Summary statistics of memory-related fMRI activity reflect dissociable neuropsychological and anatomical signatures of neurocognitive aging

Anni Richter*[1], Joram Soch[2,3], Jasmin M. Kizilirmak[2], Larissa Fischer[1], Hartmut Schütze[4,5], Anne Assmann[5], Gusalija Behnisch[1], Hannah Feldhoff[1], Lea Knopf[1], Matthias Raschick[1], Annika Schult[1], Constanze I. Seidenbecher[1,6], Renat Yakupov[4], Emrah Düzel[4,5,6] & Björn H. Schott*[1,2,6,7]

[1] Leibniz Institute for Neurobiology (LIN), Magdeburg, Germany

[2] German Center for Neurodegenerative Diseases (DZNE), Göttingen, Germany

[3] Bernstein Center for Computational Neuroscience (BCCN), Berlin, Germany

[4] German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany

[5] Otto von Guericke University, Medical Faculty, Magdeburg, Germany

[6] Center for Behavioral Brain Sciences (CBBS), Magdeburg, Germany

[7] Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany

***Address for correspondence:**

Dr. Anni Richter

Leibniz Institute for Neurobiology (LIN)

Brenneckstraße 6

39118 Magdeburg, Germany

arichter@lin-magdeburg.de

PD Dr. Dr. Björn Hendrik Schott

Leibniz Institute for Neurobiology (LIN)

Brenneckestr. 6

39118 Magdeburg, Germany

bschott@lin-magdeburg.de / bjoern-hendrik.schott@dzne.de

Key words: subsequent memory effect, episodic memory, fMRI, biomarker, hippocampus, cognitive aging, memory impairment

Abstract

Memory-related functional magnetic resonance imaging (fMRI) activations show characteristic age-related differences, but their use as biomarkers for neurocognitive aging and ultimately dementia risk states is complicated by their high dimensionality. Here, we make a step towards biomarker validation of two single-value scores reflecting deviations from prototypical fMRI activity patterns of young adults computed for the novelty and subsequent memory contrasts from an incidental visual memory encoding fMRI experiment. To optimize procedures of fMRI phenotyping we compared associations of the scores with dementia-relevant variables: cognitive performance in multiple domains and brain structure in 153 healthy older adults. While novelty-driven scores specifically correlated with hippocampus-dependent memory performance, memory-driven scores additionally correlated with global cognition and medial temporal gray matter volumes. Our results reveal complementary associations of reductionist fMRI-based scores with cognitive and structural measures relevant for future validation studies in (pre-)clinical samples, paving the way towards biomarker validation.

1. Introduction

Even healthy older adults commonly exhibit cognitive decline and brain structural alterations[1-3]. Early detection of age-related neurocognitive changes can help identify those at increased risk for neurodegenerative diseases like Alzheimer's disease (AD), paving the way for early interventions[4]. Valid and complementary markers of cognitive and functional impairment could facilitate the assessment of age-related neurocognitive changes and provide valuable information about an individual's extent of brain aging[5-8]. As suggested by Hedden et al.[9], markers that rely on age-related alterations of brain structure and function can be referred to as brain markers or, if obtained using imaging techniques, as imaging biomarkers. Examples include differences in gray matter volume (GMV)[10,11], white matter (WM) lesion load[8,12], memory-related functional magnetic resonance imaging (fMRI)[13-16], and electrophysiological measures[17]. Other indicators of healthy versus pathological brain aging are disease markers, which encompass, among others, positron emission tomography (PET) measures of beta-amyloid (Aβ) and tau deposition[18], but also neuropsychological markers like global cognition, executive function, and episodic memory as assessed with neuropsychological tests[19].

In clinical settings, neuropsychological testing is typically a gatekeeper for subsequent assessments, that is, only patients with pathological or at least suspicious test results undergo biomarker assessment[6,20]. Especially episodic memory tests of delayed free and cued recall represent highly sensitive measures of memory decline in the typical presentation of AD. Apart from cerebrospinal fluid Aβ42 concentrations[21], medial temporal lobe atrophy is most likely furthest through validation[6], but must, at present, also be considered insufficient as a single biomarker[22].

We previously found that, compared to young individuals, older adults exhibited lower activations of inferior and medial temporal structures and reduced deactivations in the Default Mode Network (DMN) during novelty processing and successful long-term memory

encoding[13]. To capture age-related deviations from the prototypical fMRI activations in younger participants, we proposed the use of reductionist fMRI-based scores:

I. The FADE score (*Functional Activity Deviations during Encoding*[14]), which reflects the *difference* of activations outside and inside a mask representing prototypical activations in a young reference sample, and

II. the SAME score (*Similarity of Activations during Memory Encoding*[13]), which reflects the *similarity* of an older adult's brain response with activation and deactivation patterns seen in young subjects, and furthermore accounts for the inter-individual variability within the reference sample.

Both markers constitute single-value scores and can be computed either from fMRI novelty (novel vs. highly familiarized images) or subsequent memory contrasts (based on a subsequent recognition memory rating of the to-be-encoded images). They thus reflect age-related processing differences in either novelty detection or successful encoding, which engage overlapping, but partly separable neural networks[13,23,24], with novelty detection not directly translating to encoding success[25]. Scores based on novelty detection versus encoding success may thus indicate age-related deviations in at least partly different cognitive domains.

The FADE and SAME scores have previously been associated with memory performance in the encoding task they were computed from[13,14], but it is yet unclear whether this relationship is also found with independent, classical neuropsychological assessments of memory. Furthermore, it is not yet known whether the scores are specifically related to hippocampus-dependent memory performance or rather global cognitive function in old age. To evaluate, which processes (hippocampus-dependent memory processes vs. other cognitive tasks) are significantly related to the four fMRI-based single-value scores (i.e., FADE vs. SAME x novelty vs. subsequent memory) and specifically to age-related differences, we performed step-wise correlational analyses for each age group. Firstly, we computed correlations between the imaging scores to assess their potential similarities. We then tested

their relationship with performance in different memory tests and additionally in other psychometric tasks covering a wide range of cognitive functions. Finally, we assessed associations between the imaging scores and brain morphometric measures (local GMV, WM lesion volume). For an overview of our approach, see Figure 1.

2. Results

2.1. Demographic data

The previously described study cohort[13,23] consisted of 259 healthy adults, including 106 young (47 male, 59 female, age range 18-35, mean age 24.12 ± 4.00 years), and 153 older (N = 153, 59 male, 94 female, age range 51-80, mean age 64.04 ± 6.74 years) participants (for details see methods section and Supplementary Table S1). Age groups did not differ significantly with respect to gender ratio, ethnic composition or ApoE genotype (χ^2 tests: all $p > .088$, see Table S1). There were significant differences regarding medication, endocrine-related surgeries (e.g. thyroidectomy and oophorectomy), and level of education: 94% of young subjects, but only about 50% of the older subjects had received the German graduation certificate qualifying for academic education ("Abitur"), most likely due to historical differences in educational systems (for a discussion, see [13], Supplementary Material). Using the Multiple-Choice Vocabulary Test (MWT-B[26]), a multiple-choice vocabulary-based screening of verbal intelligence, we could confirm that older participants had comparable or superior verbal knowledge ($z = -8.11$, $p < .001$), which did not correlate with the imaging scores (all $p > .203$).

Age groups differed significantly for all imaging scores (two-sample *t*-tests: all $p < .001$), except for the FADE score computed from the novelty contrast (see [13]; $p = .910$). All results reported below focus on the older adults. For completeness, main figures and tables also show the results from young participants.

2.2. Voxel-wise representation and inter-correlation of the imaging scores

As an initial, exploratory, analysis, we computed voxel-wise regressions of the fMRI novelty and subsequent memory contrasts with the imaging scores in the older age group. Note that this analysis is partly circular as the imaging score of each participant were computed from the

individual fMRI contrasts. These results are thus reported for illustrative purposes, to help interpreting the subsequently reported results. Figure 2 shows that the FADE score computed from the novelty contrast (hereafter: FADE novelty score) is rather specifically associated with an occipital and parahippocampal network, while the SAME scores additionally capture a wide range of processes in the default mode network (DMN; i.e., precuneus and medial prefrontal cortex), which can mainly be attributed to the score's negative components. All scores significantly correlated with the contrast they were constructed from (see Supplementary Tables S4-7 for details). Additionally, the SAME score computed from the novelty contrast (hereafter: SAME novelty score) showed a significant positive correlation with the fMRI memory effect in the striatum, precuneus, and middle occipital gyrus (see Figure 3 and Table S8).

To investigate the scores' similarity, we correlated them with each other. The scores obtained from the same contrast, that is, novelty or memory, showed significant negative correlations (all $p < .001$; see Figure 4), reflecting the fact that FADE and SAME scores were constructed in opposite ways. Importantly, neither FADE nor SAME scores obtained from the different contrasts (i.e. novelty processing vs. subsequent memory) correlated significantly with each other ($p > .768$), suggesting that they assess different constructs. The remaining correlations were not significant ($p > .092$).

We previously observed that older adults exhibited lower activations of inferior and medial temporal structures, particularly of the parahippocampal cortex, compared to younger participants, accompanied by relatively reduced deactivations in midline structures of the DMN[13]. The SAME scores can be split into separate components reflecting activations versus deactivations. *Post-hoc* correlational analysis with the SAME scores' activation and deactivation components revealed that both components contributed to the correlations with the FADE scores (novelty: activation: $r = -.646$, $p < .001$, deactivation: $r = -.160$, $p = .048$; memory: activation: $r = -.670$, $p < .001$, deactivation: $r = -.434$, $p < .001$). As expected, the correlations of the FADE scores with the activation components of the SAME scores were stronger than

those with the deactivation components (Fisher's z-test for dependent correlation coefficients: novelty: $z = -4.46$, $p < .001$, memory: $z = -2.68$, $p = .007$).

