Error-driven upregulation of memory representations

Alexander Weuthen[1,2,3,4,*], Hans Kirschner[1], Markus Ullsperger[1,3,4,5]

[1] Institute of Psychology, Otto-von-Guericke-University Magdeburg, Germany

[2] Department of Psychiatry and Psychotherapy, Jena University Hospital/Friedrich-Schiller-University, Germany

[3] German Center for Mental Health (DZPG), partner site Halle-Jena-Magdeburg

[4] Center for Intervention and Research on adaptive and maladaptive brain Circuits underlying mental health (C-I-R-C), Halle-Jena-Magdeburg

[5] Center for Behavioral Brain Sciences, Magdeburg, Germany

* Corresponding author: Alexander Weuthen, Klinik für Psychiatrie und Psychotherapie, Universitätsklinikum Jena, Philosophenweg 3, 07743 Jena; Phone: +49 (0)3641 9390230; E-Mail: Alexander.Weuthen@uni-jena.de

Abstract

Learning an association does not always succeed on the first attempt. Previous studies associated increased error signals in posterior medial frontal cortex (pMFC) with improved memory formation. However, the neurophysiological mechanisms that facilitate post-error learning remain poorly understood. To address this gap, participants performed a novel feedback-based association learning task and a 1-back localizer task. Increased hemodynamic responses in pMFC were found for internal and external origins of memory error evidence, and during post-error encoding success as quantified by subsequent recall of face-associated memories. A localizer-based machine learning model displayed a cognitive control network, including pMFC and dorsolateral prefrontal cortex, whose activity was related to face-processing evidence in the fusiform face area. Representation strength was higher during failed recall and increased during encoding when subsequent recall succeeded. These data enhance our understanding of the neurophysiological mechanisms of adaptive learning by linking the need for learning with increased processing of the relevant stimulus category.

Introduction

Forming memories and using acquired knowledge when needed is an essential cognitive capability. Imagine, for example, a teacher, who is trying to learn the names of students of a new class. For some students, the teacher will remember the names right away, but for others the teacher needs several attempts. In this study, we aim to better understand how the brain monitors learning failures and facilitates subsequent association memory formation.

Based on the assumption that successful memory recall requires successful memory encoding, previous neuroimaging studies have investigated the subsequent memory effect (SME) by determining which neurophysiological signals at time of encoding predict later recall success (Brewer et al., 1998; Uncapher & Wagner, 2009; Wagner et al., 1998). Cognitive processes and brain regions contributing to the SME have been differentiated into content-processing regions in the fusiform gyrus (FG) and left inferior frontal gyrus (IFG), attention during encoding in premotor cortex (PMC) and posterior parietal cortex (PPC), as well as storage function in medial temporal lobe regions such as hippocampus and amygdala (Kim, 2011). There is, however, a lack of studies investigating how the brain monitors failed learning attempts and implements necessary adjustments, such as increased attention and brain network states for improved memory formation. Based on the broader literature on performance monitoring in speeded choice reaction time tasks, the posterior medial frontal cortex (pMFC) has consistently been implicated in accumulating evidence of task demands and a respective signaling function indicating the need for adjustments (Gruendler et al., 2011; Kirschner & Ullsperger, 2024). For example, the magnitude of error-related functional magnetic resonance imaging (fMRI) in pMFC and frontocentral electroencephalography (EEG) signals was shown to be predictive for successful

performance adaptations (Danielmeier et al., 2011; Klein et al., 2007). Interestingly, the function of pMFC in detecting memory demands and enhancing attention has been overlooked in previous studies, although hemodynamic responses in pMFC were also found to be increased during successful learning in above meta-analysis on the SME (Kim, 2011). While error-related signals in this region have been associated with improved associative learning (de Bruijn et al., 2020; Hester et al., 2008), there is a lack of studies investigating how brain regions implicated in performance monitoring and memory formation interact.

On the one hand, clusters in pMFC have been linked to the midcingulo-insular salience network (Seeley et al., 2007; Uddin et al., 2019), encompassing mainly midcingulate cortex and anterior insula. Studies focusing on neurophysiological network interplay for successful cognitive performance highlighted that this network switches between upregulated lateral frontoparietal control network for external attention and upregulated medial frontoparietal default mode network for internal attention (Menon, 2015; Uddin et al., 2019). When learning associations of visual stimuli, it seems beneficial that brain networks for external attention and visual processing of memory stimuli are upregulated. During memory recall, hemodynamic responses in default mode network regions have been found to be increased (Shapira-Lichter et al., 2013), potentially based on attention towards internal stimulus representations. Memory performance likely also benefits from enhanced maintenance stimulus features, which is often referred to as stimulus rehearsal and may represent a combination of both, internal and external attention.

From a methodological perspective, the neurophysiological underpinnings of detected memory demands can be estimated by contrasting neural representations associated with low versus high confidence ratings and performance-based feedback

for failed versus correct recalls. According to the logic of the SME, the overall quality of memory encoding can be inferred from a participant's performance during subsequent recall. However, it is less straightforward to determine the extent to which attention is allocated to memory-relevant stimuli.

While previous studies have speculated on the mechanisms for attentional allocation of error-driven learning improvements (Gilmore et al., 2018), the current study modelled the level of evidence for processing the memory-relevant category. Improved external attention should support the extraction of to-be-learned stimulus features and increased internal attention following stimulus presentation should strengthen perceptual representations via mental rehearsal. During these memory formation epochs, neurophysiological processing of the memorized stimulus category should be increased when more attention is allocated on a stimulus. The detection of a memory error should lead to increased stimulus processing and a higher likelihood that the presented association will be remembered. Multivariate decoding may be a useful tool to capture the strength of and evidence for stimulus representations during different phases of memory formation. Previous studies have shown that the degree of behavioral relevance of a presented stimulus category can be decoded during respective cognitive tasks (Erez & Duncan, 2015; Leong et al., 2017) and that stimulus decodability is related to the degree how much attention is allocated (Nelissen et al., 2013). While most task-based fMRI studies have used multivariate pattern analyses to compare decoding accuracies for a set of stimuli, it has been suggested the decision function of multivariate models contains a more fine-grained pattern of stimulus evidence (Walther et al., 2016), which can be used to determine single-trial differences in stimulus processing and decodability. Here, we tested the hypothesis that regions associated with the monitoring of memory performance, such as pMFC, reflect

upregulated selective attention, as approximated by single-trial evidence of stimulus-processing in stimulus-specific regions. The current study investigates face-processing evidence in the ventral visual stream, because of its well described topography in the posterior and mid-fusiform gyrus (Caspers et al., 2013; Lorenz et al., 2015), often referred to as fusiform face area (FFA, (Kanwisher et al., 1997). If nodes in the midcingulo-insular salience network link memory-related demand detection, upregulated FFA-based face-processing and improved memory success, this will improve the understanding how brain networks for performance monitoring, stimulus-based attention and memory formation interact.

Results

Performance on the novel feedback-based association learning task

During continuous fMRI scanning, participants (n = 30) performed a feedback-based association learning task (FALT), in which they had to learn which of eight orientations of a gabor patch is associated to a set of unknown faces. Each trial consisted of a recall phase where only the face was presented, the choice of a presumed orientation, the selection of a low or high recall confidence, the presentation of performance-based feedback, and the display of the correct combination of the face and the associated gabor patch, followed by an inter-trial-interval (ITI), which offered a chance for stimulus rehearsal (**Fig. 1a**). Each participant performed five independent runs, in which eight new faces were learned and repeated in three more blocks. Across all runs and blocks in FALT, participants correctly remembered the associations between faces and the eight different orientations of the gabor patches in 59.35 % (*SD* = 15.20) of trials [$t(29) = 16.88$, $p < .001$, one-sample *t*-test > 12.5 % chance level]. Memory performance was improved for face repetitions in later blocks [$F(1,29) = 93.53$, $p < .001$, one-way *ANOVA*], with comparable recall success in block 3 and

block 4 [$F(1,29) = 2.23$, $p_{HSD} = .593$, Tukey's *HSD*; **Fig. 1b**]. Correct trials were associated with ratings of high confidence on 80.14 % (*SD* = 14.28) of trials, and for failed recall participants selected a low level of confidence on 76.22 % (*SD* = 17.78) of trials (**Fig. 1c**). Participants had good meta-memory performance [$d_{Prime} = 0.95$; $t(29) = 11.76$, $p < .001$, one-sample *t*-test > 0], without an indication of a bias towards under- or overconfidence [$d_{Bias} = -.03$; $t(29) = 0.24$, $p = .811$, one-sample *t*-test]. Together, these results suggest that participants learned to associate faces with tilted gabor patches and gained accurate confidence levels in a novel feedback-based association learning task.

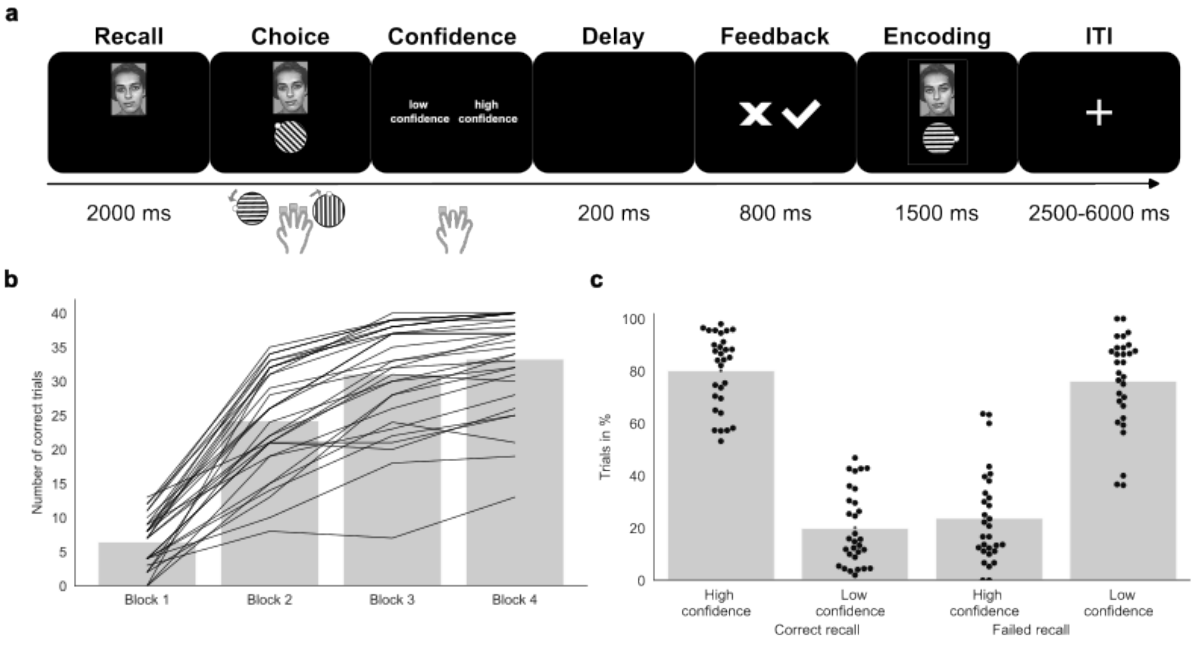

Fig. 1* | *Trial structure and behavioral results of the feedback-based association learning task (FALT). a, *In a continuous learning experiment, participants learned to associate faces and eight different orientations of gabor patches. During a trial, participants chose the presumed orientation, selected a low or high level of confidence and obtained either positive or negative feedback. Finally, the correct combination of face and gabor patch was presented as a learning opportunity for trials showing the same face in later blocks, followed by a jittered inter-trial-interval (ITI).* ***b,*** *Participants*

successfully learned the presented associations and recalled the matching orientation of the gabor patch better in later repetitions of a face. ***c,*** *In most of the trials, participants were able to distinguish successful and failed memory recall, indicating reasonable meta-memory performance.*

Implicit and explicit evidence for memory errors is represented in pMFC

In FALT, on each trial participants accumulated evidence on the quality of a current memory representation, both implicitly and explicitly. First, they attempted to recall the correct association to a face. Thereafter, they indicated their confidence in their response (binary variable low vs. high confidence). Finally, they received feedback regarding the correctness of the chosen orientation of the gabor patch. Accordingly, there were three different epochs for modelling neurophysiological correlates on the monitoring of memory errors. Univariate general linear model (GLM) analyses showed increased hemodynamic responses in pMFC at all stages of error monitoring in the task (**Fig. 2**). These epochs showed pMFC effects during failed recall [$\text{Error}_{\text{ConfidenceLow}}$ > $\text{Correct}_{\text{ConfidenceHigh}}$; $z(29)$ = 4.30, p_{FDR} < .05, x = 5, y = 22, z = 40], the selection of recall uncertainty [low > high confidence; $z(29)$ = 4.87, p_{FDR} < .05, x = 3, y = 35, z = 36] and the presentation of memory-error feedback [negative > positive feedback; $z(29)$ = 6.80, p_{FDR} < .05, x = 5, y = 17, z = 51]. Variance-inflation-factor indices were < 5 for all error monitoring regressors, indicating sufficiently low multicollinearity. The overlap in the cluster location of pMFC in all three epochs suggested that pMFC's presumed function in performance monitoring also applies to tracking internal and external evidence of currently inaccurate memory representations.

