Locus coeruleus integrity preserves memory performance across the adult life span

Martin J. Dahl*[1], Mara Mather[2], Sandra Düzel[1], Nils C. Bodammer[1], Ulman Lindenberger[1,3], Simone Kühn[1,4], & Markus Werkle-Bergner*[1]

[1]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

[2]Davis School of Gerontology, University of Southern California, Los Angeles, USA

[3]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, England, and Berlin, Germany

[4]Department of Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Hamburg, Germany

* Correspondence concerning this manuscript should be addressed to

MJD (dahl@mpib-berlin.mpg.de) or MWB (werkle@mpib-berlin.mpg.de)

Acknowledgements.
This article uses data from the Berlin Aging Study II (BASE-II), which was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under grant numbers #16SV5536 K, #16SV5537, #16SV5538, and #16SV5837, #01UW070 and #01UW0808. Additional contributions (e.g., financial, equipment, logistics, personnel) are made from each of the other participating sites, i.e., the Max Planck Institute for Human Development (MPIB), Max Planck Institute for Molecular Genetics (MPIMG), Charite-Universiätsmedizin, University Medicine, German Institute for Economic Research (DIW), Humboldt-Universität zu Berlin, all located in Berlin, Germany, and University of Lübeck in Lübeck, and University of Tübingen, Germany. For further information about the BASE-II project, see https://www.base2.mpg.de/en
MW-B received support from the German Research Foundation (DFG, WE 4269/5-1) and the Jacobs Foundation (Early Career Research Fellowship 2017–2019).
MJD is a fellow of the International Max Planck Research School on the Life Course (LIFE; http://www.imprs-life.mpg.de/en). MJD is recipient of a stipend from the Sonnenfeld-Foundation (http://www.sonnenfeld-stiftung.de/en/). MM's work was supported by an Alexander von Humboldt fellowship and by National Institutes of Health grant R01AG025340.
We thank Andrew R. Bender, Matthew Betts, and Myriam C. Sander for valuable discussions and assistance. We are grateful to Shelby Bachman and Dilara Zorbek, who performed the manual tracing of the locus coeruleus, as well as Ylva Köhncke and Yana Fandakova for statistical advice, and Michael Krause for help with cluster computing.

Abstract

For decades, the search for drivers of memory decline in human cognitive aging has focused on neocortical regions, the hippocampus, and dopaminergic neuromodulation. Recent findings indicate that the locus coeruleus (LC) and noradrenergic neuromodulation may also play an important role in shaping memory development in later life. However, technical challenges in non-invasive monitoring of LC integrity have hindered the study of LC-cognition associations in humans so far. Using high-resolution neuromelanin-sensitive magnetic resonance imaging, we found that individual differences in learning and memory were positively associated with LC integrity across a variety of memory tasks in large samples of younger (n = 66), and older adults (n = 233). Moreover, we observed spatially confined and functionally relevant age differences in rostral LC. Older adults who showed a more youth-like rostral LC also showed higher memory performance. These findings link non-invasive, in vivo indices of LC integrity to memory in human aging and extend our knowledge about the role of the LC norepinephrine system in senescent decline in human cognition.

Keywords:

Neuromodulation, cognitive aging, norepinephrine, locus coeruleus, memory

Introduction

Memory performance declines with advancing adult age[1], jeopardizing the everyday competence of older adults[2,3]. Age-associated changes in neocortical regions, the hippocampus, and dopaminergic neuromodulation have been found to contribute to age-related memory impairments[1,4–6]. More recently, findings from animal and post-mortem human research have led researchers to propose that losses in the structural integrity of the locus coeruleus (LC), the brain's primary norepinephrine (NE) source, are critically related to cognitive decline in normal aging and age-related pathologies[7–10]. However, direct in vivo evidence relating LC integrity to interindividual differences in general memory abilities is scarce. While there is a first indication of a selective role of LC integrity in the encoding of negative emotional events[11], the relation of LC integrity to memory performance in general is still an open question. Given LC's pivotal role in age-related memory disorders like Alzheimer's disease[8,9,12], this relationship, however, is of high clinical and scientific importance.

The LC is a hyperpigmented cylindrical cluster of catecholaminergic neurons located in the dorsorostral tegmentum[13]. Roughly symmetrical in both hemispheres, it extends for only about 15 mm rostrocaudally from the level of the inferior colliculi to a position in the lateral wall of the fourth ventricle[14,15]. Despite its small size, the LC has diffuse and highly arborized efferent projections throughout the brain[16]. While initial reports considered the LC a uniform structure, a spatially differentiated organization according to terminal fields is now assumed. Cells giving rise to dense hippocampal projections tend to be located in more rostral segments while those that innervate the cerebellum and spinal cord are located more caudally[17–19]. NE release by the LC modulates cognitive functions such as perception, attention, learning and memory[17,20–28]. In particular, via its action on ß-adrenoceptors in the hippocampus, the LC modulates long-term

potentiation (LTP) and long term-depression (LTD), key determinants of synaptic plasticity and memory[29,30]. Recent optogenetic work has confirmed a causal role of LC activity in hippocampal LTP and memory enhancement, potentially via co-release of dopamine[31–33].

Its location adjacent to the ventricular system and its widespread, unmyelinated projections expose LC cells to blood- and cerebrospinal-fluid bound toxins, making them a likely target for neurodegeneration[34]. Consistent with its age-related vulnerability, a prominent, rostrally accentuated decline of LC density was reported in healthy aging (reviewed by Manaye and colleagues[18]; but see Mather and colleagues[34] for a recent discussion). Based on longitudinal data, Wilson and colleagues[10] reported initial evidence pointing to the importance of LC integrity for maintaining memory abilities in old age. The authors evaluated cognitive abilities in a sample of older adults annually over a mean duration of about six years and, upon the participants' death, assessed LC integrity post mortem via autopsy. Wilson and colleagues[10] observed an attenuation of cognitive decline as well as higher baseline cognitive abilities in individuals with higher LC integrity, even after accounting for the integrity of other neuromodulatory systems and markers of neuropathology[10]. Furthermore, studies in aged mice[35], rats[36], and monkeys[37] in which NE or its agonists were manipulated indicate evidence that NE plays a causal role in learning and memory.

Non-invasive, in vivo assessments of human LC integrity are notoriously difficult, given the nucleus' small size and location deep in the brainstem[26,38,39]. Fortunately, however, a by-product of catecholamine synthesis, the dark, insoluble pigment neuromelanin, accumulates in the LC across the life span[40]. LC pigmentation increases from birth until late middle age when it starts to decline until death, probably due to selective atrophy of neuromelanin-containing cells[18,41,42]. In noradrenergic cells, neuromelanin avidly chelates metals such as copper and iron

and, as a compound, shows paramagnetic, T_1-shortening effects[43]. Sasaki and colleagues developed a neuromelanin-sensitive Turbo Spin Echo (TSE) magnetic resonance imaging (MRI) sequence that visualizes the LC as an hyperintense area adjacent to the lateral floor of the fourth ventricle[44] (for a review of neuromelanin-sensitive MRI studies, see Liu and others[45]). Keren and colleagues[46] recently validated this sequence by scanning human post-mortem samples using ultra high field MRI and then performing histological analyses of the samples. Brainstem areas showing hyperintensities on neuromelanin-sensitive scans overlapped closely with noradrenergic cells as identified by histology. Thus, neuromelanin acts as a natural contrast agent that opens the door to the non-invasive, in vivo assessment of LC integrity via MRI.

To determine the importance of structural LC integrity for the maintenance of memory functioning in vivo, we assessed individual differences in learning and memory among younger and older adults with the Rey Auditory Verbal Learning Test (RAVLT), a validated neuropsychological measure of memory functioning. In this context, specifically the analysis of individual learning trajectories, that is, the increase in recall performance across iterative item presentations, conveys valuable information about a participant's current and future cognitive status[47–50]. Using structural equation modeling (SEM) to capture the non-linear dynamics of growth in performance[51], previous studies have linked aging to lower initial memory performance[52,53] and slower learning with practice[53]. Hence, we hypothesized that the integrity of the LC-NE system as assessed by neuromelanin-sensitive MRI would be closely associated with individual differences in initial memory performance and learning rates. In sum, the goal of this study was to extend our knowledge about the role of the LC-NE system in human cognitive aging by linking non-invasive, in vivo indices of LC integrity to memory abilities in younger and older adults.

Results

Initial recall performance is reduced in older adults

We used latent growth curve modeling[54] to estimate the intercept (i.e., recall performance after the first trial) and the slope (i.e., the rate of learning in the course of additional trials) of younger and older adults' performance in an iterative verbal memory task (see Online methods and Figure 1). A multiple-group model describing the gradually increasing but negatively accelerated performance trajectories as a quadratic growth function fit the data well and outperformed competing alternative models (RMSEA < 0.06; CFI > 0.95[55]; see Supplemental results and lower part of Figure 2). Importantly, the model allows testing of group differences in factors capturing initial recall (intercept) and learning (slope) as well as for interindividual differences therein within each age group.

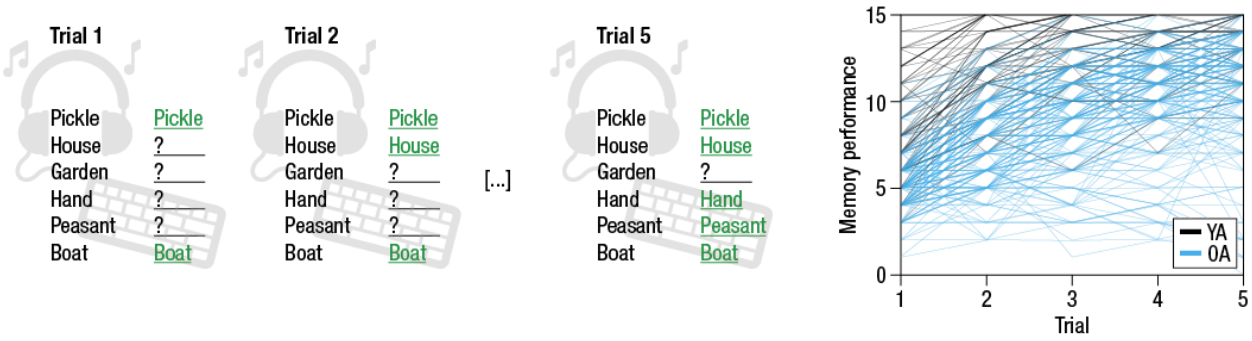


Figure 1. Schematic overview of the verbal learning and memory task (left). A word list consisting of 15 unrelated words is auditorily presented to participants over five trials. After each trial, subjects freely recall items and enter their responses via keyboard. Right: Younger (YA; black) and older adults' (OA; blue) performance increases over learning trials in a non-linear fashion.

Both the means and the variances of the intercept and slope factors differed reliably from zero in each age group (means: all $\Delta\chi^2 \geq 62.602$, $\Delta df = 1$, all $p < 0.001$; variances: all $\Delta\chi^2 \geq 11.97$, $\Delta df = 1$, all $p < 0.001$). As the sole exception, there were no reliable differences in the variance of the quadratic slope factor in younger adults, indicating similar negative acceleration of the growth curve across subjects ($\Delta\chi^2 = 0.276$, $\Delta df = 1$, $p = 0.599$). Performance differences between

younger and older adults were mainly found in intercepts ($\Delta\chi^2 = 59.533$, $\Delta df = 1$, $p < 0.001$; see Supplemental results). In sum, older adults improved their recall performance over trials at similar rates as younger adults did, but started and ended with lower scores.

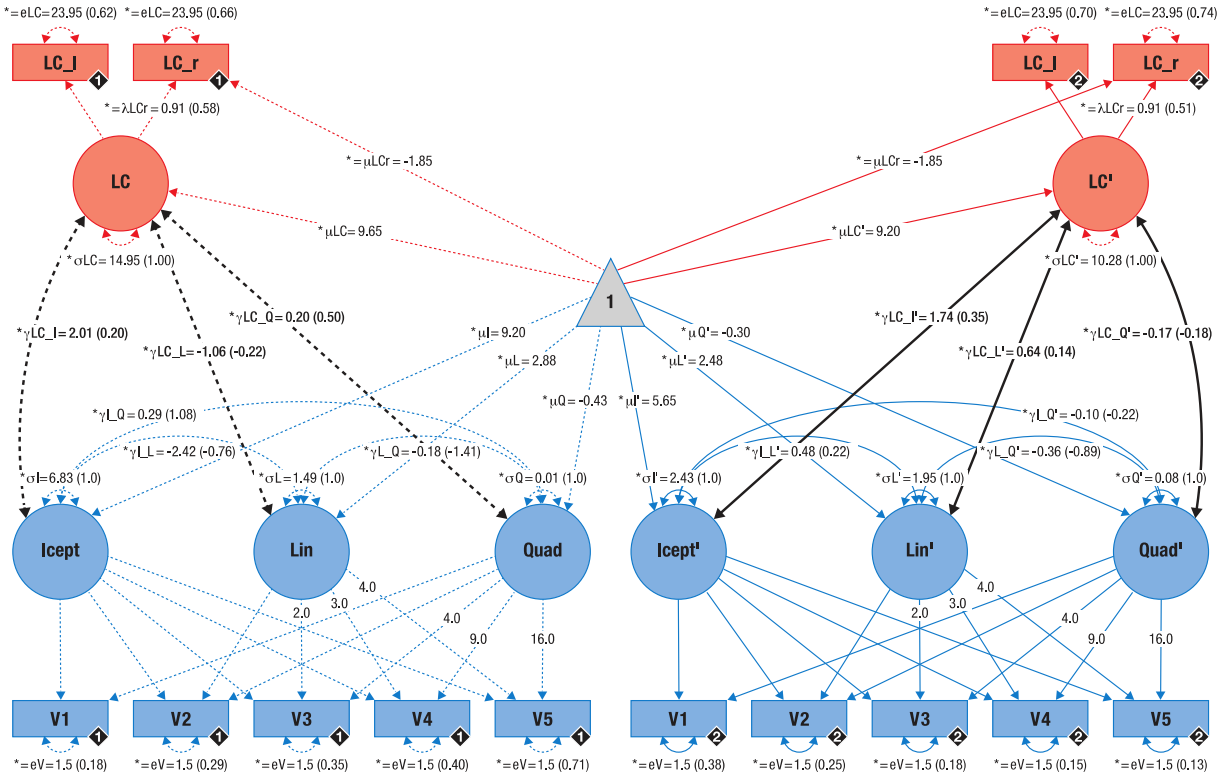


Figure 2. Pictorial rendition of the structural equation model that probes associations (black lines) between locus coeruleus integrity (LC; red) and memory performance (blue) in younger and older adults on a latent level. Cognitive manifest variables represent the iteratively assessed memory performance in a verbal learning and memory task (V1–5). Neural manifest variables are the LC intensity ratios of each hemisphere (left = LC_l; right = LC_r). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) and loadings (λ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). Icept = Intercept; Lin/Quad= linear / quadratic slope, respectively. Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle.

Locus coeruleus integrity scores are positively associated with initial recall performance in older adults

We developed a semi-automatic procedure to extract individual LC ratios[45], that is, a ratio score of peak LC MRI intensity relative to peak intensity in a dorsal pontine reference region, across the rostrocaudal extent of the nucleus (see Online methods and Figure 3a–c). The procedure evinced high reproducibility across multiple measurements (see Supplemental results) and was validated using both published LC maps (see Figure 3d) and manual intensity assessments (see Supplemental results). In order to relate memory performance to LC integrity, we integrated LC ratios over slices and hemispheres to derive a single measure reflecting LC integrity. In particular, we estimated latent scores for LC integrity by means of a multiple-group single-factor structural equation model based on each hemisphere's mean intensity ratios (see Online methods and upper part of Figure 2). The proposed model fit the data well (RMSEA < 0.06; CFI > 0.95[55]; see Supplemental results). We detected reliable average LC scores for both age groups; moreover, within each group, participants showed significant interindividual differences in LC scores (all $\Delta\chi^2 \geq 18.305$, $\Delta df = 1$, all $p < 0.001$; see Supplemental results). Younger and older adults demonstrated comparable average LC scores ($\Delta\chi^2 = 0.383$, $\Delta df = 1$, $p = 0.536$), in line with spatially confined age differences across the rostrocaudal LC axis[39,56].

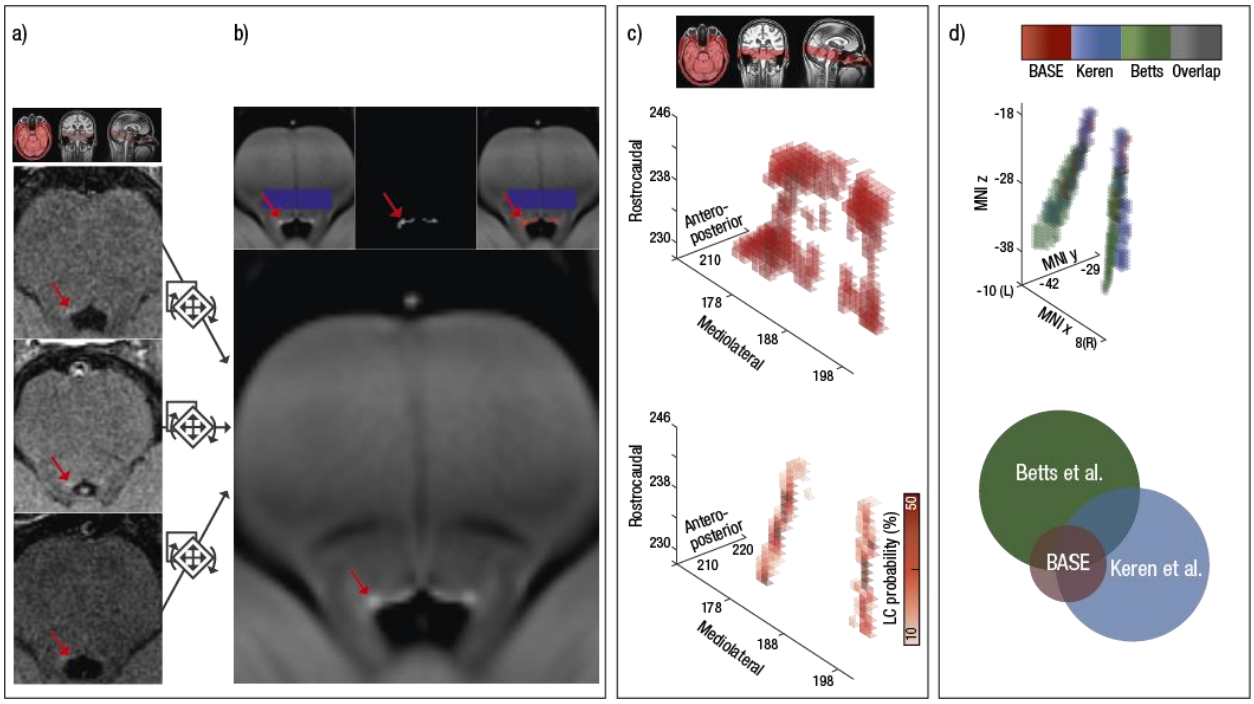


Figure 3. Schematic overview of the semi-automatic analysis procedure used to extract individual locus coeruleus (LC) intensity values across the rostrocaudal extent. (a) Native space neuromelanin-sensitive brainstem scans of three randomly selected subjects (axial slices are shown). Hyperintensities corresponding to the LC are indicated by red arrows. (b) Neuromelanin-sensitive scans were aligned and pooled across subjects in order to increase signal-to-noise ratio and facilitate LC delineation using a template-based approach. On a group level, LC location (red) was semi-automatically determined based on an intensity threshold relative to a pontine reference area (blue; see inlays). (c) Areas surviving the thresholding are grouped into a volume of interest (search space: upper plot; 3D representation) and used to restrict automatized extraction of individual peak intensities and their location. Observed peak LC locations were converted to a LC probability map (lower plot). (d) In standard space, the LC probability map was successfully validated using published maps by Keren and colleagues (2009)[39] and Betts and others (2017)[56]. Circle radius indicates map size (i.e., number of voxels). We freely share the established LC probability map with the neuroscientific community (see Online methods).