2.3.The imaging scores correlate with different tests of episodic memory

As the imaging scores were obtained from an fMRI paradigm targeting episodic memory encoding, we first tested for associations with performance in episodic memory tests. These included the recognition memory test of the fMRI experiment itself (70 minutes after onset of the experiment) as well as 30-minutes and one-day delayed recalls of the Verbal Learning and Memory Test (VLMT[27]) and the Logical Memory subtest from the Wechsler Memory Scale (WMS[28]). As expected, older participants performed significantly worse in all memory tests compared to young participants (all $p < .001$; see Table 1).

As in our previous study[13], we found significant correlations with memory performance for the pictures shown during fMRI scanning for all imaging scores (hereafter FADE A'; all $p < .001$; see Figure 5 and Supplementary Table S3), except for the FADE novelty score ($p = .372$). Owing to the construction of the scores, correlations with the FADE score (which focuses on deviations from young adults' prototypical activation patterns) were negative, while correlations with the SAME scores (which focus on similarities) were positive.

The SAME score computed from the memory contrast (hereafter: SAME memory score) was the only score that significantly correlated with the delayed recall phases of the VLMT and thus with all memory tests (all $p < .011$; see Figure 5 and Table S3). This score also showed the nominally highest correlations with most memory tests in the cohort of young subjects (see Figure 5 and Table S3). All scores correlated with the performance in the WMS logical memory test (all $p < .011$; see Figure 5 and Table S3). The highest correlations in terms of absolute values were observed with the FADE novelty score (30-minutes delayed recall: $r = -0.332$, $p < .001$; one-day delayed recall: $r = -0.326$, $p < .001$).

Next, we explored whether the observed correlations with the SAME scores were carried by their activation or deactivation components. The correlation of the SAME novelty score with FADE A' was carried by the deactivation component (activation: p = .794, deactivation: r = .267, p = .001; Table S3). This may be a reason why the FADE novelty score did not correlate with FADE A', as it did not consider deactivation differences between young and older subjects. For the SAME memory score, both components contributed to the correlation with FADE A' (activation: r = .235, p = .004, deactivation: r = .329, p < .001; Fisher's test for dependent correlation coefficients: z = -0.81, p = .421).

While correlations of the SAME memory score with VLMT delayed recalls were driven by the deactivation component (activation: all p > .246, deactivation: all p = .006; Table S3), correlations of the SAME novelty and memory scores with WMS delayed recalls were carried by the activation component (all p < .047, deactivation: all p > .161).

2.4.Correlations of the imaging scores with global cognition

To evaluate the utility of our imaging scores as potential biomarkers for neurocognitive aging beyond hippocampus-dependent memory, we performed correlational analyses with neuropsychological tests of other cognitive constructs. Compared to younger participants, older participants performed significantly worse in all neuropsychological tests (all p < .001; see Table 1). We computed a linear discriminant analysis (LDA) to reduce the number of tests and to obtain a proxy for global cognition by including the composite score gained from the discriminant function. Of our 376 subjects (including a young replication sample to increase sample size[29]), 107 could not be included in the LDA due to at least one missing value. The final LDA thus included 269 subjects (158 young and 111 older participants). Five variables significantly contributed to the discrimination between age groups as part of the discriminant function (Wilks' Λ = .348, p < .001):

I. the number of words recalled in the distractor trial of the VLMT (standardized canonical discriminant coefficient: .277),

II. the number of words recalled in the one-day delayed recall of the VLMT (.364),

III. the corrected hit rate in the 2-back task (.260),

IV. the reaction time (RT) in the flexibility task (-.478), and

V. the RT of alertness trials with tone (-.225).

90.1 % of the participants could successfully be classified as either young or old using the discriminant function (young subjects: 92.8 %; older subjects: 86.4 %). We focused our correlational analysis on these variables best discriminating between age groups, except for the VLMT one-day delayed recall, which was already considered in our analysis of episodic memory tests. The FADE novelty score showed a significant negative correlation with the recall of the VLMT distractor list (r = -.206, p = .011) and the FADE score computed from the memory contrast (hereafter: FADE memory score) showed a significant positive correlation with the RT in the flexibility task (r = .242, p = .003; see Figure 6). After Holm-Bonferroni correction for the number of variables from which correlations with the imaging scores were computed, no further correlations were significant (all other p > .044).

Regarding the discriminant function as a proxy for global cognition, both scores obtained from the memory contrast showed significant correlations (FADE memory: r = -.204, p = .019, SAME memory: r = .213, p = .014; see Figure 6). When evaluating whether the SAME memory score's correlation was carried by the activation or deactivation component, we observed a significant positive correlation with the deactivation component only (activation: p = .417, deactivation: r = .211, p = .015).

2.5.Correlations of the imaging scores with brain morphology

Next, we investigated the relationship of the imaging scores with age-related variability in brain morphology. In line with previous studies[12], older compared to young participants had

significantly lower GMV ($t = 6.89$; $p < .001$) and higher WM lesion volumes (Mann-Whitney U = 2001.00, $p < .001$).

We observed no significant correlations between the imaging scores and WM lesion volume (Spearman's ρ: all $p > .223$). For their relationship with local GMV using Voxel-based morphometry (VBM), we detected significant correlations of the memory scores with medial temporal lobe structures in older adults (see Figure 7 and Supplementary Table S9). The SAME memory score additionally showed correlations with local GMV in superior and inferior frontal gyrus, while the FADE memory score was additionally correlated with middle occipital gyrus GMV. *Post-hoc* analysis for the SAME memory score components revealed that the correlations were driven by the activation component while no correlations were observed for the deactivation component (see Supplementary Table S10). Furthermore, no correlations were observed for the novelty scores. The respective results from young participants can be found in Supplementary Table S11.

3. Discussion

Quantification of neurocognitive aging and early detection of pathological cognitive decline may help to identify individuals at increased risk and ultimately develop targeted early interventions to slow progression of cognitive impairment. Especially early lifestyle interventions, tackling physical exercise, nutrition, and to some degree cognitively demanding tasks, can be helpful to preserve healthy aging[30-33]. However, an accurate assessment of cognitive, but also neurophysiological, decline poses a major challenge due to the complexity of brain processes and functions, as well as the non-linear acceleration of cognitive decline[34].

In previous studies, comprehensive scores reflecting memory-related fMRI activations and deactivations have been constructed as potential biomarkers for neurocognitive aging (FADE and SAME scores)[13,14]. After this first step towards validation (phase 1: preclinical exploratory studies according to Frisoni et al.[6]), we now aimed to further evaluate the biological and potential clinical relevance of these scores by investigating their relationship with performance in an extensive neuropsychological testing battery as well as brain morphological measures (phase 2: assessing variables associated with biomarkers status[6]).

3.1.Neurocognitive correlates of the FADE and SAME imaging scores

While we had initially expected that, by considering both deactivation and activation deviations, the SAME score would constitute a more comprehensive or accurate measure, we found relatively few differences between the SAME and FADE scores computed from the same fMRI contrasts (i.e., novelty processing vs. subsequent memory). Instead, the fMRI contrasts had considerable influence. This already became evident from the inter-correlations of the imaging scores. We observed high correlations between the FADE and SAME scores derived from the same fMRI contrasts, while neither the FADE nor SAME scores computed from different fMRI contrasts correlated with each other. The implications are two-fold:

I. The FADE and SAME scores assess age-related deviation from (or similarity to) prototypical task-related activation patterns in younger participants to a comparable degree.

II. It is important from which functional contrast the scores are derived, as they appear to capture at least partly complementary information on age-related differences in cognitive function. The different contrasts reflect separable cognitive processes (novelty detection versus encoding strength), and they likely capture dissociable aspects of cognitive aging, as discussed below.

Imaging scores obtained from the novelty contrast could be relatively specifically associated with performance in episodic memory tasks (FADE & SAME scores: WMS; FADE only: VLMT distractor task, SAME only: FADE A'), and this association was found in older adults only. On the other hand, the imaging scores obtained from the memory contrast were significantly related to a broader set of cognitive functions in older adults, and to memory performance across age groups. Regarding neuropsychological measures, the FADE and SAME memory scores both significantly correlated with behavioral performance in the WMS, FADE A' and the global cognition score, which included measures of episodic memory, working memory, alertness, reaction speed, and cognitive flexibility. The FADE memory score was also significantly positive correlated with RTs in a flexibility task, and the SAME memory score was significantly positive associated with VLMT performance.

One explanation for the higher sensitivity of the memory scores to cognitive (behavioral) performance beyond episodic memory could be more pronounced age-related differences in the subsequent memory effect compared the novelty contrast[13]. While the subsequent memory effect is based on the participants' 5-point recognition-confidence ratings, the novelty contrast simply compares the neural responses to *de facto* novel versus highly familiarized images, not accounting for encoding performance. In fact, in our parametric design, variance attributable to successful memory formation was captured by the parametric

subsequent memory regressor[23]. Despite the overlap of brain networks involved in novelty detection and successful episodic encoding, there are differences in detail, and, importantly, the age-related between-group differences of the (parametrically modelled) subsequent memory effect are considerably more widespread than those of the novelty contrast[13]. The memory-related brain regions contributing to the scores such as the dorsolateral and ventrolateral prefrontal cortex, the parahippocampal gyrus and MTG are not only relevant for episodic encoding but also for cognitive processes like alertness[35] or working memory[36-38]. The novelty-related scores did not significantly correlate with any cognitive domain other than episodic memory and, furthermore, the correlations were exclusively observed in older adults. The robustness of the novelty-related activation patterns compared to the subsequent memory effect may be more preserved and less variable across the lifespan. When they do show age-related deviations, this may be indicative of a more pronounced age-related impairment of the hippocampus-dependent memory system. Compatible with this, attenuated hippocampal novelty responses have been linked to lower memory performance in individuals at risk for Alzheimer's disease (AD)[39].

3.2. Age-related variation in functional and structural neuroanatomy

Considering the rather specific link of the novelty-related scores with episodic memory performance in older adults, it may seem surprising that we did not observe a correlation of these scores with hippocampal GMV. One explanation for this could be that hippocampal volumes may correlate only moderately, if at all, with memory performance and fMRI indices of hippocampal functional integrity[40,41].

