF, Kokmen E. 2000. Memory and MRI-based hippocampal volumes in aging and AD. Neurology. 54:581–587.

Pudas S 2013. Brain characteristics of memory decline and stability in aging: contributions from longitudinal observations. [Stockholm (Sweden): Stockholm University Doctoral thesis].

Pudas S, Persson J, Josefsson M, de Luna X, Nilsson L-G, Nyberg L. 2013. Brain characteristics of individuals resisting age-related cognitive decline over two decades. J Neurosci. 33: 8668–8677.

Pudas S, Persson J, Nilsson L-G, Nyberg L. 2014. Midlife memory ability accounts for brain activity differences in healthy aging. Neurobiol Aging. 35:2495–2503.

Raichle ME, Macleod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. 2001. A default mode of brain function. Proc Natl Acad Sci USA. 98:676–682.

Raz N, Lindenberger U. 2013. Life-span plasticity of the brain and cognition: from questions to evidence and back. Neurosci Biobehav Rev. 37:2195–2200.

Resnick SM, Goldszal AF, Davatzikos C, Golski S, Kraut MA, Metter EJ, Nick R, Zonderman AB. 2000. One-year age changes in MRI brain volumes in older adults. Cereb Cortex. 10:464–472.

Reuter M, Schmansky NJ, Rosas HD, Fischl B. 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage. 61:1402–1418.

Reuter-Lorenz PA, Park DC. 2014. How does it STAC up? Revisiting the scaffolding theory of aging and cognition. Neuropsychol Rev. 24:355–370.

Rieckmann A, Van Dijk KRA, Sperling RA, Johnson KA, Buckner RL, Hedden T. 2016. Accelerated decline in white matter integrity in clinically normal individuals at risk for Alzheimer's disease. Neurobiol Aging. 42:177–188.

Risacher SL, Shen L, West JD, Kim S, Mcdonald BC, Beckett LA, Harvey DJ, Jack CR Jr., Weiner MW. 2010. Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. Neurobiol Aging. 31:1401–1418.

Rosen A, Prull M, O'Hara R. 2002. Variable effects of aging on frontal lobe contributions to memory. Neuroreport. 13: 2425–2428.

Rönnlund M, Nyberg L, Bäckman L, Nilsson L-G. 2005. Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. Psychol Aging. 20:3–18.

Salami A, Eriksson J, Nyberg L. 2012. Opposing effects of aging on large-scale brain systems for memory encoding and cognitive control. J Neurosci. 32:10749–10757.

Satz P. 1993. Brain reserve capacity on symptom onset after brain injury: a formulation and review of evidence for threshold theory. Neuropsychology. 7:273–295.

Schaie KW. 1994. The course of adult intellectual development. Am Psychol. 49:304–313.

Shao Z, Janse E, Visser K, Meyer AS. 2014. What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. Front Psychol. 5:1–10.

Simons JS, Spiers HJ. 2003. Prefrontal and medial temporal lobe interactions in long-term memory. Nat Rev Neurosci. 4: 637–648.

Spreng RN, Wojtowicz M, Grady CL. 2010. Reliable differences in brain activity between young and old adults: a quantitative meta-analysis across multiple cognitive domains. Neurosci Biobehav Rev. 34:1178–1194.

Steffener J, Brickman AM, Rakitin BC, Gazes Y, Stern Y. 2009. The impact of age-related changes on working memory functional activity. Brain Imaging Behav. 3:142–153.

Wechsler D. 1981. Wechsler adult intelligence scale-revised. San Antonio: The Psychological Corporation.

Zelikowsky M, Bissiere S, Hast TA, Bennett RZ, Abdipranoto A, Vissel B, Fanselow MS. 2013. Prefrontal microcircuit underlies contextual learning after hippocampal loss. Proc Natl Acad Sci USA. 110:9938–9943.