After establishing valid models for both memory performance and LC integrity in isolation, we aimed at combining the independent information from both modalities. For this, we merged both SEMs in a unified neuro-cognitive model that demonstrated good fit (RMSEA < 0.06; CFI > 0.95[55]; see Supplemental results and Figure 2). Initial recall (intercept) was positively related to latent LC scores for the older but not the younger group (for older adults: $\Delta\chi^2 = 7.939$, $\Delta df = 1$, $p = 0.005$, standardized estimate: 0.35; for younger adults: $\Delta\chi^2 = 1.181$, $\Delta df = 1$, $p = 0.147$[11]). Learning rates (slope) were not reliably associated with LC scores (all $\Delta\chi^2 \leq$

1.426, $\Delta df = 1$, all $p \geq 0.232$). These findings indicate higher initial recall performance (intercept) for older adults with high LC integrity than for those with low LC integrity (see Figure 4).

There were no reliable age group differences in the association between memory performance (intercept, slope) and LC scores (all $\Delta\chi^2 \leq 1.927$, $\Delta df = 1$, all $p \geq 0.165$). Thus, while we discovered an association between interindividual differences in memory performance and LC integrity only in older adults, the lack of reliable group differences suggests comparable associations independent of age.

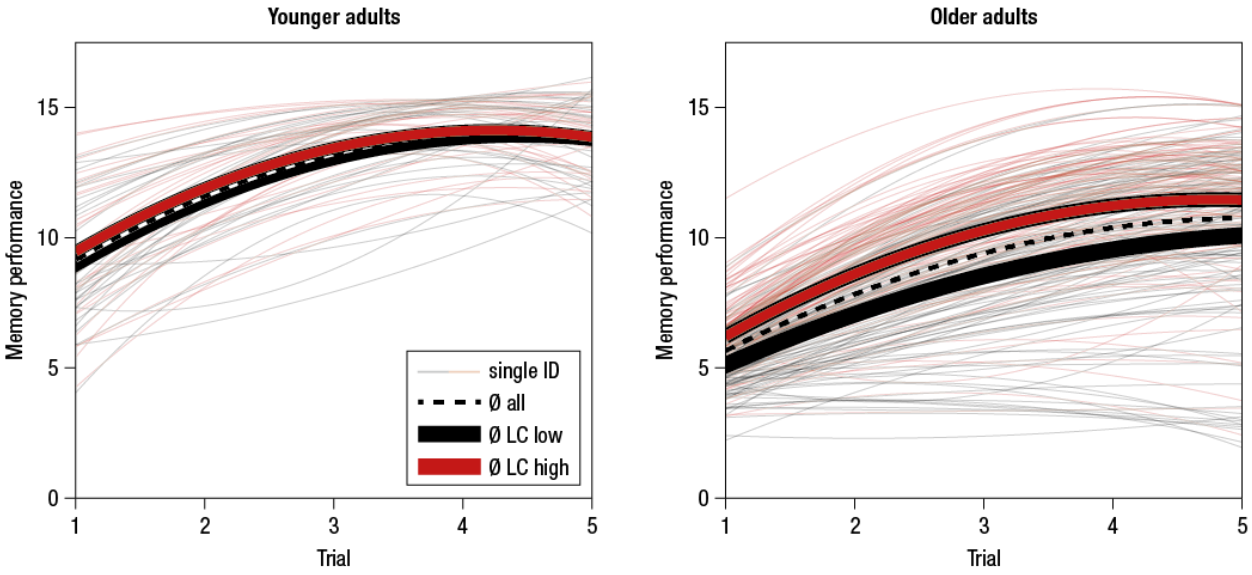


Figure 4. Estimated learning and memory performance trajectories for younger (left panel) and older adults (right panel). For visualization of the association between locus coeruleus (LC) integrity and memory performance, single subjects (thin lines) are color-coded based on LC integrity (median-split) and mean trajectories for subgroups are displayed.

Reduced age differences in rostral locus coeruleus intensity ratios relate to memory performance in older adults

Building upon reports of a spatially differentiated LC organization[17–19], we investigated age differences in LC topography and their association with memory performance by analyzing LC data slice by slice. A non-parametric cluster-based permutation test[57] (see Online methods) revealed spatially heterogeneous age differences in LC ratios along the rostrocaudal extent of the

nucleus. In comparison to younger adults, older adults showed a cluster of elevated intensity spanning caudal slices (65th–100th LC percentile; $p_{corr} < 0.001$) in line with neuromelanin accumulation across the life span[18,40,41]. In contrast, in rostral segments there was even a trend towards decreased LC intensity in older adults (29th–41st LC percentile; $p_{corr} = 0.079$; see Figure 7), in accordance with age-related decline in hippocampus-projecting LC segments[18,19]. To evaluate the functional significance of the observed topographical age differences, we related memory performance to LC ratios across all participants for each identified cluster. In caudal segments (65th–100th LC percentile), LC ratios were not reliably associated with initial recall performance (intercept; $r_s = –0.08$, $p = 0.172$; see Figure 7). However, in rostral segments (29th–41st LC percentile), higher LC ratios were significantly associated with memory performance ($r_s = 0.207$, $p < 0.001$; see Figure 7; difference between caudal and rostral correlations[59]: $Z = 3.385$, $p < 0.001$). Older adults with more youth-like intensity ratios in rostral LC segments also showed better memory performance. In sum, we observed spatially confined age differences in LC ratios that were associated with memory performance and are in line with neurodegeneration of hippocampus- and forebrain-projecting LC segments[18] in older adults.

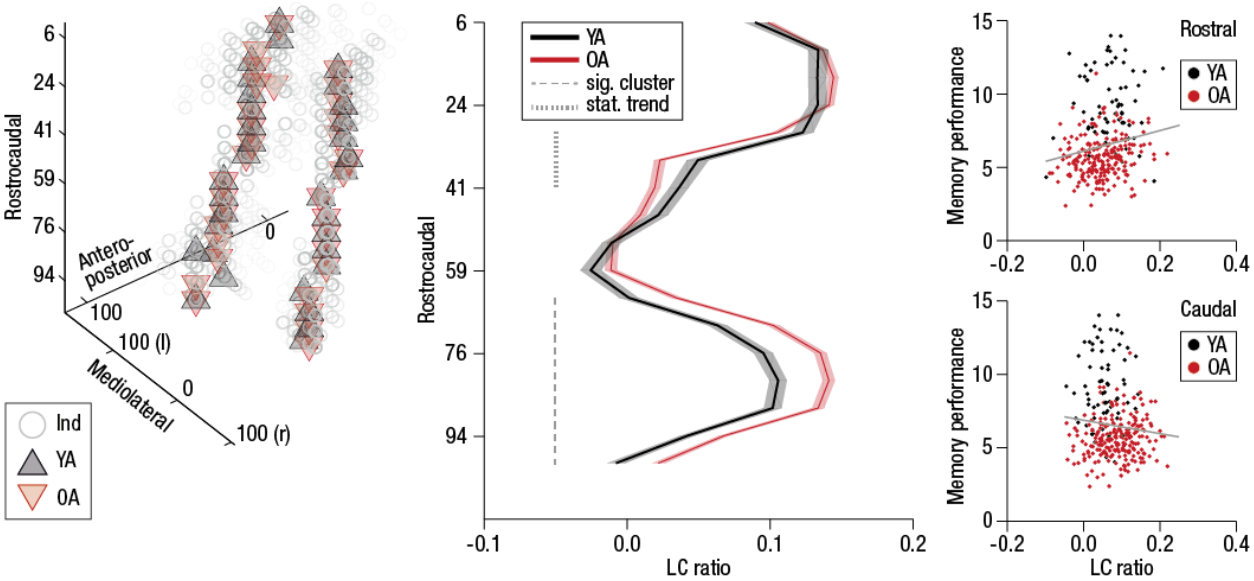


Figure 5. Topographical age differences in locus coeruleus intensity ratios and their functional significance. Left plot: On a group level, younger (YA; black) and older adults (OA; red) share highly congruent peak locus coeruleus (LC) locations. Grey circles indicate single subjects (Ind). Middle plot: A cluster-based permutation test revealed spatially confined age differences in LC intensity across the rostrocaudal axis. While older adults show significantly higher ratios in caudal segments, there is a trend for a reversed effect in rostral segments. Shaded lines represent ± 1 standard error of the mean (SEM). Rostrocaudal, anteroposterior and mediolateral positions are expressed in percentiles relative to the total size of the observed LC. Right plots: Intensity differences in the rostral but not caudal cluster are associated with memory performance (initial recall) across participants (right plot; rostral: $r_s = 0.207$, $p < 0.001$; caudal: $r_s = -0.08$, $p = 0.172$).

Discussion

Animal and post-mortem human studies suggest a link between memory performance in aging and the integrity of the central noradrenergic system[7,10]. So far, in vivo research on humans, however, has been stymied by methodological difficulties in the reliable assessment of LC integrity[26,38,39] (but see Hämmerer and colleagues[11]). Here, we took advantage of the paramagnetic properties of neuromelanin, a byproduct of NE synthesis, in T_1-weighted MRI to image the LC in vivo[44,46]. We assessed learning and memory performance in a large, healthy sample of younger and older adults along with structural MRI markers of LC integrity.

Our findings demonstrate reduced learning and memory performance in older relative to younger adults. Crucially, individual differences in learning and memory in a widely used neuropsychological test of memory functioning and, beyond that, across a variety of alternative memory tasks (see Supplemental results), were positively related to LC integrity in older adults. Moreover, we observed spatially confined age differences in LC intensity ratios. Older adults with more youth-like intensity ratios in rostral, hippocampus-projecting LC segments were better able to preserve memory performance. Making use of the longitudinal nature of this data set (see Online methods), we also examined associations between LC and memory performance using behavioral data that were obtained about 2.2 years before LC measurements were taken. We again observed that higher LC integrity was related to better memory performance among older adults, suggesting stable and lasting rather than short-lived LC–memory dependencies (see

Supplemental results). These results bridge a gap between animal and in vivo human research and provide important novel insights into the neural underpinnings of senescent cognitive decline in healthy aging.

We applied an iterative learning and memory task (RAVLT)[60] that required subjects to encode, consolidate, and retrieve verbal information and thus captures the dynamic nature of memory[51]. Our analyses allow the study of two psychologically distinct factors, namely, initial recall (i.e., performance after the first learning trial, corresponding to a standard one-trial memory assessment) and learning (i.e., changes in performance with practice). Age differences were mainly observed in initial recall performance (see Supplemental discussion).

Neurodegenerative changes have been suggested to be determinants of interindividual differences in learning and memory performance in late life[49,61]. Here, we investigated the role of the LC, one of the first targets of neurodegenerative diseases such as Alzheimer's and Parkinson's, in senescent memory decline[8,9,34]. In particular, we exploited the paramagnetic properties of neuromelanin-metal compounds in T_1-weighted MRI to index the integrity of the LC in vivo[44] (see [62,63] for recent discussions of LC contrast mechanisms). Neuromelanin accumulates non-linearly in LC cells across life with a peak concentration at around 50–60 years of age and subsequent decline, probably due to preferential loss of pigmented cells[40,41]. Thus, neuromelanin forms a reliable natural contrast agent, particularly in older adults, that can be harnessed in cell count[14] and MRI studies[46] to index LC integrity. As discussed by Manaye and colleagues[18], however, neuromelanin may constitute a less well-suited proxy for LC integrity in younger adults as compared to immunohistochemical markers (e.g., antibodies against tyrosine hydroxylase, i.e., the rate-limiting enzyme in NE synthesis). We successfully established a semi-automatic procedure that extracted individual LC intensity ratios across the rostrocaudal extent

from high-resolution neuromelanin-sensitive brainstem MRI in younger and older adults. In particular, we pooled over aligned scans to facilitate LC delineation at a group level[56], which in turn was used to restrict automatized extraction of individual peak intensities[63,64]. The LC volume of interest (search space) generated at the group level matches recent histological analyses that reported a dense packing of noradrenergic cells in a thin central LC compartment and dispersion towards rostral and especially caudal extremities[14]. We demonstrated high validity of the proposed semi-automatic procedure by comparisons to both previously published LC maps[39,56] and manual LC intensity ratings (see Supplemental results). Repeated measurements of an independent younger adult sample further confirmed high reproducibility of the intensity assessment (see Supplemental results). However, as reported earlier for neuromelanin-sensitive sequences, our analyses do not capture the complete LC extent[45,65] leading to lower volume estimates than those obtained post-mortem studies[15]. In their histological analyses, Fernandes and colleagues describe a common, central LC zone which is shared across all subjects[14]. Accordingly, the most dispersing caudal LC segments show less overlap on a group level and are thus arguably more difficult to detect using a semi-automatic approach with a conservative intensity threshold (i.e., 4*SD* above mean intensity of a reference region[63,64]). This circumstance is likely exacerbated by partial volume effects due to the relatively thick slices in most neuromelanin-sensitive sequences[65]. However, manual[56] and threshold-free automatic approaches[39] also appear to be affected by this challenge as evident from their low correspondence in caudal slices (MNI $Z \leq -30$: 6.91 % of Keren et al. in Betts et al., 4.63 % vice versa[39,56]; see Figure 2d). We thus refrained from attempts to capture the most caudal LC segments via manual segmentation, a lowering of intensity thresholds, or both, in order to effectively separate LC from the nuclei subcoerulei[13].

To investigate the link between LC integrity and memory in aging humans, we estimated latent LC integrity scores using structural equation modeling. Here, we integrated over slices and hemispheres to derive a single measure reflecting LC integrity. In line with spatially confined age differences[18,39,56], we observed comparable average LC scores in younger and older adults. In older adults, higher initial recall performance and steeper learning curves were found in subjects with high LC integrity. Even when integrating over a variety of tasks (see Supplemental results), our findings corroborate a positive dependency of memory performance on LC integrity, in line with a general mechanism beyond specific mnemonic domains (e.g., emotional memory[11]).

Coinciding with recent reports[11], there was no reliable association between LC integrity and memory performance in younger adults, potentially owing to a lower sensitivity of neuromelanin as a LC integrity proxy, as discussed above[18]. However, the association between LC integrity and memory did not differ reliably between groups. Thus, while we discovered an association between interindividual differences in memory performance and LC integrity in older but not younger adults, the lack of reliable group differences suggests comparable associations independent of age. This is in line with reports of an early ontogenetic maturation of the LC[66].

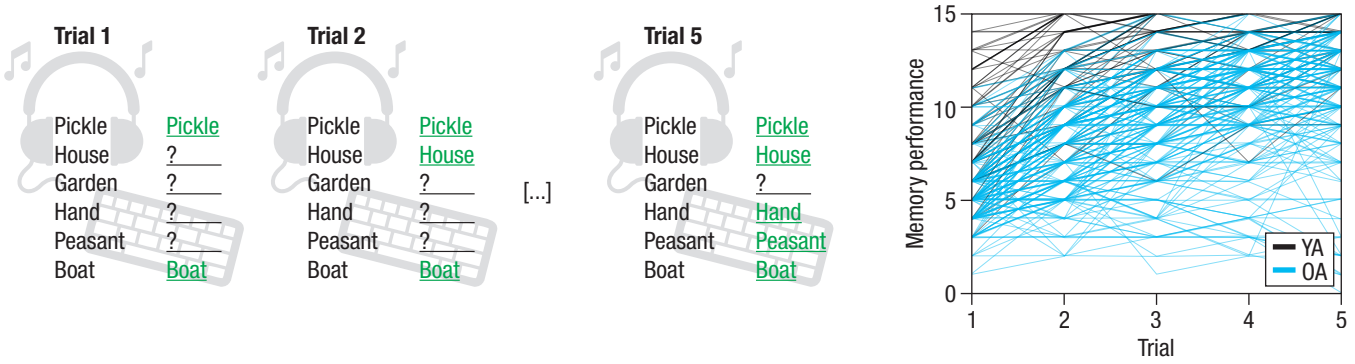
We observed spatially confined age differences in LC intensity ratios[39,56]. In older relative to younger adults, caudal LC sections showed significantly elevated intensity ratios in line with neuromelanin accumulation across the life span[40,41]. In contrast, there was an absence of increased signal in rostral LC and even a tendency towards reduced intensity ratios in older adults. This suggests neurodegeneration of rostral LC segments that are densely connected to key memory structures like the forebrain and hippocampus[17–19]. Accordingly, LC intensity ratios in rostral but not caudal segments were positively associated with memory performance. This finding is in line with a series of cell counting studies that point towards specific loss of rostral

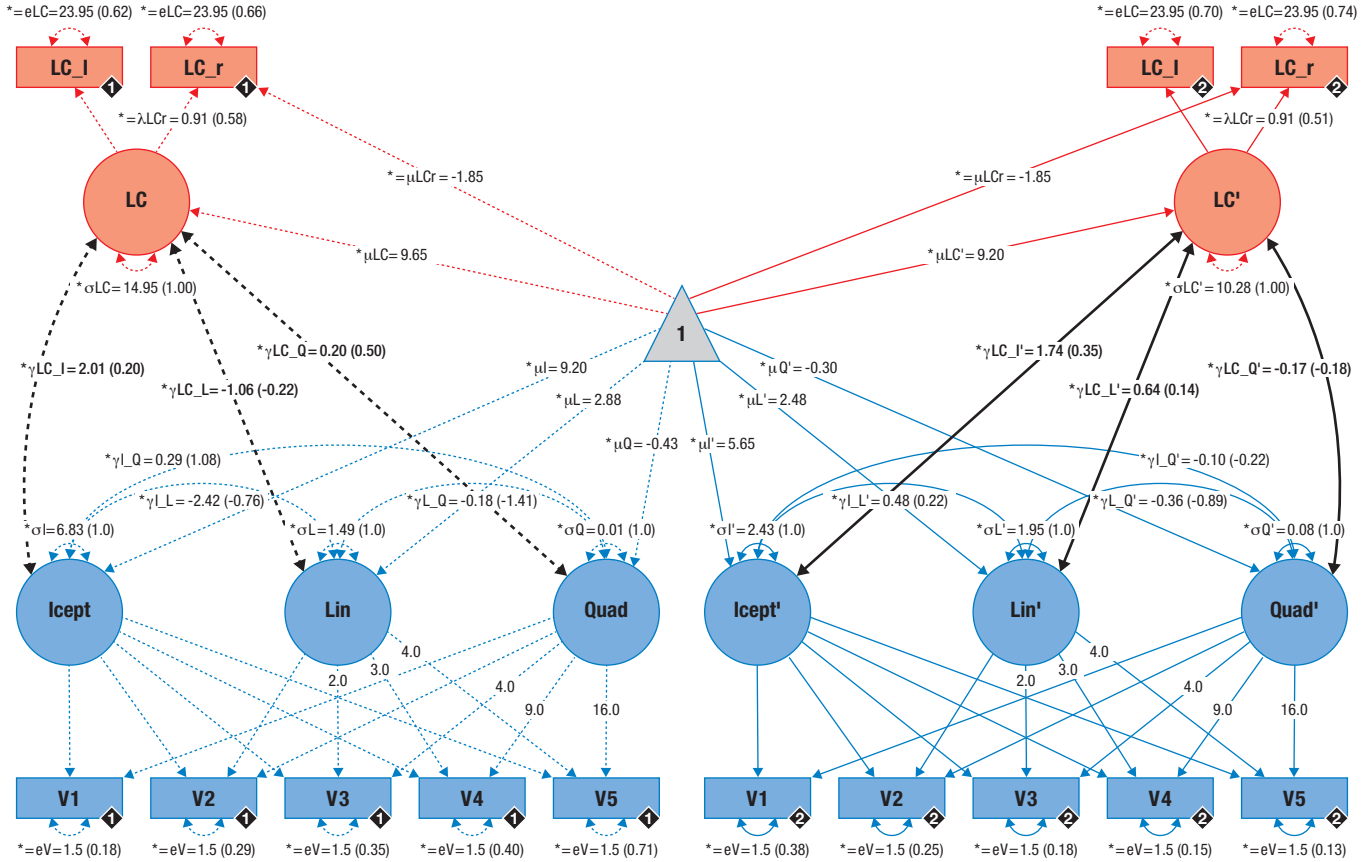
LC compartments even in healthy aging (for a review, see [18], but also see[34] for a recent discussion).