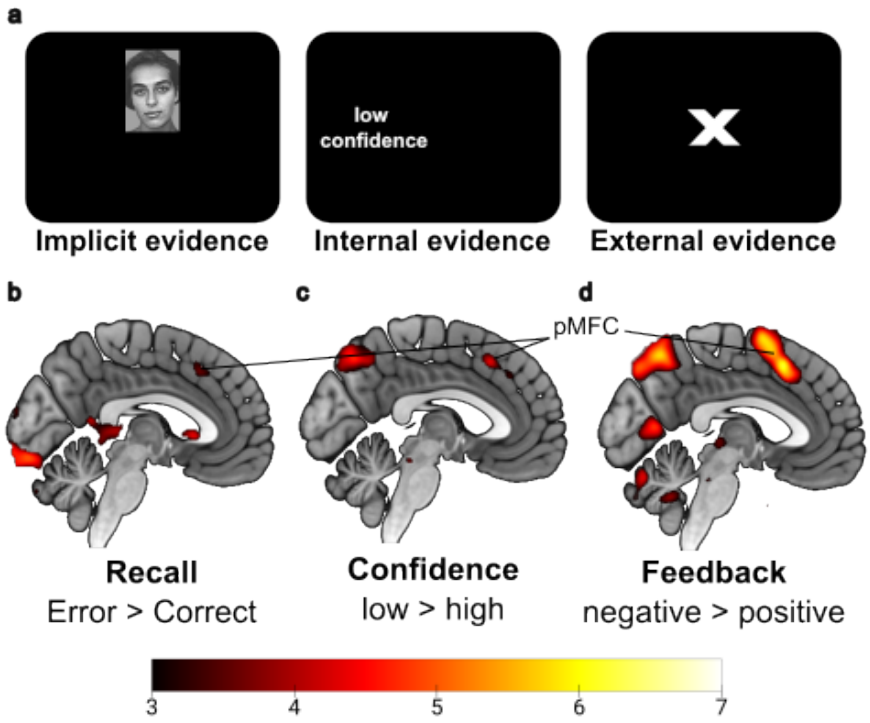

Fig. 2* | *Hemodynamic responses related to the monitoring of memory errors in the feedback-based association learning task (FALT).* *a,* *All three types of memory-error evidence, such as* ***b, *failed recall attempts,* ***c,*** *the selection of low confidence and* ***d,*** *negative feedback showed increased hemodynamic responses in posterior medial frontal cortex (pMFC) in a univariate general linear model analysis.*

Error-related pMFC activity predict successful subsequent recall

After an attempted recall in FALT, participants had another learning opportunity, in which the correct association of the presented face and gabor patch was displayed. For failed recall trials, a univariate GLM analysis determined which neurophysiological differences during post-error encoding epochs distinguish successful and failed subsequent recall. Results for the post-error SME showed increased hemodynamic responses in pMFC [$Error_{LowConfidence}Correct_{HighConfidence}$ > $Error_{LowConfidence}Error_{LowConfidence}$, $z(29) = 5.37$, $p_{FDR} < .05$, x = -3, y = 0, z = 71], and replicated regions previously reported for the SME, such as IFG, FG, PPC and PMC

(**Fig. 3** and **Supplementary table 4**). While the function of pMFC for memory formation has been neglected in a previous meta-analysis (Kim, 2011), the overlap with memory-error related regions suggested a preparatory role for an adaptive learning state. Successful post-error learning improvement has been related to increased error-related fMRI and EEG signals before. Yet, the correlational nature of these results precludes a better understanding of the underlying mechanisms. We suggest that a candidate mechanism is increased processing of information relevant to resolve the problem at hand. To test this idea, we applied a model on the strength of stimulus representations as a marker of allocated attention to the presented stimulus category.

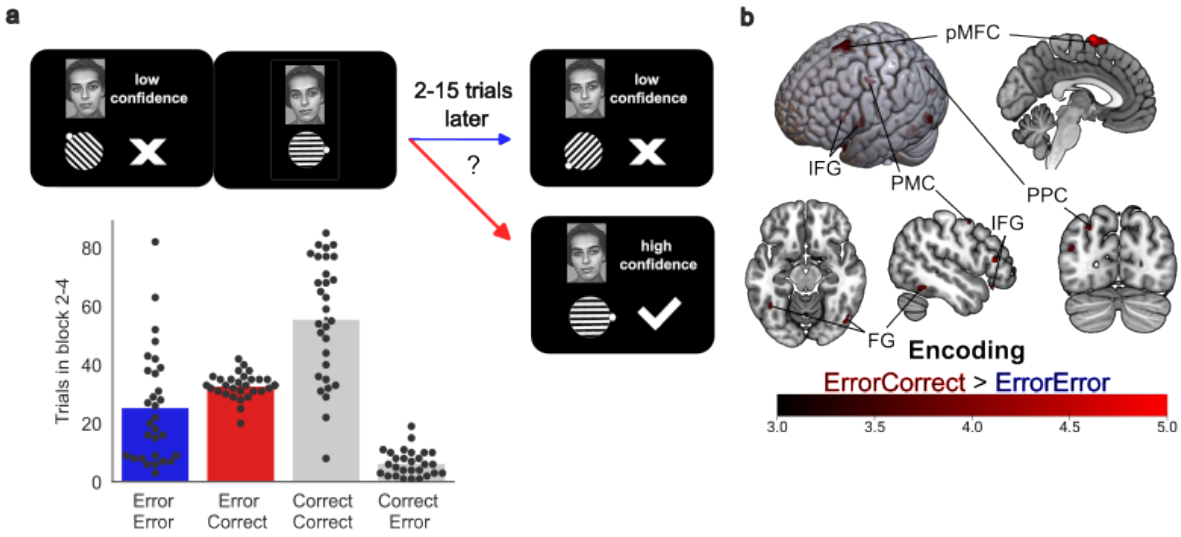

Fig. 3 | ***The post-error subsequent memory effect in a univariate functional magnetic resonance imaging (fMRI) analysis***. ***a,*** *The bar plot shows the number of trials per participant for the combination of recall success in the current trial and recall success for the next presentation of the same face. The aim was to distinguish trials with memory (re-)encoding demands which lead to successful memory formation (ErrorCorrect) from failed recall trials which did not lead to successful post-error learning adjustments (ErrorError).* ***b,*** *Univariate general linear model fMRI results*

*replicated previously described regions from a meta-analysis on the subsequent memory effect (Kim, 2011), showing increased successful recall ($Error_{LowConfidence}Correct_{HighConfidence}$) compared to trials repeatedly failed recall ($Error_{LowConfidence}Error_{LowConfidence}$).The pMFC cluster for the post-error subsequent memory effect overlapped with the cluster related to the monitoring of memory errors (**Fig. 2**), suggesting its demand-dependent upregulation has a preparatory function.*

The 1-back localizer task captured face-selective processing in a cytoarchitectonic mask of the fusiform gyrus

To build a model of stimulus representation strength, we trained a classifier on fMRI data of a 1-back localizer task, in which participants had to press the confirmation key for a repetition of the stimuli presented in the preceding trial (**Fig. 4a)**. Within each trial, participants saw either a face or a house in combination with a gabor patch. According to signal detection theory, a trial was either classified as hit, miss, correct rejection or false alarm. The 28 out of 30 participants, who performed the task either without mistakes or within two standard deviations from the group average (**Fig. 4b**), were included in the fMRI analyses and multivariate cross-classification. Univariate GLM analyses in the 1-back localizer task showed that hemodynamic responses were larger for faces than houses in FFA, as determined by a strong overlap with cytoarchitectonic probability maps of left and right FG-4 [$Face_{CorrectRejection}$ > $House_{CorrectRejection}$; $z(27) = 5.48$, $p_{FDR} < .05$, x = 44, y = -46, z = -27], but also in other regions previously described as face-selective such as superior temporal sulcus [$z(27) = 4.48$, $p_{FDR} < .05$, $x = 51$, $y = -46$, $z = 5$] and anterior temporal lobe [$z(27) = 4.87$, $p_{FDR} < .05$, x = 40, y = 19, z = -31, **Fig. 4** and **Supplementary table 5**]. Increased hemodynamic responses for houses compared to faces were found in regions among parahippocampal gyrus [$House_{CorrectRejection}$ > $Face_{CorrectRejection}$; $z(27) = 7.02$, p_{FDR} <

.05, x = -29, y = -52, z = -5; **Fig. 4** and **Supplementary table 6**]. Overall, univariate fMRI results in the localizer task displayed the classical dissociation in the ventral visual stream, displaying FFA-related hemodynamic responses being larger for faces and house-specific hemodynamic responses in parahippocampal gyrus.

In the localizer task, a machine learning model was trained, in order to predict the strength of FFA-based face-processing evidence during memory-relevant epochs in FALT. Multivariate cross-classification was performed on fMRI data within an FFA mask based on cytoarchitectonic probability maps of left and right FG-4, to quantify face-specific processing evidence in the ventral visual stream. The 14 strongest ANOVA-feature selected voxels in the FFA mask were extracted from deconvolved single-trial betaseries to train and apply the machine learning-based face-processing model. Leave-one-run-out cross-validation reached an average balanced decoding accuracy of 73.15 % [$t(27) = 12.38$, $p < .001$, one-sample t-test, > 50% chance level] on distinguishing faces and houses in trials of respective left-out runs in the localizer task. The prediction of house and face stimuli was balanced, showing no trend in the likelihood of the face representation strength models to prefer either of both categories [$t(27) = -0.18$, $p = .854$, two-sample t-test]. Taken together, the decoding accuracies and control analyses suggested that the multivariate face-processing model was able to evaluate face-processing evidence by distinguishing face and house trials.

