nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0715-2

Rostral locus coeruleus integrity is associated with better memory performance in older adults

Martin J. Dahl[1*], Mara Mather[2], Sandra Düzel[1], Nils C. Bodammer[1], Ulman Lindenberger[1,3,4], Simone Kühn[5,6] and Markus Werkle-Bergner[1*]

For decades, research into memory decline in human cognitive ageing 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 quantification of LC integrity have hindered the study of LC–cognition associations in humans. 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 both younger ($n = 66$) and older adults ($n = 228$). Moreover, we observed functionally relevant age differences confined to rostral LC. Older adults with 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 ageing and highlight the role of the LC norepinephrine system in the decline of cognition.

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 cell loss and accumulation of abnormal tau in the locus coeruleus (LC), the brain's primary norepinephrine (NE) source, are critically related to cognitive decline in normal ageing and age-related pathologies[7–10]. However, direct in vivo evidence relating LC integrity to age differences in general memory abilities in humans is scarce[11] (for evidence from animal models, see for example refs. [12–14]). While there is an initial indication of a selective role of LC integrity, as indexed by magnetic resonance imaging (MRI; see below), in the encoding of negative emotional events[15], the relation of LC integrity to memory performance in general is still an open question. Given the pivotal role of the LC in age-related memory disorders such as Alzheimer's disease[8,9,16], 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[17]. Roughly symmetrical in both hemispheres, it extends only about 15 mm rostrocaudally from the level of the inferior colliculi to a position in the lateral wall of the fourth ventricle[18,19]. Despite its small size, the LC has diffuse and highly arborized efferent projections throughout the brain[20]. While initial reports assumed that the LC is a uniform structure, current evidence indicates a spatially differentiated organization. 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[21–23]. Norepinephrine release by the LC modulates cognitive functions such as perception, attention, learning and memory[21,24–32]. In particular, via its action on β-adrenoceptors in the hippocampus, the LC modulates long-term potentiation (LTP) and long-term depression (LTD), key determinants of synaptic plasticity and memory[33,34]. Recent optogenetic work has confirmed a causal role of LC activity in hippocampal LTP and memory enhancement, potentially via co-release of dopamine[35–37].

Its location adjacent to the ventricular system and its widespread, unmyelinated projections expose LC cells to both blood- and cerebrospinal fluid-bound toxins, making them a likely target for neurodegeneration[11]. Consistent with its age-related vulnerability, a prominent, rostrally accentuated decline in LC density was reported in healthy ageing (see ref. [22] for a review of early studies, but see refs. [16,18,38,39,] and ref. [11] for a recent discussion). Based on longitudinal data, initial evidence was reported pointing to the importance of LC integrity[10], here assessed via LC neuronal density, for maintaining memory abilities in old age. Wilson and colleagues evaluated cognitive abilities in a sample of older adults annually over a mean duration of about 6 years and, following participants' death, assessed neuronal density in the LC via autopsy[10]. In that study, an attenuation of cognitive decline was found 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 (that is, Lewy bodies and neurofibrillary tangles within the brainstem)[10]. Furthermore, studies in aged mice[12], rats[13] and monkeys[14] in which NE or its agonists were manipulated indicate 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[30,40,41]. Fortunately, however, a by-product of catecholamine synthesis, the dark, insoluble pigment neuromelanin, accumulates in the LC across the lifespan[42]. Pigmentation of the LC increases from birth until late middle age[42]. While some studies observed constant neuromelanin levels in late life[38], others reported a decline from mid-life until death, probably due to selective

[1]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany. [2]Davis School of Gerontology, University of Southern California, Los Angeles, CA, USA. [3]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, UK. [4]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany. [5]Department of Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Hamburg, Germany. [6]Lise Meitner Group for Environmental Neuroscience, Max Planck Institute for Human Development, Berlin, Germany.
*e-mail: dahl@mpib-berlin.mpg.de; werkle@mpib-berlin.mpg.de

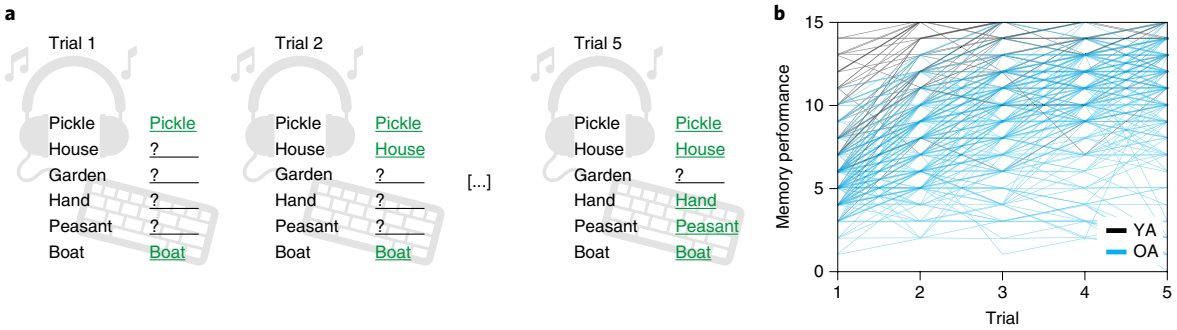

Fig. 1 | Schematic overview of the verbal learning and memory task. a, Word list consisting of 15 unrelated words was auditorily presented to participants over five trials. After each trial, participants freely recalled items and entered their responses on a keyboard. The sum of correctly recalled words provided the performance measure for each trial (range, 0–15). **b**, Performance of both younger (YA; $n = 66$, black) and older adults (OA; $n = 228$, blue) increased over learning trials in a nonlinear fashion. Credits: keyboard, adapted from needpix.com under a Creative Commons licence CC0 1.0; headphones/musical notes, Pavel Ivanov under a Creative Commons licence CC BY 4.0.

atrophy of neuromelanin-containing cells[22,43,44]. In noradrenergic cells, neuromelanin avidly chelates metals such as copper and iron and, as a compound, shows paramagnetic, T_1-shortening effects[45]. Sasaki and colleagues developed a neuromelanin-sensitive turbo-spin echo (TSE) MRI sequence that visualizes the LC as a hyperintense area adjacent to the lateral floor of the fourth ventricle[46] (for a review of neuromelanin-sensitive MRI studies, see ref. [47]). This sequence was recently validated by scanning human post-mortem samples using ultra-high-field MRI and then performing histological analyses of the samples[48]. 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 MRI-indexed, structural LC integrity in 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[49–52]. Using structural equation modelling (SEM) to capture the nonlinear dynamics of growth in performance[53], previous studies have linked ageing to lower initial memory performance[54,55] and slower learning with practice[55]. 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 ageing by linking non-invasive, in vivo indices of LC integrity to memory abilities in younger and older adults.

Results

Initial recall performance is lower in older adults. We assumed that both younger and older adults would recall a gradually increasing number of words over iterative RAVLT learning trials (see Fig. 1 and Methods (Cognitive data assessment)). 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 participants' learning performance would reach a natural (or task-induced) performance limit, we expected an increasing yet negatively accelerated slope[53,54]. Using latent growth curve modelling (a specific variant of SEM), we estimated the intercept and slope factors on a latent level[56]. To test for differences in intercept and slope parameters between age groups, we opted for a multi-group model[57,58]. In particular, we fitted a simple quadratic growth model comprising both an intercept and a quadratic and linear slope factor for each age group (see Supplementary Fig. 3). To investigate potential relations between intercept and slope terms (for example, starting out higher may be related to more rapid performance increases), we freely estimated covariances between intercept and slope parameters within age groups[55].

The adequacy of the proposed model was assessed using two frequently reported fit indices. First, the root mean square error of approximation (RMSEA) is a closeness-of-fit coefficient that expresses by how much the postulated model approaches the true model. Second, the comparative fit index (CFI) is an incremental fit index that compares the goodness of fit of the proposed model to a more restrictive nested baseline model[57,59,60]. RMSEA values close to or <0.06 and CFI values close to or >0.95 indicate good model fit[59].

A multi-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 ($\chi^2(46) = 81.764$, RMSEA = 0.052, CFI = 0.965 (ref. [59]); see Supplementary Results 2.1.1 and lower part of Fig. 2). Importantly, the model allows testing of group differences in factors capturing initial recall (intercept) and learning (slope), as well as for inter-individual differences therein within each age group. Parameters of interest were evaluated using likelihood-ratio tests. In particular, constraining parameters to be either equal (to test between-group differences) or zero (for within-group comparisons) should result in significant decrease in model fit relative to a model with unconstrained parameters in the case of reliable differences[58]. The difference in model fit between the constrained and unconstrained model (that is, $\Delta\chi^2$) under the null hypothesis follows a χ^2 distribution, with the degrees of freedom (d.f.) equivalent to the difference in numbers of constrained parameters (that is, Δd.f.)[58].