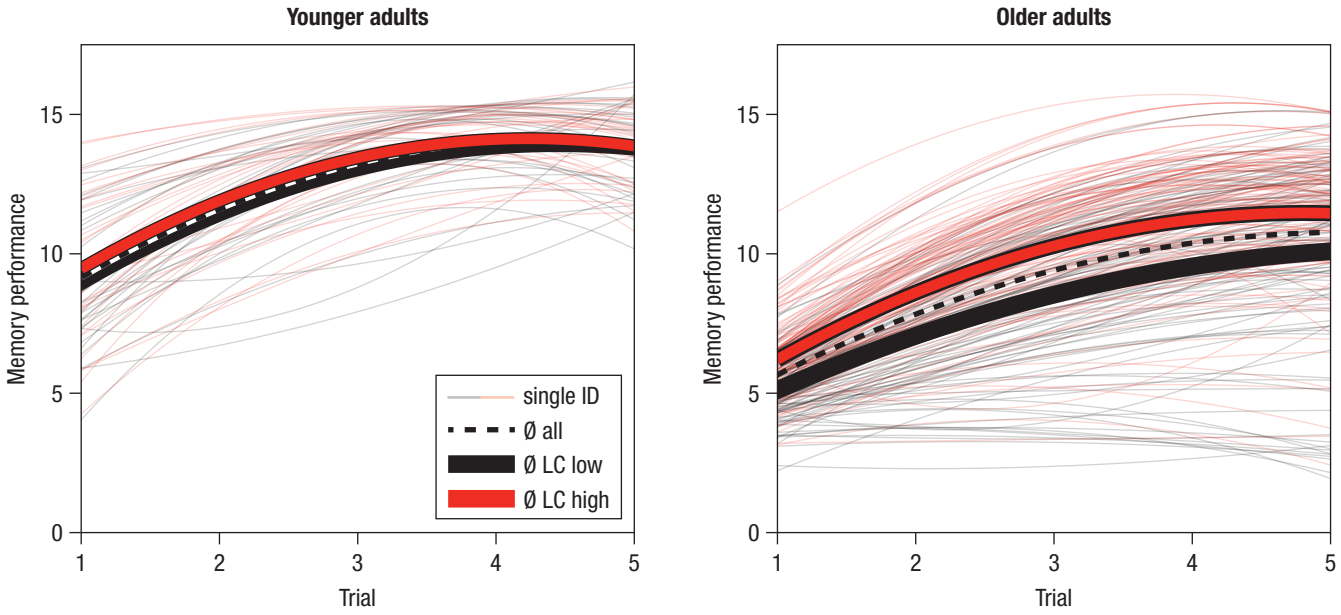
At least three well-documented mechanisms of noradrenergic action may explain the observed general memory-promoting effect of high LC integrity. First, NE release in the sensory cortices increases signal-to-noise ratios by silencing spontaneous activity while sparing or even facilitating stimulus-evoked responses[17,28], presumably in self-enhancing feedback loops with glutamate[25]. In addition, NE improves the temporal organization of neuronal responses to sensory stimulation (i.e., spike rhythmicity) and thus increases perceptual acuity[28,67,68]. Note that noisier information processing has been linked to deficient neuromodulation with effects on higher-order cognitive functions like attention and working memory[69]. Second, via action on post-synaptic α_{2A}-adrenoceptors in the prefrontal cortex, moderate levels of NE facilitate delay-related activity, which is considered a cellular analogue of working memory[22]. Accordingly, pharmacological manipulation of NE levels in aged monkeys ameliorated attention and working memory deficits[37]. Hence, LC integrity may promote attentional and control mechanisms implicated in successful episodic memory performance[70]. Third, in the amygdala and hippocampus, NE modulates synaptic strength and facilitates synaptic plasticity[27,28]. More specifically, via its action on ß-adrenoceptors, the LC modulates LTP and LTD, major determinants of long-term memory[29,30]. In sum, NE released from the LC alters perception, attention and memory at multiple cortical and subcortical sides, crucially implicating LC integrity in learning and memory in aging.

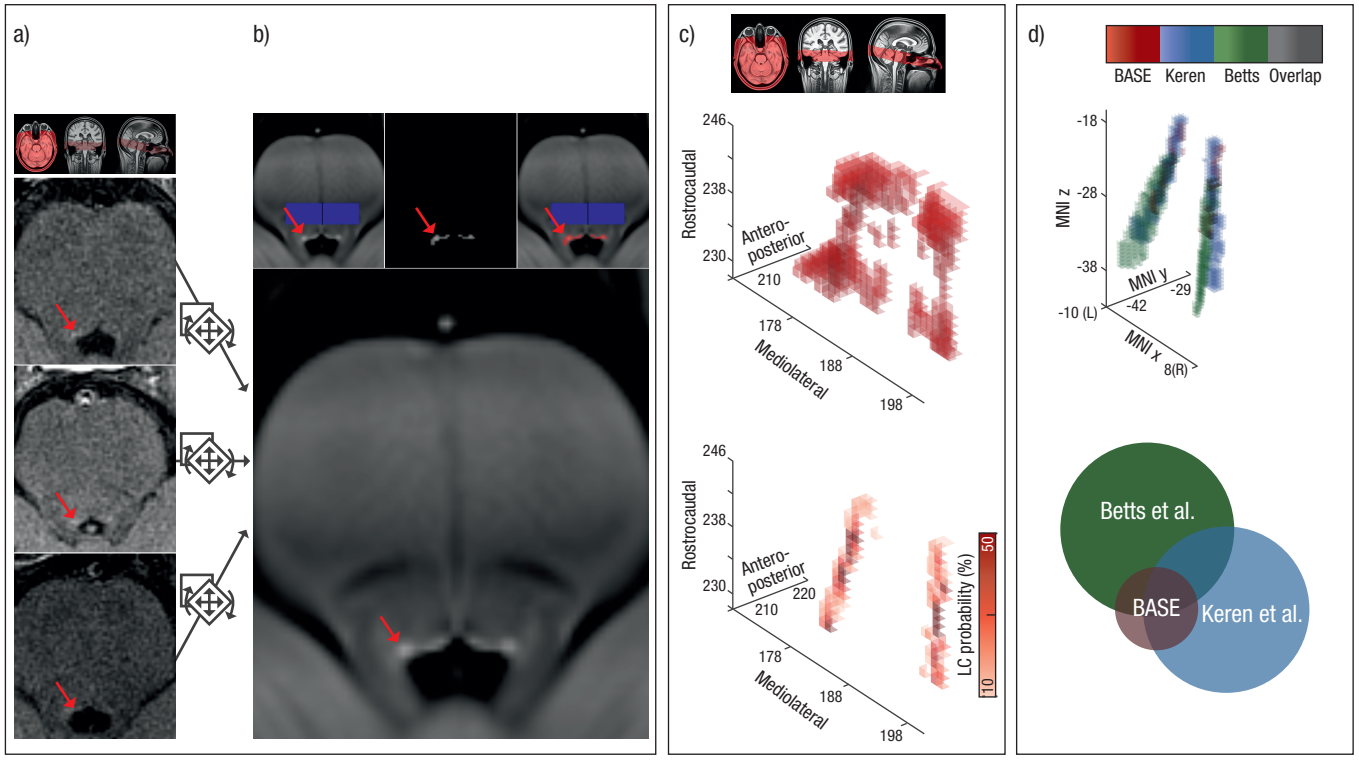
The present results underscore the utility of non-invasive, in vivo markers of LC integrity as indicators for preserved memory performance in human aging and extend our knowledge about the role of the LC-NE system in senescent decline. Using non-invasive in vivo MRI, we

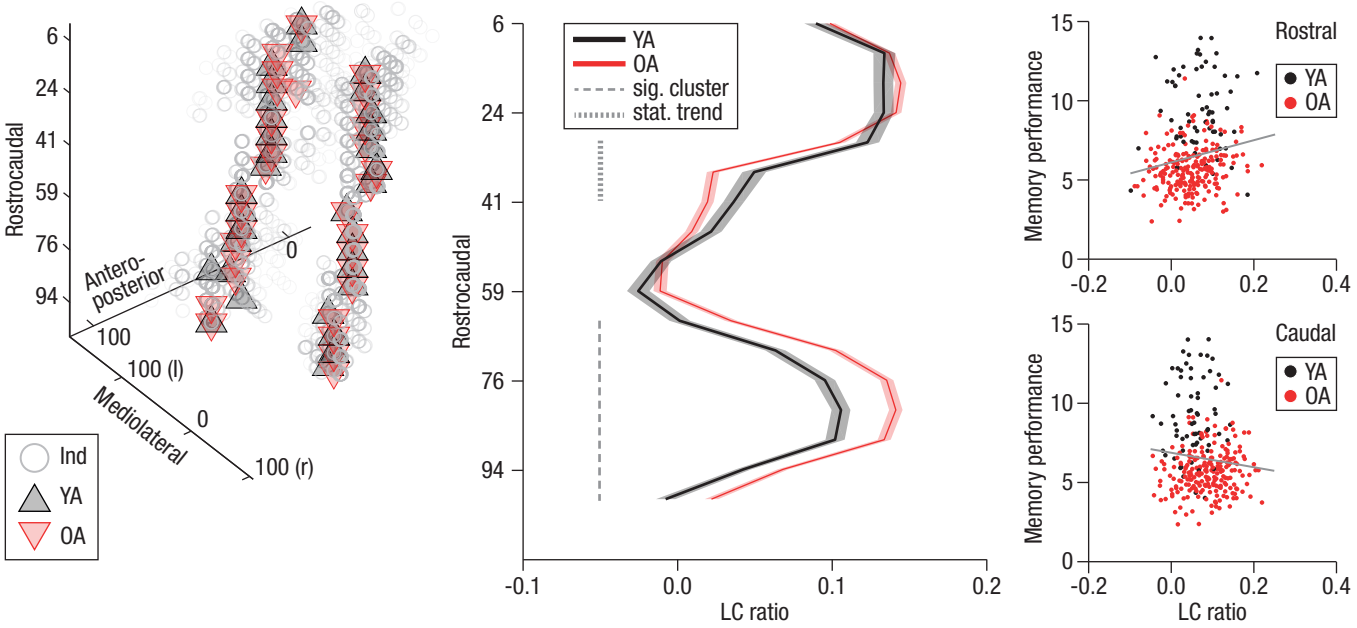
discovered spatially confined and functionally relevant differences between younger and older adults in those segments of LC that are connected to key memory structures like the hippocampus. Importantly, structural equation modeling revealed reliable and stable positive associations between LC integrity and general episodic memory among older adults. We conclude that older adults with more preserved LC integrity are equipped with more proficient episodic memory. This finding, which needs to be corroborated by long-term longitudinal evidence, adds specificity to the general proposition that brain maintenance is a key feature of successful cognitive aging[1].











Online methods

Study design and participants

Data was collected within the framework of the Berlin Aging Study-II (BASE-II[71–73]). BASE-II is a multi-disciplinary and multi-institutional longitudinal study that investigates the cognitive, neural, physical, and social conditions that are associated with successful aging. For an extensive study description, please refer to the study's website (https://www.base2.mpg.de/en) and papers[71–73].

Cognitive and neuro-imaging data of younger and older participants were collected at the Max Planck Institute for Human Development (Berlin, Germany) at two time points (T1 and T2) between 2013 and 2016. On average, data acquisitions were 2.21 years apart (standard deviation (*SD*) = 0.52, range = 0.9–3.2). Cognitive and neuro-imaging data were collected on separate occasions at each time point (mean interval = 9.16 days, *SD* = 6.32, range = –2–44; for T2). Locus coeruleus (LC) data were acquired only at the T2.

A subset of 323 BASE-II participants underwent magnetic resonance imaging (MRI), with 24 of these excluded before analysis due to missing or incomplete neural (n = 19) or cognitive (n = 5) data. After visual inspection of brainstem MRI, five additional participants (1.67 %; 0 female; mean age: 76.66 years; *SD* = 1.64; range = 74.93–78.74; at T2) were excluded from further analyses due to excessive movement artifacts (n = 2) or incorrect scan positioning (n = 3). The final MRI subsample (see Supplemental methods) included 66 younger adults (22 female) with a mean age of 32.5 years (SD = 3.53, range = 25.41–39.84; at T2) and 228 older adults (82 female) with a mean age of 72.29 years (*SD* = 4.11; range = 62.53–83.16; at T2).

Neurological and psychiatric disorders, a history of head injuries, or intake of memory-affecting medication precluded inclusion in the study. All eligible participants were MRI-compatible, right-handed, and had normal or corrected-to-normal vision.

The cognitive and MRI assessment were approved by the Ethics Committees of the Max Planck Institute for Human Development and the German Psychological Society (DGPs), respectively. Participants signed written informed consent and received monetary compensation for their participation. All experiments were performed in accordance with relevant guidelines and regulations.

Cognitive data assessment

The baseline cognitive assessment of BASE-II (T1) and its follow-up (T2) included an assessment of working memory, episodic memory and fluid intelligence in small group sessions of about 6 individuals using a computerized battery of 21 tasks (for a description of the complete battery, please refer to[74]). Each test session lasted approximately 3.5 hours. At T1, a second cognitive assessment was scheduled one week after the first session to test for long delayed recall performance.

Our hypotheses focused on the Rey Auditory Verbal Learning Test (RAVLT), a standardized and validated neuropsychological tool that provides information about participants' learning and delayed recall performance[60] (see Figure 1): Subjects first learned a 15-word list that was auditorily presented via headphones. The task was composed of five learning trials each followed by a free recall period in which participants entered the words they remembered via keyboard (trial 1–5; recall of learning list). After initial learning, a 15-word distractor list, containing semantically unrelated words was presented, followed by a free recall phase (trial 6; recall of the distractor items). Next, participants were again asked to freely recall only items

presented in the initial list (trial 7; recall of learning list after distraction). Another free recall test was administered after a delay of 30 minutes (trial 8; delayed recall of learning list). The verbal learning memory task ended with a recognition memory test that included 50 items (learning list: 15, distractor list: 15, similar lures: 20). At T1, participants were re-invited 7 days later for a final long delayed free recall test (trial 9; long delayed recall of learning list). After the initial learning cycles, the correct word list was never re-presented to subjects. The same word list was used at T1 and T2.

Making use of the comprehensive cognitive battery available for this data set, we additionally integrated performance over a variety of episodic memory tasks to retrieve a general measure of episodic memory as previously described[74] while explicitly excluding RAVLT data (see Supplemental results). In particular, we incorporated information of a visual face–profession, object–location, and scene encoding task. For detailed task descriptions, please refer to[74]. In short, the visual face–profession task (FPT) involved subjects studying the 45 pairs of face images and profession words. The tasks consisted of an incidental encoding phase, a 2-minute distraction phase, and finally a recognition memory task including old, new as well as rearranged face–profession pairs. Recognition memory (hits – false alarms) for rearranged pairs was used as the performance index. In the visual object–location memory task (OLT), participants encoded the location of 12 digital photographs of real life objects on a 6 x 6 grid. After encoding, objects reappeared next to the grid and subjects were asked to reproduce the correct location by placing the items in the correct grid position. The sum of correct placements was used as the performance index. Finally, in a visual scene encoding task (IOST), participants incidentally encoded 88 scene images by performing indoor/outdoor judgments on each image. The encoding phase was followed by an old/new recognition memory test which included confidence

judgments. Recognition memory (hits – false alarms) was tested after a delay of approximately 2.5 hours and served as the performance index.

Magnetic resonance imaging data assessment

Structural and functional MRI data were collected on both time points (T1, T2) employing a 3-Tesla Siemens Magnetom Tim Trio Scanner with a standard 12-channel head coil. Only those sequences used in the current analyses are described below.

A three-dimensional T_1-weighted magnetization prepared gradient-echo (MPRAGE) sequence with a duration of 9.2 min and the following parameters was applied: TR = 2500 ms, TE = 4.770 ms, TI = 1100 ms, flip angle = 7 °, bandwidth = 140 Hz/pixel, acquisition matrix = 256 × 256 × 192, isometric voxel size = 1 mm^3. Pre-scan normalize and 3D distortion correction options were enabled.

Based on this whole-brain MPRAGE sequence, a neuromelanin-sensitive high-resolution, two-dimensional T_1-weighted turbo-spin echo (TSE) sequence was aligned perpendicularly to the plane of the respective participant's brainstem. Acquisition of the TSE sequence took 2 × 5.9 min, and the following parameters were used: TR = 600 ms, TE = 11ms, flip angle = 120 °, bandwidth = 287 Hz/pixel, acquisition matrix = 350 × 263 × 350, voxel size = 0.5 × 0.5 × 2.5 mm^3. Each TSE scan consisted of 10 axial slices with a gap of 20 % between slices, which covered the whole extent of the pons. Pre-scan normalize and elliptical filter options were enabled. The TSE sequence yielded two brainstem MRIs per participant, each resulting from 4 (online) averages.

For some participants, specific absorption rate (SAR) limits were exceeded, which precluded further scanning with the current set of parameters (cf.[45,63]). To avoid data loss while maintaining contrast settings, for those participants the maximal number of TSE slices was

reduced. Thus, our sample consists of the following slice distribution: 243 × 10 slices, 38 × 9 slices, 27 × 8 slices, 6 × 7 slices, 1 × 6 slices, 3 × 11 slices (acquired before adjusting the number of slices in the sequence protocol). TSE scans are available only for T2.

Cognitive data analysis

We applied structural equation modeling (SEM) to analyze inter- and intra-individual differences in verbal learning and memory performance using the Ωnyx 1.0-991 software package[75] with full information maximum likelihood estimation (FIML). SEM offers a multivariate approach in which repeatedly observed (manifest) variables can be used to examine hypotheses about unobserved (latent) variables. Latent variables have the benefit of accounting for measurement error in observed scores and thereby increasing statistical power[76,77]. Here, we assumed that younger and older adults recall a gradually increasing number of words over iterative RAVLT learning trials. Formally, this can be expressed as a learning curve consisting of an initial memory performance level (intercept) and a gain over learning trials (slope). Since we hypothesized that the participants' learning performance reaches a natural (or task-induced) performance limit, we expected an increasing yet negatively accelerated slope[51,52]. Using latent growth curve modeling (LGCM, a specific variant of SEM), we estimated the intercept and slope factors on a latent level[54]. To test for differences in intercept and slope parameters between age groups, we opted for a multiple-group model[76,77]. In particular, for each time point we fitted a simple quadratic growth model comprising an intercept as well as quadratic and linear slope factor for each age group (see Figure 3 in Supplemental results). To investigate potential relations between intercept and slope terms (e.g., starting out higher may be related to faster performance increases), we freely estimated covariances between intercept and slope parameters within age groups[53].

We assessed the adequacy of the proposed growth models by testing for differences between the model-implied and empirically observed covariance matrices[78]. The χ^2-test formally tests for equity of the covariance matrices. However, it is particularly sensitive to sample size and should be interpreted with caution in large samples[55,78]. Thus, we examined two additional frequently reported fit indices: First, the root mean square error of approximation (RMSEA) is a closeness of fit coefficient that expresses how much the postulated model approaches the true model. Second, the comparative fit index (CFI) is an incremental fit index which compares the goodness of fit of the proposed model with a more restrictive nested baseline model[55,76,78]. In contrast to the χ^2-test, the CFI is not influenced by sample size. RMSEA values close to or below 0.06 and CFI values of close to 0.95 or greater indicate good model fit[55]. Finally, the adequacy of the quadratic growth model was tested against nested competing alternative models (e.g., intercept only and linear growth models) using likelihood ratio difference tests[76,78]. In the case of non-nested competing alternative models, we compared the Akaike Information Criterion (AIC), favoring the model providing the smallest AIC[76].

After establishing model fit, age differences in parameters of interest (e.g., means and (co)variances of intercept and slope factors) were tested by fixing parameters to equity between groups and comparing model fit to a model in which parameters were freely estimated using a likelihood ratio difference test [76,78].

In a similar vein, we integrated performance over a variety of memory tasks to estimate a general episodic memory score using a multiple-group single factor structural equation model as previously described[74](see Figure 4 in Supplemental results). In particular, in each age group performance in the visual face – profession, object – location, and scene encoding tasks served as manifest variables and loaded on a single latent episodic memory factor. Factor loadings (other

than the first, which was fixed to one) were estimated freely but were constrained to be equal across groups. Adequacy of the proposed model was assessed using a χ^2-test as well as two additional fit indices (RMSEA, CFI).