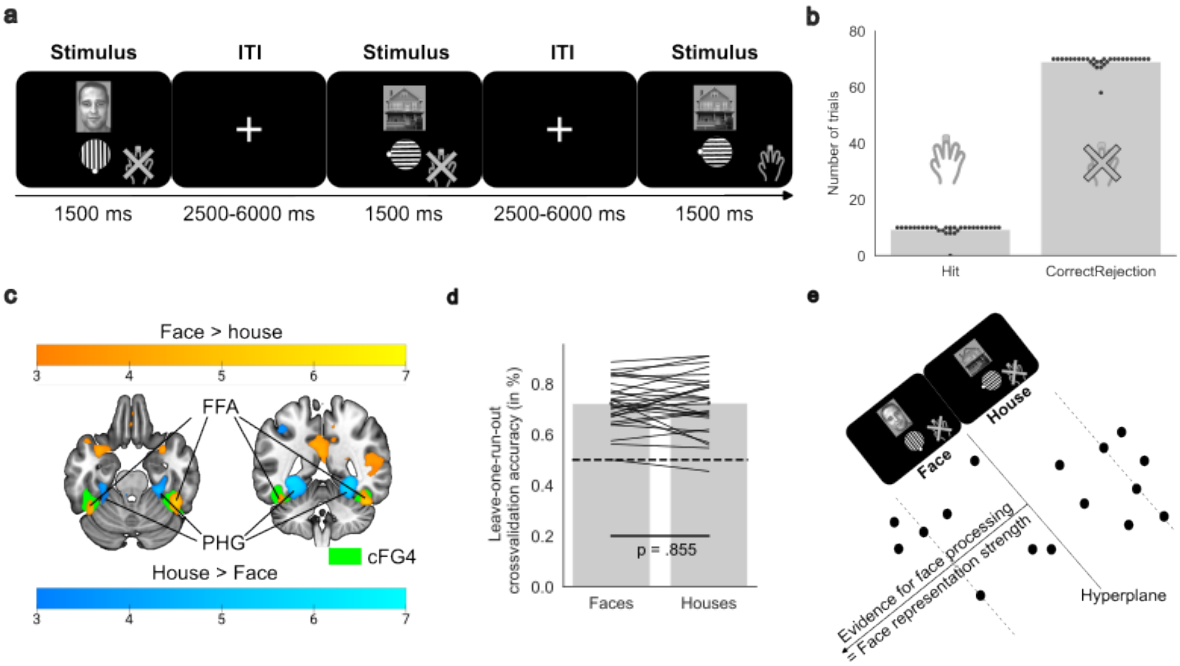

Fig. 4 | Localizer task trial structure, behavioral results, univariate fMRI analyses and training of machine learning-based face-processing model. a, *The 1-back localizer task had comparable presentation times as chosen in the feedback-based association learning task (FALT) for the stimulus presentation and the inter-trial-interval (ITI). In two of 16 trials per run, direct stimulus repetitions occurred. On these repetitions, participants were instructed to quickly press the confirmation key. The task consisted of five runs, each containing two novel faces and two novel houses.* ***b,*** *Most participants performed the task without mistakes (misses or false alarms). Two participants were excluded from further fMRI analyses of the localizer task and later multivariate pattern analyses, because task comprehension and attention to the task could not be assured.* ***c,*** *Conventional general linear model (GLM) analyses showed that hemodynamic responses were larger for faces than houses in the right fusiform gyrus, and larger for houses than faces in the parahippocampal gyrus (PHG). The fusiform gyrus cluster largely overlapped with a cytoarchitectonic probability mask for fusiform gyrus 4 (cFG4) and was identified as fusiform face area (FFA).* ***d,*** *A probability-scaled linear support vector machine was trained to distinguish faces and*

house based the 14 strongest voxels in the cFG4 mask according to ANOVA-feature selection, to quantify FFA-based face-processing evidence. Average classification accuracies during leave-one-run-out cross-validation showed a balanced prediction for faces and houses. ***e,*** *The assessed distance from the multivariate hyperplane indicates evidence for face-processing as shown in the schematic overview.*

FFA-based face-memory representations are simultaneously upregulated in a cognitive control network

In the next step, the participant-specific face-processing models, which were trained on trials in the localizer task, were applied to the single-trial betaseries of memory-relevant epochs in FALT, i.e., the presentation of the face in the recall phase, processing of the correct face-orientation association during encoding, and a rehearsal phase in the inter-trial-interval. The classifier predicted the presentation of face in 72.23 % (SD = 19.38) of recall betaseries, in 39.53 % (SD = 18.58) of encoding betaseries, and in 17.35 % (SD = 18.72) of rehearsal betaseries. The classifier's decoding accuracies were systematically related to the predicted class probability averages of a participant, during recall [$R_{Spearman}(27) = .97$, $p < .001$], encoding [$R_{Spearman}(27) = .93$, $p < .001$] and rehearsal [$R_{Spearman}(27) = .94$, $p < .001$]. This correspondence was expected and underlined that the probability measure contained information relevant to differentiate the representation strength memory-relevant stimuli on a single-trial level. Higher representation strength for face processing showed the strongest FFA-related hemodynamic responses in a region overlapping with the cytoarchitectonic mask for FG-4 as shown by separate GLMs for recall (**Supplementary table 7**), encoding (**Supplementary table 8**) and rehearsal (**Supplementary table 9**) in FALT. Higher face-representation strength was also reliably related to regions among pMFC, dlPFC and visual cortex for all three memory-

relevant epochs (**Fig. 5**). During recall and encoding, bilateral anterior insula showed increased hemodynamic responses related to face-processing evidence. During encoding and rehearsal, associations of single-trial representation strength were also overlapping with bilateral cytoarchitectonic masks of the basal forebrain and the border zone between amygdala and the nucleus basalis of Meynert. The pMFC topography during all three memory epochs overlapped with the pMFC cluster found for the monitoring of inaccurate memory representations and in the post-error SME, which suggested a role in the maintenance of upregulated memory-relevant stimulus representations. Overall, multivariate cross-classification analyses highlighted brain network nodes simultaneously upregulated with increased face-processing in face-specific regions of the ventral visual stream. This suggests that these regions work in concert to allocate attention in form of upregulated stimulus representations, which could enhance cognitive operations for association learning, such as extracting memory-relevant stimulus features or improving stimulus maintenance.

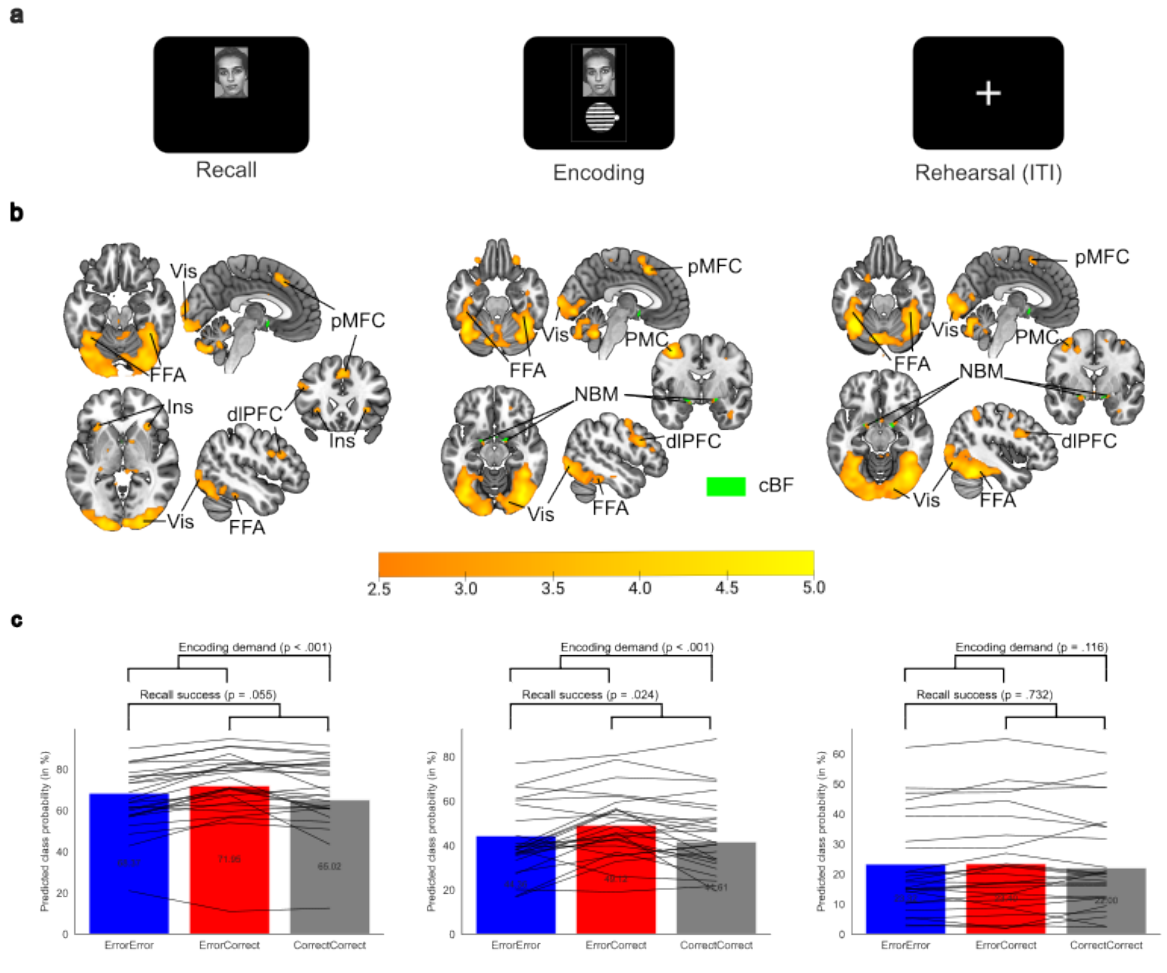

Fig. 5 | Neurophysiological and behavior underpinnings of FFA-based face-processing evidence during memory-relevant epochs in the feedback-based association learning task (FALT). *The support vector machines, which were trained on the 1-back localizer task of a participant, predicted face processing during memory encoding rehearsal and recall in FALT, based on the 14 most face-selective voxels in the cytoarchitectonic probability mask of a fusiform gyrus 4 region.* ***a,*** *GLM results on evidence for face-processing displayed increased hemodynamic response in regions among fusiform face area (FFA), posterior medial frontal cortex (pMFC), dorsolateral prefrontal cortex (dlPFC), anterior insula (Ins), premotor cortex (PMC), and a cluster overlapping with a cytoarchitectonic mask of the basal nucleus of Meynert (NBM) subregion of the cholinergic basal forebrain (cBF) and amygdala.* ***b,*** *The level of evidence for face-processing was higher when there was a demand of memory*

improvement during recall and encoding, and significantly higher for subsequent recall success during encoding epochs, as found in the linear mixed model results.

FFA-based face-processing evidence is increased after memory errors and predictive of subsequent recall success

The proxy measure for the strength of stimulus representations, as a marker of allocated attention to the presented stimulus category, was also analyzed regarding its correspondence with behavioral necessity and success on learning the presented associations. The analyses were restricted to ErrorError, ErrorCorrect and CorrectCorrect trials to understand how encoding demand and encoding success were linked to the level of stimulus representations strength. CorrectError trials were excluded since they were not present in all participants and because the quality of memory representations for a successful recall was doubtful due to its later memory failure. During memory recall, encoding demand was associated with a 3.2 % increase in face-processing evidence [$z(27) = 6.89$, $p < .001$], and subsequent recall success with a 1.1% increase [$z(27) = 1.92$, $p = .055$] in multivariate classification evidence for face representations (**Supplementary table 10)**. During encoding, encoding demand was estimated to increase the face-processing by 3.9% [$z(27) = 7.176$, $p < .001$] and subsequent recall success was related to a 1.5 % [$z(27) = 2.254$, $p = .024$] increased probability in the linear mixed model analysis (**Supplementary table 11**). During stimulus rehearsal, neither encoding demand [$z(27) = 1.571$, $p = .116$], nor subsequent recall success [$z(27) = -0.343$, $p = .732$] significantly predicted the single-trial level of evidence for face-processing (**Supplementary table 12**). The increase in stimulus representation strength during memory formation during recall and encoding suggests increased allocation of attention to the presented stimulus category according to the necessity of learning and the success in forming association memories. During a failed

recall, the need for increased attention may already become evident and increase face-processing for the following encoding attempt. During encoding, enhanced face-processing indicated a facilitation in memory formation by successful subsequent recall. Increased processing of the memory-relevant stimulus category in the ventral visual stream, therefore, links the monitoring memory errors with improved associative learning both on a behavioral level and related to cognitive control network regions.

Discussion

The present study aimed to investigate which brain regions detect memory errors and coordinate adaptation processes to improve memory formation. Different sources of memory-error evidence overlapped in a pMFC cluster, which showed increased hemodynamic responses during memory-error related events, such as during failed recall, the selection of low confidence and the presentation of negative feedback. A posterior portion of the error-related pMFC cluster further distinguished later successfully remembered memory-error trials from repeatedly failed memory formation attempts. The level of FFA-based face-processing evidence was related to increased single-trial hemodynamic responses in regions of a cognitive control network. This network encompassed pMFC, dlPFC, visual cortex, anterior insula and a cluster overlapping with basal forebrain and amygdala, for the upregulation of memory-relevant stimulus representations. Stronger face-processing evidence in the ventral visual stream was linked to the demand of improving memory formation during failed recall, and was further upregulated during improved post-error encoding epochs, as determined by subsequent recall success.