On the other hand, the FADE and SAME scores derived from the memory contrast did not only correlate with neurocognitive performance decrease, but also with morphometric changes reflecting age-related GMV loss. More specifically, we observed correlations between the memory scores and local GMV for parahippocampal gyrus, middle temporal gyrus and

prefrontal cortex using VBM. Importantly, all these correlations were observed in the older age group only, suggesting that they reflect individual differences related to aging rather than development or general cognitive ability. Concurrent brain structural alterations and lower cognitive performance in aging constitute a well-replicated finding. Hedden et al.[9] examined the relationship between age-related cognitive impairment and various brain markers (MRI and PET) and observed associations of striatal volume and WM integrity with processing speed and executive functions, and of hippocampal volume and amyloid load (as assessed with PET) with episodic memory. Considering the memory-related scores and their association with cognitive function beyond episodic memory and with brain morphology, our results are compatible with previous findings in other cohorts. Arvanitakis et al.[12] found lower whole-brain GMV to be associated with episodic memory performance and perceptual speed. Similarly, Tsapanou et al.[8] observed that age-related changes in episodic memory, processing speed and executive functions were associated with cortical thickness, WM hyperintensities and striatal volume. In a large cohort of over 3000 healthy participants, Zonneveld et al.[42] reported an association of global cognition with GMV in the left amygdala, hippocampus, parietal lobule, superior temporal gyrus, insula and posterior temporal lobe. One potential advantage of our fMRI-based scores becomes evident from the observation that correlations with memory performance were also found in young adults, whereas a relationship with variation in brain structure was only found in older adults. Future investigations should therefore explore the possibility that fMRI-based markers may be suitable as a predictor of cognitive functioning, even when age-related structural changes are not (yet) observable.

3.3. Deactivation of the Default Mode Network and cognitive function in old age

Regarding the relationships of the scores with neuropsychological assessment and neuroanatomy, a general pattern seems to emerge:

I. In most cases where an association with a FADE score was observed, we also observed a correlation with the SAME score, most often carried by its activation component.

II. In the few cases where the SAME compared to the FADE score could be associated with additional functions (e.g., FADE A' for the novelty score and VLMT delayed recall performance as well as local GMV in frontal cortex for the memory score), these associations were driven by the deactivation component of the SAME score.

This pattern can likely be attributed to the construction of the SAME score, including age-dependent differences in functional deactivation patterns, while the FADE score relies mostly on activation differences. Brain regions that showed prominent deactivations during successful memory encoding in the young participants included a network centered around the brain's midline that has previously been referred to as the DMN[43]. This observation is in line with a frequently cited meta-analysis by Maillet and Rajah[16], who found age-related differences in encoding-related processes encompassing under-recruitment of occipital, parahippocampal, and fusiform cortex, but over-recruitment of DMN regions including the medial prefrontal cortex (mPFC), precuneus, and left inferior parietal lobe in older adults. In the current study, the correlation of the SAME memory score with global cognition could be primarily accounted for by the deactivation component, which may, at least in part, reflect an older individual's general ability to suppress ongoing DMN activity during attention-demanding tasks. In line with this interpretation, reduced DMN deactivation has also been associated with lower working memory performance in older adults[38], and a meta-analysis revealed that reduced DMN deactivation in old age can be observed across a variety of cognitive tasks[44]. On the other hand, several authors discuss the role of the DMN as a potential cognitive resource in older adults[45,46], which should be further addressed in future studies (see Supplementary Discussion).

4.3 A potential role for the mesolimbic dopamine system in preserved cognition

Among the scores investigated here, the SAME novelty score stood out by showing a positive correlation with voxel-wise activations not only in novelty contrast (Figure 2), but also in the subsequent memory effect (Figure 3). Notably, the peak of this correlation was found in the striatum, a core output region of the midbrain dopaminergic nuclei. Previous studies have implicated the dopaminergic midbrain in successful encoding in young adults[47-49]. In older adults, striatal dopamine D2 receptor binding has been related to hippocampal-striatal functional connectivity and memory performance[50]. Importantly, novelty can induce midbrain activations[51,52], and structural integrity of the midbrain has been related to both midbrain and hippocampal novelty responses[53] and to memory performance in older adults[54]. Düzel et al.[55] proposed the NOMAD model, which suggests that novelty-related increase of mesolimbic dopaminergic activity promotes exploratory behavior and ultimately memory performance in older adults. In line with this framework, our results suggest that a preserved pattern of novelty-related brain activity may be related to increased activity of mesolimbic dopaminergic structures during successful memory formation in aging.

4.4 Implications for clinical research

As the present study was directed at the association between fMRI-based potential biomarkers and neurocognitive functioning in healthy older adults, the next step should be to test our scores in clinical populations or populations at risk. While some AD biomarkers are at an advanced stage of development, the validation of functional imaging markers and ultimately their implementation in clinical or research settings is in its early stages. Within the AD biomarker validation framework proposed by Frisoni et al.[6] our study covers key aspects of phase 2, as we aim to optimize fMRI phenotyping by assessing the relationship between fMRI biomarkers and AD-relevant scores from neuropsychological testing and brain morphology in healthy individuals. We particularly emphasize their potential for AD and mild cognitive impairment, the clinical condition with the greatest diagnostic uncertainty that would benefit most from

accurate diagnosis[6]. In older adults with such risk states, various biomarkers have been assessed for their potential clinical applicability. However, thus far, task-based fMRI has largely focused on dysfunctional hippocampal activity[56]. The medicore test-retest reliability of voxel-wise task-based fMRI has called into question its utility as a biomarker[57]. The reductionist summary statistics of whole-brain fMRI activation (and deactivation) patterns described here may prove more reliable and thus provide a useful additional tool to monitor and ultimately predict the transition from healthy aging to pathological cognitive decline. In this context, it is of importance that in recent studies with older participants at risk for AD, researchers have often employed novelty rather than subsequent memory contrasts, owing to the lack of successfully encoded items in individuals with pronounced memory impairment[39,40,46]. Our observation that the novelty-related scores, particularly the FADE novelty score, show relatively strong and specific correlations with tests of hippocampus-dependent memory, support the validity of this approach. It may nevertheless be of interest what the memory-related scores, and particularly the SAME memory score, signify in memory-impaired individuals. They may, for example, prove a useful tool in the assessment of cognitive impairment beyond the memory domain or in atypical presentations of pre-clinical dementia. The scores may also help to better understand and define "healthy aging" on a theoretical level and could facilitate the laborious screening of high-risk patients for pharmacological studies or may be combined with tau- or amyloid-PET[46] as a potential biomarker assessment at the clinical level.

4.4 Conclusion

Our results provide novel evidence for the validity of single-value fMRI-based scores as potential markers of cognitive ability in older adults. They further suggest that the scores provide complementary information with respect to relatively selective impairment of hippocampal function in old age versus general cognitive ability across ages and local GMV

loss in old age. Future research should address their utility and predictive value in clinical populations like AD risk states.

4. Methods

4.1. Participants

The previously described study cohort[13,23] consisted of 259 healthy adults, including 106 young (47 male, 59 female, age range 18-35, mean age 24.12 ± 4.00 years), 42 middle-aged (13 male, 29 female, age range 51-59, mean age 55.48 ± 2.57 years) and 111 older (46 male, 65 female, age range 60-80, mean age 67.28 ± 4.65 years) participants. Additionally, a replication cohort of 117 young subjects[29] (60 male, 57 female, age range 19-33, mean age 24.37 ± 2.60 years) served for outlier detection and a linear discriminant analysis (LDA). We found no significant differences for any of the imaging scores between middle-aged and older participants[13] (two-samples *t*-tests: all $p > .123$) and therefore combined middle-aged and older participants into one age group to increase the statistical power of the correlational analyses (N = 153, 59 male, 94 female, age range 51-80, mean age 64.04 ± 6.74 years).

According to self-report, all participants were right-handed, had fluent German language skills and did not take any medication for neurological or mental disorders. The Mini-International Neuropsychiatric Interview (M.I.N.I.[58,59]) was used to exclude present or past mental disorder, including alcohol or drug dependence.

Participants were recruited via flyers at the local universities (mainly the young subjects), advertisements in local newspapers (mainly the older participants) and during public outreach events of the institute (e.g., *Long Night of the Sciences*).

Data were collected at the Leibniz Institute for Neurobiology in Magdeburg in collaboration with the German Center for Neurodegenerative Diseases in Magdeburg and the Otto von Guericke University of Magdeburg as part of a project within the *Autonomy in Old Age* research alliance. All participants gave written informed consent in accordance with the Declaration of Helsinki (World Medical Association, 2013) and received financial compensation for

participation. The study was approved by the Ethics Committee of the Faculty of Medicine at the Otto von Guericke University of Magdeburg.

4.2. Neuropsychological assessment

We conducted a number of common psychometric tests that cover a wide range of psychological constructs like attention, different aspects of memory, including short- and long-term memory, working memory as well as executive functions, such as interference control and flexibility. The tests are described in detail in the Supplementary Material; the variables and psychological constructs are summarized in Table 1. Additionally, the Multiple-Choice Vocabulary Test (MWT-B[26]) was performed as a proxy for crystallized verbal intelligence. It consists of 37 items with increasing difficulty, each item containing one real word and four verbally similar but meaningless pseudo-words of which the participant has to mark the correct one.

4.3. Subsequent Memory Paradigm for fMRI

For the fMRI subsequent memory paradigm, participants performed an incidental visual memory encoding task with an indoor/outdoor judgment[40]. Subjects viewed photographs showing indoor and outdoor scenes, which were either novel at the time of presentation (44 indoor and 44 outdoor scenes) or were repetitions of two highly familiar “master” images (22 indoor and 22 outdoor trials), one indoor and one outdoor scene pre-familiarized before the actual experiment[23]. Thus, during encoding, every subject was presented with 88 unique (i.e. novel) images and 2 master images that were presented 22 times each. Participants were instructed to categorize images as “indoor” or “outdoor” via button press as the incidental encoding task (i.e., participants were unaware that their memory for the pictures would later be tested). Each picture was presented for 2.5 s, followed by a variable delay between 0.70 s and 2.65 s.