Magnetic resonance imaging data analysis

We established a multi-step analysis procedure to extract individual LC intensity values that addressed a series of methodological concerns raised in previous research: First, TSE scans of the LC typically demonstrate a low signal-to-noise ratio due to the high spatial resolution needed to image the small nucleus[63]. In addition, the LC shows amorphous borders, complicating the manual segmentation of the structure[39,65]. Thus, the use of semi-automatic threshold-based approaches was recently suggested to circumvent manual tracing of the nucleus and associated errors as well as potential biases (e.g., non-blinded raters[45,63,64]). Second, a reliable coregistration of brainstem MRI to standard space poses a major methodological challenge with standard software packages that can be addressed by the use of study-specific templates and coregistration masks[39]. Third, age- and disease-related alterations of LC integrity are not homogeneous across the rostrocaudal extent of the structure[9,39,56]. Increased specificity and sensitivity to detect such conditions may be reached by assessing intensity across the whole rostrocaudal axis[45]. Fourth, due to the use of multiple measurements and online averaging, the acquisition time for most neuromelanin-sensitive sequences exceeds 10 minutes, which increases the risk of head motion artifacts and thus reduced data quality[63]. To explicitly correct for this, images from each measurement can be saved separately and averaged after application of motion-correction methods[63].

In the proposed analysis procedure (see Figure 1 in Supplemental methods), we first pooled across subjects over aligned TSE scans to increase signal-to-noise ratio and facilitate LC

delineation using a template-based approach (cf.[39,56]; Steps 1–3). On a group level, LC location was determined semi-automatically (cf. [63,64]) to restrict fully-automatized extraction of individual peak intensity values across the rostrocaudal extent (Steps 4–5). For each participant, data from two scans (both from T2) were assessed, and extracted values were averaged to obtain a more reliable estimate of LC intensity.

Coregistration and standardization of magnetic resonance imaging data

Prior to analysis, all whole-brain and brainstem MRI scans (MPRAGE, TSE) were up-sampled to twice the initial matrix size by sinc-interpolation to improve visualization of the LC as previously described (cf.[56]).

Initially (Step 1), using Advanced Normalization Tools v. 2.1 (ANTS[79,80]), a group whole-brain template ($Template_{whole}$) was generated from all available MPRAGE scans from T2. In this iterative process, (a) individual native space scans ($MPRAGE_{native}$) were coregistered to a common group or template space and (b) coregistered scans ($MPRAGE_{template}$) were then averaged to form the group whole-brain template. An average of all input files served as an initial fixed image. One linear (rigid, then affine) registration of input files was followed by six non-linear registration iterations. Each non-linear registration was performed over three increasingly fine-grained resolutions (30 × 90 × 20 iterations). We applied a N4 bias field correction on moving images before each registration. Cross-correlation (CC) was used as the similarity metric and greedy Symmetric Normalization (SyN) as transformation model for non-linear registrations[81]. Template update steps were set to 0.25 mm.

Next (Step 2), within each subject, we coregistered the native space brainstem scan (TSE_{native}) to the whole-brain scan that had been moved to template space (i.e., $MPRAGE_{template}$; see Step 1). For this, we first performed linear (rigid, then affine) followed by nonlinear (SyN)

registration steps[81]. Histogram matching was enabled since we registered scans from the same modality (T_1-weighted). This intermediate within-subject coregistration was implemented to bring individual brainstem scans to a common (whole-brain) space and thereby correct for differences in acquisition angle that would reduce precision of the template building (see Step 3).

Finally (Step 3), the aligned brainstem scans were fed into a second template building run, resulting in (a) individual brainstem scans coregistered to template space ($TSE_{template}$) and (b) a group brainstem template image ($Template_{slab}$). The same parameters were used were used for both template building runs (Step 1, Step 3). The linear and non-linear transformation matrices of the intermediate within-subject coregistration (Step 2) and template building (Step 3) were concatenated and applied to the native space brainstem images (TSE_{native}) to achieve spatial standardization ($TSE_{template}$) within a single transformation step (cf.[56]). To correct for minor deviations between whole-brain and brainstem template space, we coregistered the two group templates using a linear (rigid, then affine) followed by nonlinear (SyN) registration. Since we registered a whole-brain image and a slab here, a coregistration (or "lesion") mask was generated that restricted registration to the brainstem, i.e. voxels falling inside the $Template_{slab}$ (cf.[39]).

Semi-automatic locus coeruleus segmentation and intensity assessment

Compared to individual scans, the group brainstem template demonstrated a high signal-to-noise ratio and markedly increased intensities in left and right brainstem sections next to the fourth ventricle corresponding to hypothesized LC locations (see Figure 3b). In order to circumvent manual segmentation on the brainstem template, we thresholded the image (Step 4) allowing only the brightest voxels to remain visible using a custom Matlab function (cf.[63,64]; Mathworks Inc., Sherborn, MA, USA). In particular, we applied a data driven approach using a threshold based on the intensity in a dorsal pontine reference region (see below) that effectively

isolated patches of high image intensity next to the fourth ventricle. The threshold was calculated for each template slice that covered the full brainstem (n = 31), separately for each hemisphere, as follows:

Eq. 1: $T_{LC} = \mu(\text{Ref}) + \sigma(\text{Ref}) \times 4$

Whereby μ and σ specify the mean and standard deviation of the intensity in the pontine reference (Ref) region, respectively. T_{LC} denotes the resulting intensity threshold. The cut-off of 4 standard deviations above the mean reference intensity was selected on the basis of previous research[63,64] and was confirmed visually as generating a volume that matched prior mapping studies of the LC[39,56]. A lateralized threshold was computed for each slice to counteract previously reported unilateral biases in brainstem signal intensity[56,65] and intensity fluctuations along the rostrocaudal axis. Next, we generated a volume of interest (VOI), subsequently referred to as search space, that included only those areas that survived thresholding. The resulting bilateral LC search space covered 17 slices and included 409 voxels in total (see Figure 3c). In addition to the LC search space, we also defined a larger rectangular dorsal pontine reference search space, spanning 35 × 10 voxels on all 17 slices on which the LC remained visible (5950 voxels). The bilateral LC and reference search spaces were then split at midline to allow for bilateral and lateralized analyses.

To obtain LC intensity ratios for all subjects (Step 5), we masked individual brainstem scans in template space (i.e., $\text{TSE}_{\text{template}}$) with the binarized LC search space. This was first done with a bilateral, followed by unilateral (left, right) search spaces. We thereby restricted the area from which intensity values could be extracted to potential LC coordinates. We then automatically selected the peak (maximal) intensity voxel within each masked scan for each slice and stored the corresponding intensity value and x/y/z coordinates[39]. In the same vein, we

masked individual scans with a binarized reference search space and extracted intensity values and coordinates. LC intensity ratios were then computed for each slice using the following formula[45]:

Eq. 2: $$LC_{ratio} = max(LC) - max(\text{Ref}) / max(\text{Ref})$$

Whereby LC denotes the intensity of the locus coeruleus and Ref indicates the intensity of the pontine reference region. Finally, for each hemisphere we computed an average LC ratio across all slices, which was used to probe LC–cognition associations in a SEM framework.

Comparison to previously published locus coeruleus masks

In order to judge the validity of the generated search space, we plotted the retrieved peak intensity coordinates for each subject's left and right LC[39]. The aggregated maximum intensity plot was then converted to express the likelihood to find the LC for any given voxel (i.e., a LC location probability map; cf.[39,65]). We thresholded the image to show only locations that were shared across participants (i.e., the upper 90 percentiles) and spatially standardized the probability map in a two-step procedure. First, we coregistered the whole-brain template (see Step 1) to 0.5 mm iso-voxel MNI space using linear (rigid, then affine), followed by nonlinear (SyN) registration. Since the brainstem template was registered with the whole-brain template (see Step 3), in a second step we were able to apply the transformation matrices ($Template_{slab} \rightarrow Template_{whole}$ and $Template_{whole} \rightarrow MNI_{0.5}$) to the LC probability map. In standard space, we calculated the overlap between the probability map and previously published LC masks[39,56]; see Figure 3d and Supplemental results). To facilitate comparability of study results, we freely share the established LC probability map with the neuroscientific community (https://www.mpib-berlin.mpg.de/LC-Map).

Comparison to manually assessed locus coeruleus intensity

The vast majority of published studies on MRI that indexed LC integrity used a manual procedure to assess intensity values (for a recent review, see[45]). To demonstrate the validity of the proposed approach, two independent, blinded raters manually assessed LC and reference intensity values for all subjects using a standard procedure[82]. On each axial slice demonstrating increased intensity in anatomically probable LC locations, raters placed three non-overlapping quadratic 1 mm² regions of interest (ROIs) in areas in which signal was most evident (cf.[82]). Raters were instructed not to place the ROIs directly adjacent to the fourth ventricle to avoid partial volume effects contaminating the signal (cf.[83]). To measure reference intensity, three rectangular non-overlapping 10 mm² ROIs were placed on the same slices in the dorsal pons, centered medially between the left and right LC ROIs (cf.[82]; see Figure 2 in Supplemental methods). Raters additionally indicated the slice with peak signal intensity. Image processing was conducted using the image j software (https://imagej.nih.gov/) in a room with constant illumination. Ratings were performed on offline-averaged, non-interpolated brainstem scans in native space. For analyses, the peak intensity for each slice was first determined for LC (left, right) and pons across ROIs, and then peak intensities were averaged across slices. We calculated the consistency of raters using intraclass correlation coefficients (ICC; two-way mixed model with absolute agreement) and then, after demonstrating high accordance, averaged ratings across hemispheres and raters to obtain a single stable intensity estimate per subject (see Supplemental results). Comparisons between manually and automatically assessed LC intensity values (see Step 5) were eventually assessed by means of ICC. For further analyses, we also calculated LC intensity ratios for both hemispheres using Eq. 2 (see Supplemental results).

Reproducibility of semi-automatically assessed locus coeruleus intensity ratios

In order to judge the temporal stability of the semi-automatic method, we repeatedly scanned a small number of younger adults who were not participating in the study over the course of 1–2 weeks (n = 3; 2 female; mean age: 23.667 years; *SD* = 1.578 years). Subjects were scanned four times resulting in eight brainstem images per participant (see MRI data assessment; 22 scans in total; missing data for n = 1 at the last time point). Data was sinc-interpolated and fed into another template-building procedure as described above (see Steps 1–3). The resulting brainstem template ($template_{slab\_reproducibility}$) was coregistered to the study's brainstem template ($template_{slab}$) using linear (rigid, then affine) followed by nonlinear (SyN) registration. This process effectively enabled alignment of each participant's scans with the group brainstem template and thus allowed us to use the established search spaces to compute LC intensity ratios as described above. Finally, ICC were calculated for bilateral and unilateral LC ratios to evaluate the stability of the proposed method. Note that we analyzed intensity ratios here since we compared scans acquired at different time points that are subject to overall scanner fluctuations.

Comparison to semi-automatically assessed locus coeruleus intensity in native space

We additionally extracted individual LC and reference intensity from native space brainstem scans (TSE_{native}) to confirm that intensity values retrieved from template space ($TSE_{template}$) are not an artifact of coregistration and template building (Steps 1–3). For this, we used the concatenated inverse transformation matrices of Steps 2 and 3 to project the search spaces back to individual subjects' coordinates. In order to compare spatial locations of native space values across subjects, however, a common frame of reference was necessary (cf.[56]). Therefore, before warping, we split the LC and reference search spaces into segments. Each but the most caudal segment contained three slices, reflecting the resolution difference between

native and template space (i.e., $1 \times TSE_{native} : 3 \times TSE_{template}$; 1.5 mm : 0.5 mm). The six resulting segments were separately transformed to native space. We then masked native space scans (TSE_{native}) with binarized LC and reference (segment) search spaces and assessed peak intensities as described above (cf. Step 5). The two native space brainstem images of each participant (see MRI data assessment) were linearly and non-linearly coregistered before masking. To judge the correspondence between automatically assessed LC intensity values from native and template space, we calculated ICC. For further analyses of native space values, we computed LC intensity ratios for both hemispheres as described above using Eq. 2 (see Supplemental results).

Analysis of age differences in the spatial distribution of locus coeruleus ratios

To investigate age differences in LC intensity ratios along the rostrocaudal extent of the nucleus[9,18,39,56], we calculated non-parametric, cluster-based, random permutation tests as implemented in the Fieldtrip toolbox. These effectively control the false alarm rate in case of multiple testing[57]. In short, first a two-sided, independent samples *t*-test is calculated for each slice. Neighboring slices with a *p*-value below 0.05 were grouped with spatially adjacent slices to form a cluster. The sum of all *t*-values within a cluster formed the respective test statistic. A reference distribution for the summed cluster-level *t*-values was computed via the Monte Carlo method. Specifically, in each of 100,000 repetitions, group membership was randomly assigned, a *t*-test computed, and the *t*-value summed for each cluster. Observed clusters whose test statistic exceeded the 97.5th percentile for its respective reference probability distribution were considered significant[57]. Cluster permutation tests were calculated first across hemispheres and then separately for each hemisphere on LC ratios assessed from template ($TSE_{template}$) and native space (TSE_{native}; see Supplemental results). Here, we used a two-sided test with a significance

level (α) of 0.025. In the following, we thus report the doubled cluster *p*-values to facilitate readability (i.e., $p \leq 0.05$ is considered significant for all statistics).

To evaluate the functional significance of identified topographical age differences, we investigated the association between memory performance and LC ratios in the obtained clusters, irrespective of age group. In particular, for each cluster, the relation between initial recall performance (intercept) and LC ratios (average over slices within the identified cluster) was assessed by means of Spearman's rank correlations.

Finally, to examine inter-hemispheric differences in LC ratios[56,65], we additionally computed related-samples Wilcoxon signed rank tests. We first calculated a statistic across age groups, which we then followed up by analyses within each age group.

Estimation of latent locus coeruleus integrity scores

As carried out for the cognitive data, we applied SEM to analyze interindividual differences in LC ratios. We generated a multiple-group model including average LC ratios of each hemisphere as observed variables (see Step 5; see Figure 6 in Supplemental results). In each group, the two observed variables loaded on a single latent LC integrity factor. Factor loadings (other than the first, which was fixed to one) were estimated freely but were constrained to be equal across groups. Adequacy of the proposed model was assessed using a χ^2-test as well as two additional fit indices (RMSEA, CFI). The analyses were first completed with automatically assessed LC ratios ($TSE_{template}$). To demonstrate the stability of the findings, we repeated the same steps with values assessed in native space (TSE_{native}) and manually assessed intensity ratios. Qualitatively similar results were obtained (see Supplemental results).

Analysis of associations between locus coeruleus integrity scores and memory performance

After generating structural equation models for our cognitive and neural measures, respectively, we set out to link both modalities. That is, we were interested in assessing the relation between interindividual as well as intra-individual differences in verbal learning and memory performance and interindividual differences in LC scores. For this, we first built a unified model merging the verbal learning and LC models described above for each time point (see Figure 2). We investigated associations between our performance and LC measures by allowing for freely estimated covariances on a latent level (shown in black, Figure 2). Models were estimated using LC ratios assessed in template space ($TSE_{template}$). To assess the stability of the obtained findings, the same analyses were repeated with LC ratios assessed in native space (TSE_{native}) and manually assessed LC ratios (see above). Qualitatively similar results were derived (see Supplemental results). In order to evaluate the generalizability of the expected LC–memory association, we repeated the analyses, this time merging the neural SEM with the general episodic memory SEM (see *Cognitive data analysis* and Figure 5).

Next, to explore change in cognition over time and its relation to neural indices, we combined the LC–verbal learning models of both time points (T1 and T2) in a latent change score model[53,77] (see Figure 10 in Supplemental results). As only cognitive performance was assessed at both time points, we calculated a univariate (multiple indicator) latent change score model. In this, we estimated the change in latent intercept and slope factors on a second order latent level, expressed by Δ-intercept and Δ-slope factors. Since the same word list was tested in all learning trials in the verbal learning memory task and an identical list was used at T1 and T2, we did not include residual covariances of error terms across time points or correlated errors (cf. [77]). This model allows conclusions about the average rate of change in cognition for each age

group (mean Δ-*) as well as about the within-group variance in change (variance Δ-*;[77]). The association between the cognitive latent change scores and the latent neural variables was assessed by means of freely estimated covariances. As before, model fit for all described models was determined using a χ^2-test in combination with two additional fit indices (RMSEA, CFI; see Supplemental results).

References

1. Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U. & Bäckman, L. Memory aging and brain maintenance. *Trends Cogn. Sci.* **16,** 292–305 (2012).
2. Crook, T. *et al.* Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change—Report of a National Institute of Mental Health Work Group. *Dev. Neuropsychol.* **2,** 261–276 (1986).
3. Prince, M. J. *et al. World Alzheimer Report 2015 - The global impact of dementia: An analysis of prevalence, incidence, cost and trends*. (London: Alzheimer's Disease International, 2015).
4. Nyberg, L. *et al.* Dopamine D2 receptor availability is linked to hippocampal–caudate functional connectivity and episodic memory. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 7918–7923 (2016).
5. Fandakova, Y., Lindenberger, U. & Shing, Y. L. Deficits in process-specific prefrontal and hippocampal activations contribute to adult age differences in episodic memory interference. *Cereb. Cortex* **24,** 1832–1844 (2014).
6. Lindenberger, U. Human cognitive aging: Corriger la fortune? *Science.* **346,** 572–578 (2014).
7. Leslie, F. M. *et al.* Noradrenergic changes and memory loss in aged mice. *Brain Res.* **359,** 292–299 (1985).
8. Braak, H., Thal, D. R., Ghebremedhin, E. & Del Tredici, K. Stages of the pathologic process in Alzheimer disease: Age categories from 1 to 100 years. *J. Neuropathol. Exp. Neurol.* **70,** 960–969 (2011).
9. Marien, M. R., Colpaert, F. C. & Rosenquist, A. C. Noradrenergic mechanisms in

neurodegenerative diseases: A theory. *Brain Res. Rev.* **45,** 38–78 (2004).

10. Wilson, R. S. *et al.* Neural reserve, neuronal density in the locus ceruleus, and cognitive decline. *Neurology* **80,** 1202–1208 (2013).
11. Hämmerer, D. *et al.* Locus coeruleus integrity in old age is selectively related to memories linked with salient negative events. *Proc. Natl. Acad. Sci. U. S. A.* **115,** 2228–2233 (2018).
12. Theofilas, P. *et al.* Locus coeruleus volume and cell population changes during Alzheimer's disease progression: A stereological study in human postmortem brains with potential implication for early-stage biomarker discovery. *Alzheimer's Dement.* **13,** 236–246 (2017).
13. Szabadi, E. Functional neuroanatomy of the central noradrenergic system. *J. Psychopharmacol.* **27,** 659–693 (2013).
14. Fernandes, P., Regala, J., Correia, F. & Gonçalves-Ferreira, A. J. The human locus coeruleus 3-D stereotactic anatomy. *Surg. Radiol. Anat.* **34,** 879–885 (2012).
15. German, D. C. *et al.* The human locus coeruleus: Computer reconstruction of cellular distribution. *J. Neurosci.* **8,** 1776–1788 (1988).
16. Aston-Jones, G. & Waterhouse, B. Locus coeruleus: From global projection system to adaptive regulation of behavior. *Brain Res.* **1645,** 75–78 (2016).
17. Berridge, C. W. & Waterhouse, B. D. The locus coeruleus-noradreneergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Rev.* **42,** 33–84 (2003).
18. Manaye, K. F., McIntire, D. D., Mann, D. M. A. & German, D. C. Locus coeruleus cell loss in the aging human brain: A non-random process. *J. Comp. Neurol.* **358,** 79–87 (1995).
19. Schwarz, L. A. & Luo, L. Organization of the locus coeruleus-norepinephrine system.

Curr. Biol. **25,** R1051–R1056 (2015).