Both the means and variances of the intercept and slope factors differed reliably from zero in each age group (likelihood ratio tests for means: all $\Delta\chi^2(\Delta\text{d.f.} = 1) \geq 62.602$, all $P < 0.001$; likelihood ratio tests for variances: all $\Delta\chi^2(\Delta\text{d.f.} = 1) \geq 11.97$, all $P < 0.001$; see Supplementary Results 2.1.2). As the sole exception, there were no reliable inter-individual differences in the variance of the quadratic slope factor in younger adults (likelihood ratio test: $\Delta\chi^2(\Delta\text{d.f.} = 1) = 0.276$, $P = 0.599$, estimate (est) = 0.011 (95% confidence interval (CI) −0.031, 0.053)). Performance differences between younger and older adults were mainly found in intercepts (likelihood ratio test: $\Delta\chi^2(\Delta\text{d.f.} = 1) = 59.533$, $P < 0.001$, $\text{est}_{\text{younger adults}} = 9.197$ (95% CI 8.508, 9.886), $\text{est}_{\text{older adults}} = 5.647$ (95% CI 5.395, 5.899); see Supplementary Results 2.1.3). In sum, learning rates did not

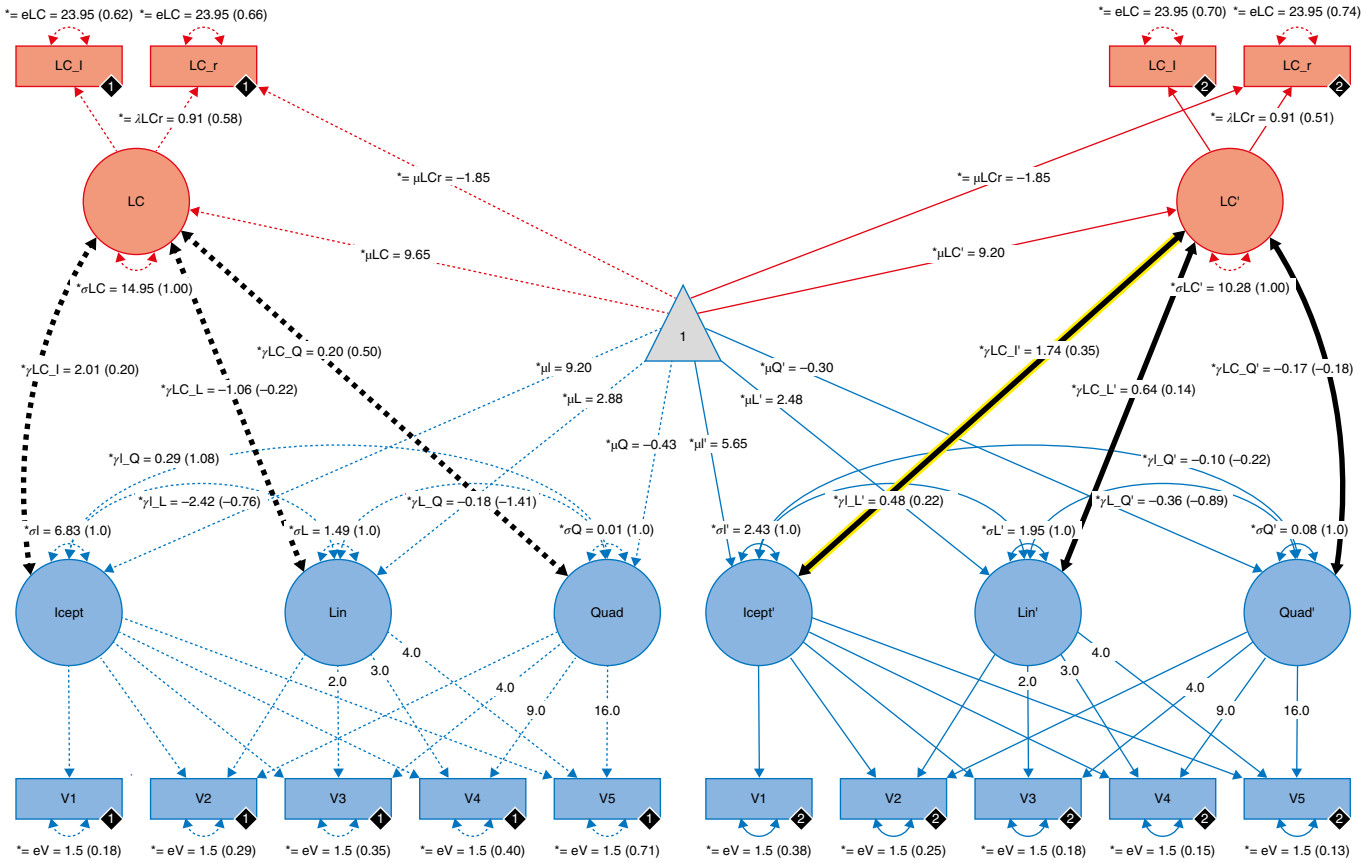

Fig. 2 | Pictorial rendition of the structural equation model probing associations between LC integrity and memory performance in younger and older adults on a latent level. Rectangles and circles indicate manifest and latent variables, respectively. Cognitive manifest variables represent the iteratively 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 ($n = 66$), broken lines; older adults, 2 ($n = 228$), solid lines). Thick black lines denote associations between LC integrity (red) and memory performance (blue) on a latent level. The significant association between LC integrity and memory performance (Icept) in older adults is highlighted (yellow shading). 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 ($n = 66$), broken lines; older adults, 2 ($n = 228$), 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 asterisks and equal signs (*=). Loadings that are freely estimated (*) for each group are indicated by a prime symbol (', in older adults). The constant is depicted by a triangle; paths from the constant to the variables indicate their mean (μ). Icept, intercept; Lin, linear slope; Quad, quadratic slope; e, error.

differ reliably between age groups, but older adults started and ended with lower recall scores compared to younger adults due to differences in initial recall.

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[47]—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 Fig. 3a–c and Methods (MRI data analysis)). The procedure evinced high reproducibility across multiple measurements (see Supplementary Results 2.2.1.3) and was validated using both published LC maps (see Fig. 3d) and manual intensity assessments (see Supplementary Results 2.2.1.2). 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 multi-group, single-factor structural equation model based on each hemisphere's mean intensity ratios (see upper part of Fig. 2 and Methods (Estimation of latent LC integrity scores)). The proposed model fit the data well ($\chi^2(7) = 1.791$, RMSEA = 0.000, CFI = 1.222 (ref. [59]); see Supplementary Results 2.2.3.1). We detected reliable average LC scores for both age groups; moreover, within each group, participants showed significant inter-individual differences in LC scores (likelihood ratio tests: all $\Delta\chi^2(\Delta\text{d.f.} = 1) \geq 18.305$, all $P < 0.001$; see Supplementary Results 2.2.3.2). Younger and older adults showed no reliable differences in average LC scores (likelihood ratio test: $\Delta\chi^2(\Delta\text{d.f.} = 1) = 0.383$, $P = 0.536$, $\text{est}_{\text{younger adults}} = 9.640$ (95% CI 8.342, 10.939), $\text{est}_{\text{older adults}} = 9.205$ (95% CI 8.484, 9.926)), in line with spatially confined age differences across the rostrocaudal LC axis[41,61].

Having established valid models for both memory performance and LC integrity in isolation, we aimed to combine the independent information from both modalities. For this, we merged both SEMs in a unified neuro-cognitive model that demonstrated good fit ($\chi^2(87) = 101.942$, RMSEA = 0.024, CFI = 0.986 (ref. [59]); see Fig. 2 and Supplementary Results 2.3). Initial recall (intercept) was positively related to latent LC scores for older adults, while the association failed to reach significance for the younger group (likelihood ratio test; for older adults: $\Delta\chi^2(\Delta\text{d.f.} = 1) = 7.939$, $P = 0.005$, $\text{est}_{\text{older adults}} = 1.737$ (95% CI 0.447, 3.027), standardized $\text{est}_{\text{older adults}} = 0.348$; for younger adults: $\Delta\chi^2(\Delta\text{d.f.} = 1) = 1.181$,