Approximately 70 minutes (70.19 ± 3.60 min) after the start of the fMRI session, subjects performed a computer-based recognition memory test outside the scanner, in which they were presented with the 88 images that were shown once during the fMRI encoding phase (*old*) and 44 images they had not seen before (*new*). Participants rated each image on a five-point Likert scale from 1 ("definitely new") over 3 ("undecided") to 5 ("definitely old"; for detailed experimental procedure, see [23,29]).

4.4. Magnetic Resonance Imaging

Structural and functional MRI data were acquired on two Siemens 3T MR tomographs (Siemens Verio: 58 young, 83 older; Siemens Skyra: 48 young, 70 older), following the exact same protocol as used in the DELCODE study[40,60].

A T1-weighted MPRAGE image (TR = 2.5 s, TE = 4.37 ms, flip-α = 7°; 192 slices, 256 x 256 in-plane resolution, voxel size = 1 x 1 x 1 mm) was acquired for co-registration and improved spatial normalization. Phase and magnitude fieldmap images were acquired to improve correction for artifacts resulting from magnetic field inhomogeneities (*unwarping*).

For functional MRI (fMRI), 206 T2*-weighted echo-planar images (TR = 2.58 s, TE = 30 ms, flip-α = 80°; 47 slices, 64 x 64 in-plane resolution, voxel size = 3.5 x 3.5 x 3.5 mm) were acquired in interleaved-ascending slice order (1, 3, …, 47, 2, 4, …, 46). The total scanning time during the task-based fMRI session was approximately 9 minutes[23].

4.4.1. Neuroimaging biomarkers (FADE and SAME scores)

Using Statistical Parametric Mapping, Version 12 (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/, University College London, UK), we generated single-subject contrast images representing effects of novelty processing (by contrasting novel with familiar images) and subsequent memory effects (by parametrically modulating the BOLD response to novel images as a function of later remembering or forgetting). Specifically, the effect of subsequent

memory on fMRI activity during encoding was quantified as the mean-centered and arcsine-transformed subject's response in a subsequent recognition memory test (ranging from 1 to 5).

As described previously[13] the FADE and SAME scores are based on:

I. computing a reference map showing significant activations (and, for the SAME score, additionally significant deactivations) on each of the two fMRI contrasts (i.e. novelty processing or subsequent memory) within young subjects, and

II. calculating summary statistics quantifying the amount of deviation (FADE score) or similarity (SAME score) for a given older subject with respect to the prototypical (de-) activations seen in young subjects.

More precisely, let J_+ be the set of voxels showing a positive effect in young subjects at an *a priori* defined significance level (here: $p < 0.05$, FWE-corrected, extent threshold k = 10 voxels), and let t_{ij} be the t-value of the i-th older subject in the j-th voxel on the same contrast. Then, the FADE score of this subject is given by

$$\mathrm{FADE}_i = \frac{1}{v}\sum_{j \notin J_+} t_{ij} - \frac{1}{v_+}\sum_{j \in J_+} t_{ij}$$

where v_+ and v is the number of voxels inside and outside J_+, respectively[13]. A larger FADE score signifies higher deviation of an older adult's memory – or novelty – response from the prototypical response seen in young adults.

Now consider J_-, the set of voxels showing a positive effect in young subjects at a given significance level. Furthermore, let $\hat{\beta}_j$ be the average contrast estimate in young subjects, let $\hat{\sigma}_j$ be the standard deviation of young subjects on a contrast at the j-th voxel, and let $\hat{\gamma}_{ij}$ be the contrast estimate of the i-th older subject at the j-th voxel. Then, the SAME score is given by

$$\mathrm{SAME}_i = \frac{1}{v_+}\sum_{j \in J_+} \frac{\hat{\gamma}_{ij} - \hat{\beta}_j}{\hat{\sigma}_j} + \frac{1}{v_-}\sum_{j \in J_-} \frac{\hat{\beta}_j - \hat{\gamma}_{ij}}{\hat{\sigma}_j}$$

where v_+ and v_- are the numbers of voxels in J_+ and J_-, respectively[13]. Note how the directions of the difference in the two sums are different, in order to accumulate reduced activations (sum over J_+) and reduced deactivations (sum over J_-). Thus, a higher SAME score indicates higher similarity of an older adult's brain response with the activation and deactivation patterns seen in young subjects. Simplified, this means that the magnitudes of the SAME (the higher the more similar) and FADE (the higher the less similar) scores have opposing meanings. As further becomes evident from the equation, the SAME score extends the concept underlying the FADE score by:

I. considering deactivation patterns in addition to activation patterns by quantifying reduced deactivations, and
II. accounting for the interindividual variability within the reference sample of young subjects via dividing by their estimated standard deviation.

As an initial, exploratory, analysis, we computed voxel-wise regressions of the fMRI novelty and subsequent memory contrasts with the imaging scores. Note that this analysis is partly circular as the imaging scores of each participant were computed from the individual fMRI contrasts. Results are reported at $p_{\text{cluster}} < 0.05$ using family-wise error rate (FWE) cluster-level correction and an uncorrected cluster-forming threshold of $p_{\text{voxel}} < 0.001$[61].

4.4.2. Brain morphometry

Voxel-based morphometry (VBM) analyses were conducted to examine morphological differences of local GMV employing CAT12 using the T1-weighted MPRAGE images. Data processing and analysis were performed as described previously[29,62,63], with minor modifications. Images were segmented into gray matter, WM and cerebrospinal fluid-filled spaces using the segmentation algorithm provided by CAT12. Segmented gray matter images were normalized to the SPM12 DARTEL template, employing a Jacobian modulation and keeping the spatial resolution at an isotropic voxel size of 1 mm^3. Normalized gray matter maps

were smoothed with an isotropic Gaussian kernel of 6 mm at FWHM. Statistical analysis was performed separately for both age groups using a regression model including total intracranial volume (TIV) as a covariate. Voxels outside the brain were excluded by employing threshold masking (relative threshold: 0.2) that removed all voxels whose intensity fell below 20% of the mean image intensity[64]. VBM results are reported at $p_{cluster} < 0.05$ using FWE cluster-level correction and an uncorrected cluster-forming threshold of $p_{voxel} < 0.001$[61].

Furthermore, we investigated individuals' brain volumes for WM lesions. Subcortical WM hyperintensities were determined via automatic segmentation in T2-weighted FLAIR images using the Lesion Segmentation Toolbox (LST v3.0.0; https://www.applied-statistics.de/lst.html) based on the Computational Anatomy Toolbox (CAT12; http://www.neuro.uni-jena.de/cat/, University Hospital Jena, Germany) as described previously[65]. For normalization purposes, WM lesion volume and GMV were divided by the estimated TIV[66].

4.5. Statistical analysis

Data were analyzed using IBM® SPSS® Statistics, Version 21. We performed step-wise correlational analyses separately for age groups. Firstly, we investigated the potential correlations of the imaging scores among each other. Secondly, we tested their relationship with performance in different memory tests. Thirdly, we correlated the scores with performance in other psychometric tasks covering a wide range of cognitive functions. Finally, we tested for associations between the imaging scores and brain morphometric measures. For an overview of our approach see Figure 1.

As the neuropsychological testing was quite extensive, we needed to reduce the number of variables to avoid excessive multiple testing. Therefore, we aimed to only include those that best separate the age groups. We thus computed a multivariate test of differences using a linear discriminant analysis (LDA). A full list of tests and variables included in our LDA can be found

in Table 1. To increase the number of young participants, we added the young replication cohort (see 2.1) to the analysis, as their neuropsychological assessment was performed in the same way. We excluded values that were classified as extreme outliers based on the interquartile range (IQR; x > 3rd quartile + 3*IQR, x < 1st quartile – 3*IQR) in the psychometric tasks separately for each age group (see Supplementary Table S2). We used the step-wise LDA method that stops including tests to the discriminant function (i.e. the linear combination of the performance in the tests that best differentiate between age groups) when there is no longer a significant change in Wilks' Lambda. With the final set of tests generated in this way, we computed correlational analyses with the SAME and FADE scores. Moreover, we used the composite score gained from the discriminant function as a proxy for global cognition.

For the memory test of the pictures shown during fMRI scanning, memory performance was quantified as A', the area under the curve (AUC) from the receiver-operating characteristic (ROC) describing the relationship between *false alarms* ("old" responses to new items) and *hits* ("old" responses to previously seen items; see [13], Appendix B).

For comparison of age groups, we used paired *t*-tests unless stated otherwise. Whenever Levene's test was significant, statistics were adjusted, but for better readability, uncorrected degrees of freedom are reported. For the correlational analysis, we used Pearson's correlations unless stated otherwise. To avoid alpha error accumulation due to multiple testing, we performed Holm-Bonferroni correction for the number of variables from which correlations with the imaging scores were computed. We compared dependent correlation coefficients as described by Meng et al.[67].

5. References

1 Anthony, M. & Lin, F. A Systematic Review for Functional Neuroimaging Studies of Cognitive Reserve Across the Cognitive Aging Spectrum. *Arch Clin Neuropsychol* **33**, 937-948, doi:10.1093/arclin/acx125 (2018).

2 Li, X. *et al.* Age-Related Decline in the Topological Efficiency of the Brain Structural Connectome and Cognitive Aging. *Cerebral cortex* **30**, 4651-4661, doi:10.1093/cercor/bhaa066 (2020).

3 Cabeza, R. *et al.* Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nat Rev Neurosci* **19**, 701-710, doi:10.1038/s41583-018-0068-2 (2018).

4 Naismith, S. L. *et al.* Early intervention for cognitive decline: is there a role for multiple medical or behavioural interventions? *Early Interv Psychiatry* **3**, 19-27, doi:10.1111/j.1751-7893.2008.00102.x (2009).

5 Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing. *Nature* **561**, 45-56, doi:10.1038/s41586-018-0457-8 (2018).

6 Frisoni, G. B. *et al.* Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *Lancet Neurol* **16**, 661-676, doi:10.1016/S1474-4422(17)30159-X (2017).

7 Jack, C. R., Jr. *et al.* Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* **12**, 207-216, doi:10.1016/S1474-4422(12)70291-0 (2013).

8 Tsapanou, A. *et al.* Brain biomarkers and cognition across adulthood. *Human brain mapping* **40**, 3832-3842, doi:10.1002/hbm.24634 (2019).