20. Waterhouse, B. D. & Chandler, D. J. Heterogeneous organization and function of the central noradrenergic system. *Brain Res.* **1641,** v–x (2016).
21. Ordway, G. A., Schwartz, M. A. & Frazer, A. *Brain norepinephrine: Neurobiology and therapeutics*. (Cambridge: Cambridge University Press, 2007). doi:10.1017/CBO9780511544156
22. Arnsten, A. F. T. & Li, B. M. Neurobiology of executive functions: Catecholamine influences on prefrontal cortical functions. *Biol. Psychiatry* **57,** 1377–1384 (2005).
23. Aston-Jones, G. & Cohen, J. D. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28,** 403–450 (2005).
24. Bouret, S. & Sara, S. J. Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci.* **28,** 574–582 (2005).
25. Mather, M., Clewett, D., Sakaki, M. & Harley, C. W. Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behav. Brain Sci.* **39,** (2016).
26. Nieuwenhuis, S. & Jepma, M. Investigating the role of the noradrenergic system in human cognition. in *Decision making, affect, and learning: Attention and performance XXIII* (eds. Delgado, M., Phelps, E. & Robbins, T.) 367–386 (Oxford: Oxford University Press, 2011). doi:10.1093/acprof:oso/9780199600434.003.0017
27. Sara, S. J. Locus coeruleus in time with the making of memories. *Curr. Opin. Neurobiol.* **35,** 87–94 (2015).
28. Sara, S. J. The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev.*

Neurosci. **10,** 211–223 (2009).

29. Hansen, N. The longevity of hippocampus-dependent memory is orchestrated by the locus coeruleus-noradrenergic system. *Neural Plast.* **2017,** (2017).

30. O'Dell, T. J., Connor, S. A., Guglietta, R. & Nguyen, P. V. β-Adrenergic receptor signaling and modulation of long-term potentiation in the mammalian hippocampus. *Learn. Mem.* **22,** 461–471 (2015).

31. Bray, N. Learning and memory: You only learn once. *Nat. Rev. Neurosci.* **19,** 59 (2018).

32. McNamara, C. G. & Dupret, D. Two sources of dopamine for the hippocampus. *Trends Neurosci.* **40,** 383–384 (2017).

33. Takeuchi, T. *et al.* Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature* **537,** 357–362 (2016).

34. Mather, M. & Harley, C. W. The locus coeruleus: Essential for maintaining cognitive function and the aging brain. *Trends Cogn. Sci.* **20,** 214–226 (2016).

35. Mei, Y. *et al.* Aging-associated formaldehyde-induced norepinephrine deficiency contributes to age-related memory decline. *Aging Cell* **19,** 659–668 (2015).

36. Luo, Y. *et al.* Reversal of aging-related emotional memory deficits by norepinephrine via regulating the stability of surface AMPA receptors. *Aging Cell* **14,** 170–179 (2015).

37. Arnsten, A. F. & Goldman-Rakic, P. S. Alpha 2-adrenergic mechanisms in prefrontal cortex associated with cognitive decline in aged nonhuman primates. *Science* **230,** 1273–1276 (1985).

38. Astafiev, S. V., Snyder, A. Z., Shulman, G. L. & Corbetta, M. Comment on 'Modafinil shifts human locus coeruleus to low-tonic, high-phasic activity during functional MRI' and 'Homeostatic sleep pressure and responses to sustained attention in the

suprachiasmatic area'. *Science.* **328,** 309 (2010).

39. Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S. & Eckert, M. A. In vivo mapping of the human locus coeruleus. *NeuroImage* **47,** 1261–1267 (2009).

40. Zecca, L. *et al.* The role of iron and copper molecules in the neuronal vulnerability of locus coeruleus and substantia nigra during aging. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 9843–9848 (2004).

41. Mann, D. M. A. & Yates, P. O. Lipoprotein pigments–their relationship to ageing in the human nervous system. II. The melanin content of pigmented nerve cells. *Brain* **97,** 489–498 (1974).

42. Marcyniuk, B., Mann, D. M. A. & Yates, P. O. The topography of nerve cell loss from the locus caeruleus in elderly persons. *Neurobiol. Aging* **10,** 5–9 (1989).

43. Sasaki, M., Shibata, E., Kudo, K. & Tohyama, K. Neuromelanin-sensitive MRI: Basics, technique, and clinical applications. *Clin. Neuroradiol.* **18,** 147–153 (2008).

44. Sasaki, M. *et al.* Neuromelanin magnetic resonance imaging of locus ceruleus and substantia nigra in Parkinson's disease. *Neuroreport* **17,** 1215–1218 (2006).

45. Liu, K. Y. *et al.* Magnetic resonance imaging of the human locus coeruleus: A systematic review. *Neurosci. Biobehav. Rev.* **83,** 325–355 (2017).

46. Keren, N. I. *et al.* Histologic validation of locus coeruleus MRI contrast in post-mortem tissue. *NeuroImage* **113,** 235–245 (2015).

47. Albert, M. S., Moss, M. B., Tanzi, R. & Jones, K. Preclinical prediction of AD using neuropsychological tests. *J. Int. Neuropsychol. Soc.* **7,** 631–639 (2001).

48. Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V. & Croteau, J. Neuropsychological measures that predict progression from mild cognitive impairment to

Alzheimer's type dementia in older adults: A systematic review and meta-analysis. *Neuropsychol. Rev.* **27,** 328–353 (2017).

49. Moradi, E., Hallikainen, I., Hänninen, T. & Tohka, J. Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease. *NeuroImage Clin.* **13,** 415–427 (2017).

50. Schoenberg, M. R. *et al.* Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Arch. Clin. Neuropsychol.* **21,** 693–703 (2006).

51. Zimprich, D., Rast, P. & Martin, M. Individual differences in verbal learning in old age. in *Handbook of cognitive aging: Interdisciplinary perspectives* (eds. Hofer, S. & Alwin, D.) 224–243 (Thousand Oaks: SAGE Publications, 2008). doi:10.4135/9781412976589.n14

52. Jones, R. N. *et al.* A growth curve model of learning acquisition among cognitively normal older adults. *Exp. Aging Res.* **31,** 291–312 (2005).

53. Zimprich, D. & Rast, P. Verbal learning changes in older adults across 18 months. *Neuropsychol. Dev. Cogn. B. Aging. Neuropsychol. Cogn.* **16,** 461–484 (2009).

54. McArdle, J. J. Dynamic but structural equation modeling of repeated measures data. in *Handbook of multivariate experimental psychology* (eds. Nesselroade, J. R. & Cattell, R. B.) 561–614 (Boston: Springer, 1988). doi:10.1007/978-1-4613-0893-5_17

55. Brown, T. A. *Confirmatory factor analysis for applied research*. (New York:Guilford Press, 2006).

56. Betts, M. J., Cardenas-Blanco, A., Kanowski, M., Jessen, F. & Düzel, E. In vivo MRI assessment of the human locus coeruleus along its rostrocaudal extent in young and older adults. *NeuroImage* **163,** 150–159 (2017).

57. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164,** 177–190 (2007).

58. Mann, D. M. A. & Yates, P. O. Lipoprotein pigments–their relationship to ageing in the human nervous system. I. The lipofuscin content of nerve cells. *Brain* **97,** 481–488 (1974).

59. Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87,** 245–251 (1980).

60. Schmidt, M. *Rey Auditory Verbal Learning Test: A handbook.* (Los Angeles: Western Psychological Services, 2004).

61. Gifford, K. A. *et al.* Associations between Verbal Learning Slope and Neuroimaging Markers across the Cognitive Aging Spectrum. *J. Int. Neuropsychol. Soc.* **21,** 455–467 (2015).

62. Priovoulos, N. *et al.* High-resolution in vivo imaging of human locus coeruleus by magnetization transfer MRI at 3T and 7T. *NeuroImage* **168,** 127–136 (2017).

63. Chen, X. *et al.* Simultaneous imaging of locus coeruleus and substantia nigra with a quantitative neuromelanin MRI approach. *Magn. Reson. Imaging* **32,** 1301–1306 (2014).

64. Langley, J., Huddleston, D. E., Liu, C. J. & Hu, X. Reproducibility of locus coeruleus and substantia nigra imaging with neuromelanin sensitive MRI. *Magn. Reson. Mater. Physics, Biol. Med.* **30,** 121–125 (2017).

65. Tona, K. D. *et al.* In vivo visualization of the locus coeruleus in humans: Quantifying the test–retest reliability. *Brain Struct. Funct.* **222,** 4203–4217 (2017).

66. Amaral, D. G. & Sinnamon, H. M. The locus coeruleus: Neurobiology of a central noradrenergic nucleus. *Prog. Neurobiol.* **9,** 147–196 (1977).

67. Bouret, S. & Sara, S. J. Locus coeruleus activation modulates firing rate and temporal

organization of odour-induced single-cell responses in rat piriform cortex. *Eur. J. Neurosci.* **16,** 2371–2382 (2002).

68. Fries, P. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **9,** 474–480 (2005).

69. Li, S.-C., Lindenberger, U. & Sikstrom, S. Aging cognittion: From neuromodulation to representation. *Trends Cogn. Sci.* **5,** 479–486 (2001).

70. Shing, Y. L. *et al.* Episodic memory across the lifespan: The contributions of associative and strategic components. *Neurosci. Biobehav. Rev.* **34,** 1080–1091 (2010).

71. Delius, J. A. M., Düzel, S., Gerstorf, D. & Lindenberger, U. Berlin Aging Studies (BASE and BASE-II). in *Encyclopedia of geropsychology* (ed. Pachana, N. A.) 386–395 (New York: Springer, 2015). doi:10.1007/978-981-287-080-3_44-1

72. Gerstorf, D. *et al. The Berlin Aging Study II – An overview*. *Gerontology* **62,** (2016).

73. Bertram, L. *et al.* Cohort profile: The Berlin Aging Study II (BASE-II). *Int. J. Epidemiol.* **43,** 703–712 (2014).

74. Düzel, S. *et al.* Supplementary material for: The Subjective Health Horizon Questionnaire (SHH-Q): Assessing future time perspectives for facets of an active lifestyle. *Gerontology* **62,** 345–353 (2016).

75. von Oertzen, T., Brandmaier, A. M. & Tsang, S. Structural equation modeling with Ωnyx. *Struct. Equ. Model.* **22,** 148–161 (2015).

76. Curran, P. J., Obeidat, K. & Losardo, D. Twelve frequently asked questions about growth curve modeling. *J. Cogn. Dev.* **11,** 121–136 (2010).

77. Kievit, R. A. *et al.* Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Dev. Cogn. Neurosci.*Advance online publication.

(2017). doi:10.1016/j.dcn.2017.11.007

78. Eid, M., Gollwitzer, M. & Schmitt, M. *Statistik und Forschungsmethoden: Lehrbuch*. *Grundlagen Psychologie* (Weinheim: Beltz, 2015).

79. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54,** 2033–2044 (2011).

80. Avants, B. B., Tustison, N. & Song, G. Advanced Normalization Tools: V1.0. *Insight J.* **2,** 1–35 (2009).

81. Klein, A. *et al.* Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage* **46,** 786–802 (2009).

82. Shibata, E. *et al.* Age-related changes in locus coeruleus on neuromelanin magnetic resonance imaging at 3 Tesla. *Magn. Reson. Med. Sci.* **5,** 197–200 (2006).

83. Clewett, D. V. *et al.* Neuromelanin marks the spot: Identifying a locus coeruleus biomarker of cognitive reserve in healthy aging. *Neurobiol. Aging* **37,** 117–126 (2016).

Locus coeruleus integrity preserves memory performance across the adult life span

– Supplemental information

Martin J. Dahl*[1], Mara Mather[2], Sandra Düzel[1], Nils C. Bodammer[1], Ulman Lindenberger[1,3], Simone Kühn[1,4], & Markus Werkle-Bergner*[1]

[1]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

[2]Davis School of Gerontology, University of Southern California, Los Angeles, USA

[3]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, England, and Berlin, Germany

[4]Department of Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Hamburg, Germany

* Correspondence concerning this manuscript should be addressed to

MJD (dahl@mpib-berlin.mpg.de) or MWB (werkle@mpib-berlin.mpg.de)

Acknowledgements.
This article uses data from the Berlin Aging Study II (BASE-II), which was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under grant numbers #16SV5536 K, #16SV5537, #16SV5538, and #16SV5837, #01UW070 and #01UW0808. Additional contributions (e.g., financial, equipment, logistics, personnel) are made from each of the other participating sites, i.e., the Max Planck Institute for Human Development (MPIB), Max Planck Institute for Molecular Genetics (MPIMG), Charite-Universiätsmedizin, University Medicine, German Institute for Economic Research (DIW), Humboldt-Universität zu Berlin, all located in Berlin, Germany, and University of Lübeck in Lübeck, and University of Tübingen, Germany. For further information about the BASE-II project, see https://www.base2.mpg.de/en
MW-B received support from the German Research Foundation (DFG, WE 4269/5-1) and the Jacobs Foundation (Early Career Research Fellowship 2017–2019).
MJD is a fellow of the International Max Planck Research School on the Life Course (LIFE; http://www.imprs-life.mpg.de/en). MJD is recipient of a stipend from the Sonnenfeld-Foundation (http://www.sonnenfeld-stiftung.de/en/). MM's work was supported by an Alexander von Humboldt fellowship and by National Institutes of Health grant R01AG025340.
We thank Andrew R. Bender, Matthew Betts, and Myriam C. Sander for valuable discussions and assistance. We are grateful to Shelby Bachman and Dilara Zorbek, who performed the manual tracing of the locus coeruleus, as well as Ylva Köhncke and Yana Fandakova for statistical advice, and Michael Krause for help with cluster computing.

Supplemental material and methods

Table S1. Summary of sample descriptives for younger (YA) and older adults (OA).

	YA (n = 66; 22 female)				OA (n = 228; 82 female)			
	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
Age (at time point 2)	32.5	3.53	25.41	39.84	72.29	4.11	62.53	83.16
BMI	23.02	3.78	11.59	31.86	26.75	3.49	19.13	37.92
Education	14.31	2.55	10	18	14.12	2.95	7	18
DSST	62.14	10.93	40	89.5	44.05	9.56	16.5	72
MMSE	-	-	-	-	28.57	1.3	22	30

Note. SD: Standard deviation; BMI: Body Mass Index; DDST: Digit Symbol Substitution Test; MMSE: Mini Mental State Examination. Age and education is expressed in years.

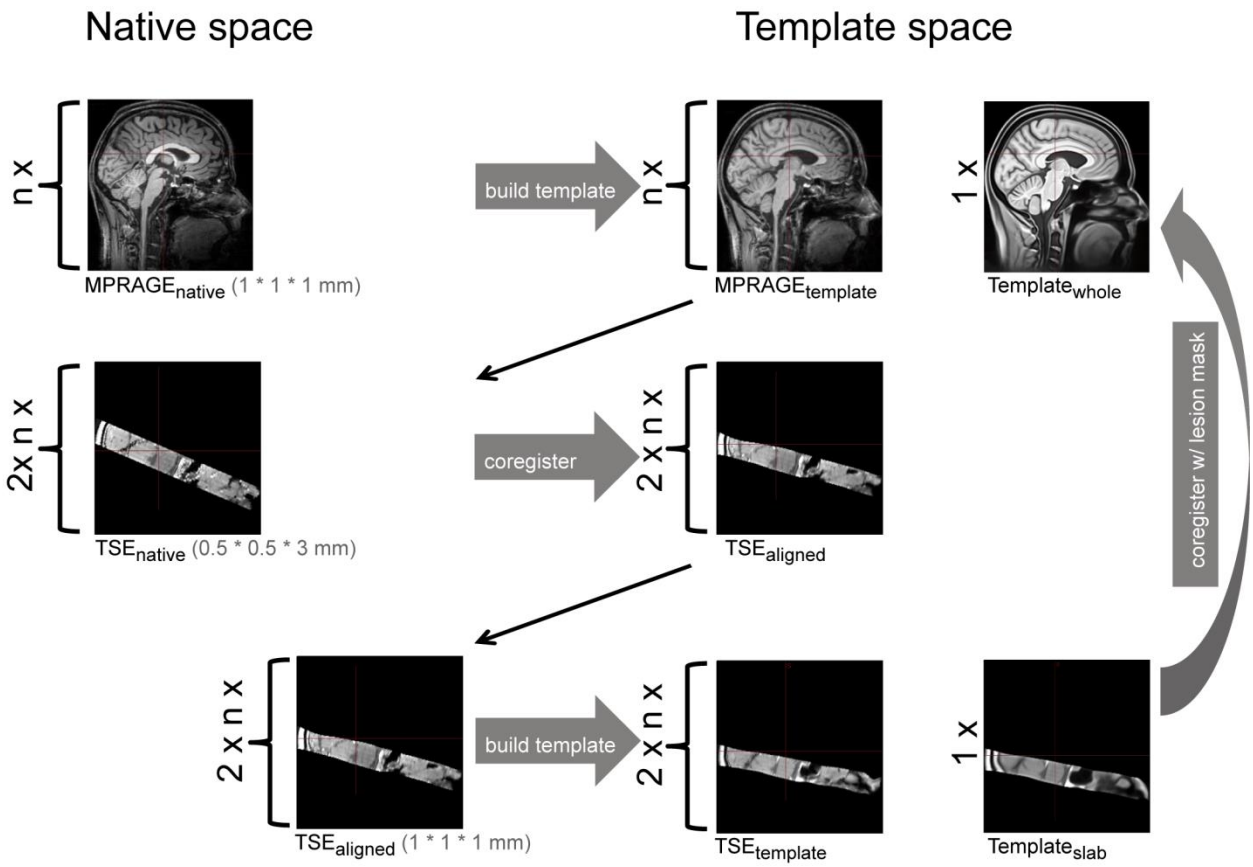


Figure S1. Overview of coregistration and template building steps used to generate a brainstem template. Step 1 (top row): Individual whole brain scans ($MPRAGE_{native}$) are aligned within a template building algorithm and used to generate a group whole brain template ($Template_{whole}$). Step 2 (middle row): Within subjects we coregister native space brainstem scans to whole brain scans ($MPRAGE_{template}$) to align scans across subjects while maintaining high coregistration accuracy. Step 3 (bottom row): Aligned brainstem scans ($TSE_{aligned}$) are used to generate a group brainstem template ($Template_{slab}$) with increased signal-to-noise ratio. Brainstem and whole brain templates are coregistered to facilitate transformation to standard space.

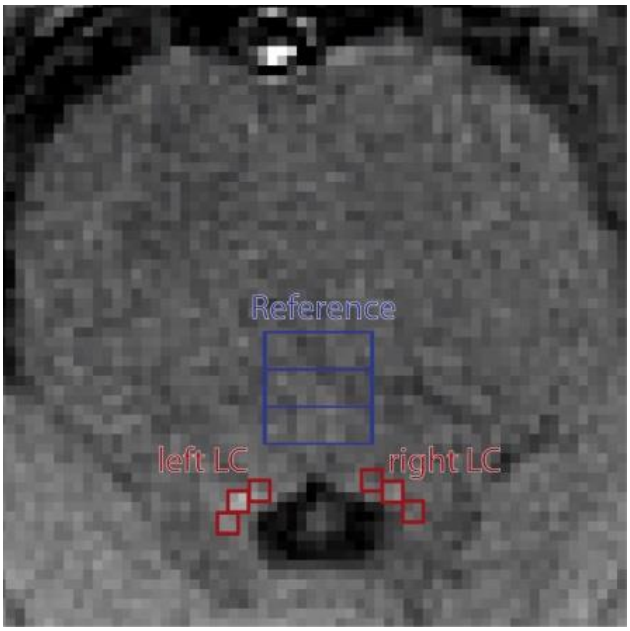


Figure S2. Depiction of the manual locus coeruleus (LC; red) and reference (blue) intensity assessment. Two independent, blinded raters place rectangular regions of interest (ROI) to capture LC-neuromelanin-related hyperintensities next to the fourth ventricle. Reference intensity is assessed in the dorsal pontine tegmentum between the LC ROI (see Online methods for a full description of the rating procedure).

Supplemental results

Cognitive results

Adequacy of the verbal learning and memory models

We captured the non-linear performance trajectories of younger and older adults' performance in the Rey Auditory Verbal Learning and Memory Task[1] (RAVLT) using a structural equation model. The proposed multiple-group quadratic growth curve model fits the data well both for time point 1 (T1) (χ^2 = 35.535, *df* = 46, RMSEA = 0.0, CFI = 1.01) and time point 2 (T2) (χ^2 = 81.764, *df* = 46, RMSEA = 0.052, CFI = 0.965 see Figure S3). Further, the quadratic growth model showed considerably better fit to the data than competing alternative models (i.e., intercept only, linear, hyperbolic[2], logarithmic[3], and exponential slope models) as confirmed by likelihood ratio difference tests (for nested models) and comparisons of the Akaike Information Criterion (AIC; for non-nested models; see Table S2). The quadratic growth curve models for T1 and T2 demonstrated strict factorial invariance (comparison of models with and without variant manifest errors (eV): all $\Delta\chi^2 \leq 2.727$; $\Delta df = 1$; all $p \geq 0.099$).