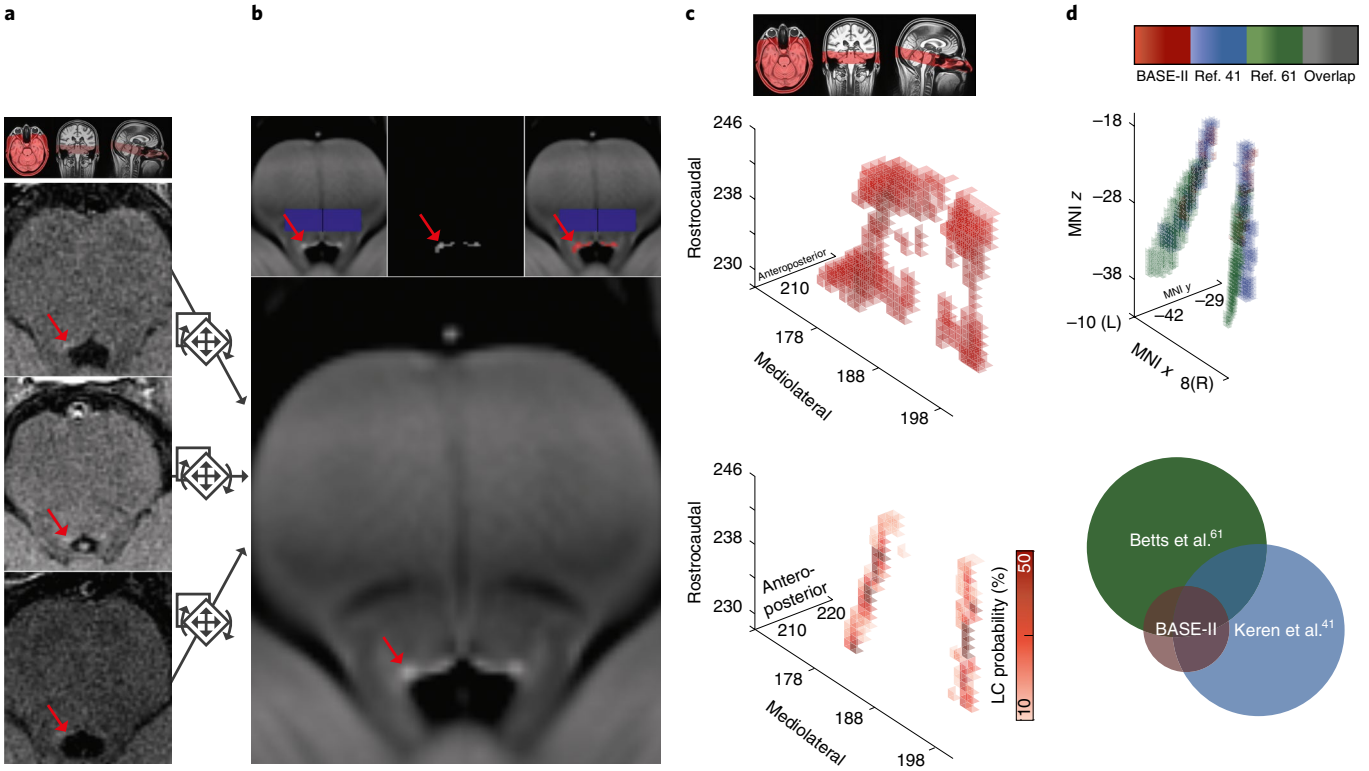

Fig. 3 | Schematic overview of the semi-automatic analysis procedure used to extract individual LC intensity values across the rostrocaudal extent. **a**, Native space neuromelanin-sensitive brainstem scans of three randomly selected participants (axial slices are shown). Hyperintensities corresponding to the LC are indicated by red arrows. **b**, Neuromelanin-sensitive scans were aligned and pooled across participants to increase the signal-to-noise ratio and to 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 previously published maps[41,61]. Circle radius indicates map size (that is, number of voxels). We freely share the resulting LC probability map at https://www.mpib-berlin.mpg.de/LC-Map.

$P = 0.147$, $est_{younger\ adults} = 2.014$ (95% CI −1.671, 5.698), standardized $est_{younger\ adults} = 0.199$; compare ref. [15]). Learning rates (slope) were not reliably associated with LC scores (likelihood ratio tests: all $\Delta\chi^2(\Delta d.f. = 1) \leq 1.426$, all $P \geq 0.232$; see Supplementary Table 8). These findings indicate higher initial recall performance (intercept) for older adults with high LC integrity than for those with low LC integrity (see Fig. 4).

There were no statistically reliable age group differences in the association between memory performance (intercept, slope) and LC scores (likelihood ratio tests: all $\Delta\chi^2(\Delta d.f. = 1) \leq 1.927$, all $P \geq 0.165$; see Supplementary Table 8). Thus, while we discovered a statistically reliable association between inter-individual differences in memory performance and LC integrity only in older adults, we observed no reliable age group differences in this association (see Supplementary Table 9 for analyses across age groups).

Making use of the longitudinal nature of this dataset (see Methods (Study design and participants)), we also examined associations between LC and memory performance using behavioural 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 Supplementary Results 2.3.1.2).

Lower age differences in rostral LC intensity ratios relate to memory performance in older adults. Building on reports of a spatially differentiated LC organization[21–23], we investigated age differences in LC topography and their association with memory performance by analysing LC data slice by slice. A non-parametric, cluster-based permutation test[62] revealed spatially heterogeneous age differences in LC ratios along the rostrocaudal extent of the nucleus (see Methods (Analysis of age differences in the spatial distribution of LC ratios)). In comparison to younger adults, older adults showed a cluster of elevated intensity spanning caudal slices (65th–100th LC percentile; cluster permutation test: P corrected for multiple comparisons (P_{corr}) < 0.001 (95% CI < 0.001, <0.001)) in line with neuromelanin accumulation across the lifespan[22,42,43]. In contrast, in rostral segments there was a trend towards decreased LC intensity in older adults (29th–41st LC percentile; cluster permutation test: $P_{corr} = 0.079$ (95% CI 0.078, 0.080); see Fig. 5), in accordance with age-related decline in hippocampus-projecting LC segments[22,23]. Age differences in LC ratios differ reliably across the rostrocaudal axis of the nucleus, as indicated by a significant age group (young/old) × topography (rostral/caudal) interaction (analysis of variance (ANOVA): $F(1, 292) = 26.650$, $P < 0.001$, $\eta_p^2 = 0.084$ (95% CI 0.033, 0.149; ref. [63]); see Fig. 5).

To evaluate the functional implications 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), we observed no reliable association between LC ratios and initial recall performance (intercept; Spearman's correlation: $r_s(292) = -0.08$ (95% CI −0.193,

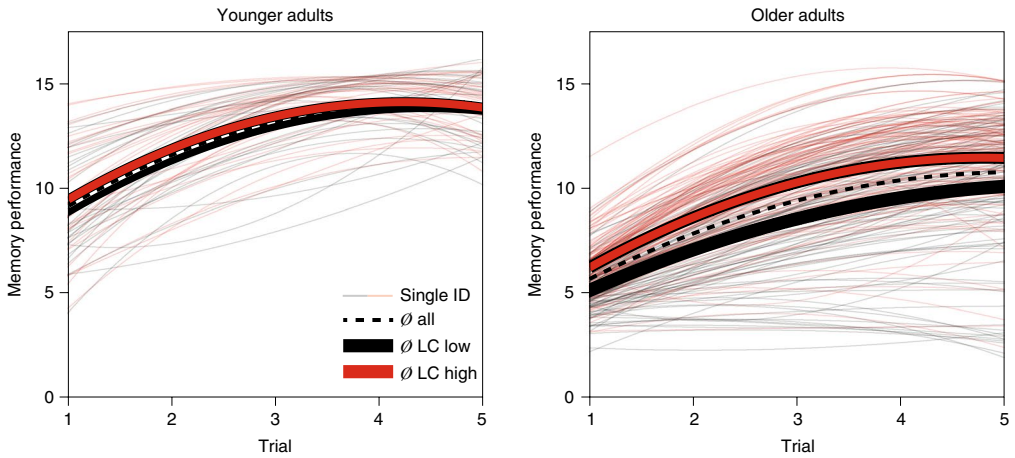

Fig. 4 | Estimated learning and memory performance trajectories for younger and older adults. To enable visualization of the association between LC integrity and memory performance, single participants (thin lines; Ind) are colour-coded based on LC integrity (median-split), and mean trajectories for subgroups are displayed (thick lines). Left: $n = 33$ younger adults in the low and high LC groups respectively; right: $n = 114$ older adults each in the low and high LC groups, respectively.

0.035), $P = 0.172$; see Fig. 5). However, in rostral segments (29th–41st LC percentile), higher LC ratios were significantly associated with memory performance (Spearman's correlation: $r_s(292) = 0.207$ (95% CI 0.095, 0.314), $P < 0.001$; see Fig. 5; difference between caudal and rostral correlations: $Z = 3.385$, $P < 0.001$ (ref. [64]); for analyses within age groups, see Supplementary Table 11). Older adults with more youth-like intensity ratios in rostral LC segments also showed better memory performance. In sum, we observed a trend towards spatially confined age differences in LC ratios that were associated with memory performance and that are in line with neurodegeneration of hippocampus- and forebrain-projecting LC segments[22] in older adults.

Discussion

Animal and post-mortem human studies suggest a link between memory performance in ageing and the integrity of the central noradrenergic system[7,10]. To date, however, in vivo research on humans has been stymied by methodological difficulties in the reliable assessment of LC integrity[30,40,41] (but see ref. [15]). Here, we took advantage of the paramagnetic properties of neuromelanin, a by-product of NE synthesis, in T_1-weighted MRI to image the LC in vivo[46,48]. 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 Supplementary Results 2.3.1.2), were positively related to LC integrity in older adults. Analyses making use of the longitudinal nature of this dataset further indicate stable and lasting rather than short-lived LC–memory dependencies (see Supplementary Results 2.3.1.2). 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. These results bridge a gap between animal and in vivo human research and provide important insights into the neural underpinnings of senescent cognitive decline in healthy ageing.

We applied an iterative learning and memory task (RAVLT)[65] that required participants to encode, consolidate and retrieve verbal information and that thus captures the dynamic nature of memory[53]. Our analyses allow the study of two psychologically distinct factors, namely, initial recall (that is, performance after the first learning trial, corresponding to a standard one-trial memory assessment) and learning (that is, changes in performance with practice). Age differences were mainly observed in initial recall performance (see Supplementary Results 2.1.3 and Supplementary Discussion).