9 Hedden, T. *et al.* Multiple Brain Markers are Linked to Age-Related Variation in Cognition. *Cerebral cortex* **26**, 1388-1400, doi:10.1093/cercor/bhu238 (2016).

10 Diaz-de-Grenu, L. Z. *et al.* MRI detection of tissue pathology beyond atrophy in Alzheimer's disease: introducing T2-VBM. *NeuroImage* **56**, 1946-1953, doi:10.1016/j.neuroimage.2011.03.082 (2011).

11 Minkova, L. *et al.* Gray matter asymmetries in aging and neurodegeneration: A review and meta-analysis. *Human brain mapping* **38**, 5890-5904, doi:10.1002/hbm.23772 (2017).

12 Arvanitakis, Z. *et al.* Association of white matter hyperintensities and gray matter volume with cognition in older individuals without cognitive impairment. *Brain Struct Funct* **221**, 2135-2146, doi:10.1007/s00429-015-1034-7 (2016).

13 Soch, J. *et al.* A comprehensive score reflecting memory-related fMRI activations and deactivations as potential biomarker for neurocognitive aging. *Human brain mapping* **42**, 4478-4496, doi:10.1002/hbm.25559 (2021).

14 Duzel, E., Schutze, H., Yonelinas, A. P. & Heinze, H. J. Functional phenotyping of successful aging in long-term memory: Preserved performance in the absence of neural compensation. *Hippocampus* **21**, 803-814, doi:10.1002/hipo.20834 (2011).

15 Grady, C. L. & Craik, F. I. Changes in memory processing with age. *Curr Opin Neurobiol* **10**, 224-231, doi:10.1016/s0959-4388(00)00073-8 (2000).

16 Maillet, D. & Rajah, M. N. Age-related differences in brain activity in the subsequent memory paradigm: a meta-analysis. *Neurosci Biobehav Rev* **45**, 246-257, doi:10.1016/j.neubiorev.2014.06.006 (2014).

17 Babiloni, C. *et al.* What electrophysiology tells us about Alzheimer's disease: a window into the synchronization and connectivity of brain neurons. *Neurobiol Aging* **85**, 58-73, doi:10.1016/j.neurobiolaging.2019.09.008 (2020).

18 Knopman, D. S. *et al.* Entorhinal cortex tau, amyloid-beta, cortical thickness and memory performance in non-demented subjects. *Brain : a journal of neurology* **142**, 1148-1160, doi:10.1093/brain/awz025 (2019).

19 Hassenstab, J. *et al.* Neuropsychological Markers of Cognitive Decline in Persons With Alzheimer Disease Neuropathology. *J Neuropathol Exp Neurol* **74**, 1086-1092, doi:10.1097/NEN.0000000000000254 (2015).
20 Cerami, C. *et al.* Clinical validity of delayed recall tests as a gateway biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. *Neurobiol Aging* **52**, 153-166, doi:10.1016/j.neurobiolaging.2016.03.034 (2017).
21 Kokkinou, M. *et al.* Plasma and cerebrospinal fluid ABeta42 for the differential diagnosis of Alzheimer's disease dementia in participants diagnosed with any dementia subtype in a specialist care setting. *Cochrane Database Syst Rev* **2**, CD010945, doi:10.1002/14651858.CD010945.pub2 (2021).
22 Lombardi, G. *et al.* Structural magnetic resonance imaging for the early diagnosis of dementia due to Alzheimer's disease in people with mild cognitive impairment. *Cochrane Database Syst Rev* **3**, CD009628, doi:10.1002/14651858.CD009628.pub2 (2020).
23 Soch, J. *et al.* Bayesian model selection favors parametric over categorical fMRI subsequent memory models in young and older adults. *NeuroImage* **230**, 117820, doi:10.1016/j.neuroimage.2021.117820 (2021).
24 Maass, A. *et al.* Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding. *Nat Commun* **5**, 5547, doi:10.1038/ncomms6547 (2014).
25 Poppenk, J., Kohler, S. & Moscovitch, M. Revisiting the novelty effect: when familiarity, not novelty, enhances memory. *J Exp Psychol Learn Mem Cogn* **36**, 1321-1330, doi:10.1037/a0019900 (2010).
26 Lehrl, S. Mehrfachwahl-Wortschatz-Intelligenztest MWT-B (5th ed.). *Spitta* (2005).
27 Helmstaedter, C., Lendt, M. & Lux, S. Verbaler Lern-und Merkfähigkeitstest (VLMT). *Göttingen: Hogrefe* (2001).
28 Härting, C. *et al.* WMS-R–Manual. *Bern: Hans Huber Verlag* (2000).
29 Assmann, A. *et al.* Neurocan genome-wide psychiatric risk variant affects explicit memory performance and hippocampal function in healthy humans. *Eur J Neurosci* **53**, 3942-3959, doi:10.1111/ejn.14872 (2021).
30 Bishop, N. A., Lu, T. & Yankner, B. A. Neural mechanisms of ageing and cognitive decline. *Nature* **464**, 529-535, doi:10.1038/nature08983 (2010).
31 Franke, K. & Gaser, C. Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Front Neurol* **10**, 789, doi:10.3389/fneur.2019.00789 (2019).
32 Stern, Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol* **11**, 1006-1012, doi:10.1016/S1474-4422(12)70191-6 (2012).
33 Whitty, E. *et al.* Efficacy of lifestyle and psychosocial interventions in reducing cognitive decline in older people: Systematic review. *Ageing Res Rev* **62**, 101113, doi:10.1016/j.arr.2020.101113 (2020).
34 Vinke, E. J. *et al.* Trajectories of imaging markers in brain aging: the Rotterdam Study. *Neurobiol Aging* **71**, 32-40, doi:10.1016/j.neurobiolaging.2018.07.001 (2018).
35 Liu, J. *et al.* Disrupted functional network in patients with temporal lobe epilepsy with impaired alertness. *Epilepsy Behav* **101**, 106573, doi:10.1016/j.yebeh.2019.106573 (2019).
36 Steiger, T. K., Herweg, N. A., Menz, M. M. & Bunzeck, N. Working memory performance in the elderly relates to theta-alpha oscillations and is predicted by parahippocampal and striatal integrity. *Sci Rep* **9**, 706, doi:10.1038/s41598-018-36793-3 (2019).
37 Steffener, J., Barulli, D. & Hill, B. Neural capacity limits on the responses to memory interference during working memory in young and old adults. *PLoS One* **15**, e0236897, doi:10.1371/journal.pone.0236897 (2020).
38 Sambataro, F. *et al.* Age-related alterations in default mode network: impact on working memory performance. *Neurobiol Aging* **31**, 839-852, doi:10.1016/j.neurobiolaging.2008.05.022 (2010).
39 Duzel, E. *et al.* Amyloid pathology but not APOE4 status is permissive for tau-related hippocampal dysfunction. *Brain : a journal of neurology* (in press).

40 Duzel, E. *et al.* CSF total tau levels are associated with hippocampal novelty irrespective of hippocampal volume. *Alzheimers Dement (Amst)* **10**, 782-790, doi:10.1016/j.dadm.2018.10.003 (2018).
41 Woodard, J. L. *et al.* Prediction of cognitive decline in healthy older adults using fMRI. *J Alzheimers Dis* **21**, 871-885, doi:10.3233/JAD-2010-091693 (2010).
42 Zonneveld, H. I. *et al.* High-Dimensional Mapping of Cognition to the Brain Using Voxel-Based Morphometry and Subcortical Shape Analysis. *J Alzheimers Dis* **71**, 141-152, doi:10.3233/JAD-181297 (2019).
43 Raichle, M. E. The brain's default mode network. *Annu Rev Neurosci* **38**, 433-447, doi:10.1146/annurev-neuro-071013-014030 (2015).
44 Li, H. J. *et al.* Putting age-related task activation into large-scale brain networks: A meta-analysis of 114 fMRI studies on healthy aging. *Neurosci Biobehav Rev* **57**, 156-174, doi:10.1016/j.neubiorev.2015.08.013 (2015).
45 Colangeli, S. *et al.* Cognitive Reserve in Healthy Aging and Alzheimer's Disease: A Meta-Analysis of fMRI Studies. *Am J Alzheimers Dis Other Demen* **31**, 443-449, doi:10.1177/1533317516653826 (2016).
46 Billette, O. V. *et al.* Novelty-related fMRI responses of precuneus and medial temporal regions in individuals at risk for Alzheimer's disease. (unpublished).
47 Wittmann, B. C. *et al.* Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron* **45**, 459-467, doi:10.1016/j.neuron.2005.01.010 (2005).
48 Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B. & Gabrieli, J. D. Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* **50**, 507-517, doi:10.1016/j.neuron.2006.03.036 (2006).
49 Schott, B. H. *et al.* The dopaminergic midbrain participates in human episodic memory formation: evidence from genetic imaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26**, 1407-1417, doi:10.1523/JNEUROSCI.3463-05.2006 (2006).
50 Nyberg, L. *et al.* Dopamine D2 receptor availability is linked to hippocampal-caudate functional connectivity and episodic memory. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 7918-7923, doi:10.1073/pnas.1606309113 (2016).
51 Bunzeck, N. & Duzel, E. Absolute coding of stimulus novelty in the human substantia nigra/VTA. *Neuron* **51**, 369-379, doi:10.1016/j.neuron.2006.06.021 (2006).
52 Schott, B. H. *et al.* Activation of midbrain structures by associative novelty and the formation of explicit memory in humans. *Learn Mem* **11**, 383-387, doi:10.1101/lm.75004 (2004).
53 Bunzeck, N. *et al.* Mesolimbic novelty processing in older adults. *Cerebral cortex* **17**, 2940-2948, doi:10.1093/cercor/bhm020 (2007).
54 Duzel, S. *et al.* A close relationship between verbal memory and SN/VTA integrity in young and older adults. *Neuropsychologia* **46**, 3042-3052, doi:10.1016/j.neuropsychologia.2008.06.001 (2008).
55 Duzel, E., Bunzeck, N., Guitart-Masip, M. & Duzel, S. NOvelty-related motivation of anticipation and exploration by dopamine (NOMAD): implications for healthy aging. *Neurosci Biobehav Rev* **34**, 660-669, doi:10.1016/j.neubiorev.2009.08.006 (2010).
56 Marquez, F. & Yassa, M. A. Neuroimaging Biomarkers for Alzheimer's Disease. *Mol Neurodegener* **14**, 21, doi:10.1186/s13024-019-0325-5 (2019).
57 Elliott, M. L. *et al.* What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. *Psychol Sci* **31**, 792-806, doi:10.1177/0956797620916786 (2020).
58 Sheehan, D. V. *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* **59 Suppl 20**, 22-33;quiz 34-57 (1998).
59 Ackenheil, M., Stotz, G., Dietz-Bauer, R. & Vossen, A. Mini International Neuropsychiatric Interview: German version 5.0.0. *Psychiatrische Universitätsklinik* (1999).