Table S2. Comparison of competing alternative growth models for time point 2

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	RMSEA	CFI	AIC	RL
Icept	1558.624	60	1476.861	14	<0.001	0.292	0.0	7407.831	--
Icept + Lin	405.838	54	324.075	8	<0.001	0.149	0.657	6267.045	--
Icept+Lin+Exp	140.126	46	--	--	--	0.084	0.908	6017.333	<0.001
Hyperbolic	90.052	46	--	--	--	0.057	0.957	5967.259	0.016
Icept+Lin+Log	82.168	46	--	--	--	0.052	0.965	5959.375	0.817
Icept+Lin+Quad	81.764	46	--	--	--	0.052	0.965	5958.97	--

Note: Icept = Intercept; Lin = Linear slope, Quad = Quadratic slope; Exp = Exponential slope; Log = Logarithmic slope; RMSEA = root mean square error of approximation; CFI =comparative fit index; AIC = Akaike Information Criterion; RL = Relative likelihood. Likelihood-ratio difference tests and relative likelihood tests compare the respective model to the Icept+Lin+Quad model.

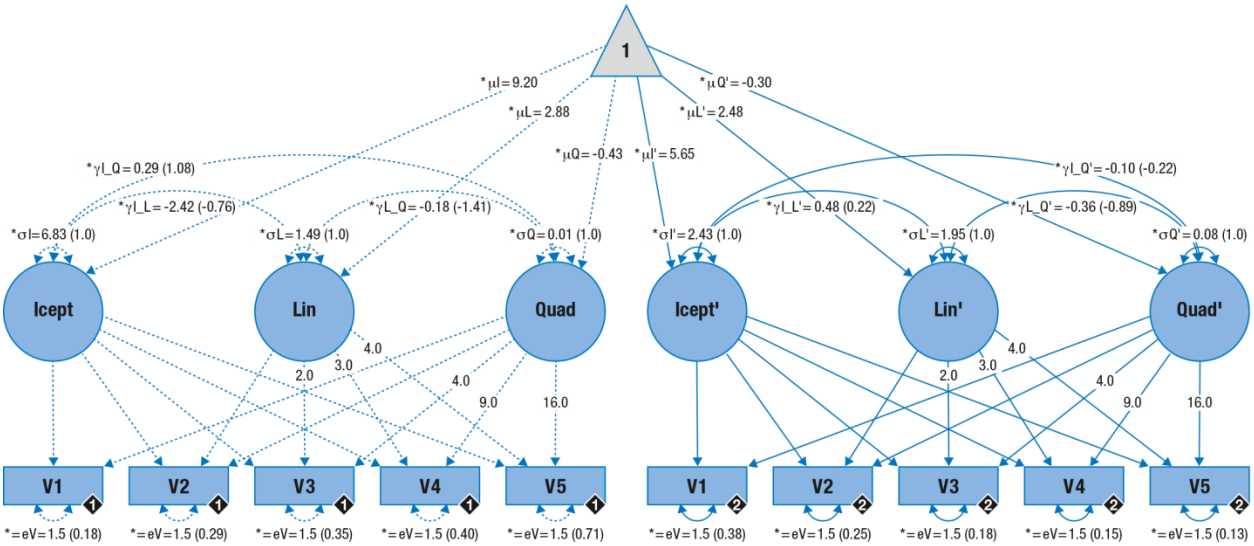


Figure S3. Graphical depiction of the structural equation model that approximates participants' learning curves with a quadratic function consisting of an initial memory performance level (intercept) and a gain over learning trials (linear and quadratic slope). Cognitive manifest variables represent the iteratively assessed memory performance in the verbal learning and memory task (V1–5). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). Icept = Intercept; Lin/Quad= linear / quadratic slope, respectively. Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle.

Verbal learning and memory within younger and older adults

For both time points and age groups, we observed reliable average intercept and slope factors (all $\Delta\chi^2 \geq 24.39$, $df = 1$, all $p < 0.001$). Further, there were consistent interindividual differences in initial recall (intercept) and learning (quadratic and linear slope) parameters. This applied to both age groups at both time points as indicated by significant variances of the latent factors (all $\Delta\chi^2 \geq 11.97$, $df = 1$, all $p < 0.001$). Only the variance of the quadratic slope factor failed to reach significance in younger adults at T2, indicating a similar negative acceleration of the growth curve across subjects ($\Delta\chi^2 = 0.276$, $df = 1$, $p = 0.599$).

The covariances between initial recall and learning rate (e.g., starting out higher is related to faster performance increases) demonstrated a mixed pattern. While for T1, there was no significant relation between initial recall memory and learning rate (i.e., the covariance between

intercept and quadratic/linear slope factors; all $\Delta\chi^2 \leq 1.603$, $df = 1$, all $p \geq 0.205$), there was a trend for an association at T2 (covariance intercept and quadratic slope in younger adults: $\Delta\chi^2 = 6.656$, $df = 1$, $p = 0.01$; in older adults: $\Delta\chi^2 = 2.593$, $df = 1$, $p = 0.107$. Covariance intercept and linear slope in younger adults: $\Delta\chi^2 = 19.076$, $df = 1$, $p < 0.001$; in older adults: $\Delta\chi^2 = 3.049$, $df = 1$, $p = 0.081$). Further, a steeper linear increase in memory performance was reliably associated with a more negative acceleration of growth at T1 for younger and older adults (i.e., the covariance between linear and quadratic slope factors; all $\Delta\chi^2 \geq 20.453$, $df = 1$, all $p \leq 0.001$) and marginally at T2 (for younger adults: $\Delta\chi^2 = 3.788$, $df = 1$, $p = 0.052$; for older adults: $\Delta\chi^2 = 30.484$, $df = 1$, $p < 0.001$).

Age differences in verbal learning and memory

At T1, likelihood ratio difference tests revealed significant age differences in the intercept factors ($\Delta\chi^2 = 45.391$, $df = 1$, $p < 0.001$), pointing to higher initial recall performance in younger adults. However, differences in the slope factors failed to reach significance (linear: $\Delta\chi^2 = 2.274$, $df = 1$, $p = 0.132$; quadratic: $\Delta\chi^2 = 1.819$, $df = 1$, $p = 0.177$). Hence, the rate of learning increases did not differ reliably between age groups. Also the covariances between the latent intercept and slope factors did not differ between groups (all $\Delta\chi^2 \leq 1.191$, $df = 1$, all $p \geq 0.275$).

For T2, likelihood ratio difference tests also indicated age differences in the intercept factors ($\Delta\chi^2 = 59.533$, $df = 1$, $p < 0.001$). Again, the linear slope factors did not differ between groups ($\Delta\chi^2 = 2.381$, $df = 1$, $p = 0.123$), however, the difference in the quadratic slope term turned out to be significant ($\Delta\chi^2 = 6.579$, $df = 1$, $p = 0.01$). This finding points to age differences in the negative acceleration of the learning curves. Note however, that ceiling effects may play a role here (see main text, Figure 1). The associations between initial recall and learning factors differed reliably among younger and older adults (covariance intercept and quadratic slope: $\Delta\chi^2 =$

9.454, $df = 1$, $p = 0.002$; covariance intercept and linear slope: $\Delta\chi^2 = 21.839$, $df = 1$, $p < 0.001$). However, there were no reliable age differences in the relation of the slope factors ($\Delta\chi^2 = 1.571$, $df = 1$, $p = 0.21$).

In sum, a quadratic growth model adequately described the observed learning curves of younger and older adults' performance in the RAVLT for both time points. Within groups, we observed reliable average initial recall (intercept) and learning (slope) factors. Individuals in both age groups differed in their initial recall level and the rate of learning. Age differences were mainly observed in initial recall performance.

Adequacy of the general episodic memory model

To obtain a single measure for general episodic memory performance, i.e., independent of the specific task used, we made use of the comprehensive cognitive battery available for this data set (see Online methods). In particular, using an established episodic memory factor structure[4] we repeated our analyses (T2 only), this time integrating performance across multiple memory tasks while explicitly excluding information from the RAVLT (see Figure S4). The model adequately fit the observed data ($\chi^2 = 1.991$, $df = 14$, RMSEA = 0.0, CFI = 1.312). However, the model only demonstrated weak factorial invariance (i.e., it required variant manifest intercepts across groups) and thus precluded an interpretation of age group differences in the means of latent factors[5,6] (comparison of models with and without variant manifest intercepts (e.g., λFPT): $\Delta\chi^2 = 19.799$; $\Delta df = 2$; $p < 0.001$). For both age groups, we observed reliable average episodic memory factors (all $\Delta\chi^2 \geq 134.486$; all $\Delta df = 1$; all $p < 0.001$) as well as interindividual differences therein (all $\Delta\chi^2 \geq 29.606$; all $\Delta df = 1$; all $p < 0.001$). In sum, the general episodic memory model adequately described the observed data. Within groups, we detected reliable latent factors and individual differences therein.

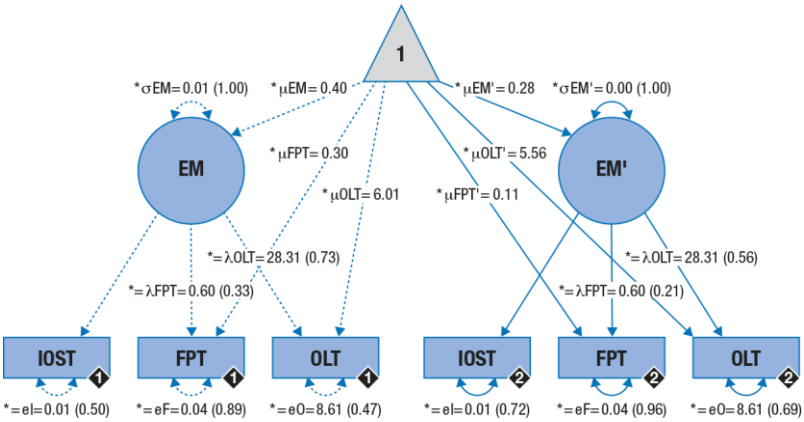


Figure S4. Graphical depiction of the structural equation model that captures participants' general episodic memory performance in a single latent factor. Cognitive manifest variables represent memory performance in the indoor/outdoor scene encoding task (IOST), face–profession task (FPT), and object–location task (OLT). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) and loadings (λ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle.

Magnetic resonance imaging results

Validity of automatically assessed locus coeruleus intensity ratios

Before the investigation of age differences in automatically assessed locus coeruleus (LC) intensity and their relation to learning and memory performance, we sought to establish the validity of the proposed approach. Unless otherwise stated, all following reports concern LC measures automatically assessed in template space ($TSE_{template}$; see Online methods).

Comparison to previously published locus coeruleus masks

We semi-automatically extracted individual peak intensities across the rostrocaudal extent of the LC. Peak intensity coordinates were converted to a probability map and warped into 0.5 mm iso-voxel standard space where we calculated the overlap with previously published LC maps[7,8]. The thresholded probability map contains 175 voxels equivalent to a volume of 26 mm^3. It includes two mostly symmetrical cylindrical structures corresponding to left and right LC that shift somewhat more laterally and posteriorly, following the wall of the fourth ventricle, along the rostrocaudal axis. The voxels of the proposed map overlap largely with the map published by Betts and colleagues[8] (57.71 %) as well as the one by Keren and others[7] (69.14 %; see main text, Figure 03 d; we used Keren's 1 SD LC map for comparisons). About half of the map lies within the intersection of all three maps (40 %; compared to 9.41 % and 8.53 % for Keren et al.[7] and Betts et al.[8], respectively). Thus, taken together, 86.86 % of the probability map is in line with previously published maps. However, our probability map extends less caudally compared to the other two, which explains why proportionally fewer segments of their maps fall within its boundaries (12.30 % and 16.26 % of Betts et al.[8] and Keren et al.[7], respectively). The two reference maps themselves overlap moderately (22.78 % of Betts et al. in Keren et al., 25.13 % vice versa[7,8]).

Comparison to manually assessed locus coeruleus intensity

Two independent, blinded raters manually assessed LC intensity on all axial slices that showed elevated signal in anatomically plausible LC locations. Average manual peak intensities showed high accordance between raters as demonstrated by means of intraclass correlation coefficients (ICC; left LC: ICC = 0.991, $p < 0.001$; right LC: ICC = 0.990, $p < 0.001$), so we collapsed values across raters. On average, elevated signal intensity was detected on 3.028 slices (*SD* = 0.122, range = 3–4). Manually and automatically assessed LC intensities were closely related (left LC: ICC = 0.855, $p < 0.001$; right LC: ICC = 0.814, $p < 0.001$; mean over left and right LC: ICC = 0.845, $p < 0.001$) highlighting the validity of the proposed method.

Reproducibility of semi-automatically assessed locus coeruleus intensity ratios

In order to judge the temporal stability of the semi-automatic method, we repeatedly over the course of several weeks scanned a small number of younger adults that did not participate in the main study. We applied the same coregistration and template building steps as in the main study (see Online methods). After alignment with the study brainstem template ($Template_{slab}$), we used the established search spaces to extract intensity ratios. Across six measurements, we observed high reproducibility of automatically assessed peak LC ratios (mean over left and right LC: ICC = 0.878, $p = 0.012$; left LC: ICC = 0.896, $p = 0.008$; right LC: ICC = 0.730, $p = 0.056$) compared to manual assessments[9]. In addition, these results demonstrate the applicability of the semi-automatic procedure for small samples.

Comparison to automatically assessed locus coeruleus intensity in native space

We warped the LC and reference search spaces to individual subject's coordinates (i.e., native space) and extracted intensity values for all participants. Values automatically assessed in template and native space demonstrated high accordance (mean over left and right LC: ICC =

0.906, $p < 0.001$; left LC: ICC = 0.909, $p < 0.001$; right LC: ICC = 0.896, $p < 0.001$) indicating the stability of the proposed method and its relative unaffectedness by the entailed warping and registration steps.

Age differences in the spatial distribution of locus coeruleus intensity ratios

A non-parametric cluster-based permutation test[10] (see Online methods) revealed spatially heterogeneous age differences in LC ratios along the rostrocaudal extent of the nucleus (see Figure S5a). Relative to younger adults, older adults showed a cluster of elevated intensity spanning caudal slices (left LC: 59th–100th LC percentile; $p_{corr} < 0.001$; right LC: 65th–94th LC percentile; $p_{corr} = 0.001$). In contrast, older adults demonstrated a tendency towards reduced intensity values in rostral LC segments that, however, did not reach significance in the unilateral analyses (left LC: 29th–35th LC percentile; $p_{corr} = 0.139$; right LC: 29th–35th LC percentile; $p_{corr} = 0.107$). In general, these hemisphere-specific analyses replicate the bilateral findings reported in the main text and indicate spatially confined age differences in LC intensity[11].

Previous studies reported lateralized differences in LC ratios (left > right; cf.[8,9]) that we replicate here by means of Wilcoxon signed rank tests. In particular, we computed tests (collapsing over slices) across age groups, and then for each group separately (younger and older adults: $W(294) = 9397$, $Z = -8.421$, $p < 0.001$; younger adults: $W(66) = 628$, $Z = -3.050$, $p = 0.002$; older adults: $W(228) = 5143$, $Z = -7.933$, $p < 0.001$).

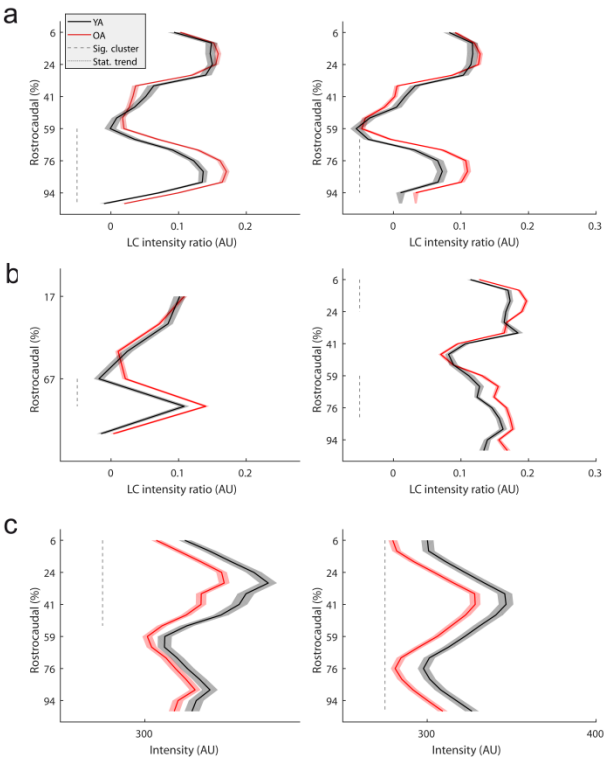


Figure S5. Analyses of age differences in locus coeruleus (LC) intensity ratios (a–b) and raw intensity (c) across the rostrocaudal axis by means of cluster-based permutation tests (see Online methods).(a) Peak intensity ratios automatically assessed in template space were analyzed separately for the left and right hemisphere (left, right, respectively). (b) Comparisons based on peak native space (left plot) and mean template space intensity (right plot) LC measures. (c) Analysis of peak raw intensity values split for LC and reference (left, right, respectively). Shaded areas represent ± 1 standard error of the mean.

Estimation of latent locus coeruleus integrity scores

Adequacy of the latent locus coeruleus integrity model

We used a multiple-group single factor SEM to estimate overall LC integrity scores on a latent level while accounting for measurement error in the observed variables (see Figure S6). For each age group, peak LC ratios (averaged across slices; see Online methods) of the left and right hemispheres loaded on a single LC factor. Before model estimation, LC ratios were scaled (multiplied × 100) to facilitate model estimation. In the model, error variances, manifest means, and factor loadings were constrained to be equal across age groups (i.e., strict factor invariance; comparison of models with and without variant manifest errors, means, and factor loadings: $\Delta\chi^2$

= 1.791; $\Delta df = 3$; $p = 0.617$). The proposed model fit the data well ($\chi^2 = 1.791$, $df = 7$, RMSEA = 0.000, CFI = 1.222).

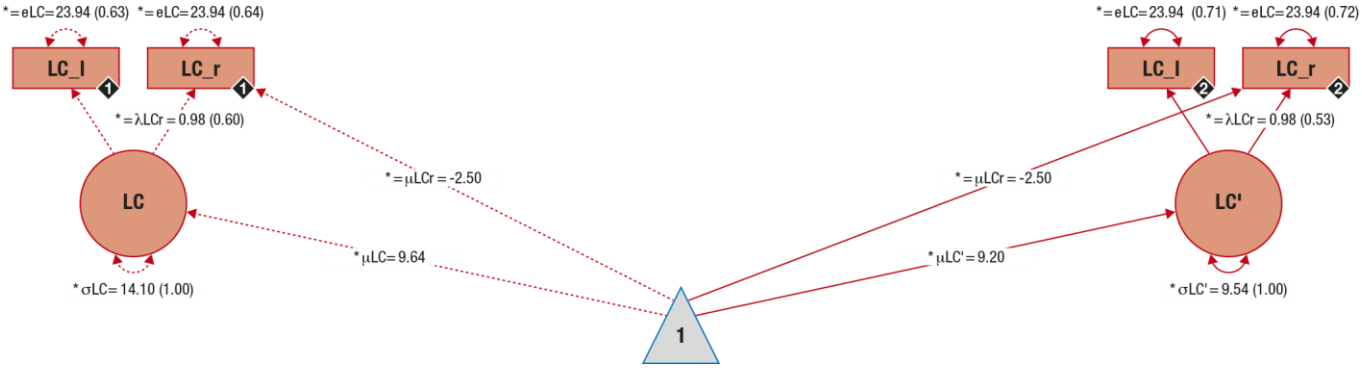


Figure S6. Graphical depiction of the structural equation model that integrates over locus coeruleus intensity data to estimate a single locus coeruleus integrity factor (LC; red) on a latent level for younger and older adults. Neural manifest variables are the LC intensity ratios of each hemisphere (left = LC_l; right = LC_r). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) and loadings (λ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle.