Neurodegenerative changes have been suggested to be determinants of inter-individual differences in learning and memory performance in later life[51,66]. Here, we investigated the role of the LC, one of the first targets in neurodegenerative diseases such as Alzheimer's and Parkinson's, in senescent memory decline[8,9,11]. 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[46] (see refs. [67,68] for recent discussions on LC contrast mechanisms). Neuromelanin accumulates nonlinearly in LC cells across the lifespan, with a peak concentration at around 50–60 years of age, and subsequently stabilizes[38] or declines, probably due to preferential loss of pigmented cells[42,43]. Thus, neuromelanin forms a reliable natural contrast agent, particularly in older adults, that can be harnessed in cell-count[18] and MRI studies[48] to index LC integrity. However, neuromelanin may be a less effective indicator of LC structure in younger adults due to the many LC neurons that do not yet contain neuromelanin, based on post-mortem analyses using tyrosine hydroxylase (that is, the rate-limiting enzyme in NE synthesis) as a marker of LC neurons[22]. From the sixth decade onwards the numbers of LC neurons identified by tyrosine hydroxylase and neuromelanin markers were similar[22]. 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[61], which in turn was used to restrict automatized extraction of individual peak intensities[68,69]. 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[18]. We demonstrated high validity of the proposed semi-automatic procedure by comparison to both previously published LC maps[41,61] and manual LC intensity ratings (see Supplementary Results 2.2.1.1–2.2.1.2). Repeated measurements of an independent younger adult sample

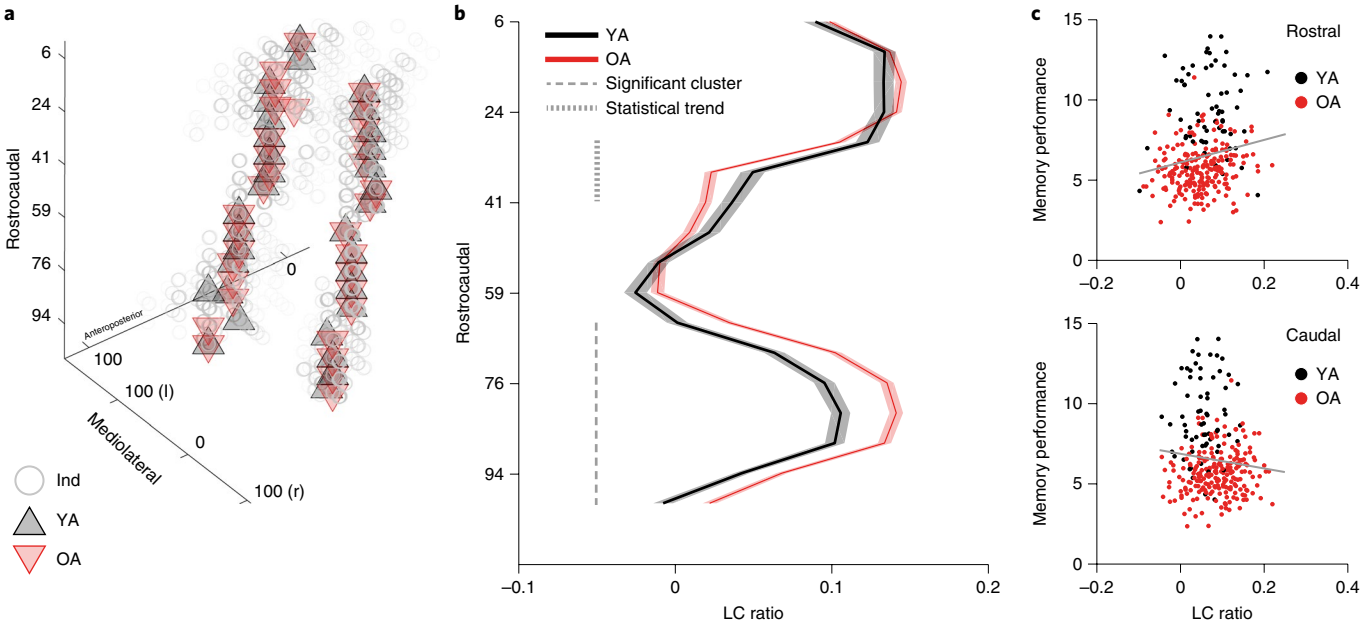

Fig. 5 | Topographical age differences in LC intensity ratios and their functional implications. a, On a group level, younger (YA; $n = 66$, black) and older adults (OA; $n = 228$, red) share highly congruent peak LC locations. Grey circles indicate single participants (Ind). **b**, A cluster-based permutation test revealed spatially confined age differences in LC intensity across the rostrocaudal axis. While older adults show significantly higher ratios in the caudal segments (65th–100th LC percentile; cluster permutation test: $P_{corr} < 0.001$ (95% CI <0.001, <0.001)), there is a trend for a reversed effect in rostral segments (29th–41st LC percentile; cluster permutation test: $P_{corr} = 0.079$ (95% CI 0.078, 0.080)). Age differences in LC ratios differ reliably across the rostrocaudal axis of the nucleus, as indicated by a significant age group (young/old) × topography (rostral/caudal) interaction (ANOVA: $F(1, 292) = 26.650$, $P < 0.001$, partial η^2 (η_p^2) = 0.084 (90% CI 0.040, 0.137))[63]. Shaded lines represent ±1 s.e.m. (see Supplementary Fig. 12 for the distribution of younger and older adults' LC ratios along the rostrocaudal axis). Rostrocaudal, anteroposterior and mediolateral positions are expressed in percentiles relative to the total size of the observed LC. **c**, Correlation between memory performance and LC ratios, calculated across YA and OA ($n = 294$), reaching statistical significance in the rostral cluster ($r_s(292) = 0.207$ (95% CI 0.095, 0.314), $P < 0.001$); it fails to reach statistical significance in the caudal cluster ($r_s(292) = -0.08$ (95% CI −0.193, 0.035, $P = 0.172$); difference between rostral and caudal correlations: $Z = 3.385$, $P < 0.001$.

further confirmed high reproducibility of the intensity assessment (see Supplementary Results 2.2.1.3). However, as reported earlier for neuromelanin-sensitive sequences, our analyses do not capture the complete LC extent[47,70] leading to lower volume estimates than those obtained from post-mortem studies[19]. In their histological analyses, Fernandes and colleagues describe a common, central LC zone that is shared across all participants[18]. Accordingly, the segments with the most dispersed noradrenergic cells in the caudal LC show less overlap at a group level and are thus arguably more difficult to detect using a semi-automatic approach with a conservative intensity threshold (that is, 4 s.d. above the mean intensity of a reference region[68,69]). This circumstance is probably exacerbated by partial volume effects due to the relatively thick slices in most neuromelanin-sensitive sequences[70]. However, manual[61] and threshold-free automatic approaches[41] also appear to be affected by this challenge, as evident from their low correspondence in caudal slices (Montreal Neurological Institute (MNI) $Z \leq -30$: 6.91% of ref. [41] in ref. [61], 4.63% vice versa[41,61]; see Fig. 2d). We thus refrained from attempts to capture the most caudal LC segments via either manual segmentation, a lowering of intensity thresholds or both, to effectively separate LC from the nuclei subcoerulei[17].

To investigate the link between LC integrity and memory in ageing humans, we estimated latent LC integrity scores using structural equation modelling. Here, we integrated over slices and hemispheres to derive a single measure reflecting LC integrity. Compatible with spatially confined age differences[22,41,61], differences between younger and older adults in average LC scores failed to reach statistical significance. In older adults, higher initial recall performance and steeper learning curves were found in participants with high LC integrity. Even when integrating over a variety of tasks, our findings corroborate a positive dependency of memory performance on LC integrity (see Supplementary Results 2.3.1.2), in line with a general mechanism beyond specific mnemonic domains (for example, emotional memory[15]).

Besides regulating healthy cognition[21,26,29,31,32], the LC subserves critical neuroprotective functions[11,71,72]. For instance, NE protects against neuro-inflammation via the regulation of inflammatory gene expression and directly enhances clearance of aggregated β-amyloid via activation of microglia[11,72], two major threats to the ageing brain[73]. This led Robertson and colleagues[74] to suggest that repeated activation of the LC-NE system (for example, by exposure to novelty[37]) helps maintain cognitive functionality in ageing despite underlying neuropathological changes. At the same time, the LC is susceptible to neurodegeneration in ageing. Its vulnerable anatomical location next to the fourth ventricle exposes it to toxins, and its high energy demand increases the risk of oxidative stress over time[71]. Thus, while in ageing, the LC-NE system's neuroprotective function is especially important, the loss of noradrenergic cells may have wide-ranging consequences for both cognition and brain health. Whereas we observed a positive association between LC integrity and memory performance in older adults, we did not find a robust relation in younger adults. However, the association between LC integrity and memory did not differ reliably between groups. Thus, while we discovered an association between inter-individual differences in memory performance and LC integrity in older but not in younger adults, we did not observe reliable group differences, which may hint at comparable associations independent of age. This appears to contrast with the notion of an increasingly important

role of the LC-NE system in ageing[11]. However, to constitute a reliable proxy for neuronal density within the LC[48], neuromelanin-sensitive MRI requires a sufficient neuromelanin saturation of the nucleus. Post-mortem studies indicate that neuromelanin concentration within the LC peaks around late middle adulthood (~50 years of age[43,75]). Thus, the lower sensitivity of neuromelanin as a LC integrity proxy in younger adults may have impeded the detection of reliable associations within this subgroup[15,76] as well as differences between age groups.

We observed spatially confined age differences in LC intensity ratios[41,61]. In older relative to younger adults, caudal LC sections showed significantly elevated intensity ratios in line with neuromelanin accumulation across the lifespan[42,43]. In contrast, there was an absence of increased signal in rostral LC and, descriptively, even a tendency towards reduced intensity ratios in older adults. This may suggest neurodegeneration of rostral LC segments that are densely connected to key memory structures, such as the forebrain and hippocampus[21–23]. Consistent with this possibility, 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 ageing (as covered by a review of early studies[22] but, as recently discussed[11], there are also recent observations[16,18,38,39] of stable LC cell counts in healthy ageing).

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[21,32], presumably in self-enhancing feedback loops with glutamate[29]. In addition, NE improves the temporal organization of neuronal responses to sensory stimulation (that is, spike rhythmicity) and thus increases perceptual acuity[32,77,78]. Note that noisier information processing has been linked to deficient neuromodulation with effects on higher-order cognitive functions such as attention and working memory[79]. 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[26]. Accordingly, pharmacological manipulation of NE levels in aged monkeys ameliorated attention and working memory deficits[14]. Hence, LC integrity may promote attentional and control mechanisms implicated in successful episodic memory performance[80]. Third, in the amygdala and hippocampus, NE modulates synaptic strength and facilitates synaptic plasticity[31,32]. More specifically, via its action on β-adrenoceptors, the LC modulates LTP and LTD, major determinants of long-term memory[33,34]. In sum, NE released by the LC alters perception, attention and memory at multiple cortical and subcortical sites, crucially implicating LC integrity in learning and memory in ageing.