60 Jessen, F. *et al.* Design and first baseline data of the DZNE multicenter observational study on predementia Alzheimer's disease (DELCODE). *Alzheimers Res Ther* **10**, 15, doi:10.1186/s13195-017-0314-2 (2018).
61 Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 7900-7905, doi:10.1073/pnas.1602413113 (2016).
62 Gvozdanovic, G., Stampfli, P., Seifritz, E. & Rasch, B. Structural brain differences predict early traumatic memory processing. *Psychophysiology* **57**, e13354, doi:10.1111/psyp.13354 (2020).
63 Weise, C. M., Bachmann, T., Schroeter, M. L. & Saur, D. When less is more: Structural correlates of core executive functions in young adults - A VBM and cortical thickness study. *NeuroImage* **189**, 896-903, doi:10.1016/j.neuroimage.2019.01.070 (2019).
64 Scarpazza, C., Tognin, S., Frisciata, S., Sartori, G. & Mechelli, A. False positive rates in Voxel-based Morphometry studies of the human brain: should we be worried? *Neurosci Biobehav Rev* **52**, 49-55, doi:10.1016/j.neubiorev.2015.02.008 (2015).
65 Schmidt, P. *et al.* An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage* **59**, 3774-3783, doi:10.1016/j.neuroimage.2011.11.032 (2012).
66 Guo, C., Ferreira, D., Fink, K., Westman, E. & Granberg, T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. *Eur Radiol* **29**, 1355-1364, doi:10.1007/s00330-018-5710-x (2019).
67 Meng, X.-L., Rosenthal, R. & Rubin, D. B. Comparing correlated correlation coefficients. *Psychological Bulletin* **111**, 172-175 (1992).
68 Zimmermann, P. & Fimm, B. Testbatterie zur Aufmerksamkeitsprüfung (TAP). *Herzogenrath: Psytest* (1993).

6. Statements

6.1. Acknowledgments

We are grateful to Herta Flor and Shari Wiseman for valuable comments on the manuscript. We thank Adriana Barman, Marieke Klein, Kerstin Möhring, Katja Neumann, Ilona Wiedenhöft, and Claus Tempelmann for assistance with MRI data acquisition.

6.2. Data Availability Statement

Due to data protection regulations, sharing of the entire data set underlying this study in a public repository is not possible. We have previously provided GLM contrast images as a NeuroVault collection (https://neurovault.org/collections/QBHNSRVW/) and MATLAB code for imaging scores as a GitHub repository (https://github.com/JoramSoch/FADE_SAME) for an earlier article using the same dataset [13]. Access to de-identified raw data will be provided by the authors upon reasonable request.

6.3. Funding and Conflict of Interest declaration

This study was supported by the State of Saxony-Anhalt and the European Union (Research Alliance "Autonomy in Old Age") and by the Deutsche Forschungsgemeinschaft (SFB 1436/ A05 to C.S. and B.H.S.; RI 2964-1 to A.R.). The funding agencies had no role in the design or analysis of the study. The authors have no conflict of interest, financial or otherwise, to declare.

7. Figures

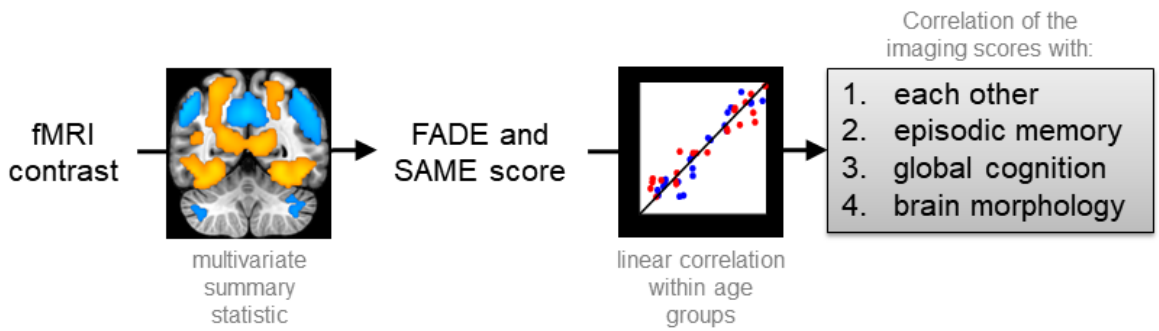

Figure 1. Overview of our approach to investigate the construct validity of single-value fMRI-based scores as potential biomarkers of cognitive ability in older adults. Imaging scores were calculated from a voxel-wise fMRI contrast map (warm colors indicate positive effects and cool colors indicate negative effects) and correlated firstly with each other, secondly with neuropsychological test performance in episodic memory, thirdly with global cognition, and lastly with measures of brain morphology separately for each age group (red: young, blue: older subjects). All activation maps are superimposed on the MNI template brain provided by MRIcroGL (https://www.nitrc.org/projects/mricrogl/). Figure adapted from Soch et al.[13].

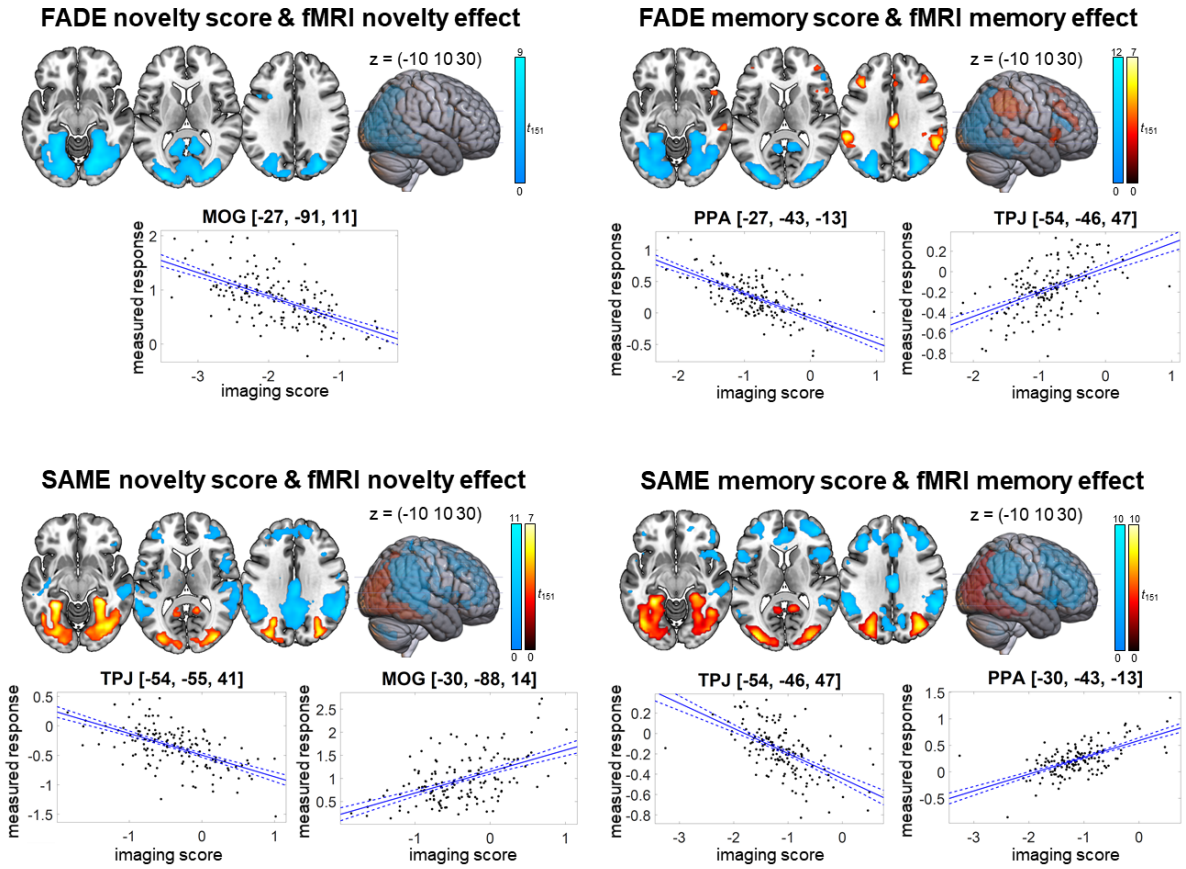

Figure 2. Imaging scores and fMRI effects (novelty effect and subsequent memory effect) in older participants. Warm colors indicate positive effects and cool colors indicate negative effects. $p < .05$, family-wise error-corrected at cluster level, cluster-defining threshold $p < .001$, uncorrected. MOG: Middle occipital gyrus, PPA: Parahippocampal place area, TPJ: Temporoparietal junction. All activation maps are superimposed on the MNI template brain provided by MRIcroGL (https://www.nitrc.org/projects/mricrogl/).