The results favor the perspective that pMFC is involved in monitoring incorrect and low-confident memory representations and that it orchestrates brain networks involved in allocating attention to the relevant stimulus category for error-driven

363 improvements in memory formation. Previous studies have shown that the monitoring
364 of behaviorally relevant events is associated with hemodynamic responses and
365 electrophysiological signals in pMFC (for a review, see Kirschner & Ullsperger, 2024).
366 It has been an open question whether pMFC's function in performance monitoring also
367 applies to evaluating the quality of memory representations. The current study
368 demonstrates that increased hemodynamic responses in pMFC are related to the
369 processing of negative feedback, as has been described for failed associative recall
370 in previous fMRI (Hester et al., 2008) and EEG (de Bruijn et al., 2020) studies.
371 Furthermore, overlapping clusters in pMFC were also found for hemodynamic
372 responses during failed recall attempts and upon reporting low confidence, which
373 suggests a more general role of pMFC in accumulating evidence of memory errors
374 beyond the processing external error evidence, such as during the presentation of
375 negative feedback. If increased fMRI signals in pMFC are relevant for recognizing the
376 insufficiency of current memory representations, pMFC's involvement in the post-error
377 SME suggests a role in driving error-following adjustments in associative learning.
378 Consistently, error-related signals in pMFC have shown to be predictive for later recall
379 success (de Bruijn et al., 2020; Hester et al., 2008) and enhanced performance in
380 other cognitive tasks (Danielmeier et al., 2011; Klein et al., 2007). Results of the
381 current study complement a meta-analysis of previous fMRI studies on the SME (Kim,
382 2011), which has shown consistent involvement of pMFC in the SME but has not
383 described its function for memory formation. In this regard, the results demonstrate
384 that pMFC is not only related to successful encoding but its hemodynamic responses
385 are already increased upon monitoring error evidence, which is closely linked to
386 encoding demand and emphases a preparatory function for following learning
387 attempts.

While previous studies have shown increased pMFC-based error signals for improved performance, it has been an open question how post-error learning improvements are implemented. One of the speculated mechanisms how failed recall leads to enhanced memory formation, has been increased attentional allocation (Gilmore et al., 2018). The current study tested the hypothesis that detected recall errors increase the processing of memory-relevant stimulus representations to facilitate association learning. The current study used FFA-based fMRI evidence for face-category processing as a proxy measure for stimulus-based attention and showed that hemodynamic responses in regions such as pMFC, dlPFC, anterior insula and the basal forebrain increase as a function stimulus-specific processing evidence. These regions may interact to enhance attention for following learning attempts. While multivariate pattern analyses have been used to estimate levels of attention, it remained to be shown that a marker for allocated attention provides a link between memory-error detection and improved learning. Previous studies have used multivariate fMRI analyses to show that decoding accuracies and classification probabilities are increased for attended objects. More specifically, the highest decoding accuracies of occipitotemporal stimulus representations have been found for stimuli in the focus of attention (Nelissen et al., 2013) and when they are behaviorally relevant (Erez & Duncan, 2015). Another study used a combination of multivariate classification probability and eye tracking to develop a marker for how much attention was allocated (Leong et al., 2017). By using single-trial decoding probabilities instead of binary classification accuracies (Walther et al., 2016), the relationship between neurophysiological processing strength of memory-relevant stimulus representations and their behavioral correspondence to encoding demand and subsequent learning success became apparent. This suggests that multivariate evidence for stimulus-

processing during associative learning can be used as a marker for stimulus-based attention and represents a link between performance monitoring and improved memory formation. The current study aligns with previous studies linking multivariate stimulus models with behavior, by showing that single-trial evidence for face-processing in face-selective ventral visual stream regions is associated with increased hemodynamic responses in pMFC. This suggests a systematic relationship between the neurophysiological underpinnings of enhanced stimulus representations, the detection of memory errors and following encoding success.

Assuming that, upon the detection of respective task demands, pMFC upregulates stimulus-selective regions such as FFA for face processing, direct or indirect synaptic connections between these regions could mediate error-driven adaptations on visual attention (Ullsperger & Stork, 2021). Rodent studies suggested that direct connections between midfrontal and visual regions underly post-error upregulation of visual attention (Norman et al., 2021). Other studies emphasized that lateral frontoparietal network regions, such as dlPFC, are responsible for maintaining stimulus representations for memory formation (Curtis & D'Esposito, 2003; Nelissen et al., 2013). In the current study, representation strength was also associated with increased hemodynamic responses in dlPFC, suggesting it as an important node of a control network for attentional allocation. In this regard, effects between stimulus-specific regions, such as FFA in the ventral visual stream, and the midcingulo-insular salience network regions, such as pMFC and anterior insula, may be mediated by lateral frontoparietal control network upregulation (Menon, 2015). During encoding and rehearsal, FFA-based evidence for face-processing was, however, also related to a cluster at the border zone to the basal forebrain, a region important for modulating arousal (Liu et al., 2018; Turchi et al., 2018) and releasing the neuromodulator

acetylcholine. The cholinergic system has shown to mediate post-error upregulation of visual attention in a pharmacological fMRI study (Danielmeier et al., 2015). Further work is needed to determine to which degree these different pathways are exclusive or working in concert, to understand whether and when an error-driven increase of stimulus processing is caused by direct pMFC connections to stimulus-specific regions, mediated by lateral frontoparietal network regions such as dlPFC and/or modulated by the basal forebrain cholinergic system.

In conclusion, the current study showed that higher hemodynamic responses in pMFC are not only related to improved encoding but are already increased when there is evidence of currently-insufficient memory representations. Higher FFA-based face-processing evidence was accompanied by a systematic increase of hemodynamic responses in regions among pMFC, dlPFC, visual cortex, anterior insula, basal forebrain and amygdala. When sufficient evidence on memory errors has been detected, these regions may interact to increase attention during encoding and improve following learning attempts. In the past years, multivariate fMRI analyses have gained popularity and decoding accuracies of brain regions have been used as estimate for how much stimulus information is represented in neurophysiological data. In this regard, the current study highlights how multivariate stimulus-based models vary in correspondence with hemodynamic responses of a midcingulo-insular network node in pMFC, which may monitor task demands and detect memory errors. The results help explain in correspondence with which brain regions stimulus representations are enhanced for improved memory formation, and emphasize memory-error detection as a basis for adaptive task performance and associative learning. Future studies may implement single-trial analyses and investigate

multivariate processing evidence, to explain why memory formation fails or succeeds from time to time.

Methods

Participants

30 young adults (15 male, age 18-35 years) participated in the current fMRI study after checking inclusion criteria (body mass index between 20 and 30 kg/m^2, non-smokers, no history of psychiatric or neurological disorders, no metal implants) via phone interview. Participants gave written informed consent before the study began and were compensated with study credits or money (10 EUR per hour) for their time. They obtained written instructions on the behavioral tasks and task comprehension was checked within a practice session outside the scanner. Next, participants were positioned in the MRI scanner. The keyboard was placed under the right hand, a photoplethysmography sensor on the left middle finger and a breathing belt around the chest on the position of the highest elevation. The study was approved by the ethics committee of the medical faculty at Otto-von-Guericke University Magdeburg, Germany.

Stimuli

Publicly available images of emotionally neutral faces from the Picture Database of Morphed Faces (Jäger et al., 2005) and house images from the DalHouses sample (Filliter et al., 2016) were used. The background color of the house images was replaced with the same grey scale as in the face images. The tasks also contained eight differently tilted gabor patch stimuli with an orientation point in extension of the middle white stripe rendered with Psychtoolbox 3 with Matlab 2018a on a Windows 10 computer.

Behavioral tasks

In FALT (see **Fig. 1a**), participants learned to associate faces with gabor patches in eight possible orientations. Each trial began with an inter-trial-interval (ITI) showing a fixation cross in the middle of the screen for a jittered duration between 2500 and 6000 milliseconds (ms). Then, a face stimulus was presented and 1000 ms later a gabor patch appeared in a random but incorrect orientation. Participants had to choose the matching orientation with their right index finger for a left directed rotation and right ring finger for a right directed rotation. If they saw a face for the first time, they were instructed to make a guess. On subsequent encounters of the face, they should recall the associated orientation from their memory. After confirming their choice with the right middle finger, low and high confidence options were presented on screen, such that participants could indicate their recall certainty with respective index and ring finger presses. The side of presentation for low-confidence and high-confidence ratings was altered for each trial. After a 200 ms delay period, based on recall success, either positive or negative feedback was presented for 800 ms. At the end of each trial, the correct combination of face and gabor patch was presented for 1500 ms for (re-) encoding. Each face was presented four times, with at least two and a maximum of 15 trials until the next trial with the same face. The task consisted of five independent runs with eight new faces each, summing up to 160 trials in total. Between runs, participants were presented with a pause screen on which the relative number of correct trials was displayed. The next run with eight new face stimuli was resumed with a confirmation button press.

In the 1-back localizer task, on each trial, participants were presented a face or a house together with a gabor patch in one of eight possible orientations. They were instructed to attend and compare both stimuli with the stimulus combination shown in

the directly preceding trials, and to press the confirmation key as fast as possible when the presented stimulus combination was a direct repetition (see **Fig. 4a**). Presentation times were analog to the durations of encoding with 1500 ms and the ITI jittered between 2500 and 6000 ms as in FALT. Within each run, two new face and two new house stimuli were presented four times each, summing up to 80 trials for five runs in total. Direct repetitions occurred in two of 16 trials per run to keep participants engaged with attending, encoding and rehearsing the presented stimuli.

Data acquisition

Magnetic resonance imaging (MRI) data were obtained by a 3 Tesla Siemens Prisma scanner with a 64-channel head coil. After brief anatomical scout images, structural MRI data were assessed using a magnetization prepared rapid gradient echo sequence in sagittal slices (voxel size = 1 x 1 x 1 mm, matrix size = 192 x 256 x 256, repetition time = 2.5 s, echo time = 0.00282 s, flip angle = 7°, multi band factor = 2). While participants performed the FALT and the localizer task, fMRI scans were recorded with a field of view aligned to anterior and posterior commissures (voxel size = 2.2 x 2.2 x 2.2 mm, matrix size = 100 x 100 x 66, repetition time = 2.0 s, echo time = 0.03 s, flip angle = 80°, multi band factor = 2, interleaved order, no interslice gap). Single band reference images were recorded on the first and field maps after the last functional scan. Due to technical issues, one participant lacked the single-band reference image and two participants lacked peripheral physiological recordings.

fMRI preprocessing

MRI data were converted using dcm2niix (version v1.0.20190902), and renamed in accordance with Brain-Imaging-Data-Structure format (Gorgolewski et al., 2016). Data were analyzed on a high-performance computing cluster using Linux Debian (version 4.9.0-16-amd64). For preprocessing, fMRIPrep version 23.2.2

(Esteban et al., 2019) was run with a singularity image (version 2.6.1-dist) wrapped around a docker container. Preprocessing encompassed slice time correction, susceptibility distortion correction, boundary-based registration and spatial normalization to obtain images in *MNI152NLin2009cAsym* output space, keeping the size of 2.2 mm^3 voxels. Further details on fMRIPrep-based preprocessing pipeline can be found in the section **Supplementary Methods**. Physiological regressors for retrospective image correction of respiratory and cardiac confounds were obtained from the PhysIO package in the TAPAS toolbox (Kasper et al., 2017). For simultaneous denoising and fitting of event-related hemodynamic response functions, GLMs on the preprocessed images contained following confound regressors: 24 motion parameters (six rigid body motion parameters, six derivatives, and respective twelve squared motion parameters), 18 physiological regressors (six cardiac, eight respiration, four combined cardiac and respiration), ten anatomical component correction regressors (five white matter, five cerebrospinal fluid), the global signal, a cosine drift model and a constant intercept.