Latent locus coeruleus integrity scores within younger and older adults

We detected reliable average latent LC factors for both age groups (all $\Delta\chi^2 \geq 90.454$, $\Delta df = 1$, all $p < 0.001$), and within each group, subjects showed significant interindividual differences in LC scores (all $\Delta\chi^2 \geq 18.305$, $\Delta df = 1$, all $p < 0.001$).

Age differences in latent locus coeruleus integrity scores

Younger and older adults demonstrated comparable average LC scores ($\Delta\chi^2 = 0.383$, $\Delta df = 1$, $p = 0.536$) in line with spatially confined age differences in LC integrity[7,8]. In sum, a multiple-group, single factor model adequately captures the interindividual differences in LC intensity ratios in the data. Within each age group, we detected reliable latent LC integrity factors as well as interindividual differences within them.

Combined cognitive and magnetic resonance imaging results

Associations between latent locus coeruleus integrity scores and memory performance

After establishing valid models for memory performance (see Figures S3 and S4) and for LC integrity (see Figure S6) in isolation, we sought to link the two modalities.

We first investigated the association between LC integrity and memory performance as assessed by the RAVLT. Next, to determine whether LC integrity is associated with episodic memory performance more generally, i.e., independent of the specific task used, we exploited the comprehensive cognitive battery available for this data set (see Online methods). Specifically, we repeated our analyses, using an established episodic memory factor structure[4] this time integrating performance across multiple memory tasks while explicitly excluding information of the RAVLT.

Model adequacy

We merged both modalities in unified neuro-cognitive models that demonstrated good fit for both of the cognitive measures used (i.e., RAVLT and general episodic memory; cf. Table S3; see also main text, Figure 2, and Figure S7).

Table S3. Fit statistics for unified neural and cognitive models

Model		χ^2	*df*	RMSEA	CFI
LC–RAVLT	(Time point 1)	47.41	87	0.0	1.037
LC–RAVLT	(Time point 2)	101.942	87	0.024	0.986
LC–EM	(Time point 2)	14.727	43	0.0	1.515

Note: LC: Locus Coeruleus; RAVLT: Rey Auditory Verbal Learning Task; EM: General Episodic Memory; df: Degrees of freedom; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index.

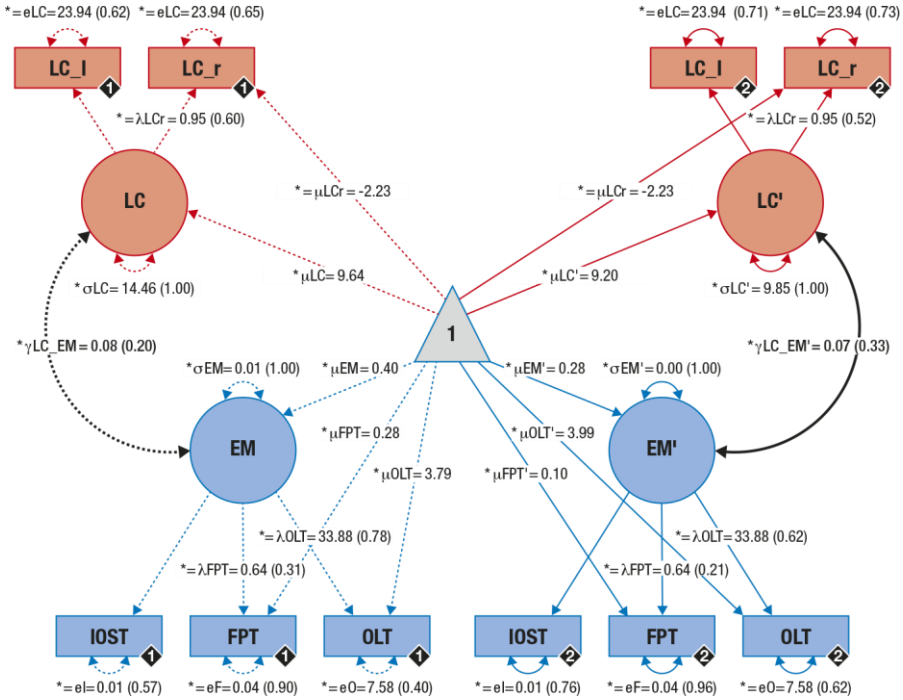


Figure S7. Graphical depiction of the structural equation model that probes associations (black lines) between locus coeruleus integrity (LC; red) and general memory performance (blue) in younger and older adults on a latent level. Cognitive manifest variables represent the memory performance across three memory tasks (Indoor/Outdoor Scene Task (IOST); Face–Profession Task (FPT); Object–Location Task (OLT)). Neural manifest variables are the LC intensity ratios of each hemisphere (left = LC_l; right = LC_r). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) and loadings (λ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). EM = General Episodic Memory. Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle.

Latent locus coeruleus integrity scores and memory performance within younger and older adults

We observed a positive association between LC integrity and memory performance in older adults, irrespective of the cognitive measure used (i.e., RAVLT or general episodic memory) and time point analyzed (i.e., T1 or T2).

In particular, at T1, older adults with higher LC integrity (assessed at T2) demonstrated higher initial recall performance ($\Delta\chi^2 = 3.993$, $\Delta df = 1$, $p = 0.046$; standardized estimate: 0.25) and steeper learning curves (linear slope: $\Delta\chi^2 = 5.589$, $\Delta df = 1$, $p = 0.018$; standardized estimate: 0.3; quadratic slope: $\Delta\chi^2 = 5.612$, $\Delta df = 1$, $p = 0.018$; standardized estimate: –0.34) in the RAVLT

(see Figure S8, left). While we did not observe a reliable association between LC integrity and memory performance in younger adults[12] (all $\Delta\chi^2 \leq 2.691$, $\Delta df = 1$, all $p \geq 0.101$), the lack of reliable age group differences in the LC–memory association (all $\Delta\chi^2 \leq 0.417$, $\Delta df = 1$, all $p \geq 0.519$) suggests comparable relations independent of age.

For findings concerning the association between LC integrity and memory performance as assessed by the RAVLT at T2, please refer to the main text and Figure S8 (middle).

Relating LC integrity to general episodic memory scores, we again observed a positive association in older but not younger adults while there were no reliable differences in this association between groups (see Figure S9; older adults: $\Delta\chi^2 = 5.108$, $\Delta df = 1$, $p = 0.024$, standardized estimate: 0.33; younger adults: $\Delta\chi^2 = 0.969$, $\Delta df = 1$, $p = 0.325$[12]; age differences: $\Delta\chi^2 = 0.01$, $\Delta df = 1$, $p = 0.919$).

Together, these additional analyses corroborate our finding that interindividual differences in learning and memory are positively related to LC integrity in older adults (see main text) and, beyond that, suggest a stable and lasting general association.

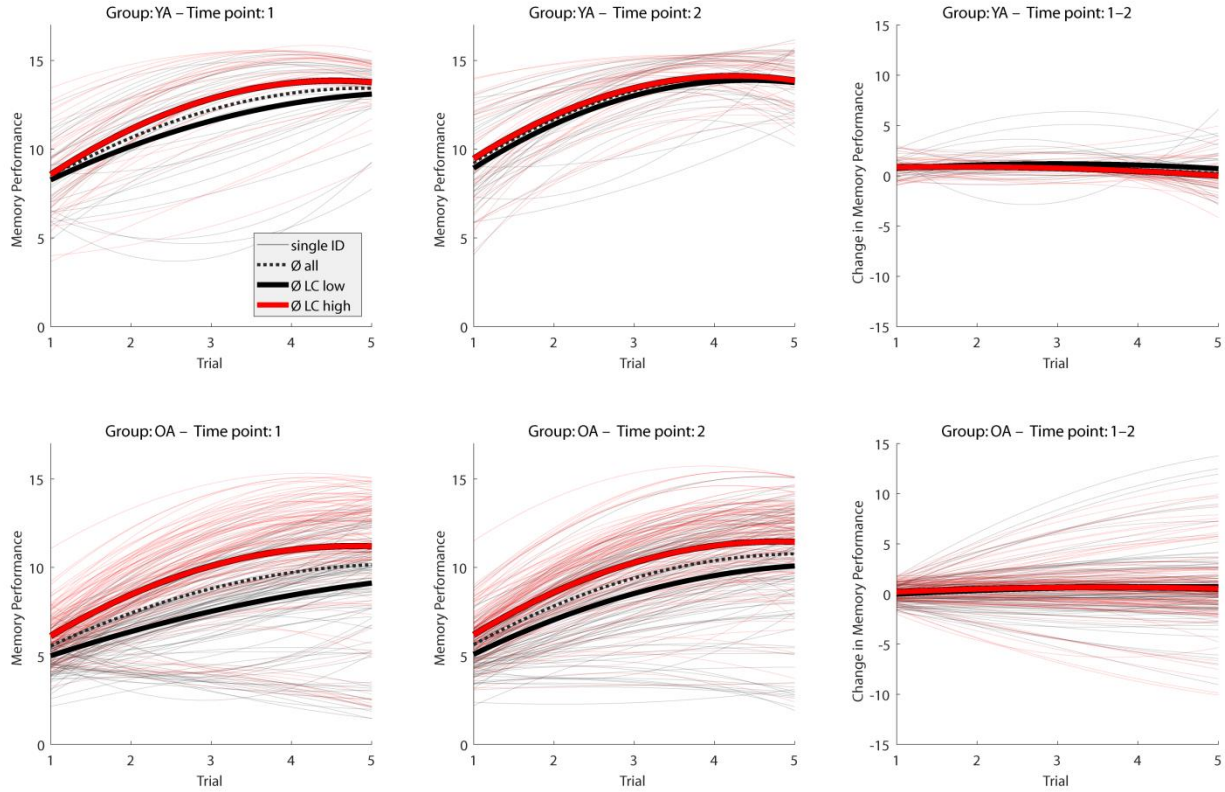


*Figure S8.*Estimated learning and memory performance trajectories (RAVLT) for younger (YA) and older (OA) adults (upper and lower plots, respectively) for time points 1 and 2 (left and middle, respectively) and change across time points (right). For visualization of the association between locus coeruleus (LC) integrity and memory performance, single subjects (ID, thin lines) are color-coded based on LC integrity (median-split) and mean trajectories for subgroups are included. Here,we used LC values automatically assessed in template space (see Online methods).

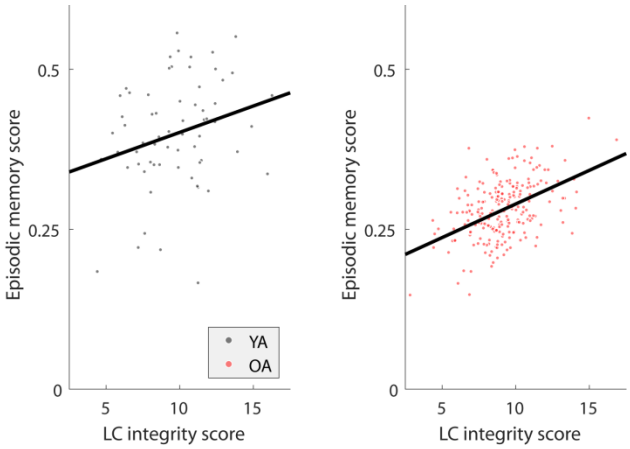


Figure S9. Association of locus coeruleus integrity and general episodic memory performance for younger (YA; black; left; standardized estimate = 0.2; p = 0.325) and older adults (OA; red; right; standardized estimate = 0.33; p = 0.024) based on latent estimates (see Figure S7).

Associations between latent locus coeruleus integrity scores and change in memory performance

Model adequacy

The combined neural and cognitive models of each time point were then merged in a univariate latent change score model[13] (see Figure S10). Here we only report change analyses for verbal learning and memory data (RAVLT). In doing so, we attempted to answer the question whether intraindividual (longitudinal) change in memory is related to LC scores. The model adequately represents the observed covariance matrix ($\chi^2 = 304.414$, $df = 268$, RMSEA = 0.022; CFI = 0.983).

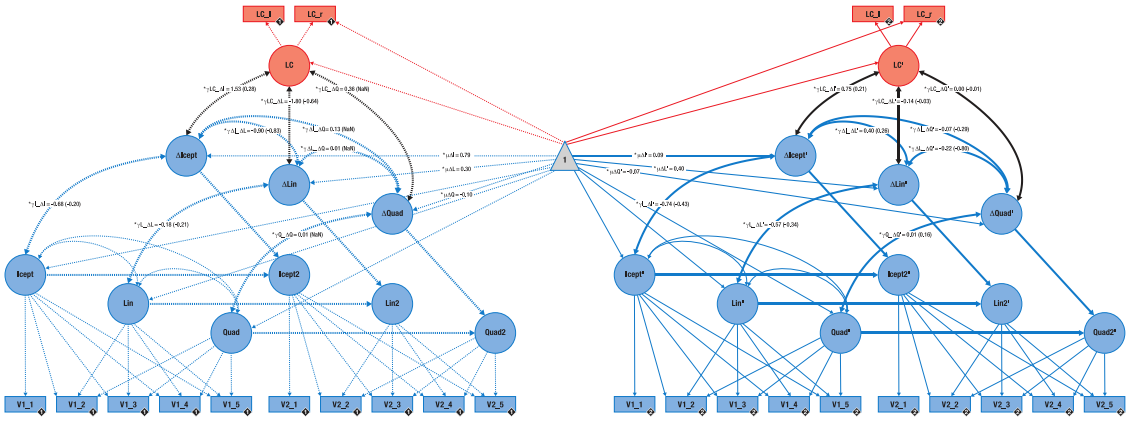


Figure S10. Graphical depiction of the structural equation model that probes associations (black lines) between locus coeruleus integrity (LC; red) and change in memory performance (blue) in younger and older adults on a latent level. Cognitive manifest variables represent the iteratively assessed memory performance in the verbal learning and memory task (V1–5) for time point 1 (V1_*) and time point 2 (V2_*). Neural manifest variables are the LC intensity ratios of each hemisphere (left = LC_l; right = LC_r). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) and loadings (λ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). Icept = Intercept; Lin/Quad= linear/quadratic slope, respectively. Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle. Variances on manifest and latent variables are omitted to facilitate readability.

Change in memory performance within younger and older adults

In younger adults, we detected reliable average change in initial recall performance (intercept) over time (T1 → T2) as well as significant interindividual differences therein (mean Δ-intercept: $\Delta\chi^2 = 7.171$, $\Delta df = 1$, $p = 0.007$; variance Δ-intercept: $\Delta\chi^2 = 11.986$, $\Delta df = 1$, $p \leq 0.001$; see Figure S8, right plots). Further, there was a trend towards change in the quadratic growth factor on the group level (mean Δ-quadratic slope: $\Delta\chi^2 = 3.13$, $\Delta df = 1$, $p = 0.077$) but no reliable interindividual differences therein ($\Delta\chi^2 = 2.041$, $\Delta df = 1$, $p = 0.153$). We did not find any mean changes in the linear slope factor ($\Delta\chi^2 = 1.164$, $\Delta df = 1$, $p = 0.281$) or interindividual differences in change ($\Delta\chi^2 = 1.068$, $\Delta\, df = 1$, $p = 0.301$).

In older adults, we observed no reliable intraindividual change (T1 → T2) in initial recall performance ($\Delta\chi^2 = 0.409$, $\Delta df = 1$, $p = 0.522$) on a group level but substantial interindividual differences in the amount of change ($\Delta\chi^2 = 17.541$, $\Delta df = 1$, $p < 0.001$). On average, older adults showed a trend towards intraindividual change in the quadratic slope factor ($\Delta\chi^2 = 3.255$, $\Delta\, df = 1$, $p = 0.071$) as well as reliable interindividual differences therein ($\Delta\chi^2 = 5.131$, $\Delta\, df = 1$, $p = 0.024$). Finally, older adults demonstrated reliable intraindividual change in linear growth and differed reliably in the amount of change they showed ($\Delta\chi^2 = 5.772$, $\Delta\, df = 1$, $p = 0.016$ and $\Delta\chi^2 = 31.277$, $\Delta\, df = 1$, $p < 0.001$, respectively).

In sum, we detected reliable intraindividual change in initial recall in younger and linear slope in older adults, indicating slightly better memory performance at the T2 in both age groups. Within the group of younger and older adults, we observed significant interindividual differences in the amount of change in initial recall and learning.

However, as indicated by Figure S8 (right), performance changes across the ~ two year interval between assessment waves tended to be negligible.

Associations between latent locus coeruleus integrity scores and change in memory performance

We did not detect associations between LC integrity and change in memory performance in any of the age groups (all $\Delta\chi^2 \leq 2.518$, $\Delta df = 1$, all $p \geq 0.113$). Accordingly, we do not report age group differences in these associations.

Association between age differences in the spatial distribution of locus coeruleus intensity ratios and memory performance

To evaluate the functional significance of the observed topographical age differences (see main text, Figure 7), we related memory performance to LC ratios for each identified cluster. In the main text, we report analyses across all participants relying on memory performance as assessed by the RAVLT. Note that the general episodic memory model, that integrated performance over multiple memory tasks, only demonstrated weak factorial invariance (i.e., requires variant manifest intercepts across groups) and thus precluded an interpretation of age group differences in the means of latent factors[5,6]. We thus only report the association between topographical age differences and initial recall performance.

Figure S11 shows LC ratios for younger and older adults across the rostrocaudal axis, split by performance (median split based on initial recall performance within each age group). High and low performing older adults (light and dark red lines) demonstrated differential intensity ratios in rostral LC segments that largely overlapped with the cluster identified by the age comparisons (see above for cluster results). In contrast, LC ratios of high and low performing older adults largely overlapped in caudal segments.

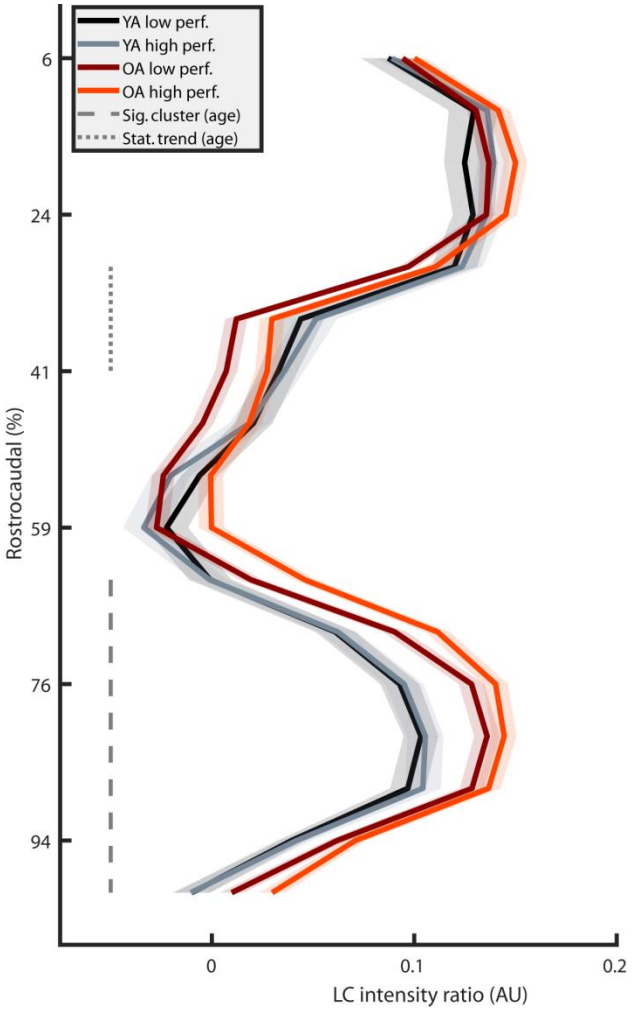


Figure S11. Cluster-based permutation tests revealed spatially confined age differences in locus coeruleus (LC) intensity across the rostrocaudal axis. While older adults (OA) show significantly higher ratios in caudal segments, there is a trend for a reversed effect in rostral segments. Age groups are split by performance (median split based on initial recall performance within each age group). High and low performing older adults (light and dark red lines) demonstrated differential intensity ratios in rostral LC segments that overlapped largely with the cluster identified by the age comparisons (see broken lines). In contrast, LC ratios of high and low performing older adults largely overlapped in caudal segments. Shading represents ± 1 standard error of the mean.

Additional magnetic resonance imaging results

In the following, we briefly describe replications of key analyses reported above or in the main text. However, we use different LC measures here (i.e., LC values automatically extracted from native space) in order to evaluate the stability of the observed findings.