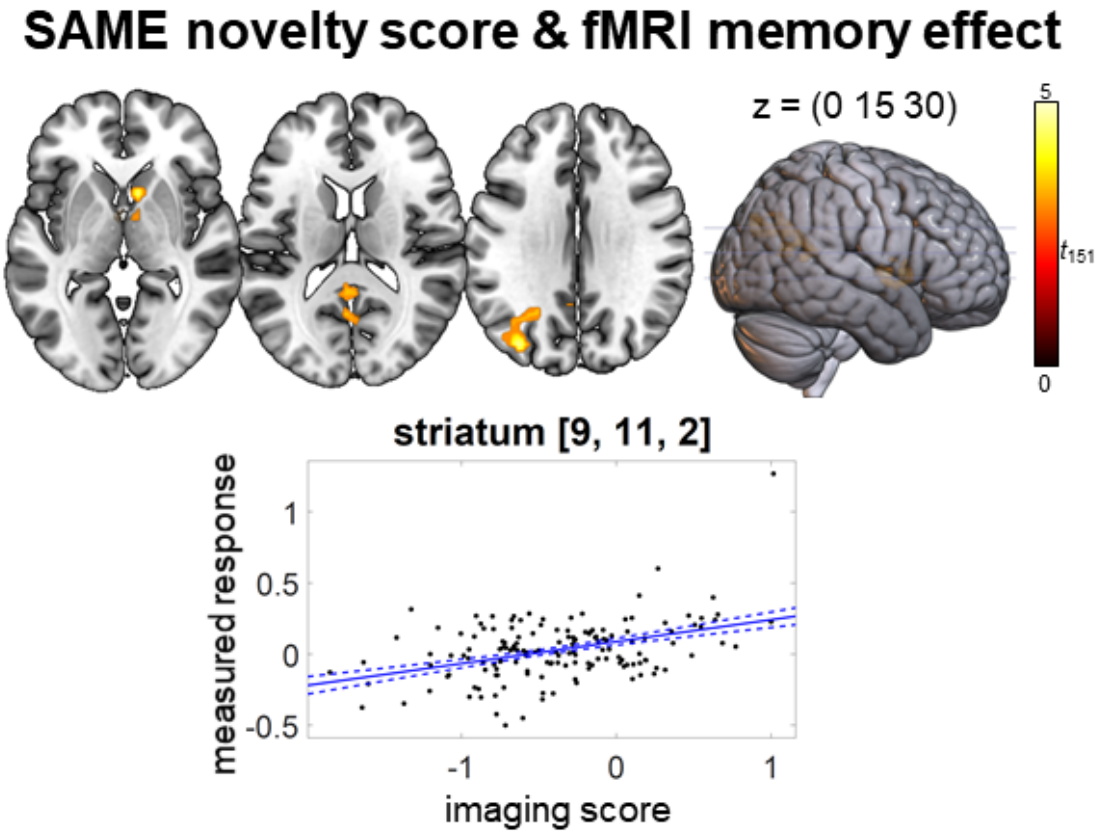

Figure 3. SAME novelty score and fMRI memory effect (positive effect). $p < .05$, family-wise error-corrected at cluster level, cluster-defining threshold $p < .001$, uncorrected. All activation maps are superimposed on the MNI template brain provided by MRIcroGL (https://www.nitrc.org/projects/mricrogl/).

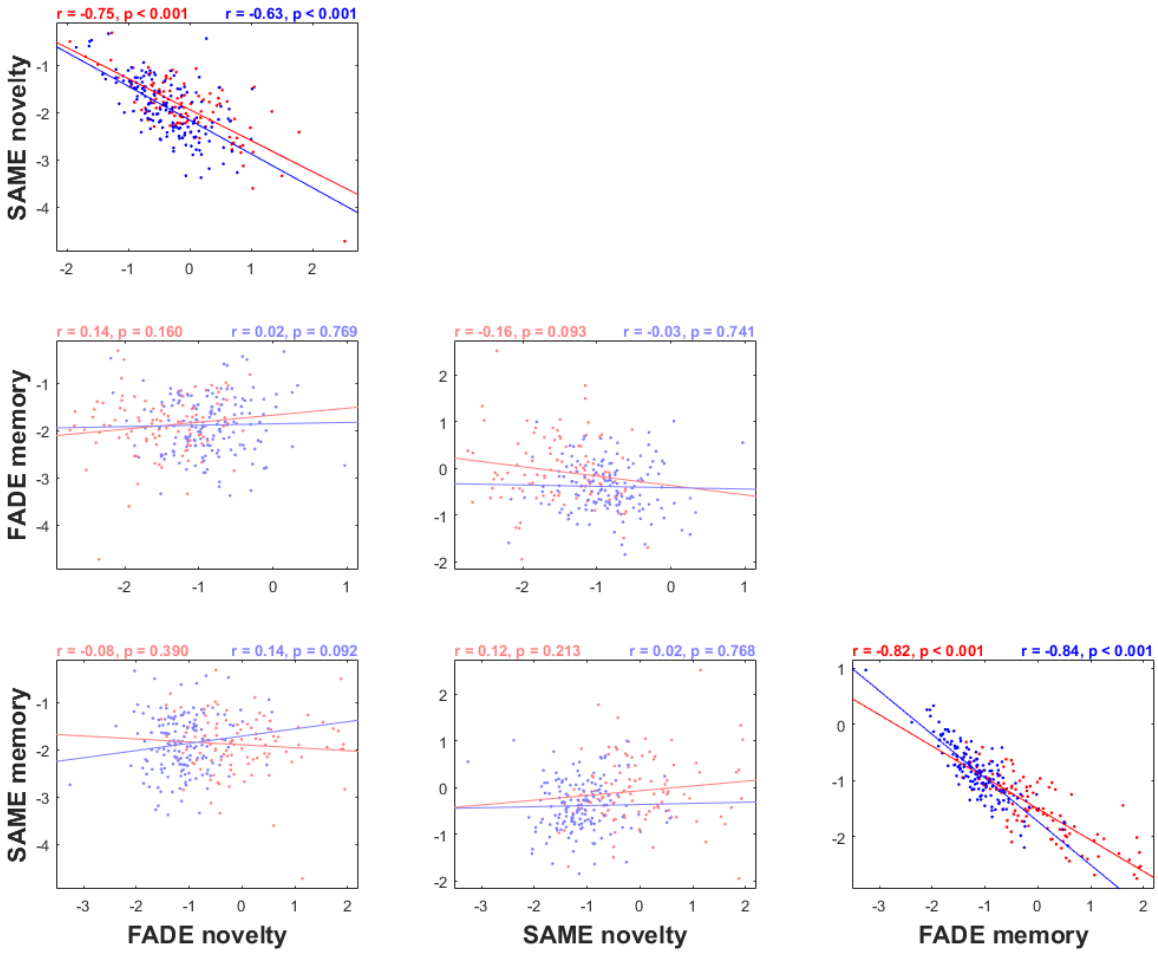

Figure 4. Pearson correlations between the FADE and SAME imaging scores conducted from the novelty and memory fMRI contrasts, separated by age group (red: young, blue: older subjects). Each dot represents one participant. Highlighted: correlation is significant at the 0.05 level (two-tailed).

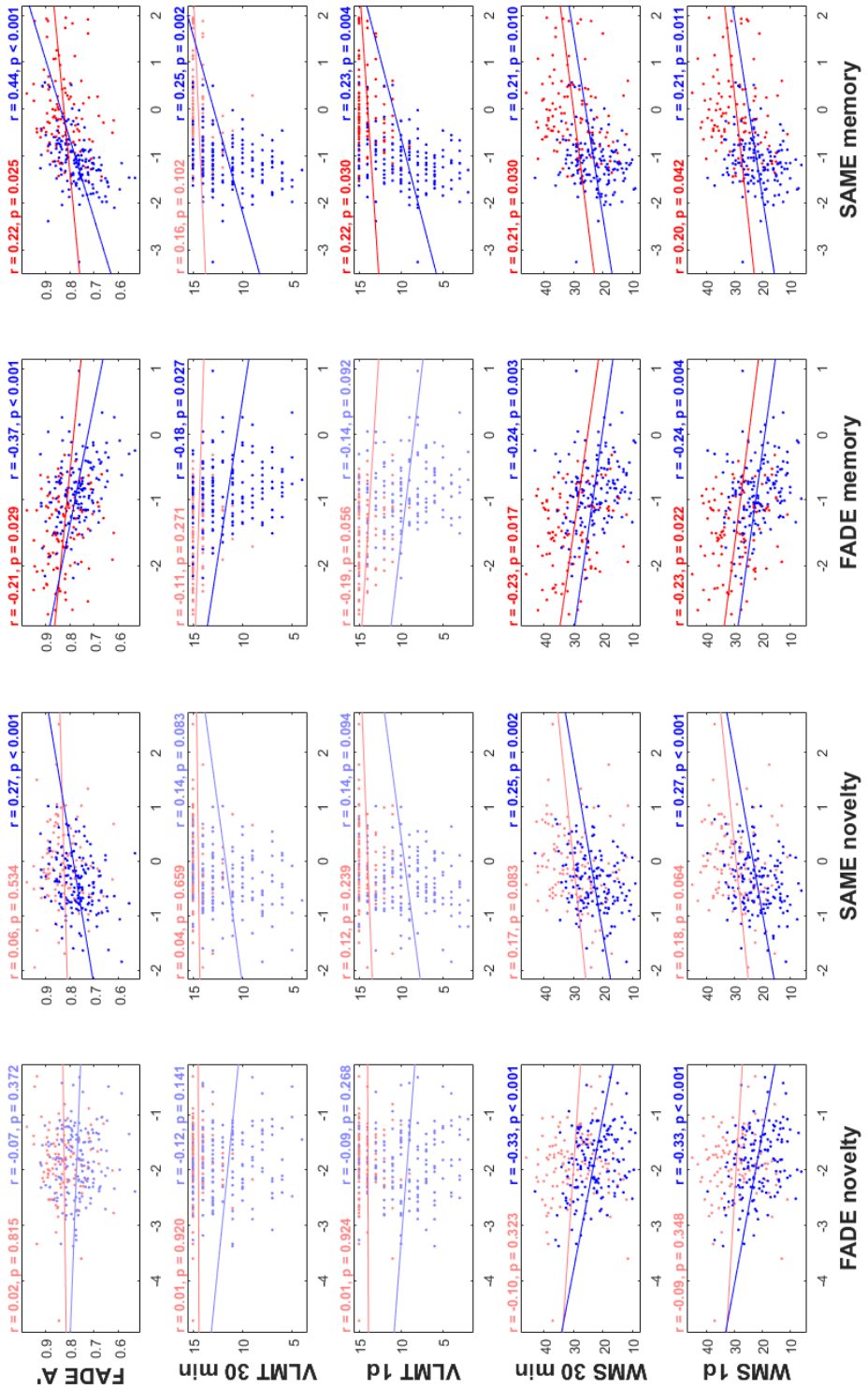

Figure 5. Pearson correlations of the FADE and SAME imaging scores conducted from the novelty and memory fMRI contrasts with performance in different memory tests, separated by age group (red: young, blue: older subjects). Highlighted: correlation is significant at the 0.05 level (2-sided).