Behavioral analyses

In FALT, for in total 160 trials in four blocks and because participants had to guess in the first block, there were maximally 120 trials in which participants could remember the correct orientation of the associated gabor patch from a past learning opportunity. A one-sample *t*-test against chance level of 12.5 % was performed for the relative number of correct trials per participant, to determine whether the presented face and gabor patch associations were successfully learned. The performance increase between different blocks was assessed with a one-way analysis of variances (ANOVA) and post-hoc tests with Tukey's honestly significant differences (HSD). Participant's meta-memory performance (d_{Prime}) was assessed as the average of the

probability distribution between the proportion of high-confidence selections upon successful recall (sensitivity) and the proportion of low-confidence selections in failed recall trials (specificity). Sensitivity and specificity probability distributions functions were adjusted for infinite values by subtracting the proportion of one correct or incorrect trial respectively. Meta-memory performance d_{Prime} and bias d_{Bias} were tested for significance with a one-sample *t*-test.

In the 1-back localizer task, there were ten repetition trials on which participants had to press the confirmation key and 70 non-repetition trials where they were instructed to attend and encode the presented stimuli but not to press. According to signal detection theory, trial types were distinguished into hits for a correct press on a repetition, misses for a non-press on a repetition, correct rejections for a non-press on a non-repetition and false alarms for a press on a non-repetition. Task performance was evaluated based on hit and correct rejection rates and significance was tested using one-sample *t*-tests. Insufficient task comprehension of a participant was assumed for outliers, which were defined by a task performance being two standard deviations (*SD*) lower than the average (*M*) performance of all participants.

fMRI analyses

In FALT, univariate GLM fMRI analyses were conducted by simultaneously fitting a hemodynamic response function using the Glover model convolved with respective event regressors during memory recall ($Error_{ConfidenceLow}$ or $Error_{ConfidenceHigh}$ or $Correct_{ConfidenceLow}$ or $Correct_{ConfidenceHigh}$), confidence selection (low or high), feedback presentation (positive or negative), encoding as determined by current and subsequent recall success (*ErrorError* or *ErrorCorrect* or *CorrectCorrect* or *CorrectError in combination with respective confidence levels*). Two GLM analyses

were performed, one for the post-error subsequent memory effect and one for memory-error detection.

In the first GLM, neurophysiological signals related to recall, confidence and feedback were assessed and convolved as separate regressors with respective onset times, such that the shared variance is encompassed in the residual variance of the model. The encoding-related regressors were excluded because of the statistical and hemodynamic overlap with the feedback-related regressors. Multicollinearity between convolved regressors was examined using the variance-inflation-factor index, assuming moderate multicollinearity for values > 5 and < 10, and high multicollinearity for a variance-inflation-factor > 10. To determine the brain regions associated with performance monitoring of memory errors, fMRI contrasts were calculated for hemodynamic responses upon failed recall ($Error_{LowConfidence} > Correct_{HighConfidence}$) as implicit indication for a detected demand of better memory formation, the selection of recall uncertainty (low > high confidence) as discrete internal memory error evidence, and the presentation of memory error feedback (negative > positive) as external evidence.

In the second GLM, encoding regressors were used together with regressors for recall and for confidence while the feedback-related regressors were excluded because of the redundancy and temporal overlap with encoding regressors. To ensure that participants were aware of required memory demands before successful re-learning, the post-error SME was calculated between low-confident error trials which were later remembered with a high level of confidence and those error trials with subsequent failed recall and low confidence ($Error_{LowConfidence}Correct_{HighConfidence} > Error_{LowConfidence}Error_{LowConfidence}$).

In the 1-back localizer task, the univariate GLM analysis consisted of hemodynamic response functions convolved for faces and houses which were further differentiated into eight regressors based on four different signal detection theory trial types (hit, miss, correct rejection, false alarm), and denoising parameters as described in the section fMRI preprocessing. To determine which brain regions are systematically related to face-processing, a contrast on correct non-press trials ($Face_{CorrectRejection}$ > $House_{CorrectRejection}$) was calculated. The topography of significant clusters in FG was visually compared regarding its overlap with probabilistic cytoarchitectonic maps for right FG-2 and FG-4 regions (Eickhoff et al., 2005).

Upon statistical testing of the group results in a second level GLM, contrasts maps were smoothed with an 8 mm kernel and a voxel-wise false-discovery rate threshold was applied, removing clusters with an extent of less than five continuous voxels (equivalent to clusters of at least 53.24 mm^3).

Multivariate cross-classification

A key aim of the current study was to develop a quantitative proxy measure for stimulus-based attention as a link between error-driven demand detection and encoding success. For each participant, a multivariate model on face-processing was trained in the localizer task and later applied to memory-related epochs in FALT, such as memory recall, encoding and rehearsal (i.e., the intertrial interval). First, the univariate GLMs described in the previous sections were adapted for single-trial deconvolution according to the least-squares separate approach (Mumford et al., 2012) to obtain a series of beta-maps. In this regard, all correct rejection face and house trials in the localizer task were determined and stimulus presentation of each trial was once defined as target event in an additional, independent GLM. The target trial was convolved with a hemodynamic response function as a separate regressor,

while controlling for all other events and denoising parameters such as in conventional GLM analyses. In case participants showed optimal performance in the localizer task (i.e., they correctly identified all repetitions and did not display false alarms) a total of 70 single-trial (M = 69.13, SD = 2.29) beta-maps could be derived.

Based on the univariate fMRI results in the localizer task, bilateral cytoarchitectonic probability masks for FG-4 showed a strong overlap with increased hemodynamic responses during face processing. After smoothing with a 6 mm full-width at half-maximum kernel, deconvolved betaseries of voxels within the FG-4 mask were extracted and trials were labeled for five folds according to the presented run in the task. A balanced, probability-scaled linear support vector machine (C = 1) with a squared penalty function was trained on four of the five runs to predict whether trials from the left-out run were either faces or houses. Within the five-fold leave-one-run-out cross-validation a standard scaler (M = 0, SD = 1) was fit to the four training runs and applied to the left-out run. Univariate feature selection was applied by maintaining only the beta-weights of the 14 voxels with the strongest positive ANOVA effects, to obtain results for participant-specific FFA voxels and to reach a feature-to-sample ration of approximately 1:5 before fitting the support vector machine. Feature selection was only based on the training samples, both during cross-validation and cross-classification, to prevent leakage and overfitting. Decoding accuracies were evaluated by testing whether the average accuracies of the five runs per participant exceeded a chance level of 50 % with a one-sample t-test. Face and house trials were tested for equal decoding accuracies with a t-test for dependent samples to ensure that the FFA-based face-processing model was balanced and did not prefer either of the two categories.

After leave-one-run-out cross-validated model evaluation, trials from all five folds were included in model training. A full model was fit on correct rejection trials of all localizer task runs of a participant with the same scaling procedure and feature selection as during cross-validation. The support vector machine was fit on the 14 selected voxels of up to 70 correct rejection trials of all localizer runs of a participant, to be applied to the memory epochs for trials in FALT. Single-trial deconvolution and selection of FFA voxels was repeated for the 160 trials in FALT and the three memory-relevant epochs of stimulus recall, encoding and rehearsal (ITI). The machine learning model for evaluating the strength of stimulus representations then predicted the presented class and estimated the probability of face-processing for each trial in each of the epochs. To evaluate the validity of the single-trial face processing model, the correspondence between the average probability the face-class and the absolute number of predicted face-class trials was controlled by significance tests for the Spearman correlation coefficient. To assess which other regions are potentially involved in allocating attention to the presented stimulus category, the classifiers single-trial decoding probability of FFA-based face-processing evidence was then fit to all other voxels in a whole-brain GLM analysis on the single-trial betaseries for each of the memory-relevant epochs (recall, encoding, rehearsal), respectively. The predicted face class-probability parameter was compared between different behaviorally-assessed trial types during stimulus recall, encoding and rehearsal, to determine whether the proxy measure for allocated stimulus-based attention is increased after memory errors and related to a higher likelihood of successful memory formation, as determined by later recall success. For each of the three memory epochs, in a linear mixed model the representation strength measure was fit to the regressors encoding demand and subsequent recall success, while restricting the

683 analysis to ErrorError, ErrorCorrect and CorrectCorrect trials, and controlling within-
684 participant dependencies by using participant as group factor.

685 **Data and code accessibility**

686 Behavioral and fMRI analyses were based on custom Python code within
687 Jupyter Lab, using plotting functions from Matplotlib and Seaborn, numerical
688 processing and statistical testing with Numpy, Scipy, Pandas and Statsmodels, and
689 decoding tools from Scikitlearn and Nilearn (version 0.10.0; Abraham et al., 2014).
690 Visualization of fMRI results was based on MRIcroGL (version 1.2.20220720b). The
691 code used for behavioral and fMRI analyses, and the unthresholded statistical fMRI
692 maps will be uploaded on respective public repositories upon publication of the
693 manuscript.

694 **References**

695 Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi,
696 J.,…Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn.
697 *Frontiers in Neuroinformatics*, *8*, 14. https://doi.org/10.3389/fninf.2014.00014
698 Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making
699 memories: brain activity that predicts how well visual experience will be
700 remembered. *Science*, *281*(5380), 1185-1187.
701 https://doi.org/10.1126/science.281.5380.1185
702 Caspers, J., Zilles, K., Eickhoff, S. B., Schleicher, A., Mohlberg, H., & Amunts, K.
703 (2013). Cytoarchitectonical analysis and probabilistic mapping of two
704 extrastriate areas of the human posterior fusiform gyrus. *Brain Structure and*
705 *Function*, *218*(2), 511-526.

706 Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during
707 working memory. *Trends Cogn Sci*, *7*(9), 415-423.
708 https://doi.org/10.1016/s1364-6613(03)00197-9
709 Danielmeier, C., Allen, E. A., Jocham, G., Onur, O. A., Eichele, T., & Ullsperger, M.
710 (2015). Acetylcholine mediates behavioral and neural post-error control.
711 *Current biology : CB*, *25*(11), 1461–1468.
712 https://doi.org/10.1016/j.cub.2015.04.022
713 Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M.
714 (2011). Posterior medial frontal cortex activity predicts post-error adaptations in
715 task-related visual and motor areas. *J Neurosci*, *31*(5), 1780-1789.
716 https://doi.org/10.1523/jneurosci.4299-10.2011
717 de Bruijn, E. R. A., Mars, R. B., & Hester, R. (2020). Processing of performance errors
718 predicts memory formation: Enhanced feedback-related negativities for
719 corrected versus repeated errors in an associative learning paradigm.
720 *European Journal of Neuroscience*, *51*(3), 881-890.
721 https://doi.org/https://doi.org/10.1111/ejn.14566
722 Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., &
723 Zilles, K. (2005). A new SPM toolbox for combining probabilistic
724 cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*(4), 1325-
725 1335. https://doi.org/10.1016/j.neuroimage.2004.12.034
726 Erez, Y., & Duncan, J. (2015). Discrimination of Visual Categories Based on
727 Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *J*
728 *Neurosci*, *35*(36), 12383-12393. https://doi.org/10.1523/jneurosci.1134-
729 15.2015

730 Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe,
731 A.,…Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for
732 functional MRI. *Nat Methods*, *16*(1), 111-116. https://doi.org/10.1038/s41592-
733 018-0235-4
734 Filliter, J. H., Glover, J. M., McMullen, P. A., Salmon, J. P., & Johnson, S. A. (2016).
735 The DalHouses: 100 new photographs of houses with ratings of typicality,
736 familiarity, and degree of similarity to faces. *Behavior Research Methods*, *48*(1),
737 178-183. https://doi.org/10.3758/s13428-015-0561-8
738 Gilmore, A. W., Nelson, S. M., Naaz, F., Shaffer, R. A., & McDermott, K. B. (2018).
739 BOLD Activity During Correct-Answer Feedback in Cued Recall Predicts
740 Subsequent Retrieval Performance: An fMRI Investigation Using a Partial Trial
741 Design. *Cereb Cortex*, *28*(11), 4008-4022.
742 https://doi.org/10.1093/cercor/bhx264
743 Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E.
744 P.,…Poldrack, R. A. (2016). The brain imaging data structure, a format for
745 organizing and describing outputs of neuroimaging experiments. *Scientific*
746 *Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44
747 Gruendler, T. O. J., Ullsperger, M., & Huster, R. J. (2011). Event-related potential
748 correlates of performance-monitoring in a lateralized time-estimation task. *PloS*
749 *one*, *6*(10), e25591-e25591. https://doi.org/10.1371/journal.pone.0025591
750 Hester, R., Barre, N., Murphy, K., Silk, T. J., & Mattingley, J. B. (2008). Human medial
751 frontal cortex activity predicts learning from errors. *Cerebral cortex (New York,*
752 *N.Y. : 1991)*, *18*(8), 1933–1940. https://doi.org/10.1093/cercor/bhm219
753 Jäger, T., Seiler, K. H., & Mecklinger, A. (2005). Picture database of morphed faces
754 (MoFa): technical report. https://doi.org/10.23668/psycharchives.9011