Recent analyses of large post-mortem samples have indicated that the LC is among the earliest brain regions to show hyperphosphorylated tau, a hallmark of Alzheimer's disease, which accumulates in the nucleus linearly with the progression of the disease[8,16,81] (but see ref. [82] for a different observation and ref. [83] for a commentary). In tauopathies such as Alzheimer's, misfolded tau proteins are thought to spread in a stereotypical, transcellular propagation pattern through neural networks[8,82], which may lead to tau pathology in prominent noradrenergic projection targets such as the medial temporal lobe[72]. Elevated tau levels are associated with more aggregated amyloid-β, the second hallmark of Alzheimer's disease, and worse cognition[73]. A mere accumulation of subcortical tau could be harmless without evidence for neurodegenerative consequences. However, LC volume decreases by about 8% per Braak stage increment (a classification system that allows staging of the progression of tau spread) followed by death of LC cells (from middle Braak stages in refs. [16,84]) that correlates with cognitive decline (for a review on the impact of hyperphosphorylated tau on LC, see ref. [72]). Notably, LC cell loss in Alzheimer's disease reaches profound levels (~50–80% (ref. [71])) and even surpasses neurodegeneration of the cholinergic nucleus basalis[85]. In animal models of Alzheimer's disease, LC lesions exacerbated both neural and behavioural decline[71]. In particular, in a mouse model of tauopathology, LC-ablated mice showed impaired memory and increased neuronal loss in the hippocampal formation compared to non-ablated animals[72,81,82,86]. Furthermore, in a rat model with both β-amyloid and tau pathology, chemogenic activation of the LC[87] rescued hippocampus-dependent behaviour[84,88]. Together, this indicates that, in Alzheimer's disease, LC neurodegeneration acts synergistically with tau pathology to drive pathological changes and does not merely constitute an epiphenomenon[87]. The threshold between healthy ageing and early manifestations of neurodegenerative diseases such as Alzheimer's is yet to be defined[84]. Most likely, both LC neurodegeneration and the transition from healthy cognition in ageing to pathological conditions (that is, mild cognitive impairment, dementia) occur along a continuum[76]. While all but two of the older adults tested in this study scored above threshold on a dementia screening (that is, Mini-Mental State Examination[89]) and demonstrated a high level of functionality, neither our cognitive nor neural data can rule out the existence of early levels of neuropathology (for example, subcortical accumulation of pretangle tau[8]). We thus cannot rule out that both normal ageing and neuropathological changes affected LC integrity in our older adult sample.

The present results underscore the utility of non-invasive, in vivo markers of LC integrity as indicators for preserved memory performance in human ageing, 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 the LC that are connected to key memory structures such as the hippocampus. Importantly, SEM revealed reliable and stable positive associations between LC integrity and general episodic memory among older adults. We conclude that older adults with better 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 ageing[1].

Methods

Study design and participants. Data were collected within the framework of the Berlin Aging Study II (BASE-II (refs. [90–92])). BASE-II is a multidisciplinary, multi-institutional longitudinal study that investigates the cognitive[93], neural, physical and social conditions associated with successful ageing. For an extensive study description, please refer to the study's website (https://www.base2.mpg.de/en) and papers[90–92].

Cognitive and neuroimaging 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 (s.d. = 0.52, range = 0.9–3.2). Cognitive and neuroimaging data were collected on separate occasions at each time point (mean interval = 9.16 d, s.d. = 6.32, range = −2–44; for T2); LC data were acquired only at T2.

A subset of 323 BASE-II participants underwent 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 (0 females, mean age = 76.66 years, s.d. = 1.64, range = 74.93–78.74; at T2) were excluded from further analyses due to either excessive movement artefacts ($n = 2$) or incorrect scan positioning ($n = 3$). The final MRI subsample (see Supplementary Table 1) included 66 younger adults (22 female) with a mean age of 32.5 years (s.d. = 3.53, range = 25.41–39.84; at T2) and 228 older adults (82 female) with a mean age of 72.29 years (s.d. = 4.11, range = 62.53–83.16; at T2). All participants with full cognitive and MRI data were included in the study—that is, we did not conduct a formal power calculation given that there was no available previous evidence on the phenomena studied. 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 assessments were approved by the ethics committees of the Max Planck Institute for Human Development and the German Psychological Society, 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 six individuals using a computerized battery of 21 tasks (for a description of the complete battery, see ref. [93]). Each test session lasted approximately 3.5 h. 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 RAVLT, a standardized and validated neuropsychological tool that provides information about participants' learning and delayed recall performance[65] (see Fig. 1); participants 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 a keyboard (trials 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 min (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 d 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 participants. The same word list was used at T1 and T2.

Making use of the comprehensive cognitive battery available for this dataset, we additionally integrated performance over a variety of episodic memory tasks to retrieve a general measure of episodic memory as previously described[93] while explicitly excluding RAVLT data (see Supplementary Results 2.1.4). In particular, we incorporated information derived from a visual face–profession, object–location and scene-encoding task. For detailed task descriptions, see ref. [93]. In short, the visual face–profession task involved participants studying 45 pairs of face images and profession descriptors. The tasks consisted of an incidental encoding phase, a 2-min distraction phase and finally a recognition memory task including old, new and 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, participants encoded the location of 12 digital photographs of real-life objects on a 6×6 grid. After encoding, objects reappeared next to the grid and participants 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, participants incidentally encoded 88 scene images by performing indoor/outdoor judgements on each image. The encoding phase was followed by an old/new recognition memory test that included confidence judgements. Recognition memory (hits – false alarms) was tested after a delay of approximately 2.5 h and served as the performance index.

MRI data assessment. Structural and functional MRI data were collected at 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 (3D) T_1-weighted magnetization prepared gradient-echo (MPRAGE) sequence with a duration of 9.2 min and the following parameters was applied: repetition time (TR) = 2,500 ms; echo time (TE) = 4.770 ms; inversion time (TI) = 1,100 ms; flip angle = 7°; bandwidth = 140 Hz pixel^{-1}; acquisition matrix = 256 × 256 × 192; isometric voxel size = 1 mm^3. Pre-scan normalization and 3D distortion correction options were enabled.

Based on this whole-brain MPRAGE sequence, a neuromelanin-sensitive high-resolution, two-dimensional T_1-weighted 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 = 11 ms; flip angle = 120°; bandwidth = 287 Hz pixel^{-1}; acquisition matrix = 350 × 263 × 350; voxel size = 0.5 × 0.5 × 2.5 mm^3. Each TSE scan consisted of ten axial slices with a gap of 20% between slices, which covered the whole extent of the pons. Pre-scan normalization and elliptical filter options were enabled. The TSE sequence yielded two brainstem MRIs per participant, each resulting from four (online) averages.

For some participants, specific absorption rate limits were exceeded, which precluded further scanning with the current set of parameters (see refs. [47,68]). To avoid data loss while maintaining contrast settings, the maximal number of TSE slices was reduced for those participants. Thus, our sample consists of the following slice distribution: 243 × 10 slices, 38 × 9 slices, 27 × 8 slices, 6 × 7 slices, 1 × 6 slices and 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 SEM to analyse inter- and intra-individual differences in verbal learning and memory performance using the Ωnyx 1.0–991 software package[94] with full information maximum likelihood estimation. SEM offers a multivariate approach in which repeatedly observed (manifest) variables can be used to examine hypotheses about unobserved (latent) variables. Latent variables provide the benefit of accounting for measurement error in observed scores, thereby increasing statistical power[57,58].

We assessed the adequacy of the proposed growth models (that is, a specific variant of SEM) by testing for differences between the model-implied and empirically observed covariance matrices[60]. The χ^2-test formally tests for equity of covariance matrices; however, it is particularly sensitive to sample size and should be interpreted with caution in large samples[59,60]. Thus, we examined two additional frequently reported fit indices (RMSEA, CFI). In contrast to the χ^2-test, the CFI is not influenced by sample size. Finally, the adequacy of the quadratic growth model was tested against nested competing alternative models (for example, intercept-only and linear growth models) using likelihood ratio difference tests[57,60]. In the case of non-nested competing alternative models, we compared the Akaike information criterion, favouring the model providing the lowest value[57].

After establishing model fit, age differences in parameters of interest (for example, 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[57,60] (see Results (Initial recall performance is lower in older adults)). In addition, for all models, parameters of interest are listed with their point estimate and the confidence interval, which is based on the s.e.m. of the estimate (that is, point estimate ± s.e.m. × 1.96), computed by the inverse of the Hessian[94]. The confidence interval allows a rapid indication of whether parameters differ reliably from zero. Note, however, that the estimation of confidence intervals focuses on a single parameter and does not take into account the correlation with the remaining parameters, thus leading to lower statistical power[94]. Statistical inferences are therefore based on likelihood ratio tests as previously suggested[57,58,94]. When applicable, we also added standardized parameter estimates (that is, correlation coefficients for instance for associations between latent LC and memory estimates).

In a similar vein, we integrated performance over a variety of memory tasks to estimate a general episodic memory score using a multi-group single-factor structural equation model as previously described[93] (see Supplementary Fig. 4). 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).