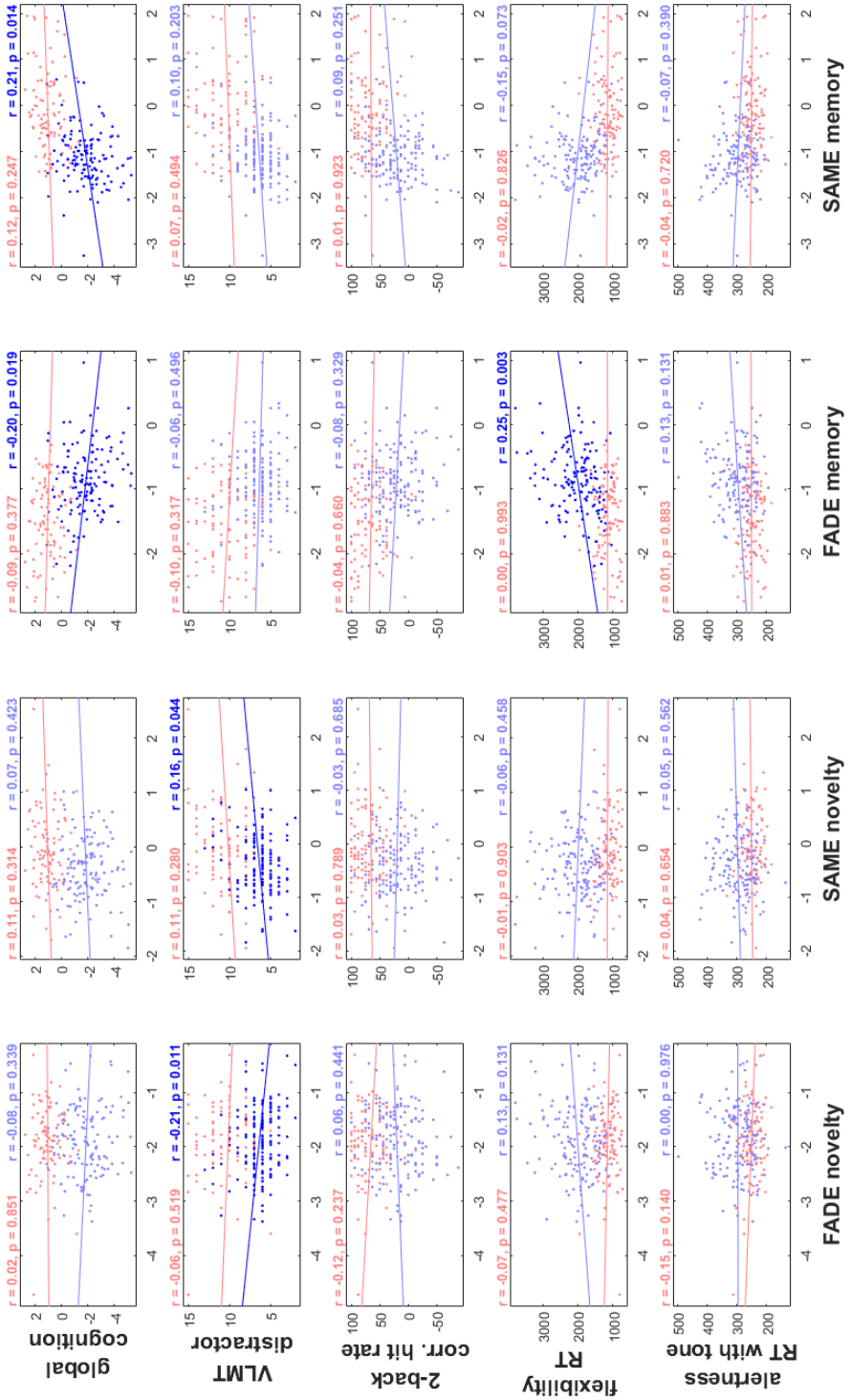

Figure 6. Pearson correlations of the FADE and SAME imaging scores conducted from the novelty and memory fMRI contrasts with performance in different neuropsychological tests and a composite score (global cognition), separated by age group (red: young, blue: older subjects). Highlighted: correlation is significant at the 0.05 level (2-sided).

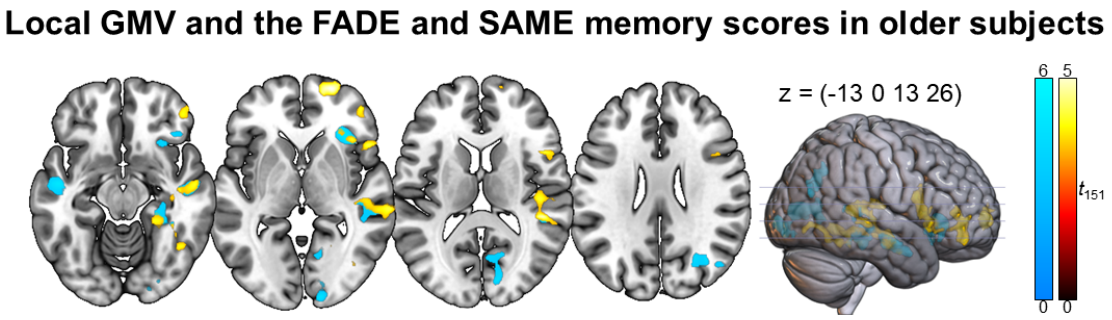

Figure 7. Imaging scores computed from the memory contrast and GMV using VBM. Warm colors indicate positive effects of the SAME memory score and cool colors indicate negative effects of the FADE memory score. $p < .05$, family-wise error-corrected at cluster level, cluster-defining threshold $p < .001$, uncorrected. All activation maps are superimposed on the MNI template brain provided by MRIcroGL (https://www.nitrc.org/projects/mricrogl/).

8. Tables

Table 1. Tests and variables of the neuropsychological testing battery

test	variables	psychological construct	young subjects M ± SD (N)	older subjects M ± SD (N)	statistics
Verbal Learning and Memory Test (VLMT)	number of correctly named words of:				
	-repetitions of list A (sum score)	learning ability	67.02 ± 6.09 (102)	53.42 ± 9.38 (152)	$t = 14.01$, $p < .001$
	-distractor list B	pro-active inhibition	10.14 ± 2.68 (103)	6.36 ± 2.02 (152)	$t = 12.17$, $p < .001$
	-recall of list A	retro-active inhibition	14.47 ± 1.02 (103)	11.32 ± 2.84 (152)	$t = 12.53$, $p < .001$
	-30-min delayed recall of list A	episodic memory	14.44 ± 1.09 (104)	11.43 ± 2.88 (152)	$t = 11.74$, $p < .001$
	-one-day delayed recall of list A	episodic memory	13.94 ± 1.49 (100)	9.26 ± 3.43 (148)	$t = 14.70$, $p < .001$
Logical Memory subtest from the WMS	number of story details retrieved at:				
	-immediate recall	learning ability	31.35 ± 7.32 (103)	25.45 ± 6.27 (149)	$t = 6.66$, $p < .001$
	-30-min delayed recall	episodic memory	29.85 ± 7.99 (103)	22.99 ± 6.58 (148	$t = 7.18$, $p < .001$
	-one-day delayed recall	episodic memory	29.21 ± 7.77 (102)	21.93 ± 6.83 (146)	$t = 7.63$, $p < .001$
Alertness subtest from the TAP	reaction on the appearance of a cross:				
	-RT in trials with cue tone	tonic alertness	249.91 ± 29.71 (102)	295.54 ± 54.87 (144)	$t = -8.39$, $p < .001$
	-RT in trials without cue tone	phasic alertness	276.28 ± 30.40 (102)	329.74 ± 58.40 (144)	$t = -9.34$, $p < .001$
Flexibility subtest from the TAP	switching attention between targets:				
	-error rate	flexibility	4.42 ± 4.62 (102)	11.25 ± 13.19 (147)	$t = -5.78$, $p < .001$
	-RT	flexibility	1146.73 ± 264.59 (101)	2006.76 ± 575.52 (147)	$t = -15.84$, $p < .001$
Flanker task	incongruent vs. congruent trials:				
	-RT difference	interference processing	111.26 ± 52.52 (103)	213.37 ± 133.95 (140)	$t = -8.20$, $p < .001$
N-Back task	responses on reoccurring letters:				
	-1-back corrected hit rate	working memory	97.45 ± 4.63 (100)	89.65 ± 18.16 (139)	$t = 4.85$, $p < .001$
	-1-back RT	working memory	433.17 ± 54.23 (100)	490.50 ± 86.10 (139)	$t = -6.30$, $p < .001$
	-2-back corrected hit rate	working memory	65.29 ± 28.39 (104)	20.35 ± 37.47 (150)	$t = 10.86$, $p < .001$
	-2-back RT	working memory	588.91 ± 100.65 (104)	663.39 ± 108.36 (150)	$t = -5.55$, $p < .001$
	-3-back corrected hit rate	working memory	23.67 ± 34.40 (103)	-11.77 ± 31.03 (149)	$t = 8.52$, $p < .001$
	-3-back RT	working memory	630.91 ± 118.27 (103)	708.45 ± 150.52 (149)	$t = -4.57$, $p < .001$

Bold type: variables that best discriminate between age groups (see the linear discriminant analysis). RT: reaction time. WMS: Wechsler Memory Scale ([28]). TAP: Test Battery for Attention ([68]). VLMT: Verbal Learning and Memory Test ([27]).