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, *17*(11), 4302-4311. https://doi.org/10.1523/jneurosci.17-11-04302.1997

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S.,…Stephan, K. E. (2017). The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *Journal of Neuroscience Methods*, *276*, 56-72. https://doi.org/https://doi.org/10.1016/j.jneumeth.2016.10.019

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *Neuroimage*, *54*(3), 2446-2461. https://doi.org/10.1016/j.neuroimage.2010.09.045

Kirschner, H., & Ullsperger, M. (2024). The medial frontal cortex, performance monitoring, cognitive control, and decision making. In *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-820480-1.00048-6

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *Neuroimage*, *34*(4), 1774-1781. https://doi.org/10.1016/j.neuroimage.2006.11.014

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451-463. https://doi.org/10.1016/j.neuron.2016.12.040

Liu, X., de Zwart, J. A., Scholvinck, M. L., Chang, C., Ye, F. Q., Leopold, D. A., & Duyn, J. H. (2018). Subcortical evidence for a contribution of arousal to fMRI studies of brain activity. *Nat Commun*, *9*(1), 395. https://doi.org/10.1038/s41467-017-02815-3

780 Lorenz, S., Weiner, K. S., Caspers, J., Mohlberg, H., Schleicher, A., Bludau,
781 S.,…Amunts, K. (2015). Two New Cytoarchitectonic Areas on the Human Mid-
782 Fusiform Gyrus. *Cerebral Cortex*, *27*(1), 373-385.
783 https://doi.org/10.1093/cercor/bhv225
784 Menon, V. (2015). Salience Network. In A. W. Toga (Ed.), *Brain Mapping* (pp. 597-
785 611). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-
786 397025-1.00052-X
787 Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving
788 BOLD activation in event-related designs for multivoxel pattern classification
789 analyses. *Neuroimage*, *59*(3), 2636-2643.
790 https://doi.org/10.1016/j.neuroimage.2011.08.076
791 Nelissen, N., Stokes, M., Nobre, A. C., & Rushworth, M. F. (2013). Frontal and parietal
792 cortical interactions with distributed visual representations during selective
793 attention and action selection. *J Neurosci*, *33*(42), 16443-16458.
794 https://doi.org/10.1523/jneurosci.2625-13.2013
795 Norman, K. J., Riceberg, J. S., Koike, H., Bateh, J., McCraney, S. E., Caro,
796 K.,…Morishita, H. (2021). Post-error recruitment of frontal sensory cortical
797 projections promotes attention in mice. *Neuron*, *109*(7), 1202-1213.e1205.
798 https://doi.org/10.1016/j.neuron.2021.02.001
799 Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna,
800 H.,…Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for
801 Salience Processing and Executive Control. *The Journal of Neuroscience*,
802 *27*(9), 2349. https://doi.org/10.1523/JNEUROSCI.5587-06.2007
803 Shapira-Lichter, I., Oren, N., Jacob, Y., Gruberger, M., & Hendler, T. (2013).
804 Portraying the unique contribution of the default mode network to internally

driven mnemonic processes. *Proc Natl Acad Sci U S A*, *110*(13), 4950-4955. https://doi.org/10.1073/pnas.1209888110

Turchi, J., Chang, C., Ye, F. Q., Russ, B. E., Yu, D. K., Cortes, C. R.,…Leopold, D. A. (2018). The Basal Forebrain Regulates Global Resting-State fMRI Fluctuations. *Neuron*, *97*(4), 940-952.e944. https://doi.org/https://doi.org/10.1016/j.neuron.2018.01.032

Uddin, L. Q., Yeo, B. T. T., & Spreng, R. N. (2019). Towards a Universal Taxonomy of Macro-scale Functional Human Brain Networks. *Brain topography*, *32*(6), 926-942. https://doi.org/10.1007/s10548-019-00744-6

Ullsperger, M., & Stork, O. (2021). To err is (not only) human: Mechanisms of post-error attentional regulation illuminated in mice. *Neuron*, *109*(7), 1074-1076. https://doi.org/10.1016/j.neuron.2021.03.014

Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiol Learn Mem*, *91*(2), 139-154. https://doi.org/10.1016/j.nlm.2008.10.011

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M.,…Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*(5380), 1188-1191. https://doi.org/10.1126/science.281.5380.1188

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, *137*, 188-200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Supplementary Tables

Supplementary table 1 | Implicit memory error evidence. *Significant clusters during failed recall ($Error_{LowConfidence}$ > $Correct_{HighConfidence}$) in the feedback-based association learning task (FALT) for voxels with a $p_{FDR} < .05$ and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Occipital Inf R	27	-90	-9	6.298	65112
Lingual L	-36	-85	-16	6.149	
Occipital Inf L	-23	-96	-5	5.817	
Lingual L	-10	-92	-16	5.780	
Frontal Inf Tri R	51	28	20	5.167	6995
Frontal Inf Oper R	55	17	36	3.762	
unspecified in AAL	-1	17	1	4.961	11957
Caudate R	11	-2	14	4.797	
unspecified in AAL	-18	-24	20	4.694	
unspecified in AAL	0	-37	5	4.583	
Fusiform L	-29	-48	-9	4.909	2257
Frontal Inf Tri L	-45	28	23	4.867	6165
Frontal Inf Tri L	-51	22	29	4.675	
Precentral L	-38	4	34	3.762	
Cingulum Mid R	5	22	40	4.299	3002
Supp Motor Area L	-7	19	45	3.910	
Frontal Sup Medial R	5	33	42	3.297	
Supp Motor Area R	9	15	56	3.090	
unspecified in AAL	-23	-46	20	4.103	841
unspecified in AAL	-21	-57	25	3.627	
Cerebelum 10 L	-23	-35	-42	3.928	883
Cerebelum 4 5 L	-25	-28	-31	3.451	
Temporal Pole Sup R	66	6	-1	3.818	191
Supp Motor Area L	-1	22	67	3.595	202
Precentral R	18	-24	78	3.535	330
Temporal Sup R	49	-8	-3	3.530	489
Temporal Sup R	47	-6	-12	3.000	
unspecified in AAL	38	0	-14	3.496	308
Hippocampus R	36	-19	-9	3.446	223
Frontal Mid R	27	13	56	3.392	447
Heschl R	51	-17	9	3.343	404
Frontal Sup L	-23	66	7	3.334	255
unspecified in AAL	22	-32	53	3.322	95
unspecified in AAL	14	-15	23	3.290	117
Caudate R	20	-21	20	3.286	234

unspecified in AAL	-1	50	51	3.266	63
unspecified in AAL	29	-26	-1	3.266	117
Paracentral Lobule R	3	-32	73	3.089	53
Supp Motor Area R	9	8	69	3.083	63
Calcarine L	-14	-54	12	3.069	53
Temporal Sup R	69	-13	7	3.009	63
unspecified in AAL	27	-26	29	2.949	53
Heschl L	-45	-15	5	2.887	63

835

Supplementary table 2 | Internal memory error evidence. *Increased hemodynamic responses for the selection of low compared to high recall certainty during confidence selection in the feedback-based association learning task (FALT), for voxels with a* $p_{FDR} < .05$ *and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Frontal Mid R	36	48	20	5.774	33338
Frontal Sup R	22	13	65	5.572	
Frontal Mid R	42	35	29	4.952	
Frontal Mid R	44	28	42	4.853	
Precuneus R	9	-70	45	5.291	18506
Precuneus L	-7	-68	49	5.056	
Precuneus R	5	-70	56	4.983	
Precuneus R	14	-61	29	4.779	
SupraMarginal R	55	-46	29	5.038	20955
Parietal Inf R	51	-41	53	4.685	
Parietal Sup R	38	-61	56	4.397	
Parietal Inf R	47	-57	49	4.301	
Frontal Inf Tri L	-45	28	27	4.933	13416
Frontal Inf Tri L	-34	24	29	4.137	
Frontal Mid L	-38	55	12	3.951	
unspecified in AAL	-38	59	1	3.762	
Cingulum Mid R	3	35	36	4.866	3652
Supp Motor Area L	-3	19	47	4.369	
Frontal Sup Medial R	3	41	51	2.857	
Temporal Mid R	69	-21	-3	4.571	5292
Temporal Mid R	66	-30	-5	4.219	
Temporal Inf R	55	-30	-23	3.694	
Temporal Mid R	64	-50	-5	3.571	
Parietal Inf L	-51	-43	49	4.541	9040
Parietal Inf L	-32	-74	49	4.480	
Parietal Inf L	-36	-54	38	3.956	
Angular L	-43	-61	47	3.636	
unspecified in AAL	-25	-87	-45	3.835	489
unspecified in AAL	-36	-83	-45	2.906	
unspecified in AAL	-5	-32	-16	3.776	170
unspecified in AAL	0	-24	27	3.610	500
Cingulum Mid R	9	-26	31	2.938	
Precuneus R	11	-37	3	3.579	383
unspecified in AAL	0	-30	7	2.922	
Insula L	-29	26	-5	3.481	468

Supp Motor Area L	-12	6	67	3.439	1331
Frontal Mid L	-21	11	62	3.218	
unspecified in AAL	0	-13	-27	3.328	149
Frontal Sup R	22	59	29	3.315	117
unspecified in AAL	9	-26	-18	3.168	106
unspecified in AAL	-25	37	-25	3.151	138
Frontal Sup R	20	70	3	3.052	63
unspecified in AAL	-23	-26	29	3.036	85
Frontal Mid Orb L	-18	50	-16	2.978	138
unspecified in AAL	-18	-52	27	2.916	74

842

***Supplementary table 3 | External memory error evidence**. Hemodynamic responses increased for negative compared to positive feedback in the feedback-based association learning task (FALT), for voxels with a $p_{FDR} < .05$ and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Occipital Mid R	36	-76	34	7.618	187085
Parietal Inf L	-34	-54	45	7.360	
Occipital Mid L	-27	-70	25	6.818	
Parietal Inf L	-27	-79	42	6.782	
Insula R	31	26	-1	6.958	33605
Frontal Sup R	29	4	60	6.312	
Frontal Inf Oper R	51	11	29	5.522	
Precentral R	27	-4	47	5.077	
Insula L	-29	24	1	6.881	60416
Frontal Inf Tri L	-49	26	27	6.806	
Supp Motor Area R	5	17	51	6.804	
Frontal Mid L	-25	0	58	6.694	
Cerebelum 6 R	11	-74	-25	6.314	13459
Cerebelum 7b L	-29	-70	-47	6.020	
Cerebelum Crus1 L	-7	-74	-25	5.194	
Cerebelum Crus2 R	5	-81	-36	4.717	
Cerebelum 9 R	9	-52	-51	5.677	2214
Cerebelum 8 R	29	-72	-49	5.070	3322
unspecified in AAL	-1	-32	-25	3.799	819
ParaHippocampal L	-29	-43	-7	3.778	393
unspecified in AAL	-3	6	-42	3.546	85
Cerebelum 8 R	31	-41	-47	3.468	170
Lingual R	5	-85	-7	3.170	181
Frontal Sup Orb R	25	52	-7	3.151	351
unspecified in AAL	25	37	-25	3.080	117
unspecified in AAL	5	-6	-3	3.016	106
unspecified in AAL	27	33	-27	2.913	53
unspecified in AAL	-25	33	-27	2.890	223
Fusiform R	36	-15	-36	2.870	340
Fusiform R	33	-6	-34	2.727	
unspecified in AAL	-16	8	9	2.775	223
Fusiform L	-40	-32	-20	2.619	95
Temporal Sup R	44	-28	3	2.564	85
Frontal Mid Orb L	-38	44	-3	2.549	95