MRI 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[68]. In addition, the LC shows amorphous borders, complicating the manual segmentation of the structure[41,70]. 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 (for example, non-blinded raters[47,68,69]). 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[41]. Third, age- and disease-related alterations of LC integrity are not homogeneous across the rostrocaudal extent of the structure[9,41,61]. Increased specificity and sensitivity for detection of such conditions may be achieved by assessing intensity across the whole rostrocaudal axis[47]. Fourth, due to the use of multiple measurements and online averaging, the acquisition time for most neuromelanin-sensitive sequences exceeds 10 min, which increases the risk of head motion artefacts and thus reduced data quality[68]. To explicitly correct for this, images from each measurement can be saved separately and averaged after application of motion-correction methods[68].

In the implemented analysis procedure (see Supplementary Fig. 1), we first pooled across participants over aligned TSE scans to increase the signal-to-noise ratio and facilitate LC delineation using a template-based approach[41,61] (Steps 1–3). On a group level, LC location was determined semi-automatically[68,69] 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 MRI data. Before 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[61].

Initially (Step 1), using Advanced Normalization Tools v.2.1 (ANTS[95,96]), a group whole-brain template ($Template_{whole}$) was generated from all available MPRAGE scans from T2. In this iterative process, (1) individual native space scans

($MPRAGE_{native}$) were coregistered to a common group or template space and (2) 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 nonlinear registration iterations. Each nonlinear 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 was used as the similarity metric and greedy symmetric normalization (SyN) as the transformation model for nonlinear registrations[97]. Template update steps were set to 0.25 mm.

Next (Step 2), within each participant, we coregistered the native space brainstem scan (TSE_{native}) to the whole-brain scan that had been moved to template space (that is, $MPRAGE_{template}$; see Step 1). For this, we first performed linear (rigid, then affine), followed by nonlinear (SyN), registration steps[97]. Histogram matching was enabled since we registered scans from the same modality (T_1-weighted). This intermediate, within-participant coregistration was implemented to bring individual brainstem scans into 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 (1) individual brainstem scans coregistered to template space ($TSE_{template}$) and (2) a group brainstem template image ($Template_{slab}$). The same parameters were used for both template-building runs (Steps 1 and 3). The linear and nonlinear transformation matrices of the intermediate within-participant 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[61]. 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—that is, voxels falling within the $Template_{slab}$ (ref. [41]).

Semi-automatic LC 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 Fig. 3b). 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[68,69] (Mathworks Inc., including the SPM software toolkit[98]). 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:

$$T_{LC} = \mu(\mathrm{Ref}) + \sigma(\mathrm{Ref}) \times 4 \quad (1)$$

where μ and σ specify the mean and s.d. of the intensity in the pontine reference (Ref) region, respectively. T_{LC} denotes the resulting intensity threshold. The cut-off of four standard deviations above the mean reference intensity was selected on the basis of previous research[68,69] and was confirmed visually as generating a volume that matched previous mapping studies of the LC[41,61]. A lateralized threshold was computed for each slice to counteract previously reported unilateral biases in brainstem signal intensity[61,70] and intensity fluctuations along the rostrocaudal axis. Next, we generated a volume of interest, 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 Fig. 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 (5,950 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 participants (Step 5), we masked individual brainstem scans in template space (that is, $TSE_{template}$) with the binarized LC search space. This was first done with 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[41]. 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[47]:

$$\mathrm{LC}_{\mathrm{ratio}} = (\max(\mathrm{LC}) - \max(\mathrm{Ref}))/\max(\mathrm{Ref}) \quad (2)$$

where LC denotes the intensity of the LC and Ref indicates the intensity of the pontine reference region. Finally, for each hemisphere we computed a LC ratio across all slices, which was used to probe LC–cognition associations in a SEM framework (that is, we calculated a ratio score of peak LC intensity relative to peak reference intensity).

Comparison to previously published LC masks. To judge the validity of the generated search space, we plotted the retrieved peak intensity coordinates for each participant's left and right LC[41]. The aggregated maximum intensity plot was then converted to express the likelihood of finding the LC for any given voxel (that is, a LC location probability map[41,70]). We thresholded the image to show only locations that were shared across participants (that is, 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[41,61]; see Fig. 3d and Supplementary Results 2.2.1.1). To facilitate comparability of study results, we freely share the resulting LC probability map with the neuroscientific community (https://www.mpib-berlin.mpg.de/LC-Map).

Comparison to manually assessed LC intensity. Almost all published MRI studies that indexed LC integrity used a manual procedure to assess intensity values (for a recent review, see ref. [47]). To demonstrate the validity of our approach, two independent, blinded raters (research assistants) manually assessed LC and reference intensity values for all participants using a standard procedure[75]. On each axial slice demonstrating increased intensity in anatomically probable LC locations, raters placed three non-overlapping quadratic 1-mm^2 regions of interest (ROIs) in areas in which signal was most evident[75]. Raters were instructed not to place the ROIs directly adjacent to the fourth ventricle, to avoid partial volume effects contaminating the signal[76]. To measure reference intensity, three rectangular, non-overlapping 10-mm^2 ROIs were placed on the same slices in the dorsal pons, centred medially between the left and right LC ROIs[75] (see Supplementary Fig. 2). Raters additionally indicated the slice with peak signal intensity. Image processing was conducted using the ImageJ 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 four participants manual assessments were not available from both raters, and those participants were thus removed from the evaluation. For analyses, the peak intensity for each slice was first determined for LC (left, right) and pons across ROIs, and then the peak intensities were selected across slices. We calculated the consistency of raters using intraclass correlation coefficients (ICC; two-way mixed model with absolute agreement, using IBM SPSS Statistics v.24) and then, after demonstrating high accordance, averaged ratings across hemispheres and raters to obtain a single stable intensity estimate per participant (see Supplementary Results 2.2.1.2). 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 equation (2) (see semi-automatic LC intensity ratio calculation; see Supplementary Results 2.2.1.2 and 2.5).

Reproducibility of semi-automatically assessed LC intensity ratios. 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; s.d. = 1.578 years). Participants 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 were sinc-interpolated and fed into another template-building procedure as described above (see Steps 1–3). The resulting brainstem template ($template_{slab\text{—}reproducibility}$) 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 averaged left and right LC ratios to evaluate the stability of the proposed method. Note that we analysed intensity ratios here since we compared scans acquired at different time points that are subject to overall scanner fluctuations.

Comparison to semi-automatically assessed LC 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 ($TSE_{template}$) were not an artefact 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 participants' coordinates. To compare spatial locations of native space values across participants, however, a common frame of reference was necessary[61]. Therefore, before warping, we split the LC and reference search spaces into segments. All but the most caudal segment contained three slices, reflecting the resolution difference between native and template space (that is, 1×TSE_{native}: 3×$TSE_{template}$; 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 (Step 5). The two native

space brainstem images of each participant (see MRI data assessment) were linearly and nonlinearly coregistered before masking. To judge the correspondence between automatically assessed LC intensity values from native and template space, we calculated ICC (see Supplementary Results 2.2.1.4). For further analyses of native space values, we computed LC intensity ratios for both hemispheres as described above using equation (2) (see Supplementary Results 2.4).

Analysis of age differences in the spatial distribution of LC ratios. To investigate age differences in LC intensity ratios along the rostrocaudal extent of the nucleus[9,22,41,61], we calculated non-parametric, cluster-based, random permutation tests as implemented in the Fieldtrip toolbox (see ref. [99]). These effectively control the false-alarm rate in the case of multiple testing[62]. In short, a two-sided, independent samples *t*-test is first calculated for each slice. Neighbouring slices with $P<0.05$ 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[62]. 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 Supplementary Results 2.2.2 and 2.4.1). 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 (that is, $P\leq0.05$ is considered significant for all statistics). Reliable clusters were followed up by means of a mixed-effects ANOVA with the factors age (YA; OA) and topography (rostral/caudal) to evaluate spatially confined age differences in LC ratios. For this, LC intensity ratios were averaged within participants within the ranges of the observed clusters.

To judge the functional implications of identified topographical age differences, we investigated the association between memory performance and LC ratios in the obtained clusters—first, irrespective of age group and then within each 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 interhemispheric differences in LC ratios[61,70], 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 LC integrity scores. As done for the cognitive data, we applied SEM to analyse inter-individual differences in LC ratios. We generated a multi-group model including LC ratios of each hemisphere as observed variables (see Step 5 and Supplementary Fig. 6). 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 Supplementary Results 2.5.1.1).

Analysis of associations between LC integrity scores and memory performance. After generating structural equation models for our cognitive and neural measures, respectively, we set out to link these modalities. That is, we were interested in assessing the relation between both inter- and intra-individual differences in verbal learning and memory performance, and inter-individual 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 Fig. 2). We investigated associations between our performance and LC measures by allowing for freely estimated covariances on a latent level (shown in black, Fig. 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 Supplementary Results 2.5.1.2). 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 Methods (Cognitive data analysis) and Supplementary Fig. 7).

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[55,58] (see Supplementary Fig. 10). 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[58]. 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−*; ref. [58]). The association between cognitive latent change scores and 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 Supplementary Results 2.3.2).

Statistical parameters. All reported results are based on two-sided statistical tests. We used an alpha level of 0.05 for all statistical tests. Statistical results with $P=0.05$–0.1 are described as statistical trends. When indicated, tests were adjusted for multiple comparisons. The applied statistical tests did not include covariates, unless noted differently. Data distribution was assumed to be normal (see Supplementary Fig. 12), but this was not formally tested.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data on which our results are based are available from the BASE-II steering committee following approval of a research proposal (https://www.base2.mpg.de/en/project-information/data-documentation). For inquiries please contact L. Müller, BASE-II project coordinator (lmueller@mpib-berlin.mpg.de). 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).