Age differences in the spatial distribution of locus coeruleus intensity ratios

The semi-automatic procedure that we used to extract LC intensity ratios incorporates several coregistration and transformation steps (see Online methods) which might influence image intensity and thereby constitute a confounding factor. To exclude this possibility, we first warped the LC and reference search spaces from template space back to individual subjects' coordinates (i.e., native space) and extracted intensity values for all participants. We then calculated non-parametric cluster permutation tests[10] (see Online methods) to test for age differences in LC ratios (mean over left and right LC) across the rostrocaudal extent of the nucleus. The obtained native space results largely replicate the pattern observed in template space (see Figure S5b, left plot): Older adults show significantly higher ratios in caudal segments (67th–83rd LC percentile; $p_{corr} = 0.002$), in line with the accumulation of neuromelanin across the life span[14]. In contrast, numerically, there was a reversed pattern in rostral, hippocampus-projecting segments[15]. Thus, the reported topographic analyses seem to be largely independent of transformation steps.

Studies using neuromelanin-sensitive MRI as an index for LC integrity typically reference LC signal to dorsal pontine intensity to allow for comparisons across participants[16]. However, different measures have previously been used as a basis to calculate the intensity ratio (i.e., mean, median, or max (peak) intensity)[7,8,16]. Figure S5b (right plot) replicates the topographical analyses of age differences using mean LC and reference intensity. Compared to

younger adults, older adults demonstrate elevated intensity ratios in caudal and some rostral slices (59th–82nd LC percentile; $p_{corr} = 0.008$; 6th–24th LC percentile; $p_{corr} = 0.017$) in line with age-related neuromelanin accumulation[14]. However, as observed before (see Figure S5a), this effect is numerically reversed in the second rostral LC quartile. Thus, the mean intensity analyses largely replicated the peak intensity findings with some variations in cluster extent. Together, they point to a reduced integrity of more rostral, hippocampus-projecting LC segments[15]. Note that the mean-based LC measures produce larger (and non-negative) intensity ratios overall. Since both Keren and colleagues (2009)[7] and Betts and others (2017)[8], the two previous studies that reported local age differences in LC integrity, used peak intensity we also adopt this measure for the sake of comparability.

Finally, since intensity ratios as used here may obscure whether differences in the region of interest (LC) or reference (pons) produce the age effect, we separately assessed raw peak LC and reference intensity (see Figure S5c). While age differences were uniformly distributed across the rostrocaudal extent ($p_{corr} = 0.002$) in the reference area (right plot), age differences were confined to the rostral segments (6th–53rd LC percentile; $p_{corr} = 0.019$; left plot) in the LC. This underscores our interpretation of reduced rostral LC integrity in late life[11].

Additional combined cognitive and magnetic resonance imaging results

In the following, we briefly report replications of key analyses described above/in the main text. However, we use different LC measures here (i.e., LC values manually or automatically extracted from native space) as the basis and test their association with learning and memory (i.e., performance in the RAVLT at T2) in older adults.

Model adequacy

As described above (see Online methods), we applied SEM to estimate LC integrity scores and their relation to learning and memory. Here, however, LC integrity was calculated based on manually assessed values and then on the basis of values automatically extracted from native space (see Figures S12 and S13, respectively). Both models demonstrate good fit for both time points (all $\chi^2 \leq 103.297$, all $df = 87$, all RMSEA ≤ 0.025, all CFI ≥ 0.986).

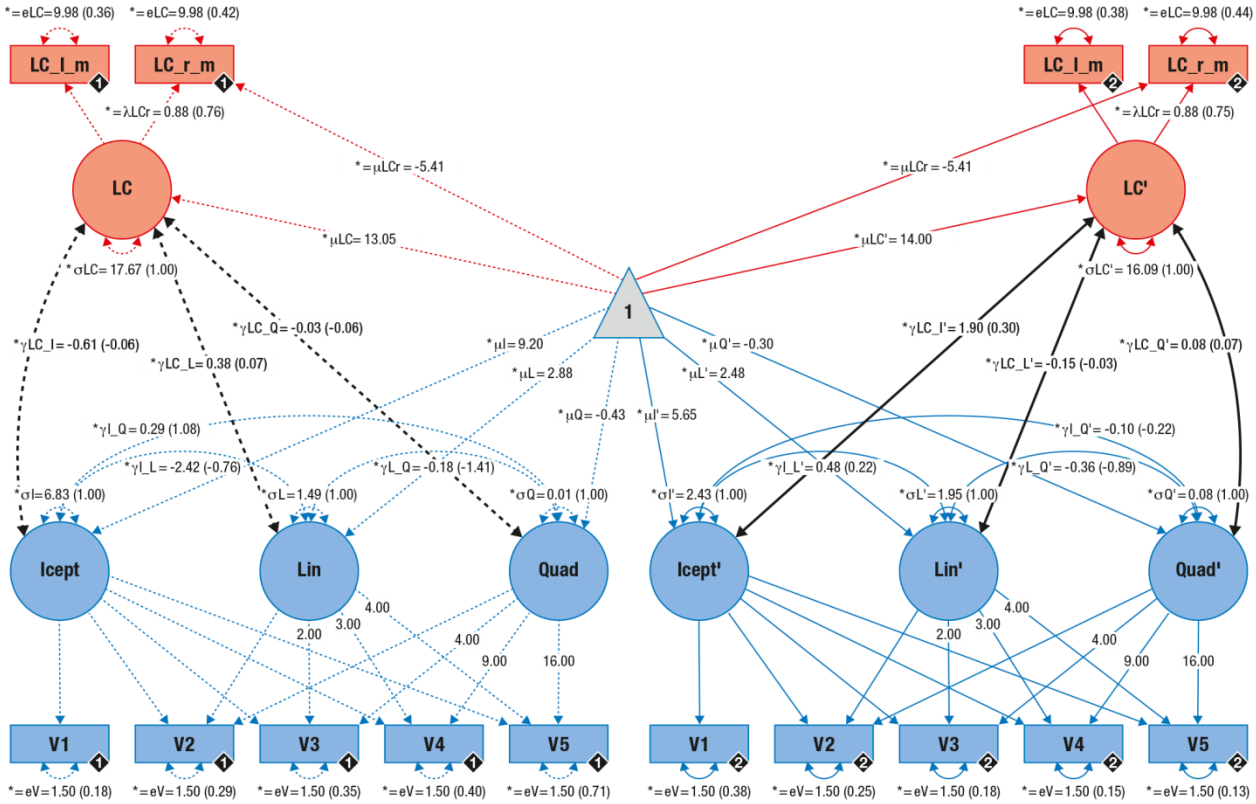


Figure S12. Graphical depiction of the structural equation model that probes associations (black lines) between locus coeruleus integrity (LC; red) and learning and memory performance (blue) in younger and older adults on a latent level. Cognitive manifest variables represent the iteratively assessed memory performance in the verbal learning and memory task (V1–5). Neural manifest variables are the manually assessed LC intensity ratios of each hemisphere (left = LC_l, right = LC_r; see Online methods). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) and loadings (λ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). Icept = Intercept; Lin/Quad= linear/quadratic slope, respectively. Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle.

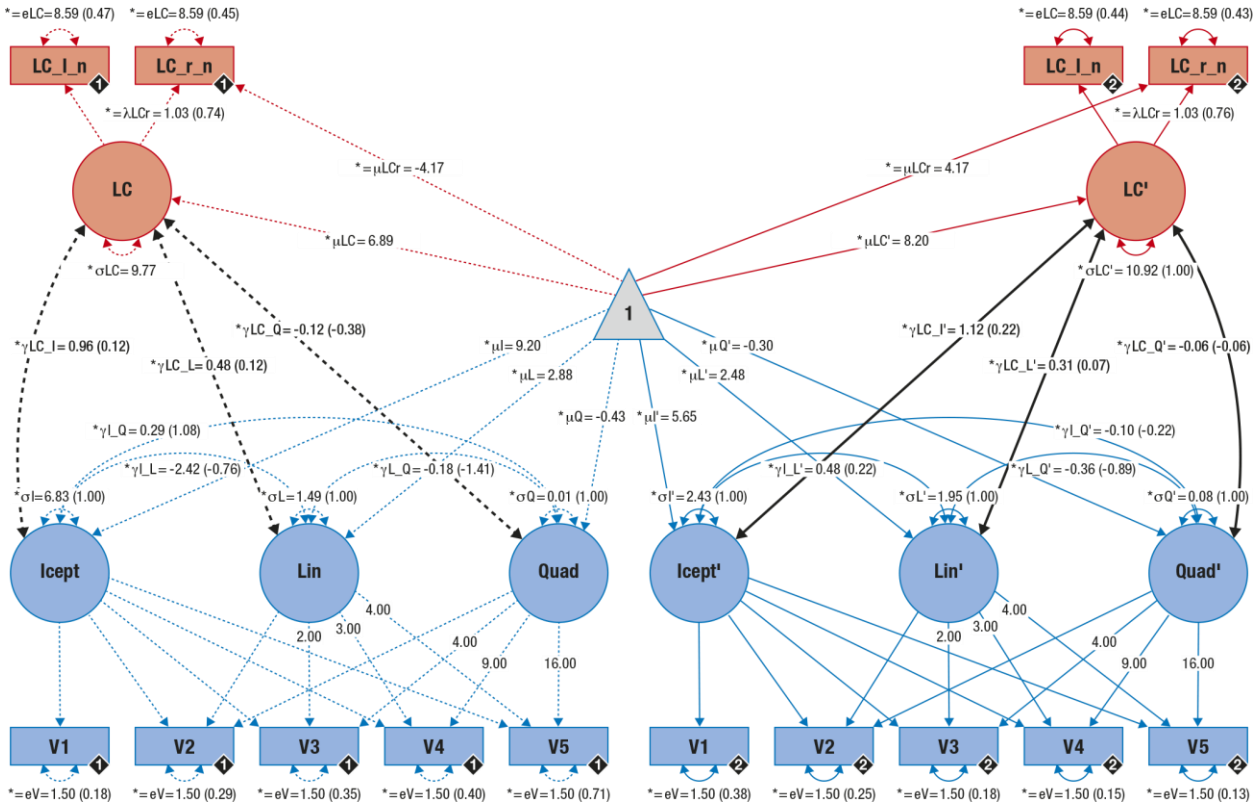


Figure S13. Graphical depiction of the structural equation model that probes associations (black lines) between locus coeruleus integrity (LC; red) and learning and memory performance (blue) in younger and older adults on a latent level. Cognitive manifest variables represent the iteratively assessed memory performance in the verbal learning and memory task (V1–5). Neural manifest variables are the LC intensity ratios of each hemisphere (left = LC_l; right = LC_r), automatically assessed in native space (see Online methods). Black diamonds on manifest variables indicate the age group (younger adults = 1, broken lines; older adults = 2, solid lines). (Co)Variances (γ, σ) and loadings (λ) in brackets indicate standardized estimates. Loadings that are freely estimated (*) but constrained to be equal across age groups (=) are indicated by both asterisk and equal signs (*=). Icept = Intercept; Lin/Quad= linear/quadratic slope, respectively. Rectangles and circles indicate manifest and latent variables, respectively. The constant is depicted by a triangle.

Latent locus coeruleus integrity scores and memory performance within older adults

Manually segmented LC values were related to better memory performance in older adults (see Figure S14, left). In particular, as observed before (see main results), the initial recall (intercept) factor was related to LC integrity (intercept: $\Delta\chi^2 = 10.282$, $\Delta df = 1$, $p = 0.001$, standardized estimate: 0.3).

Further, LC integrity values automatically assessed from native space showed a positive association with memory performance (see Figure S14, right). Again, higher LC integrity was

related to a higher initial recall performance (intercept; $\Delta\chi^2 = 5.135$, $\Delta df = 1$, $p = 0.024$; standardized estimate: 0.22). Taken together, these supplementary findings corroborate the positive link between LC integrity and memory in older adults across multiple analysis methods.

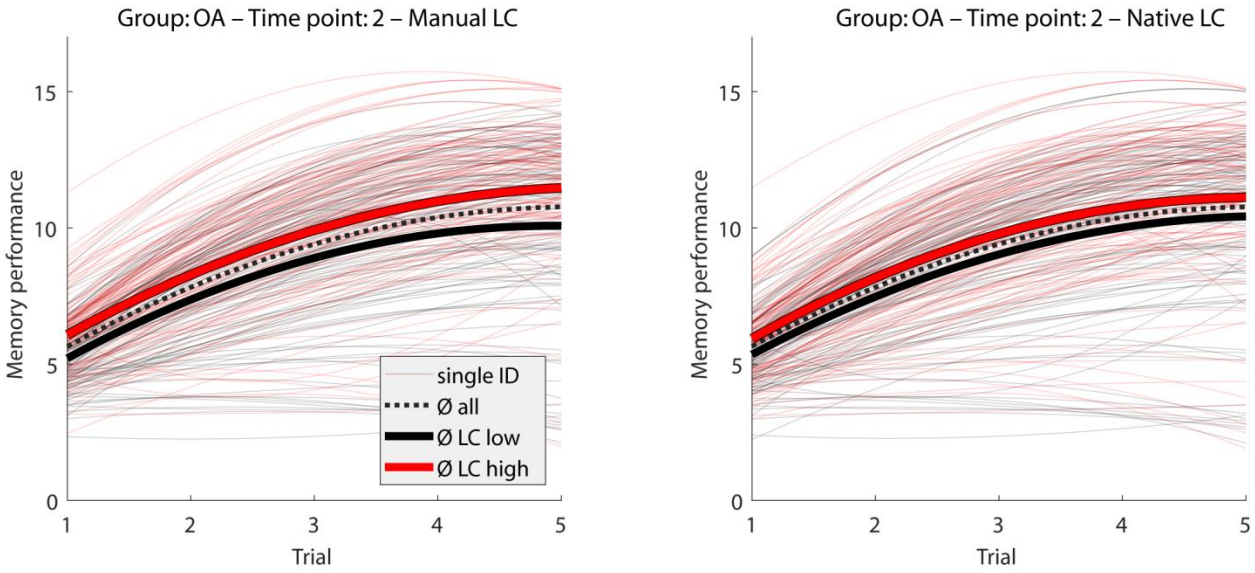


Figure S14. Estimated learning and memory performance trajectories (RAVLT) for older adults (OA) for time point 2. For visualization of the association between locus coeruleus (LC) integrity and memory performance, single subjects (ID; thin lines) are color-coded based on LC integrity (median split) and mean trajectories for subgroups are included. Left: LC integrity is calculated based on manually assessed intensity values. Right: LC integrity is estimated on the basis of automatically assessed values from native space (see Online methods).

Supplemental discussion

We applied an iterative learning and memory task (RAVLT)[1] that required subjects to encode, consolidate, and retrieve verbal information and thus captures memory in its dynamic nature[2]. The non-linear performance trajectories of younger and older adults were modeled by a quadratic growth function[17] that provides a good tradeoff between parsimony and model adequacy. We used a SEM approach that structures individual differences in initial recall performance and learning. The proposed model closely approximated the observed data (i.e., excellent fit). Our analyses allowed the study of two psychologically distinct factors, namely, initial recall (i.e., performance after the first learning trial, corresponding to a standard one-trial memory assessment) and learning (i.e., changes in performance with practice). In line with previous research[3,18–20], we observed reduced initial recall performance in older adults. Age differences in the rate of learning were less consistent than in some previous reports[19], likely due to a different statistical conceptualization of the parameter (i.e., a linear and quadratic slope factor vs. a single hyperbolic factor). However, a polynomial decomposition more accurately captured the observed performance trajectories in the present data set (see *Supplemental results*).

Exploiting the longitudinal nature of this data set, we evaluated performance changes over the ~ two year interval between assessment waves. However, changes in initial recall and learning factors tended to be negligible. This indicates a relative stability of memory, even in older adults. Using a comparable paradigm and population, a similar pattern emerged in the Zürich Longitudinal Study on Cognitive Aging over a delay of 1.5 years[19]. The authors mainly attributed the high stability of memory performance to retest effects resulting from a relatively short follow-up interval (see also[21], Figure 1 therein referring to[22]). Here, retest effects may even

be exacerbated since identical verbal material was used at both time points. Accordingly, we mainly restricted our analyses to the neural basis of learning and memory within time points.

References

1. Schmidt, M. *Rey Auditory Verbal Learning Test: A handbook.* (Los Angeles: Western Psychological Services, 2004).

2. Zimprich, D., Rast, P. & Martin, M. Individual differences in verbal learning in old age. in *Handbook of cognitive aging: Interdisciplinary perspectives* (eds. Hofer, S. & Alwin, D.) 224–243 (Thousand Oaks: SAGE Publications, 2008). doi:10.4135/9781412976589.n14

3. Jones, R. N. *et al.* A growth curve model of learning acquisition among cognitively normal older adults. *Exp. Aging Res.* **31,** 291–312 (2005).

4. Düzel, S. *et al.* Supplementary material for: The Subjective Health Horizon Questionnaire (SHH-Q): Assessing future time perspectives for facets of an active lifestyle. *Gerontology* **62,** 345–353 (2016).

5. Meredith, W. & Teresi, J. A. An essay on measurement and factorial invariance. *Med. Care* **44,** S69–S77 (2006).

6. Schwab, S. & Helm, C. Überprüfung von Messinvarianz mittels CFA und DIF-Analysen. *Empirische Sonderpädagogik* **4845,** 175–193 (2015).

7. Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S. & Eckert, M. A. In vivo mapping of the human locus coeruleus. *NeuroImage* **47,** 1261–1267 (2009).

8. Betts, M. J., Cardenas-Blanco, A., Kanowski, M., Jessen, F. & Düzel, E. In vivo MRI assessment of the human locus coeruleus along its rostrocaudal extent in young and older adults. *NeuroImage* **163,** 150–159 (2017).

9. Tona, K. D. *et al.* In vivo visualization of the locus coeruleus in humans: Quantifying the test–retest reliability. *Brain Struct. Funct.* **222,** 4203–4217 (2017).

10. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J.*

Neurosci. Methods **164,** 177–190 (2007).

11. Manaye, K. F., McIntire, D. D., Mann, D. M. A. & German, D. C. Locus coeruleus cell loss in the aging human brain: A non-random process. *J. Comp. Neurol.* **358,** 79–87 (1995).
12. Hämmerer, D. *et al.* Locus coeruleus integrity in old age is selectively related to memories linked with salient negative events. *Proc. Natl. Acad. Sci. U. S. A.* **115,** 2228–2233 (2018).
13. Kievit, R. A. *et al.* Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Dev. Cogn. Neurosci.*Advance online publication. (2017). doi:10.1016/j.dcn.2017.11.007
14. Mann, D. M. A. & Yates, P. O. Lipoprotein pigments–their relationship to ageing in the human nervous system. II. The melanin content of pigmented nerve cells. *Brain* **97,** 489–498 (1974).
15. Schwarz, L. A. & Luo, L. Organization of the locus coeruleus-norepinephrine system. *Curr. Biol.* **25,** R1051–R1056 (2015).
16. Liu, K. Y. *et al.* Magnetic resonance imaging of the human locus coeruleus: A systematic review. *Neurosci. Biobehav. Rev.* **83,** 325–355 (2017).
17. McArdle, J. J. Dynamic but structural equation modeling of repeated measures data. in *Handbook of multivariate experimental psychology* (eds. Nesselroade, J. R. & Cattell, R. B.) 561–614 (Boston: Springer, 1988). doi:10.1007/978-1-4613-0893-5_17
18. Zhang, Z., Davis, H. P., Salthouse, T. A. & Tucker-Drob, E. M. Correlates of individual, and age-related, differences in short-term learning. *Learn. Individ. Differ.* **17,** 231–240 (2007).
19. Zimprich, D. & Rast, P. Verbal learning changes in older adults across 18 months. *Neuropsychol. Dev. Cogn. B. Aging. Neuropsychol. Cogn.* **16,** 461–484 (2009).

20. Poreh, A. Analysis of mean learning of normal participants on the Rey Auditory-Verbal Learning Test. *Psychol. Assess.* **17,** 191–199 (2005).

21. Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U. & Bäckman, L. Memory aging and brain maintenance. *Trends Cogn. Sci.* **16,** 292–305 (2012).

22. Rönnlund, M., Nyberg, L., Bäckman, L. & Nilsson, L.-G. 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 (2005).