Supplementary table 4 | The post-error subsequent memory effect. *The table shows regions with increased hemodynamic responses for later recall success during the encoding epochs which followed memory errors (*$Error_{LowConfidence}Correct_{HighConfidence}$ *>* $Error_{LowConfidence}Error_{LowConfidence}$*), for voxels with a* $p_{FDR} < .05$ *and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Supp Motor Area L	-3	0	71	5.370	2299
Supp Motor Area L	-10	4	73	5.149	
Supp Motor Area L	-1	13	69	4.387	
Occipital Mid L	-27	-79	40	4.374	191
Frontal Inf Tri L	-49	28	12	4.323	372
Frontal Inf Tri L	-56	24	23	3.799	
Fusiform L	-40	-54	-18	4.294	1224
Temporal Inf L	-49	-50	-18	4.192	
unspecified in AAL	-62	-63	-1	4.293	692
Temporal Mid L	-43	-74	20	4.250	202
Frontal Inf Oper L	-56	8	9	4.040	628
Frontal Inf Tri L	-54	15	5	3.845	
Occipital Inf R	47	-72	-14	3.991	340
Temporal Inf R	51	-65	-12	3.912	
unspecified in AAL	-51	26	-20	3.979	106
unspecified in AAL	-47	-35	-1	3.945	63
Cerebelum Crus1 R	49	-61	-25	3.943	170
Temporal Pole Sup L	-27	13	-27	3.855	53
Precentral L	-54	0	49	3.830	298
Fusiform L	-25	-57	-16	3.748	74

Supplementary table 5* | *Face-selective regions. *The table displays regions found to show increased hemodynamic responses for faces compared to houses in the 1-back localizer task, for voxels with a $p_{FDR} < .05$ and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Precuneus R	3	-65	38	6.401	22424
Precuneus R	3	-52	20	5.282	
Calcarine L	-14	-79	12	4.483	
Cingulum Post L	-7	-46	31	4.447	
Cerebelum 6 R	44	-46	-27	5.477	3726
Temporal Mid R	53	-65	5	5.246	18357
Temporal Mid R	53	-57	7	5.129	
Temporal Mid R	51	-39	5	4.482	
unspecified in AAL	44	-41	16	4.370	
Frontal Mid R	22	28	40	5.137	4099
Hippocampus R	18	-6	-16	5.115	1054
Frontal Med Orb L	-3	46	-14	5.036	25597
Rectus R	3	44	-18	4.945	
Frontal Med Orb R	0	52	-12	4.921	
Frontal Sup L	-27	66	9	4.582	
Temporal Pole Sup R	40	19	-31	4.873	5717
Temporal Pole Sup R	31	8	-27	4.565	
unspecified in AAL	31	8	-18	4.321	
Temporal Pole Mid R	51	13	-34	3.617	
Temporal Pole Mid L	-45	17	-29	4.720	5206
Temporal Pole Sup L	-32	8	-25	4.486	
Amygdala L	-18	-4	-14	3.924	
unspecified in AAL	-21	0	-7	3.367	
Fusiform L	-40	-52	-23	4.683	1533
Temporal Mid L	-62	-10	-12	4.307	2374
Temporal Sup L	-45	-15	-12	4.125	
Temporal Mid L	-54	-15	-9	4.101	
Temporal Mid L	-54	-4	-18	2.941	
Frontal Inf Tri R	38	28	12	4.048	340
Temporal Mid R	51	-6	-18	4.036	3162
Temporal Sup R	66	-8	-9	3.94	
Temporal Sup R	55	-8	-7	3.638	
Temporal Mid R	64	-2	-18	3.282	
Cingulum Mid L	-1	-19	40	3.966	1693
Cingulum Mid L	0	-4	38	3.156	
unspecified in AAL	0	6	-12	3.956	1341
unspecified in AAL	-51	-74	12	3.861	4525

Temporal Mid L	-49	-59	9	3.619	
Angular L	-49	-68	25	3.312	
unspecified in AAL	-51	-70	38	3.160	
Cerebelum Crus2 L	-7	-85	-40	3.816	585
Parietal Sup R	20	-52	62	3.689	266
unspecified in AAL	16	-43	58	3.111	
Precentral R	27	-17	78	3.642	234
unspecified in AAL	33	-21	73	3.091	
SupraMarginal R	58	-28	20	3.635	947
Cerebelum Crus1 R	49	-74	-38	3.575	298
Cerebelum Crus1 R	40	-79	-36	2.911	
unspecified in AAL	-25	-50	16	3.568	415
unspecified in AAL	-23	-41	18	3.013	
Insula L	-45	8	-5	3.561	681
Frontal Inf Orb L	-45	17	-12	3.372	
Postcentral L	-43	-24	29	3.525	181
Frontal Mid L	-27	26	58	3.502	2587
Frontal Mid L	-25	35	49	3.349	
Frontal Mid L	-21	24	51	3.306	
Frontal Sup L	-14	24	65	3.204	
unspecified in AAL	-21	41	7	3.475	436
Precentral L	-23	-24	80	3.454	170
Postcentral L	-25	-41	60	3.429	212
Thalamus R	11	-30	9	3.400	191
Frontal Sup L	-18	13	67	3.199	266
Supp Motor Area R	7	-15	58	3.197	53
Hippocampus L	-14	-39	5	3.189	63
Parietal Sup L	-23	-48	76	3.146	212
unspecified in AAL	-62	-17	-29	3.143	74
Cerebelum Crus1 R	51	-61	-38	3.114	127
unspecified in AAL	-7	-6	12	3.110	74
Frontal Mid R	25	17	56	3.074	95
Temporal Mid R	69	-37	1	3.033	63
Precentral R	42	-13	67	3.024	85
Frontal Inf Orb L	-36	28	-18	2.996	74
unspecified in AAL	-34	-61	9	2.993	63
Frontal Inf Orb L	-23	28	-16	2.987	95
Cingulum Ant R	3	41	5	2.965	223
Cerebelum 7b R	42	-59	-49	2.857	63

863

Supplementary table 6 | House-selective regions. *The table shows which regions displayed increased hemodynamic responses for houses compared to faces in the 1-back localizer task, for voxels with a* $p_{FDR} < .05$ *and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Lingual L	-29	-52	-5	7.019	111399
Fusiform R	29	-52	-3	6.809	
Occipital Mid R	36	-85	16	6.527	
Occipital Mid L	-32	-90	12	6.445	
Cerebelum 7b L	-29	-74	-51	4.298	1086
unspecified in AAL	7	-68	-58	4.064	564
Precentral L	-51	4	40	3.937	1788
Frontal Inf Orb L	-29	28	-7	3.767	244
Cerebelum 8 R	27	-72	-51	3.759	457
Temporal Inf R	55	-50	-9	3.577	468
unspecified in AAL	3	-8	-25	3.433	127
unspecified in AAL	-5	-37	-25	3.371	138
unspecified in AAL	-16	30	-31	3.160	74
Cerebelum Crus2 R	5	-76	-36	3.080	63
Temporal Inf R	49	-61	-9	3.025	95
Cerebelum Crus1 L	-56	-57	-34	3.015	74
Supp Motor Area R	9	15	47	2.997	53
Parietal Inf R	44	-37	53	2.879	53

Supplementary table 7 | Recall-related face processing regions. *Significant clusters related to higher evidence for face-processing during single-trial recall epochs in the feedback-based association learning task (FALT), for voxels with a* $p_{FDR} < .05$ *and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Cuneus R	14	-96	7	5.563	89794
Occipital Inf L	-23	-92	-5	5.008	
Occipital Inf R	31	-85	-16	4.986	
unspecified in AAL	38	-92	-16	4.944	
Cerebelum 4 5 R	7	-46	-16	4.813	2683
Cerebelum 4 5 L	-5	-52	-14	3.761	
Cerebelum 4 5 L	-7	-48	-3	3.498	
Supp Motor Area R	3	17	49	4.805	3481
Supp Motor Area L	-7	17	49	3.602	
unspecified in AAL	25	37	-25	4.403	745
unspecified in AAL	20	30	-29	3.949	
Frontal Inf Oper L	-38	2	27	4.303	3854
Frontal Inf Tri L	-43	17	27	3.529	
Frontal Inf Oper L	-54	22	31	3.411	
Angular R	27	-59	42	4.262	3907
Parietal Sup R	20	-68	47	3.519	
Parietal Sup R	31	-72	53	3.478	
Parietal Sup R	29	-63	51	3.167	
Cerebelum Crus1 L	-12	-68	-29	4.012	1299
unspecified in AAL	-14	-54	-36	3.174	
Insula R	29	24	-1	3.947	500
unspecified in AAL	18	8	29	3.934	287
Cerebelum 10 R	22	-37	-47	3.869	1181
unspecified in AAL	11	4	1	3.854	255
Parietal Inf L	-38	-54	42	3.852	2108
Thalamus L	-14	-13	5	3.818	181
Cerebelum 8 R	33	-65	-56	3.743	1181
unspecified in AAL	-5	-26	29	3.743	170
Paracentral Lobule R	9	-35	65	3.687	117
Pallidum L	-18	-2	-3	3.623	181
unspecified in AAL	-10	-2	25	3.598	436
Cingulum Ant L	-5	4	27	3.073	
Cerebelum 10 L	-25	-35	-40	3.576	851
Insula L	-32	24	-1	3.530	404
unspecified in AAL	36	17	20	3.524	255

Cerebelum 8 L	-27	-68	-53	3.517	383
unspecified in AAL	-23	-32	14	3.404	85
unspecified in AAL	-27	-32	1	3.347	255
unspecified in AAL	5	-2	23	3.341	53
Caudate L	-18	-19	23	3.273	74
unspecified in AAL	16	-30	-3	3.225	543
unspecified in AAL	5	-30	-1	3.103	
ParaHippocampal R	36	-13	-29	3.207	223
unspecified in AAL	25	-52	-34	3.175	63
Postcentral L	-21	-50	53	3.151	53
unspecified in AAL	-16	28	-9	3.128	53
Frontal Inf Tri R	53	33	25	3.077	340
Caudate L	-7	8	1	3.014	85
Precuneus R	16	-61	27	3.000	85
unspecified in AAL	-1	-30	-18	2.984	53
unspecified in AAL	25	-32	31	2.975	85
Parietal Sup L	-21	-63	53	2.975	149

876

Supplementary table 8 | Encoding-related face processing regions. *Significant clusters related to higher evidence for face-processing during single-trial encoding epochs in the feedback-based association learning task (FALT), for voxels with a* p_{FDR} *< .05 and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Precentral L	-38	-4	53	6.245	11915
Frontal Inf Tri L	-43	13	25	4.556	
Frontal Inf Tri L	-49	33	16	2.928	
Vermis 9	-1	-57	-38	6.080	121834
Occipital Mid L	-10	-103	1	5.575	
Occipital Inf R	38	-70	-9	5.345	
Cerebelum 6 L	-40	-52	-25	5.290	
Supp Motor Area R	5	13	47	5.240	4940
Supp Motor Area L	-5	6	60	3.700	
Parietal Sup R	29	-72	51	4.614	17643
unspecified in AAL	31	-50	36	4.523	
Occipital Mid R	36	-68	25	4.214	
Angular R	25	-63	47	4.153	
Parietal Inf L	-34	-50	51	4.301	16504
Parietal Inf L	-38	-54	42	4.238	
Parietal Sup L	-27	-61	45	4.154	
Parietal Sup L	-36	-63	53	4.016	
unspecified in AAL	25	35	-23	4.080	1341
Fusiform R	40	-8	-34	3.949	1235
Cerebelum 9 R	20	-39	-47	3.802	287
unspecified in AAL	-27	35	-25	3.767	383
Temporal Sup R	44	-35	3	3.656	425
Insula L	-34	17	-3	3.517	873
Frontal Inf Tri L	-32	30	1	3.264	
Cerebelum 8 L	-25	-70	-49	3.506	734
unspecified in AAL	25	-26	-1	3.479	617
Frontal Inf Tri R	49	30	16	3.465	2683
Frontal Inf Tri R	42	15	23	3.159	
unspecified in AAL	36	11	20	3.139	
Paracentral Lobule R	9	-39	65	3.386	287
Cerebelum 9 L	-18	-39	-47	3.368	138
unspecified in AAL	16	-4	-9	3.196	106
Temporal Pole Sup L	-25	6	-25	3.195	276
Lingual R	22	-50	-1	3.168	117
Thalamus R	16	-15	14	3.146	223

unspecified in AAL	-5	-46	-23	3.126	138
Hippocampus L	-14	-6	-14	3.114	85
Precuneus L	-7	-79	47	3.073	159
unspecified in AAL	-40	-21	-42	3.006	53
Supp Motor Area R	5	-19	53	2.986	63
Calcarine R	16	-72	9	2.958	170
unspecified in AAL	29	26	-3	2.901	212
Frontal Mid R	38	-2	58	2.796	95
Paracentral Lobule L	-1	-35	62	2.730	85
unspecified in AAL	5	6	-31	2.724	53
unspecified in AAL	-5	-30	27	2.709	53
Precentral R	33	-2	45	2.658	63

883

Supplementary table 9 | Rehearsal-related face processing regions. *Significant clusters related to higher evidence for face-processing during single-trial fixation epochs, which may have been used by the participants for stimulus rehearsal in the feedback-based association learning task (FALT), for voxels with a* $p_{FDR} < .05$ *and clusters of at least 5 continuous voxels according to automated anatomical labeling (AAL) atlas. Clusters without specified cluster size represent subclusters of above specified regions.*

Region	X	Y	Z	Peak statistic	Cluster size
Cerebelum 6 L	-36	-57	-23	5.999	129266
Occipital Inf R	36	-70	-9	5.752	
unspecified in AAL	-7	-107	5	5.483	
Occipital Mid L	-16	-96	3	5.119	
Postcentral R	11	-37	69	4.749	1022
Postcentral R	18	-32	73	3.146	
Paracentral Lobule L	-1	-35	60	2.917	
Cerebelum 8 R	27	-70	-51	4.281	2236
Cerebelum 8 R	29	-59	-51	3.477	
Parietal Sup R	29	-65	56	4.242	15386
Angular R	25	-61	47	4.230	
Parietal Sup R	31	-70	56	4.220	
Parietal Sup R	22	-65	58	4.137	
Frontal Inf Oper L	-38	8	27	3.903	2651
Frontal Inf Tri L	-40	19	29	3.558	
Precentral L	-36	-4	58	3.891	2491
Precentral L	-40	-2	51	3.794	
Frontal Sup L	-21	-2	47	3.546	
Frontal Mid L	-27	2	51	2.937	
Parietal Sup L	-18	-63	53	3.886	15056
Parietal Sup L	-34	-65	53	3.784	
Parietal Sup L	-27	-68	47	3.765	
Parietal Inf L	-34	-54	51	3.672	
Postcentral L	-21	-48	56	3.869	181
unspecified in AAL	-23	39	-27	3.793	255
unspecified in AAL	-23	-30	9	3.714	1405
unspecified in AAL	-32	-30	7	3.456	
Thalamus L	-10	-13	9	3.452	
Cerebelum 8 L	-29	-70	-51	3.690	1213
Frontal Sup L	-12	11	47	3.672	351
Supp Motor Area L	-1	11	56	3.620	702
unspecified in AAL	-16	-39	65	3.592	276
Calcarine L	-3	-72	7	3.580	819

Frontal Mid R	38	-2	58	3.497	202
Frontal Inf Oper R	42	13	25	3.477	873
ParaHippocampal L	-18	4	-25	3.475	543
Fusiform R	33	-4	-34	3.466	1288
Fusiform R	40	-10	-31	3.364	
Temporal Inf R	62	-39	-23	3.464	212
Supp Motor Area R	14	0	62	3.419	149
Vermis 3	5	-43	-18	3.365	447
unspecified in AAL	3	2	-14	3.326	319
unspecified in AAL	-1	-2	-7	2.862	
Hippocampus L	-16	-6	-14	3.276	468
unspecified in AAL	25	-15	36	3.229	106
Occipital Sup L	-10	-83	47	3.223	170
unspecified in AAL	5	-28	-3	3.069	149
Cerebelum 10 L	-23	-41	-42	3.050	266
Precuneus L	-12	-52	71	2.994	149
Amygdala R	20	0	-12	2.992	181
Calcarine R	20	-52	3	2.990	202
unspecified in AAL	-29	-28	-47	2.875	74
unspecified in AAL	-49	-26	65	2.854	53
Thalamus L	-7	-19	18	2.831	63

891

Supplementary table 10 | Recall-related face representation strength. *Mixed linear model regression results fit to the probability-scaled likelihood of face-processing during deconvolved single-trial recall epochs during the feedback-based association learning task (FALT).*

Model:	MixedLM	Dependent variable		Face Probability (Recall)		
Number of observations:	3183	Method:		REML		
Number of groups:	28	Scale:		0.0480		
Minimal group size:	101	Log-Likelihood:		248.0065		
Maximal group size:	120	Converged:		Yes		
Mean group size:	113.7					
	Coefficient	**Standard error**	***Z***	***p***	**[0.025**	**0.975]**
Intercept	0.664	0.030	22.235	<0.001	0.606	0.723
Encoding demand	0.032	0.005	6.894	<0.001	0.023	0.041
Subsequent recall success	0.011	0.006	1.922	0.055	-0.000	0.022
Group variable	0.024	0.031	-	-	-	-

Supplementary table 11 | Encoding-related face representation strength. *Mixed linear model regression results fit to the probability-scaled likelihood of face-processing during deconvolved single-trial encoding epochs.*

Model:	MixedLM	Dependent variable	Face Probability (Encoding)
Number of observations:	3183	Method:	REML
Number of groups:	28	Scale:	0.0654
Minimal group size:	101	Log-Likelihood:	-235.6007
Maximal group size:	120	Converged:	Yes
Mean group size:	113.7		

	Coefficient	**Standard error**	***z***	***p***	**[0.025**	**0.975]**
Intercept	0.435	0.026	16.423	<0.001	0.383	0.487
Encoding demand	0.039	0.005	7.176	<0.001	0.028	0.049
Subsequent recall success	0.015	0.007	2.254	0.024	0.002	0.028
Group variable	0.019	0.021	-	-	-	-

Supplementary table 12 | Rehearsal-related face representation strength. *Mixed linear model regression results fit to the probability-scaled likelihood of face-processing during deconvolved single-trial rehearsal epochs during the presentation of the fixation cross in the feedback-based association learning task (FALT).*

Model:	MixedLM	Dependent variable	Face Probability (Rehearsal)
Number of observations:	3183	Method:	REML
Number of groups:	28	Scale:	0.0382
Minimal group size:	101	Log-Likelihood:	606.6500
Maximal group size:	120	Converged:	Yes
Mean group size:	113.7		

	Coefficient	**Standard error**	***z***	***p***	**[0.025**	**0.975]**
Intercept	0.244	0.031	7.752	<0.001	0.182	0.305
Encoding demand	0.006	0.004	1.571	0.116	-0.002	0.015
Subsequent recall success	-0.002	0.005	-0.343	0.732	-0.012	0.008
Group variable	0.027	0.038	-	-	-	-

Supplementary Methods

MRI data preprocessing

Results included in this manuscript come from preprocessing performed using fMRIPrep 23.2.2 (Esteban et al. (2019); Esteban et al. (2018); RRID:SCR_016216), which is based on Nipype 1.8.6 (K. Gorgolewski et al. (2011); K. J. Gorgolewski et al. (2018); RRID:SCR_002502).

Preprocessing of B0 inhomogeneity mappings

A total of 1 fieldmaps were found available within the input BIDS structure for this particular subject. A B0 nonuniformity map (or fieldmap) was estimated from the phase-drift map(s) measure with two consecutive GRE (gradient-recalled echo) acquisitions. The corresponding phase-map(s) were phase-unwrapped with prelude (FSL None).

Anatomical data preprocessing

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1w image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.5.0 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL (version unknown), RRID:SCR_002823, Zhang, Brady, and Smith 2001). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.5.0), using brain-extracted versions of both T1w reference and the T1w template. The following

template was were selected for spatial normalization and accessed with TemplateFlow (23.1.0, Ciric et al. 2022): ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

For each of the 3 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume was generated, using a custom methodology of fMRIPrep, for use in head motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL , Jenkinson et al. 2002). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. The BOLD reference was then co-registered to the T1w reference using mri_coreg (FreeSurfer) followed by flirt (FSL , Jenkinson and Smith 2001) with the boundary-based registration (Greve and Fischl 2009) cost-function. Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are

estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat, Reynolds, and Birn 2017). All resamplings can be performed with a single interpolation step by composing all the

pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using nitransforms, configured with cubic B-spline interpolation.

Many internal operations of fMRIPrep use Nilearn 0.10.2 (Abraham et al. 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts unchanged. It is released under the CC0 license.

References

Abraham, Alexandre, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. 2014. "Machine Learning for Neuroimaging with Scikit-Learn." *Frontiers in Neuroinformatics 8*. https://doi.org/10.3389/fninf.2014.00014.

Avants, B. B., C. L. Epstein, M. Grossman, and J. C. Gee. 2008. "Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain." *Medical Image Analysis 12* (1): 26–41. https://doi.org/10.1016/j.media.2007.06.004.

Behzadi, Yashar, Khaled Restom, Joy Liau, and Thomas T. Liu. 2007. "A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI." *NeuroImage 37* (1): 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Ciric, R., William H. Thompson, R. Lorenz, M. Goncalves, E. MacNicol, C. J. Markiewicz, Y. O. Halchenko, et al. 2022. “TemplateFlow: FAIR-Sharing of Multi-Scale, Multi-Species Brain Models.” *Nature Methods 19*: 1568–71. https://doi.org/10.1038/s41592-022-01681-2.

Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, et al. 2018. “fMRIPrep 23.2.2.” *Software*. https://doi.org/10.5281/zenodo.852659.

Esteban, Oscar, Christopher Markiewicz, Ross W Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, et al. 2019. “fMRIPrep: A Robust Preprocessing Pipeline for Functional MRI.” *Nature Methods 16*: 111–16. https://doi.org/10.1038/s41592-018-0235-4.

Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins. 2009. “Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood.” *NeuroImage 47*, Supplement 1: S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

Gorgolewski, K., C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. 2011. “Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python.” *Frontiers in Neuroinformatics 5*: 13. https://doi.org/10.3389/fninf.2011.00013.

Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, et al. 2018. “Nipype.” *Software*. https://doi.org/10.5281/zenodo.596855.

Greve, Douglas N, and Bruce Fischl. 2009. “Accurate and Robust Brain Image Alignment Using Boundary-Based Registration.” *NeuroImage 48* (1): 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith. 2002. "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images." *NeuroImage 17* (2): 825–41. https://doi.org/10.1006/nimg.2002.1132.

Jenkinson, Mark, and Stephen Smith. 2001. "A Global Optimisation Method for Robust Affine Registration of Brain Images." Medical Image Analysis 5 (2): 143–56. https://doi.org/10.1016/S1361-8415(01)00036-6.

Patriat, Rémi, Richard C. Reynolds, and Rasmus M. Birn. 2017. "An Improved Model of Motion-Related Signal Changes in fMRI." *NeuroImage 144*, Part A (January): 74–82. https://doi.org/10.1016/j.neuroimage.2016.08.051.

Power, Jonathan D., Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2014. "Methods to Detect, Characterize, and Remove Motion Artifact in Resting State fMRI." *NeuroImage 84* (Supplement C): 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048.

Satterthwaite, Theodore D., Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, et al. 2013. "An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data." *NeuroImage 64* (1): 240–56. https://doi.org/10.1016/j.neuroimage.2012.08.052.

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. "N4ITK: Improved N3 Bias Correction." *IEEE Transactions on Medical Imaging 29* (6): 1310–20. https://doi.org/10.1109/TMI.2010.2046908.

Zhang, Y., M. Brady, and S. Smith. 2001. "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization

1055 Algorithm." *IEEE Transactions on Medical Imaging 20* (1): 45–57.
1056 https://doi.org/10.1109/42.906424.