Code availability

The custom code used for these analyses is available from the corresponding authors upon request.

Received: 6 June 2018; Accepted: 29 July 2019;
Published online: 9 September 2019

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* (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. USA* **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 coeruleus, and cognitive decline. *Neurology* **80**, 1202–1208 (2013).
11. Mather, M. & Harley, C. W. The locus coeruleus: essential for maintaining cognitive function and the aging brain. *Trends Cogn. Sci.* **20**, 214–226 (2016).
12. Mei, Y. et al. Aging-associated formaldehyde-induced norepinephrine deficiency contributes to age-related memory decline. *Aging Cell* **19**, 659–668 (2015).
13. 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).
14. Arnsten, A. F. T. & Goldman-Rakic, P. S. Alpha 2-adrenergic mechanisms in prefrontal cortex associated with cognitive decline in aged nonhuman primates. *Science* **230**, 1273–1276 (1985).
15. Hämmerer, D. et al. Locus coeruleus integrity in old age is selectively related to memories linked with salient negative events. *Proc. Natl Acad. Sci. USA* **115**, 2228–2233 (2018).
16. 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. *Alzheimers Dement.* **13**, 236–246 (2017).

17. Szabadi, E. Functional neuroanatomy of the central noradrenergic system. *J. Psychopharmacol.* **27**, 659–693 (2013).
18. 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).
19. German, D. C. et al. The human locus coeruleus: computer reconstruction of cellular distribution. *J. Neurosci.* **8**, 1776–1788 (1988).
20. Aston-Jones, G. & Waterhouse, B. Locus coeruleus: from global projection system to adaptive regulation of behavior. *Brain Res.* **1645**, 75–78 (2016).
21. Berridge, C. W. & Waterhouse, B. D. The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Rev.* **42**, 33–84 (2003).
22. 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).
23. Schwarz, L. A. & Luo, L. Organization of the locus coeruleus-norepinephrine system. *Curr. Biol.* **25**, R1051–R1056 (2015).
24. Waterhouse, B. D. & Chandler, D. J. Heterogeneous organization and function of the central noradrenergic system. *Brain Res.* **1641**, v–x (2016).
25. Ordway, G. A., Schwartz, M. A. & Frazer, A. *Brain Norepinephrine: Neurobiology and Therapeutics* (Cambridge Univ. Press, 2007).
26. Arnsten, A. F. T. & Li, B. M. Neurobiology of executive functions: catecholamine influences on prefrontal cortical functions. *Biol. Psychiatry* **57**, 1377–1384 (2005).
27. 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).
28. Bouret, S. & Sara, S. J. Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci.* **28**, 574–582 (2005).
29. 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**, e200 (2016).
30. 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., et al.) 367–386 (Oxford Univ. Press, 2011).
31. Sara, S. J. Locus coeruleus in time with the making of memories. *Curr. Opin. Neurobiol.* **35**, 87–94 (2015).
32. Sara, S. J. The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.* **10**, 211–223 (2009).
33. Hansen, N. The longevity of hippocampus-dependent memory is orchestrated by the locus coeruleus-noradrenergic system. *Neural Plast.* **2017**, 2727602 (2017).
34. 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).
35. Bray, N. Learning and memory: you only learn once. *Nat. Rev. Neurosci.* **19**, 59 (2018).
36. McNamara, C. G. & Dupret, D. Two sources of dopamine for the hippocampus. *Trends Neurosci.* **40**, 383–384 (2017).
37. Takeuchi, T. et al. Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature* **537**, 357–362 (2016).
38. Ohm, T. G., Busch, C. & Bohl, J. Unbiased estimation of neuronal numbers in the human nucleus coeruleus during aging. *Neurobiol. Aging* **18**, 393–399 (1997).
39. Mouton, P. R., Pakkenberg, B., Gundersen, H. J. G. & Price, D. L. Absolute number and size of pigmented locus coeruleus neurons in young and aged individuals. *J. Chem. Neuroanat.* **7**, 185–190 (1994).
40. 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 'Jomeostatic sleep pressure and responses to sustained attention in the suprachiasmatic area'. *Science* **328**, 309 (2010).
41. 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).
42. 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. USA* **101**, 9843–9848 (2004).
43. 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).
44. 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).
45. Sasaki, M., Shibata, E., Kudo, K. & Tohyama, K. Neuromelanin-sensitive MRI: basics, technique, and clinical applications. *Clin. Neuroradiol.* **18**, 147–153 (2008).
46. Sasaki, M. et al. Neuromelanin magnetic resonance imaging of locus ceruleus and substantia nigra in Parkinson's disease. *Neuroreport* **17**, 1215–1218 (2006).
47. Liu, K. Y. et al. Magnetic resonance imaging of the human locus coeruleus: a systematic review. *Neurosci. Biobehav. Rev.* **83**, 325–355 (2017).
48. Keren, N. I. et al. Histologic validation of locus coeruleus MRI contrast in post-mortem tissue. *Neuroimage* **113**, 235–245 (2015).
49. 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).
50. 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).
51. 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).
52. 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).
53. 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 (SAGE Publications, 2008).
54. Jones, R. N. et al. A growth curve model of learning acquisition among cognitively normal older adults. *Exp. Aging Res.* **31**, 291–312 (2005).
55. Zimprich, D. & Rast, P. Verbal learning changes in older adults across 18 months. *Neuropsychol. Dev. Cogn. B* **16**, 461–484 (2009).
56. 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 (Springer, 1988).
57. Curran, P. J., Obeidat, K. & Losardo, D. Twelve frequently asked questions about growth curve modeling. *J. Cogn. Dev.* **11**, 121–136 (2010).
58. Kievit, R. A. et al. Developmental cognitive neuroscience using latent change score models: a tutorial and applications. *Dev. Cogn. Neurosci.* **33**, 99–117 (2018).
59. Brown, T. A. *Confirmatory Factor Analysis for Applied Research.* (Guilford Press, 2006).
60. Eid, M., Gollwitzer, M. & Schmitt, M. *Statistik und Forschungsmethoden: Lehrbuch. Grundlagen Psychologie* (Beltz, 2015).
61. 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).
62. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
63. Steiger, J. H. Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychol. Methods* **9**, 164–182 (2004).
64. Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87**, 245–251 (1980).
65. Schmidt, M. *Rey Auditory Verbal Learning Test: A Handbook* (Western Psychological Services, 2004).
66. 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).
67. 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).
68. 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).
69. 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. Phys. Biol. Med.* **30**, 121–125 (2017).
70. 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).
71. Weinshenker, D. Long road to ruin: noradrenergic dysfunction in neurodegenerative disease. *Trends Neurosci.* **41**, 211–223 (2018).
72. Chalermpalanupap, T., Weinshenker, D. & Rorabaugh, J. M. Down but not out: the consequences of pretangle tau in the locus coeruleus. *Neural Plast.* https://doi.org/10.1155/2017/7829507 (2017).
73. Jagust, W. Imaging the evolution and pathophysiology of Alzheimer disease. *Nat. Rev. Neurosci.* **19**, 687–700 (2018).
74. Robertson, I. H. A noradrenergic theory of cognitive reserve: implications for Alzheimer's disease. *Neurobiol. Aging* **34**, 298–308 (2013).
75. 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).
76. 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).
77. 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).

78. Fries, P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **9**, 474–480 (2005).
79. Li, S.-C., Lindenberger, U. & Sikstrom, S. Aging cognition: from neuromodulation to representation. *Trends Cogn. Sci.* **5**, 479–486 (2001).
80. Shing, Y. L. et al. Episodic memory across the lifespan: the contributions of associative and strategic components. *Neurosci. Biobehav. Rev.* **34**, 1080–1091 (2010).
81. Ehrenberg, A. J. et al. Quantifying the accretion of hyperphosphorylated tau in the locus coeruleus and dorsal raphe nucleus: the pathological building blocks of early Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **43**, 393–408 (2017).
82. Kaufman, S. K., Del Tredici, K., Thomas, T. L., Braak, H. & Diamond, M. I. Tau seeding activity begins in the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in Alzheimer's disease and PART. *Acta Neuropathol.* **136**, 57 (2018).
83. Heinsen, H. & Grinberg, L. T. On the origin of tau seeding activity in Alzheimer's disease. *Acta Neuropathol.* **136**, 815–817 (2018).
84. Grinberg, L. T. & Heinsen, H. Light at the beginning of the tunnel? Investigating early mechanistic changes in Alzheimer's disease. *Brain* **140**, 2770–2773 (2017).
85. Zarow, C., Lyness, S. A., Mortimer, J. A. & Chui, H. C. Neuronal loss is greater in the locus coeruleus than nucleus basalis and substantia nigra in Alzheimer and Parkinson diseases. *Arch. Neurol.* **60**, 337 (2003).
86. Chalermpalanupap, T. et al. Locus coeruleus ablation exacerbates cognitive deficits, neuropathology, and lethality in P301S tau transgenic mice. *J. Neurosci.* **38**, 74–92 (2018).
87. Betts, M. J., Ehrenberg, A. J., Hämmerer, D. & Düzel, E. Commentary: locus coeruleus ablation exacerbates cognitive deficits, neuropathology, and lethality in P301S tau transgenic mice. *Front. Neurosci.* **12**, 401 (2018).
88. Rorabaugh, J. M. et al. Chemogenetic locus coeruleus activation restores reversal learning in a rat model of Alzheimer's disease. *Brain* **140**, 3023–3038 (2017).
89. Folstein, M. F., Folstein, S. E. & McHugh, P. R. Mini-Mental State. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
90. 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 (Springer, 2015).
91. Gerstorf, D. et al. The Berlin Aging Study II – an overview. *Gerontology* **62**, 311–315 (2016).
92. Bertram, L. et al. Cohort profile: the Berlin Aging Study II (BASE-II). *Int. J. Epidemiol.* **43**, 703–712 (2014).
93. 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).
94. von Oertzen, T., Brandmaier, A. M. & Tsang, S. Structural equation modeling with Ωnyx. *Struct. Equ. Modeling* **22**, 148–161 (2015).
95. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
96. Avants, B. B., Tustison, N. & Song, G. Advanced normalization tools: V1.0. *Insight J.* **2**, 1–35 (2009).
97. Klein, A. et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46**, 786–802 (2009).
98. Friston, K. J. Experimental design and Statistical Parametric Mapping. in *Human Brain Function* (eds Frackowiak, R. S. J. et al.) 599–632 (Academic Press, 2004).
99. Dahl, M. J., Ilg, L., Li, S.-C., Passow, S. & Werkle-Bergner, M. Diminished pre-stimulus alpha-lateralization suggests compromised self-initiated attentional control of auditory processing in old age. *Neuroimage* **197**, 414–424 (2019).

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) under grant nos. 16SV5536 K, 16SV5537, 16SV5538, 16SV5837, 01UW070 and 01UW0808. Additional contributions (for example, financial, equipment, logistics, personnel) were provided by each of the other participating sites—that is, the Max Planck Institute for Human Development, Max Planck Institute for Molecular Genetics, Charité-Universitätsmedizin, German Institute for Economic Research and Humboldt-Universität zu Berlin, all located in Berlin, Germany; and the Universities of Lübeck and Tübingen, Germany. For further information about the BASE-II project, see https://www.base2.mpg.de/en. M.W.-B. received support from the German Research Foundation (grant no. WE 4269/5-1) and the Jacobs Foundation (Early Career Research Fellowship 2017–2019). M.J.D. recieves support from a fellowship with the International Max Planck Research School on the Life Course (http://www.imprs-life.mpg.de/en/). M.J.D. is recipient of a stipend from the Sonnenfeld-Foundation (http://www.sonnenfeld-stiftung.de/en/). M.M. was supported by an Alexander von Humboldt fellowship and by National Institutes of Health (grant no. R01AG025340). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank A. Bender, M. Betts and M. Sander for valuable discussions and assistance. We are grateful to S. Bachman and D. Zorbek, who performed the manual tracing of the LC, as well as Y. Köhncke and Y. Fandakova for statistical advice and M. Krause for help with cluster computing.

Author contributions

U.L. and S.K. designed the broader BASE-II study. M.J.D., M.W.-B., M.M., S.K. and N.C.B. designed the additional LC component. S.D. performed the experiments. M.J.D. and M.W.-B. analysed the data. M.J.D. and M.W.-B. wrote the manuscript. U.L., S.K., M.M., S.D. and N.C.B. gave conceptual advice. All authors revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0715-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.J.D. or M.W.-B.

Peer Review Information: Primary Handling Editor: Marike Schiffer

Corresponding author(s): Markus Werkle-Bergner
Martin J. Dahl

Last updated by author(s): Jul 3, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Cognitive data was collected using a computerized battery based on custom code (available upon reasonable request). MRI data was collected using a 3-Tesla Siemens Magnetom Tim Trio Scanner running standard Siemens operating software.
Data analysis	Structural equation models were calculated using Onyx v.1.0-991. MRI data was analyzed using Matlab including the SPM 12 and fieldtrip toolboxes. For MRI template generation and coregistration, ANTs v.2.10 and MRIcron v.1 was employed. Manual locus coeruleus assessments were performed using ImageJ v.1.49. Correlations and ICC were computed using IBM SPSS v.24

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that our results are based on are available from the BASE-II steering committee (https://www.base2.mpg.de/en/project-information/team) upon approved research proposal. For inquiries please contact Dr. Katrin Schaar, BASE-II project coordinator, schaar@mpib-berlin.mpg.de. 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).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All participants of the Berlin Aging Study-II (BASE-II), an ongoing longitudinal aging study, with full cognitive and magnetic resonance imaging (MRI) data were included in the study, i.e., we did not conduct a formal power calculation, given that there was no available prior evidence on the studied phenomena.
Data exclusions	A subset of 323 BASE-II participants underwent 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 (0 female; mean age: 76.66 years; SD = 1.64; range = 74.93–78.74; at time point (T) 2) were excluded from further analyses due to excessive movement artifacts (n = 2) or incorrect scan positioning (n = 3). The final MRI subsample (see Supplementary Table 1) 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).
Replication	Making use of the comprehensive cognitive battery available for this data set, we integrated performance over a variety of episodic memory tasks to retrieve a general memory factor and related this to locus coeruleus integrity. The obtained findings replicate those reported for the Rey Auditory Verbal Learning Task and indicate a task-independent association between locus coeruleus integrity and memory performance. In addition, locus coeruleus integrity was assessed both manually and automatically. Both approaches produce qualitatively similar results (positive association between locus coeruleus integrity and memory). Finally, we repeated our analyses (relating memory performance and locus coeruleus integrity) using cognitive data acquired at an earlier time point (about 2 years before MRI assessments took place) and retrieve qualitatively similar results.
Randomization	Randomization between experimental groups was not relevant in the current study as it did not include different experimental conditions. Participants were allocated to groups based on their age (i.e., non-random).
Blinding	Manual locus coeruleus intensity assessment was performed blinded. Beyond this, blinding was not relevant in the current study as it did not include different experimental conditions. Identical computer code was used to perform the analyses in both age groups. Staff involved in data collection was not involved in data analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The final MRI subsample (see Supplementary Table 1) 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.
Recruitment	Data of the Berlin Aging Study-II was analyzed. For details, please refer to: Bertram, L. et al. Cohort profile: The Berlin Aging Study II (BASE-II). Int. J. Epidemiol. 43, 703–712 (2014).

In short:
Only residents of the greater metropolitan area of Berlin, Germany, were eligible for participation in BASE-II. Potential participants were drawn from a pool of individuals originally recruited at the Max-Planck-Institute for Human Development as part of a number of earlier projects with a focus on neurocognition (a detailed description of these projects can be found at: http://www.mpib-berlin.mpg.de/sites/default/files/media/pdf/25/lip_report_11.pdf).
Briefly, participant recruitment for these and other studies was based on advertisements in local newspapers and the public commuter transport system. This led to approximately 10 000 responders of whom 2875 were invited for an additional screening (either in-house or by telephone), leading to 2262 individuals eligible for inclusion in BASE-II, i.e. 79% of those who were initially invited. From these, we selected 2200 individuals to represent the BASE-II baseline cohort based on their age and sex as follows. A total of 1600 participants were assigned to an older subgroup aged between 60 and 80 years, whereas the remaining 600 individuals were assigned to a younger subgroup (serving as a reference population) aged between 20 and 35 years. By design, each age subgroup contains equal numbers of males and females. See Bertram et al., 2014 for other socio-demographic details of the BASE-II baseline cohort. In the present study, only the subset of participants with full cognitive and MRI data are included.

Ethics oversight

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.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type

We analyzed structural MRI data only.

Design specifications

We analyzed structural MRI data only.

Behavioral performance measures

We analyzed structural MRI data only. Behavioral (cognitive) data was not assessed within the scanner.

Acquisition

Imaging type(s)

structural

Field strength

3T

Sequence & imaging parameters

Structural 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 T1-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 mm3. Pre-scan normalize and 3D distortion correction options were enabled.
Based on this whole-brain MPRAGE sequence, a neuromelanin-sensitive high-resolution, two-dimensional T1-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 mm3. 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.

Area of acquisition

Whole brain (MPRAGE), Brainstem, covering the complete pons (TSE, see above)

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software

MRI data was analyzed using Matlab including the SPM 12 and fieldtrip toolboxes. For MRI template generation and coregistration, ANTs v.2.10 and MRIcron v.1 was employed. Manual locus coeruleus assessments were performed using ImageJ v.1.49.

Normalization

Prior to analysis, all whole-brain and brainstem MRI scans (MPRAGE, TSE (see above)) were up-sampled to twice the initial matrix size by sinc-interpolation to improve visualization of the LC as previously described.
Initially (Step 1), using Advanced Normalization Tools v. 2.1 (ANTS), 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. 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. Histogram matching was enabled since we registered scans from the same modality (T1-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. 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. 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 → Template_whole and Template_whole → MNI0.5) to the LC probability map.

Normalization template: 0.5 mm iso-voxel MNI 152

Noise and artifact removal: No correction for motion parameters, physiological signals etc. was applied.

Volume censoring: No volume censoring was applied.

Statistical modeling & inference

Model type and settings: 1) The association between cognitive and MRI factors was assessed using a multiple indicator structural equation model (SEM).
2) Age differences in locus coeruleus intensity ratios along the rostrocaudal extent of the nucleus were evaluated based on non-parametric, cluster-based, random permutation tests.

Effect(s) tested: 1) Covariances between memory performance and locus coeruleus integrity on a latent level.
2) Main effects of age group (using cluster tests), followed up by a Topography x Age group interaction analysis (using ANOVA)

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s): We produced a search space and probabilistic map for the region of interest (locus coeruleus).

Statistic type for inference (See Eklund et al. 2016): 2) Cluster-wise (see below).

Correction: To investigate age differences in LC intensity ratios along the rostrocaudal extent of the nucleus, 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. 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.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis