Cerebral Cortex, October 2019; 4154–4168

doi: 10.1093/cercor/bhy298
Advance Access Publication Date: 7 December 2018
Original Article

ORIGINAL ARTICLE

Dorsolateral Prefrontal Cortex Enables Updating of Established Memories

Lisa Marieke Kluen[1], Lisa Catherine Dandolo[1], Gerhard Jocham[2,3] and Lars Schwabe[1]

[1]Department of Cognitive Psychology, Institute of Psychology, University of Hamburg, 20146 Hamburg, Germany, [2]Institute of Experimental Psychology, Heinrich Heine University Düsseldorf, Germany and [3]Center for Behavioral Brain Sciences, Otto von Guericke University Magdeburg, Germany

Address correspondence to Lars Schwabe, Department of Cognitive Psychology, Institute of Psychology, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany. Email: lars.schwabe@uni-hamburg.de

Abstract

Updating established memories in light of new information is fundamental for memory to guide future behavior. However, little is known about the brain mechanisms by which existing memories can be updated. Here, we combined functional magnetic resonance imaging and multivariate representational similarity analysis to elucidate the neural mechanisms underlying the updating of consolidated memories. To this end, participants first learned face–city name pairs. Twenty-four hours later, while lying in the MRI scanner, participants were required to update some of these associations, but not others, and to encode entirely new pairs. Updating success was tested again 24 h later. Our results showed increased activity of the dorsolateral prefrontal cortex (dlPFC) specifically during the updating of existing associations that was significantly stronger than when simple retrieval or new encoding was required. The updating-related activity of the dlPFC and its functional connectivity with the hippocampus were directly linked to updating success. Furthermore, neural similarity for updated items was markedly higher in the dlPFC and this increase in dlPFC neural similarity distinguished individuals with high updating performance from those with low updating performance. Together, these findings suggest a key role of the dlPFC, presumably in interaction with the hippocampus, in the updating of established memories.

Key words: hippocampus, memory, prefrontal cortex

Episodic memories allow us to mentally travel back in time and relive events from our past (Tulving 2002). Beyond remembering the past, these mnemonic representations support future survival. They enable us to imagine and simulate upcoming events (Schacter et al. 2007; Jing et al. 2017), to guide our attention and current decision-making (Chun and Turk-Browne 2007; Wimmer and Shohamy 2012). In order to accomplish these prospective functions, it is fundamental that memories are updated in light of new information. Indeed, it is increasingly acknowledged that memories are highly dynamic entities (Dudai 2012; Kroes and Fernandez 2012; Nadel et al. 2012; Schwabe et al. 2014) and there is considerable evidence that consolidated memories can be modified as a function of current experience (Baddeley and Dale 1966; Loftus 1975; Schiller et al. 2010; Zeithamova et al. 2012). However, although the updating of established memories is essential for our adaptation to changing environments, the neural mechanisms underlying the updating process of consolidated memories are not well understood.

A prime candidate that may contribute to the updating of established memories is the dorsolateral prefrontal cortex (dlPFC), a brain region that is thought to support relational encoding (Murray and Ranganath 2007; Blumenfeld et al. 2011) and strategic aspects of memory retrieval (Simons and Spiers

2003; Manenti et al. 2010). Decades of research have linked the dlPFC to cognitive control processes, such as monitoring and inhibition (Egner and Hirsch 2005; Cole and Schneider 2007), which are critical in the context of memory updating. Indeed, the dlPFC appears to play a causal role in the strengthening of memory after reactivation (Sandrini et al. 2013, 2014). Moreover, the dlPFC is directly connected to medial temporal lobe (MTL) areas, such as the hippocampus (Bilek et al. 2013; Preston and Eichenbaum 2013), that are crucial for memory formation and storage (Squire and Zola-Morgan 1991; Alvarez and Squire 1994). Through its interaction with the hippocampus, the dlPFC may orchestrate encoding and incorporation of updated information and suppress the reactivation of old memory representations (Anderson et al. 2004; Depue et al. 2007), thus representing a potential mechanism that enables updating of established memories.

Updating processes have recently been investigated in short-term memory (Kuhl et al. 2012; Schlichting and Preston 2016) and there is evidence for an important role of the dlPFC in working memory updating (D'Ardenne et al. 2012). However, the timescales of working memory processes and consolidated long-term memories are clearly distinct. While the dlPFC is known to be crucial for the maintenance of working memory representations (Fuster and Alexander 1971; Cohen et al. 1997), this is not the case for long-term memories, which are stored (at least transiently) in MTL areas (Squire and Zola-Morgan 1991; Burgess et al. 2002). It remains unclear to what extent updating processes in long-term memory resemble those in working memory and whether the dlPFC is implicated in the updating of established memories is completely unknown.

In the present study, we combined functional magnetic resonance imaging (fMRI) with multivariate representational similarity analysis (RSA) to elucidate the brain mechanisms supporting the updating of consolidated long-term memories. Participants were tested in a novel experimental paradigm on 3 consecutive days. On day 1, they learned a number of face–city name associations (Fig. 1A). Twenty-four hours later, some of the learned faces were paired with new cities, requiring participants to update the encoded associations. This updating phase was performed in the MRI scanner and included also face–city pairs that were not updated as well as entirely new face–city pairs, thus allowing us to control for simple retrieval as well as new learning, respectively. On day 3, memory updating success was assessed in a recognition test. We hypothesized that the dlPFC would be critically involved in successful memory updating, presumably in interaction with the hippocampus. In addition to the dlPFC and MTL areas, we focused on structures implicated in the representation of prior knowledge, such as the ventromedial prefrontal cortex (vmPFC) and angular gyrus, that have also been shown to be critically implicated in schema-based learning processes (van Kesteren et al. 2012; Wagner et al. 2015). Activity of these structures may therefore hamper memory updating.

Methods

Participants

We tested 49 healthy, right-handed adults in a 3-day study design, including an MRI scanning session and behavioral testing. One participant was excluded because of minimal performance in the task (more than 2 standard deviations below average performance in trial types where information did not change), thus leaving 48 participants (25 women; mean age 24.58 years, range: 19–32 years) for analyses. An a priori sample size calculation using G*POWER 3.19.2. showed that this sample size is sufficient to detect small to medium effects with a power of 0.95. Exclusion criteria were checked in a standardized telephone screening and comprised any current physical illnesses or medication intake, a lifetime history of any neurological or mental disorders, as well as any contraindications for MRI measurements, such as non-removable metal parts, pacemaker, pregnancy, or claustrophobia. All participants provided written informed consent before the beginning of testing and received a moderate monetary compensation. The study protocol was approved by the local ethics committee.

Stimulus Materials

In the memory-updating task, we used 100 pictures of faces and 400 well-known (non-German) city names. Pictures of faces were taken from the Radboud Faces Database (RaFD, Langner et al. 2010) and the Karolinska Directed Emotional Faces Database (KDEF, Lundqvist et al. 1998). Only Caucasian faces (50 men and 50 women) with a neutral expression were included. Pictures were resized (762 × 562 pixels) and formatted using Adobe Photoshop CS6 (64 bit) so that all faces had a white background. City names were checked for ease of spelling and possible similarity in their name with other city names. If a city was considered too difficult to spell or too similar to another city used, it was replaced by another city. The chosen city names were rated in an independent pilot study ($n = 15$) as to whether the city name was commonly known or not. The rating sale ranged from 0 ("never heard of") to 10 ("very well known"). Cities used in the task had an average familiarity rating of 6.1 and can thus be considered as well known.

Experimental Design and Procedure

To investigate the updating of already established memories, participants were tested in a novel memory-updating paradigm over the course of 3 consecutive days.

On experimental day 1, after arriving at the lab, participants were welcomed by the experimenter and informed about the general procedure, as well as the MRI procedure on the next day. Informed consent was obtained and any questions the participants had were answered. Participants then completed the "learning phase" (Fig. 1A) of the memory-updating paradigm, which took up to 100 min. They were informed that they would see a face of a person and next to it the name of this person's last vacation destination. The task of the participants was to memorize all face–city pairs they were presented with, 75 in total. Each pair was presented 4 times, each time for 3.5 s. In the first encoding run, participants were presented with 15 blocks of 5 face–city pairs and performed a cued recall test after each block for the 5 pairs shown. During the cued recall, they saw the face and had to type in the name of the city. Participants had to start typing in the word within the first 5 s of its presentation, after which they had enough time to finish typing the word. After their response, the correct city name was presented in green (until termination by the participant) above the city name typed in by the participant. In the second encoding run, participants saw 5 blocks of 15 face–city pairs and in the third run 3 blocks of 25 face–city pairs. In both runs, again each block was followed by a cued recall test for the pairs presented. Finally, in the fourth encoding run, all 75 pairs were presented one after the other, which was followed by a cued recall test for all face–city pairs. This encoding procedure

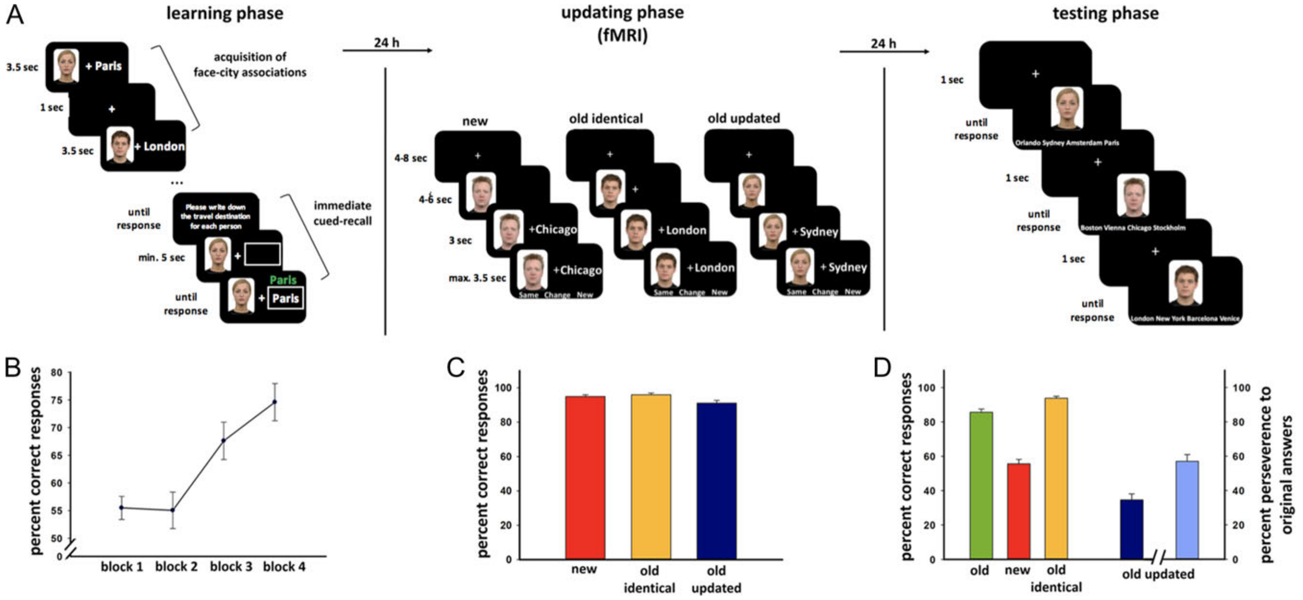


Figure 1. Memory updating paradigm and behavioral performance. (*A*) Memory updating paradigm. Participants performed the memory updating paradigm over the course of 3 consecutive days. During the learning phase on day 1, participants observed 75 face–city pairs, each presented 4 times and completed an immediate cued recall test followed by feedback showing the correct answer. On day 2, about 24 h after learning, participants completed the updating phase in the MRI scanner. In this phase, participants were presented with trials containing the same face–city pairs as on day 1 ("old identical" trials), entirely "new" trials (both face and city had not been shown before) and trials in which a known face was paired with a new city ("old updated" trials). Participants were asked to explicitly indicate whether the pair was the same as on day 1, whether the city was changed, or whether the pair was completely new. During the testing phase on day 3, participants completed a recognition task, comprising trials of face–city pairs that were only shown during the learning phase ("old" trials), "old identical" pairs, the "new" pairs learned on day 2 as well as trials for which the city was updated on day 2. Participants were presented with a face and had to select the correct city out of 4 alternatives. In the "old updated" trials, these 4 alternatives included both the city learned on day 1 and the updated city learned on day 2. Participants were explicitly instructed that if the face–city pair was updated on day 2, the city shown during the updating phase was the correct one. (*B*) Learning phase. Memory performance showed a significant increase across the learning blocks on day 1, reaching a final performance of 74.60 (SD ± 23.32) percent. (*C*) Updating phase. In the updating phase (day 2), performance differed significantly between trial types. More specifically, performance in "old updated" pairs was significantly lower (91.17 (SD ± 10.47) percent) than in "old identical" pairs (96.00 (SD ± 7.05) percent) as well as in "new" pairs (94.92 (SD ± 7.46) percent) while performance scores in "old identical" and "new" trials did not differ. However, despite the significant difference in performance, it is to note that participants were able to correctly identify more than 91% of trials as updated on day 2. (*D*) Testing phase. In the testing phase (day 3), performance between trial types differed significantly. Most importantly, however, participants named the updated city in about 34.7 (SD ± 24.21) percent of trials, whereas in 57.04 (SD ± 27.00) percent of the trials, the city associated with the face chosen on day 1. Error bars represent standard error of the mean. Error bars indicate standard error of the mean.

should help the participants to learn the 75 face–city pairs well. For each participant, the face–city associations as well as the cities that were updated on day 2 were predetermined in a face–city map. This predetermination ensured that cities with different familiarity scores from the pilot study were balanced across trial types. Trial order was also computed in advance for all phases for each participant and semi-randomized to avoid showing a trial type more than 3 times in a row in the later phases of the task. Participants were explicitly instructed to memorize the face–city pairs and they were informed that they could earn extra money if they perform well in the subsequent memory test

On experimental day 2, participants completed the "updating phase" of the paradigm in the fMRI scanner. Participants were instructed that they would see again face–city pairs. They were informed that some of the pairs would be the same as those learned on day 1. For other pairs, however, participants were told that some of the people they saw the faces of the day before made an error, and their last travel destination was actually a different city. Thus, in those pairs, a known face would be paired with an entirely novel city. Participants were asked to memorize this updated face–city association as this updated association would be the correct one (and the one learned on day 1 would be incorrect). Moreover, they were told, that they would also see new pairs in which both the face and the city were novel and they were also instructed to memorize these new pairs. Participants saw first a face (for a duration jittered 4–6 s), then a city appeared next to the face (for 3 s), after which 3 alternatives ("same", "changed", "new", for max 3.5 s) appeared below the face–city pair (see Fig. 1A). Participants were instructed to use a button box to indicate whether the pair was the same as the day before, whether the city changed and the face stayed the same or whether both face and city were new. Twenty-five of the pairs presented in this phase were identical as on day 1 ("old identical trials"), 25 pairs involved a face that had been presented on day 1 now paired with a new city ("old updated trials"), and 25 pairs were completely new ("new trials"). The face–city associations were taken from the face–city map that was computed at the beginning of the learning phase. Participants were explicitly told that they should again memorize the pairings. Once participants had pressed a button, the chosen alternative was highlighted for 0.1 s, which was then followed by a fixation cross shown for 6–8 s before a new trial started.

On experimental day 3, participants completed the "testing phase" (Fig. 1A). Participants saw all the faces they had seen on the previous 2 days, one after another in randomized order. Importantly, together with the face, they were shown 4 city names and were required to select the city that was correctly associated with the face. They were instructed that if the city was changed for a face on day 2, they should select the city that was associated with the face on day 2. The face together with the alternatives were shown until participants made a choice (no time limit). Once participants made a selection, a fixation cross appeared (for 1 s) to signal the beginning of a new trial. Participants completed a total of 100 test trials, comprising 25 “old” trials that involved face–city pairs that were only presented during the learning phase on day 1, 25 “old identical” trials (neither face nor city changed) that were presented during the learning phase on day 1 and the updating phase on day 2, 25 “old updated” trials for which the face was associated with a different city name during the updating phase on day 2 than during learning on day 1, and 25 “new” trials that involved the faces and cities that appeared for the first time during the updating phase. Trial presentation was again semi-randomized as described above and the face–city map was used. Faces in “old updated” trials were presented with the city name from the learning phase, the 1 from the updating phase, and 2 other completely new city names. All other faces were presented with the city they had been associated with on days 1 and 2, respectively, together with 3 other cities that had never appeared before in the task.

Statistical Analysis

Performance (measured as percentage of correct trials) in the learning, updating, and testing phases was subjected to repeated measure ANOVAs, using block (learning phase) or trial type (“old identical,” “old updated,” “new”—in the updating phase; “old”, “old identical”, “old updated correctly”, “old updated” perseverance, “new”—in the testing phase) as within subject factors.

Adjustment of recognition data (day 3). When an updated city was not correctly identified as updated on day 2, we assumed that participants would not be able to update their memory for these trials. We therefore removed these trials from the analysis of the day 3 recognition data and calculated the percent correctly updated over the reduced number of trials. This was also done for “old identical” and “new” trials. On average, we removed one trial from the “old identical” trial types, 2 trials from the “old updated” trial types and one trial from the “new” trial types.

Behavioral data were analyzed using SPSS (IBM 22). Significant main or interaction effects were pursued by appropriate post hoc tests if indicated. In the case of violations of sphericity, Greenhouse–Geisser correction was applied. All reported *P*-values are 2-tailed. In the case of multiple comparisons, we performed a Bonferroni correction where appropriate.

MRI Data Acquisition

MRI data were collected on a 3 T Siemens Skyra MRI Scanner, using a 32-channel head coil. A magnetic (B0) fieldmap for later unwarping of the functional images was recorded with a TR of 421 ms, a TE of 4.92 ms, and a voxel size of 3 × 3 × 3 mm. For the functional images, an echoplanar imaging sequence (476 volumes) was recorded with 3 × 3 × 3 mm voxel size (36 slices), a TR of 2.5 s and TE of 30 ms, and a flipangle of 90°. The slices were tilted 30° from the AC/PC line in order to minimize dropout artifacts in medial temporal and orbitofrontal regions. A T1 structural image was acquired with a voxel size of 0.8 × 0.8 × 0.9 mm, a TR of 2.5 s, and a TE of 2.12 ms, with 256 slices.

MRI Data Preprocessing

The functional MRI data were preprocessed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The first 3 images were discarded to ensure T1 equilibration. We acquired a magnetic (B0) fieldmap to use the realign and unwarp function in SPM12. Using the FieldMap Toolbox in SPM12, a voxel displacement map was created as required for the unwarp function, which utilizes a combined static and dynamic distortion correction. The realignment function then realigns the acquired time series for each subject using a least squares approach and a 6-parameter (rigid body) spatial transformation. All scans are realigned to the first functional image that is used as a reference. Images were the coregistered to the structural image by using a rigid body model based on the work by Collignon et al. (1995). Data were then spatially normalized to fit the MNI space. To estimate the deformation, template data are deformed to match an individual scan. In the last step, data were spatially smoothed with an 8-mm full-width half-maximum Gaussian kernel.

Univariate fMRI Analyses

Data were analyzed using general linear modeling (GLM) as implemented in SPM12. Three separate regressors for each trial type (“old identical,” “old updated,” and “new”) were modeled using the duration in seconds of the individual events. A high-pass filter of 128 s was used to eliminate low-frequency drifts and serial correlations in the time series were accounted for using an autoregressive AR(1) model. For the second-level models, contrast files for the contrasts of interest (“old updated” > “old identical”; “old identical” > “old updated”; “old updated” > “new”) were tested using one-sample *t*-tests. Note that due to the large interindividual variation in updating success, an analysis with separate regressors for successfully versus not successfully updated items was not feasible as this would have resulted in a rather high number of participants for which one of the regressors had a low number of events and therefore had to be excluded from this analysis, resulting in a significant reduction of statistical power. We then performed region of interest (ROI) analyses that focused on brain areas that have been implicated in episodic long-term memory, the representation of prior knowledge, as this is critical during schema-based learning and memory, or cognitive control (Eichenbaum 1999; van Veen and Carter 2002; van Kesteren et al. 2012; Wagner et al. 2015; Gilboa and Marlatte 2017). From the Harvard–Oxford Atlas, we selected masks for the hippocampus, the anterior cingulate gyrus, medial frontal cortex as well as angular gyrus with a probability threshold of 50% as well as the left and right vmPFC and dlPFC masks, created using MARINA software (http://www.bion.de/eng/MARINA.php). Subsequently, we applied a small volume correction (svc) for the areas of interest. The svc was applied on voxel level. Voxels were regarded as significant, when falling below a corrected voxel threshold of 0.05 (family wise error (FWE) corrected) adjusted for the small volume. Only clusters within an ROI comprising $k \geq 5$ significant voxels are reported.

Furthermore, a second-level model was run that included the adjusted memory updating performance (i.e., percentage correct updated adjusted for false choices on day 2) that was measured on day 3 as a covariate. This way it was possible to account for the interindividual variability in updating performance. We then performed ROI analysis, with the above described ROIs and applied small volume correction, to obtain significantly activated clusters of voxels in our ROIs.

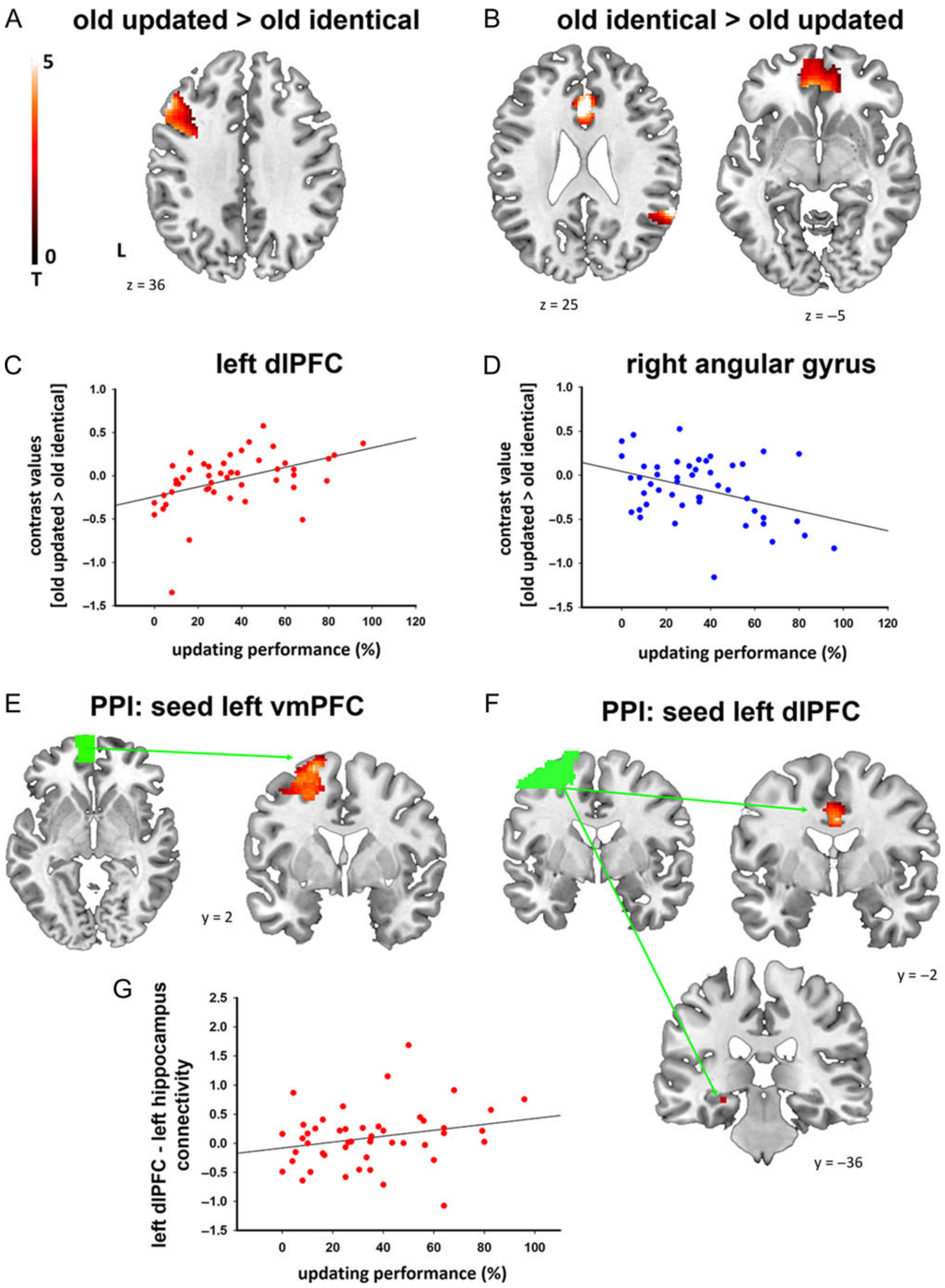


Figure 2. Neural underpinnings of memory updating. (A) The comparison of "old updated" and "old identical" face–city pairs during the updating phase revealed significant activation in the left dlPFC. (B) Activity in the anterior cingulate gyrus, bilateral ventromedial prefrontal cortex (vmPFC), and right angular gyrus was higher for "old identical" than "old updated" pairs. (C) Activity in the left dlPFC was positively correlated ($r = 0.434$) with updating success as assessed on day 3. For visualization purposes, we created activation masks using SPM Imcalc and the anatomical mask of the specific ROI. These masks were then converted to binary files compatible with the Marsbar toolbox (http://marsbar.sourceforge.net/). Marsbar was then used to extract the contrast values for each ROI that were above zero. These contrast values were then correlated with the adjusted memory updating scores (percent correctly updated in the testing phase adjusted for performance in the updating phase). (D) Updating performance was negatively correlated with updating success ($r = -0.386$) in the right angular gyrus. (E) A functional connectivity analysis revealed increased connectivity between the left vmPFC (seed region, indicated in green) and the left dlPFC for updated (vs. identical) face–city pairs. (F) When we used the left dlPFC as a seed region (green), we obtained increased functional connectivity for "old updated" versus "old identical" trials in the hippocampal dentate gyrus and the anterior cingulate gyrus. (G) Left dlPFC–left hippocampus connectivity was directly linked to memory updating success. For visualization purposes, we performed a correlation with the contrast values and the adjusted updating performance. Results indicate a positive correlation of the contrast values and updating performance ($r = 0.249$).

In addition, we performed a generalized form of context-dependent psychophysiological interaction (gPPI, https://www.nitrc.org/projects/gppi) to assess task-dependent connectivity with those ROIs that were most relevant in the previous analyses as seed regions. gPPI has the advantage over standard PPI, as implemented in SPM12, that it allows the inclusion of more than 2 task conditions (McLaren et al. 2012). We used the previously described regressors ("old identical," "old updated," and "new") plus a PPI

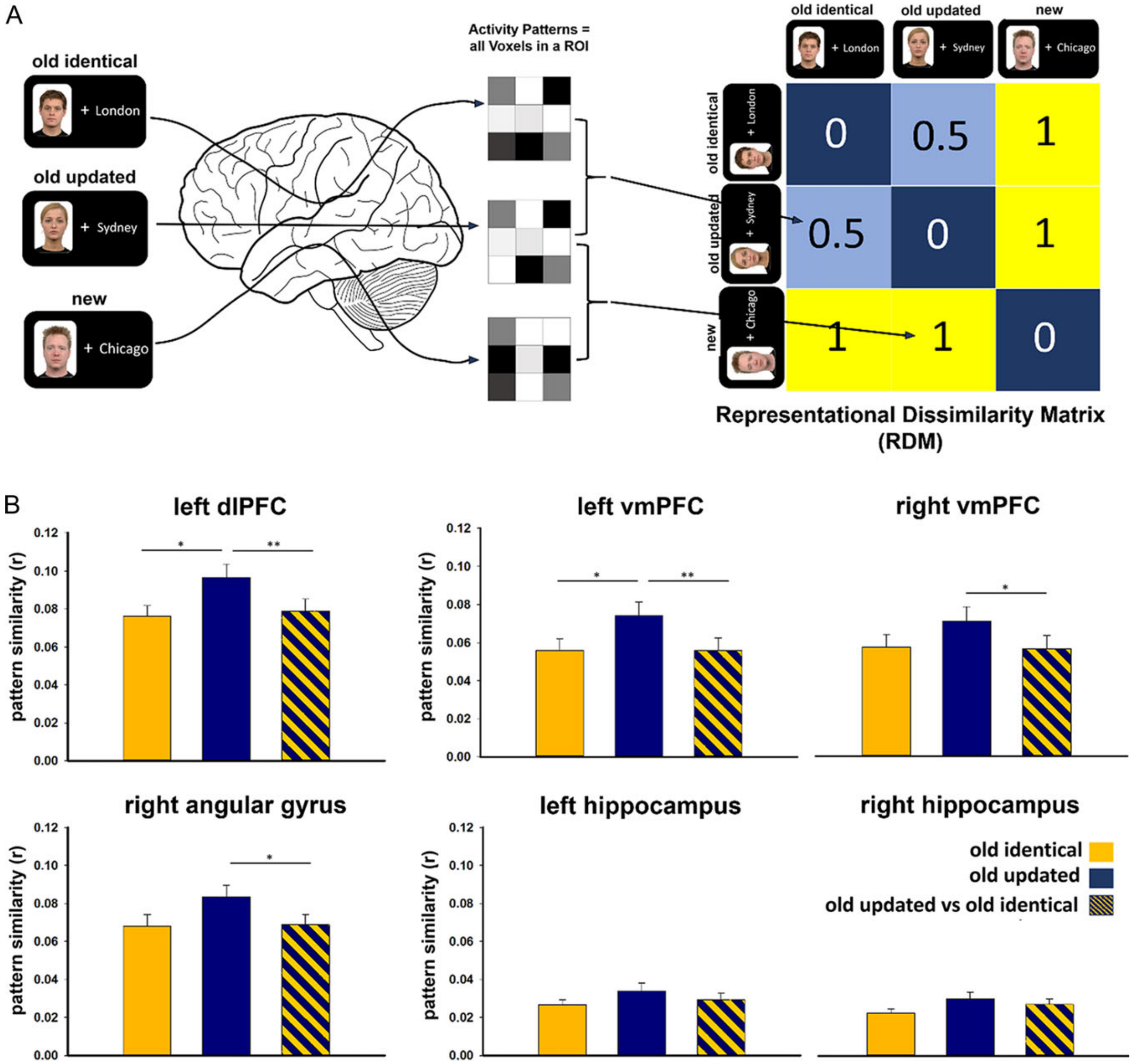
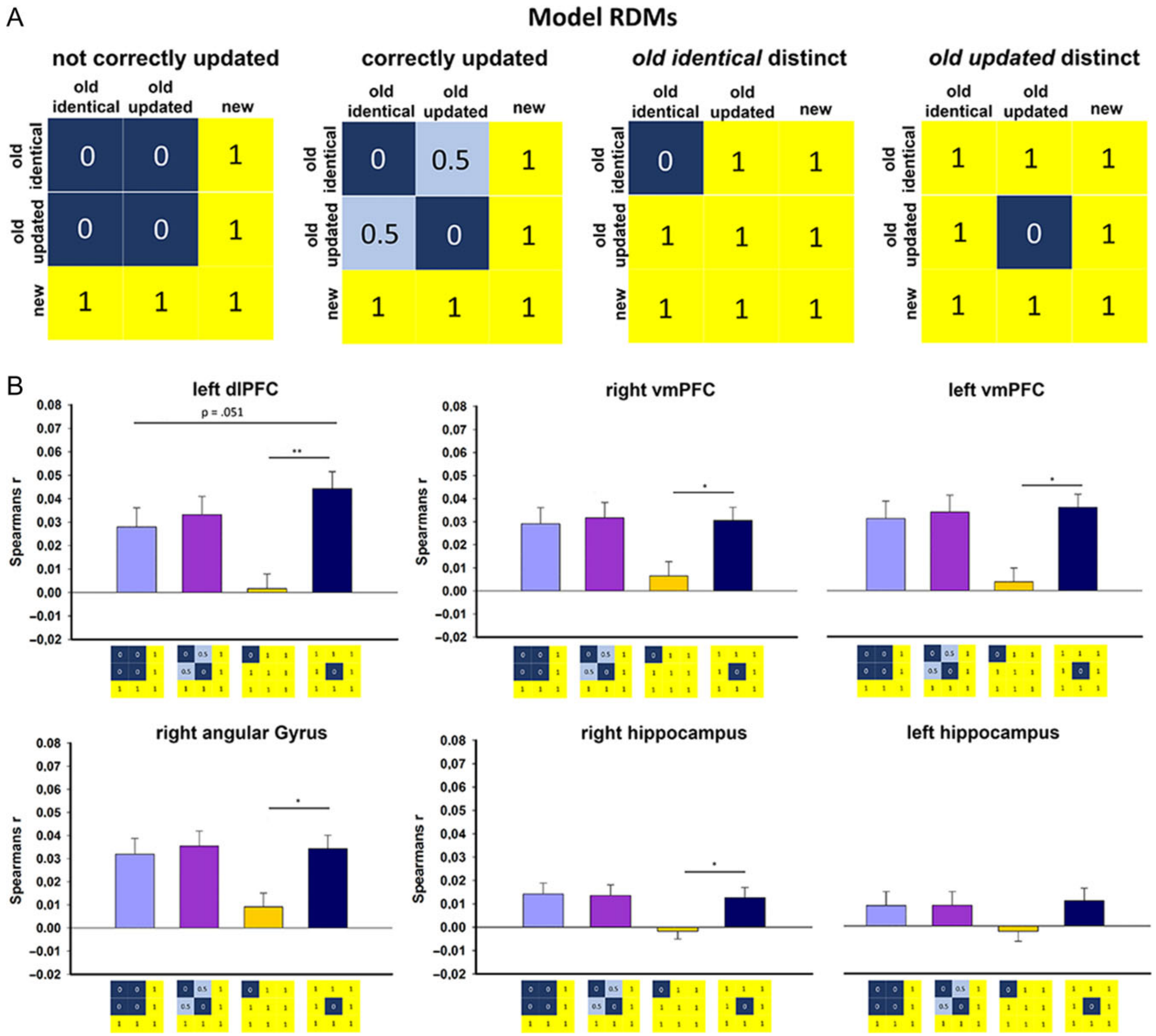


Figure 3. Neural pattern similarity within and between ROIs during memory updating. (A) Representational dissimilarity matrices (RDMs) used in the in the following analyses. The figure provides a schematic overview of the creation of an RDM for tree trial types (adapted from Nili et al.'s 2014). Quadrants representing the individual trial types and their hypothesized similarity with other trial types are indicated. (B) Pattern similarity in the left and right vmPFC indicated a significant difference between the quadrant "old updated" and the quadrant "old identical–old updated." In the left vmPFC the pattern similarity within the "old updated" quadrant was significantly higher compared with the "old identical" trials. In the right angular gyrus, pattern similarity in the "old updated" quadrant was again higher than in the "old updated–old identical" quadrant. Pattern similarity in the left dlPFC differed significantly between the "old updated" and "old identical" quadrants as well as the "old updated" and "old updated–old identical" quadrant. Figure shows results as Pearson correlations, to aid interpretability, but statistical tests show results performed after Fisher transformation was applied. Error bars represent standard error of the mean. *$P < 0.05$, **$P < 0.001$.

Interaction term for each of these regressors, plus the time course from the respective seed region in our first-level PPI model. For the second-level models, contrast files from the contrast PPI "old updated" > PPI "old identical," the contrast PPI "old identical" > PPI "old updated," the contrasts PPI "old identical" > PPI "old updated," PPI "old updated" > PPI "new," and PPI "new" > PPI "old identical" were tested, using a one sample *t*-test. Subsequently, we applied an svc for the other ROIs, to determine a difference in connectivity between respective regions depending on the trial types.

In order to analyze associations between the functional connectivity of ROIs and behavioral performance, we ran a second-level gPPI model with the contrasts PPI "old updated" > PPI "old identical" and PPI "old identical" > PPI "old updated," and the adjusted updating performance (i.e., percent correctly updated adjusted for performance on day 2) as covariate, and subsequently an svc that allowed us to obtain significant clusters of voxels within an ROI. For visualization purposes and only for these ROIs, we used SPM Imcalc and MarsBar to create binary activation maps including all voxels within an ROI with values above zero for the second-level contrast. For each of these activation maps, we then extracted an average time series for each participant using MarsBar and estimated the model on these,



not correctly updated

	old identical	old updated	new
old identical	0	0	1
old updated	0	0	1
new	1	1	1

correctly updated

	old identical	old updated	new
old identical	0	0.5	1
old updated	0.5	0	1
new	1	1	1

old identical distinct

	old identical	old updated	new
old identical	0	1	1
old updated	1	1	1
new	1	1	1

old updated distinct

	old identical	old updated	new
old identical	1	1	1
old updated	1	0	1
new	1	1	1

Figure 4. Distinct representation of updated items in the dlPFC. (A) Model RDMs. We performed a model fit analysis, comparing 4 model RDMs (not correctly updated, correctly updated, “old identical” distinct, “old updated” distinct) with the representational patterns observed in the individual subject RDMs. The model “correctly updated” represents the expected similarities for the “old identical” and “old updated” trials, suggesting a high similarity measure in the “old identical–old identical” and “old updated–old updated” quadrants (indicated by a blue color), and less similarity in the “old identical–old updated” quadrants (light blue color). The model “not correctly updated” does not distinguish between the “old identical” and “old updated” trials, that is, the highest similarity in the quadrants of “old identical,” “old updated,” and “old identical–old updated.” The models “old identical distinct” and “old updated distinct” assume discrete activity patterns for the “old identical” and “old updated” trials. Model fit over all participants was assessed in 4 ROIs, the left and right vmPFC, left and right hippocampus, and the right angular gyrus and left dlPFC. The highest value of the Spearmans r indicates the best model fit. (B) The “old updated distinct” model shows a good fit in the left dlPFC, while this was less pronounced in the remaining ROIs. (B) The right angular gyrus and vmPFC showed similar fits for all models, while in the hippocampus the observed fits for all models were low (all Spearmans r < 0.015). Error bars indicate standard error of the mean. *P < 0.05, **P < 0.001, n.s. = nonsignificant.

resulting in one contrast value for each participant per ROI. The values from the first-level contrast “PPI old updated > PPI old identical” were correlated with the percentage correctly updated (i.e., the adjusted memory updating scores) from the testing phase.

Representational Similarity Analysis

In addition to the univariate analysis, we performed an RSA (Kriegeskorte et al. 2008; Nili et al. 2014) in those ROIs that turned out to be most relevant in the univariate analyses, that is, the left and right vmPFC, left and right hippocampus, right angular gyrus, and left dlPFC. For each of these ROIs and each subject, we computed representational dissimilarity matrices (RDMs) that were based on a single-trial univariate GLM (one regressor for each trial) that was estimated on native space functional images. We have utilized the data from the updating phase (day 2, in the MRI Scanner) when participants were presented with the cities, that were either new, known or updated. Due to technical difficulties, certain participants did not have all 25 trials for each trial type (“old identical”, “old updated”, and “new”). We therefore removed the last 2 trials of each trial type to ensure that each subject RDM comprised 23 trials for each trial type, that is, 69 trials in total. We utilized the spm T-files for the single trial regressors, to create vectors of activity patterns for each trial, but separately for each ROI. The activity patterns were then utilized to compute the dissimilarity between 2 trials by correlation distances (1–r, Pearson linear correlation). After that, the dissimilarities based on each combination of trials were positioned into the corresponding cells of the 69 × 69 RDMs (see Fig. 3A).

Comparison of Pattern Similarities Between Trial Types

We extracted the mean pattern similarity (r, Pearson Correlation) out of specific RDM quadrants (within trial type: "old identical–old identical," "old updated–old updated," and between trial type: "old identical–old updated") from each single-subject RDM (Wolosin et al. 2013; Ritchey et al. 2015; Aly and Turk-Browne 2016). We then z-transformed the extracted mean pattern similarity values and compared 1) the within-trial-type similarity and 2) the within-trial-type similarity with the between-trial-type similarity; paired t-tests were performed using SPSS 22.

Comparison with Model RDMs

In addition, we compared the RDMs obtained to 4 model RDMs (Fig. 4A) that were constructed based on the expected similarities of the different trial types. The model "correctly updated" expects a high similarity between "old identical" and "old identical" trials and between "old updated" and "old updated" trials but less similarity between "old identical" and "old updated" trials, while all trials have the least similarity with "new" trials. The model "not correctly updated" does not distinguish between the "old identical" and "old updated" trials, hence it expects the highest similarity in the quadrants of "old identical," "old updated," and "old identical–old updated." The models "old identical distinct" and "old updated distinct" assume distinct activity patterns for the "old identical" and "old updated" trials, respectively, with a high similarity for trials within the respective quadrants. We used Spearman's rank correlation coefficient to obtain the correlation between the brain RDMs of the ROIs with the respective model RDMs, thus the pattern similarity values of the brain RDMs were rank transformed before calculating the correlation between the respective matrices (Nili et al. 2014). We then used 2 measures to test for the relatedness of each of the 4 model RDMs to the respective brain RDMs across subjects within each of the ROIs: 1) we first used the default one-tailed Wilcoxon signed-rank subject random-effects tests. Thus, first computing the Spearman's rank correlations of the respective brain RDM to the model RDM for each subject and then performing a Wilcoxon signed-rank test against the null hypothesis of a correlation of zero across all subjects (one-sided to test for a positive correlation only). 2) As an alternative, we further reanalyzed the data using the rather conservative stimulus-label randomization test. Here, the relatedness of the models and the brain RDMs are tested by randomizing the condition labels of the brain RDM and then calculating the Spearman's rank correlations of this randomized matrix to the model RDM. This randomization process is repeated 10 000 times to obtain a distribution of correlations simulating the null hypothesis that the brain RDMs and model RDMs are unrelated. Next one tests if the actual correlation of the (not randomized) brain RDM falls within the top 5% of the simulated distribution (Kriegeskorte et al. 2008).

In the next step, we compared the model fits of the 4 models to determine which of the models fits best in the respective ROIs, and if this differs across ROIs. For this, we extracted the Spearman's rank correlation coefficients for each model per subject in each ROI and then performed repeated measures ANOVA and post hoc t-tests using SPSS 22. The ANOVA contained the within subject factors ROI and model.

Moreover, we distinguished between successful and poor updaters based on whether the individual updating performance was above or below the median of all participants (median: 32% correctly updated). We again utilized a repeated measures ANOVA with the within subject factors ROI and model and the between subject factor Group (successful vs. poor updaters). Post hoc t-tests were performed when adequate using SPSS 22.

Results

Successful Learning and Memory Updating

During the learning phase on experimental day 1, participants (n = 48) learned a total of 75 face–city pairs. Each pair was presented 4 times and tested in 4 immediate cued recall tests (Fig. 1A). Memory performance increased significantly over the course of these 4 learning blocks ($F(1.718, 80.729) = 73.623$, $P < 0.001$, $\eta^2 = 0.610$), with a final cued recall performance of 74.60 (SD ± 23.32) percent (Fig. 1B), indicating that participants learned the face–city pairs well. There was no difference in performance between items that were subsequently updated or not (i.e., items used in later "old," "old identical," and "old updated" trials; $F(2,94) = 1.435$, $P = 0.243$, $\eta^2 = 0.030$).

On experimental day 2, about 24 h after initial learning, participants completed the memory updating phase in the MRI scanner (Fig. 1A). During this updating phase, participants saw 25 pairs that were identical to those presented during the learning phase, controlling for retrieval processes ("old identical"), 25 completely "new" pairs, that controlled for new learning (i.e., new face and new city name), and 25 pairs that included a known face now paired with a new city name ("old updated"), requiring participants to update the previously learned face–city associations. Participants were explicitly asked to memorize all pairs they were shown and, for "old updated" pairs, to retain the updated city as the correct one. Furthermore, participants were requested to indicate by button press whether the city paired with a particular face had changed from day 1 to day 2, whether it stayed the same, or whether the face–city pair was completely new (Fig. 1A). Participants were very well able to distinguish between "old identical," "old updated," and entirely "new" face–city pairs with an average performance of 94.03 (SD ± 6.03) percent correctly identified trials. Performance differed between the trial types though ($F(2, 94) = 5.838$, $P = 0.004$, $\eta^2 = 0.110$): performance for "old updated" pairs (91.17 (SD ± 10.47) percent correct, indicating that in over 91% of the trials participants were able to correctly identify a city as updated), was slightly lower than for "old identical" (96.00 (SD ± 7.05) percent correct, $t(47) = -3.705$, $P = 0.001$), and "new" pairs (94.92 (SD ± 7.46) percent correct, $t(47) = -2.186$, $P = 0.034$), while performance in "old identical" and "new" trials was comparable ($t(47) = 0.772$, $P = 0.444$; Fig. 1C).

Updating success was then tested on day 3 (testing phase), again about 24 h later, in a recognition test that included "old" pairs that were learned on day 1 but not shown on day 2, as well as "old identical," "old updated," and "new" pairs from day 2 (Fig. 1A). On each trial, participants were asked to choose the correct city name associated with a given face from 4 alternatives. For "old updated" trials, these alternatives included both the original and the updated city name. Performance differed significantly between trial types ($F(1.314, 61.735) = 68.422$, $P < 0.001$, $\eta^2 = 0.593$). Performance was highest in "old identical" pairs (94% correct) compared with both "old" (i.e., pairs that were only shown during the learning phase, 86% correct) and "new" pairs (about 56% correct, all $P < 0.001$). The lower performance in new pairs was expected given that those "new" pairs were presented only once on day 2. Most importantly, for the face–city pairs that were updated on day 2, the updated city name was correctly chosen in 34.7 (SD ± 24.21) percent of the trials, while in 57.04 (SD ± 27.00) percent of the "old updated" trials participants incorrectly persevered with the city name that was paired with the respective face during the learning phase ($t(47) = -3.055$, $P = 0.04$, Fig. 1*D*). Notably, updating performance or perseverance was not dependent on initial learning performance (comparison of day 1 memory for subsequently correctly updated vs. not correctly updated memories: $t(47) = -0.869$, $P = 0.389$). Perseverance with the old city names may well be explained by the fact that participants saw the updated items only once during the updating phase, while the old items were presented 4 times during the learning phase, each time followed by a cued recall test, which is known to boost subsequent memory (Karpicke and Roediger 2008). City names that were not presented on either day 1 or day 2 were endorsed only in about 4% of the trials each, that is, significantly less often than the updated city names from day 2. The finding that participants correctly chose, despite the differential strength of memory for the original and the updated information, the updated city name in more than a third of the trials clearly demonstrates participants' capability for updating established memory traces, which raises the question how the brain implements this memory updating capacity.

Neural Signature of Successful Memory Updating

To determine the neural mechanisms that facilitate successful memory updating, we assessed brain activity during the updating phase using fMRI. As a first step, we investigated the differences in brain activation during the presentation of a face–city pair that was updated ("old updated") compared with when face–city pairs were presented that did not require updating ("old identical"). The left dlPFC showed significantly increased activity when "old updated" face–city pairs were shown (peak coordinate: $xyz = -50, 26, 34$; $P_{svc} = 0.0026$; $t = 5.398$, $k = 32$; Fig. 2*A*), suggesting a crucial involvement of this region in the memory updating process. Moreover, the left and right dlPFC showed also significantly increased activity when comparing the activity during the presentation of "old updated" face–city pairs to the activity during the presentation of completely new face–city pairs (left peak coordinate: $xyz = -46, 22, 42$; $P_{svc} = 0.004365$; $k = 127$; right peak coordinate: $xyz = 46, 30, 36$; $P_{svc} = 0.01665$; $k = 38$), indicating that the activation of the (left) dlPFC in the contrast "old updated" versus "old identical" does not simply reflect the encoding of new items but is indeed linked to the updating of established associations. In contrast to the dlPFC, the vmPFC (left peak coordinate: $xyz = -8, 54, -2$; $P_{svc} = 0.00446$; $t = 4.6643$; $k = 16$; right peak coordinate: $xyz = 2, 44, -2$; $P_{svc} = 0.00113$; $t = 5.1508$, $k = 29$), the angular gyrus (peak coordinate: $x = 62$, $y = -48$, $z = 20$; $P_{svc} = 0.00002$; $t = 6.1211$, $k = 139$), and the anterior cingulate cortex (peak coordinate: $xyz = 0, 28, 22$; $P_{svc} = 0.00002$; $t = 6.5494$; $k = 930$), areas implicated in the representation of prior knowledge as well as selective attention and the recall of remote memories (Lenartowicz and McIntosh 2005; Weible 2013; Wagner et al. 2015; Gilboa and Marlatte 2017) were significantly less activated when "old updated" face–city pairs were presented, compared with when "old identical" items were shown (Fig. 2*B*).

In order to assess whether altered activity in those areas was directly related to the actual updating success, we correlated the neural activity during the presentation of "old updated" (vs. "old identical") face–city pairs with the percentage of correct choices of the updated city name on experimental day 3 (individual updating success) as a behavioral measure of successful memory updating. Activity in the left dlPFC was positively correlated with updating performance (peak coordinate: $xyz = -48, 12, 36$; $P_{svc} = 0.00813$; $t = 5.0338$, $k = 12$; Fig. 2*C*), indicating that this area was crucially involved in successful memory updating. In contrast, activity in the right angular gyrus was negatively correlated with updating success (peak coordinate: $xyz = 54, -48, 32$; $P_{svc} = 0.02026$; $t = 3.7943$, $k = 17$, Fig. 2*D*), suggesting that the recruitment of this area impedes the updating of established associations. Notably, parameter estimates for the contrast "old updated" versus "old identical" for the dlPFC and angular gyrus were not correlated ($r = 0.053$, $P = 0.719$), thus ruling out the possibility that the opposite correlation between updating success and activity in the left dlPFC and angular gyrus was simply due to a negative correlation between left dlPFC and angular gyrus activity. Additional activations observed in an exploratory whole-brain analyses are shown in Supplementary tables S1 and S2.

Crosstalk Between Left dlPFC and Hippocampus Supports Successful Memory Updating

In order to identify the networks that support memory updating, we next investigated which regions showed increased functional connectivity to those areas that were linked to successful memory updating. A gPPI analysis (McLaren et al. 2012) revealed that the attempt to update previously learned face–city pairs was associated with an increase in connectivity among specific prefrontal areas (i.e., the left dlFC, ACC, and left vmPFC; Fig. 2*E*,*F*). More precisely, taking the left vmPFC as a seed region, there was increased connectivity with the left dlPFC (peak coordinate: $xyz = -36, 36, 34$; $P_{svc} = 0.00587$; $t = 5.14074$, $k = 24$, Fig. 2*E*) for "old updated" (vs. "old identical") face–city pairs. When we used the left dlPFC as a seed region, we obtained increased functional connectivity for "old updated" vs. "old identical" trials with the anterior cingulate gyrus (peak coordinate: $xyz = 2, -2, 34$; $P_{svc} = 0.00354$; $t = 4.83124$, $k = 34$, Fig. 2*F*). Even more interestingly, however, when established associations were updated (vs. "old identical"), there was a significant increase in connectivity between the left dlPFC and the left hippocampus (peak coordinate in the dentate gyrus: $xyz = -26, -36, -2$; $P_{svc} = 0.02326$; $t = 3.19508$, $k = 6$; Fig. 2*F*), a critical hub in memory formation (Schacter et al. 1996; Eichenbaum 1999). When we correlated the connectivity between these areas with the individual updating success as assessed on experimental day 3, we found that specifically the crosstalk between the left dlPFC and the left hippocampus was associated with successful memory updating (peak coordinate: $xyz = -30, -26, -10$; $P_{svc} = 0.03967$; $t = 3.5807$, $k = 5$, Fig. 2*G*). To further investigate whether the dlPFC–hippocampus connectivity reflects rather new encoding or updating processes in particular, we ran an additional analysis investigating the contrast of PPI "old updated" versus PPI "new" face–city pairs. Results after a small volume correction did not reveal any difference for the "old updated" trials compared with the "new" trials in functional connectivity between the dlPFC and any of our other ROIs. This might suggest that the dlPFC–hippocampus connectivity reflects mainly the encoding of new information, as a part of the updating process. However, this should then also be reflected in the contrast PPI "new" > PPI "old identical." However, our results showed no significant connectivity between the dlPFC and any other ROIs in this contrast. Thus, these results suggest a very specific process that is required when one is presented with the updated information (vs. "old identical"). More specifically, the familiar face in "old updated" trials may recruit a specific retrieval process, in combination with the detection of new information and, possibly, an attempt to integrate old and new information. For additional exploratory whole brain connectivity analyses, please see Supplementary Tables S3 and S4.

Neural Representations of Memory Updating

Our univariate fMRI data indicate a key role for the left dlPFC—and its functional connectivity with the hippocampus—for successful memory updating. If the left dlPFC indeed initiates specific processes that are crucial for the updating of established memories, then these processes should be paralleled by similar neural activity patterns in the left dlPFC specifically for updated information. To test this prediction, we examined in a next step the neural activity patterns of trials in which information was attempted to be updated and trials on which no updating was required. Specifically, we used multivariate RSA (Kriegeskorte et al. 2008; Nili et al. 2014) and analyzed neural activity pattern similarity for "old identical" and "old updated" (and "new") face–city pairs as well as the overlap between the activity patterns of "old updated" and "old identical" pairs in our ROIs (Wolosin et al. 2013; Ritchey et al. 2015; Aly and Turk-Browne 2016) (Fig. 3A; see Methods for details). In line with a proposed key role of the left dlPFC in memory updating, we observed a higher value for neural similarity for "old updated" face–city pairs in the left dlPFC, compared with both "old identical" pairs ($t(47) = -3.636$, $P = 0.001$) and the overlap between "old updated" and "old identical" pairs ($t(47) = 3.936$, $P < 0.001$, Fig. 3B). Comparable results were observed in the angular gyrus and bilateral vmPFC. More specifically, in the left vmPFC the pattern similarity within the "old updated" quadrant was significantly higher compared with both the "old identical" and "old updated–old identical" quadrants (both $P \leq 0.001$). In the right vmPFC, similarity within the old updated quadrant was significantly higher compared with the "old updated–old identical" similarity ($t(47) = 3.082$, $P = 0.003$), while the comparison between the "old updated" and "old identical" quadrant did not survive correction for multiple comparisons ($P = 0.018$). In the right angular gyrus, pattern similarity in the "old updated" quadrant was again higher than in the "old updated–old identical" quadrant ($t(47) = 2.87$, $P = 0.006$), yet the comparison between the "old identical" and "old updated" quadrants did not survive correction for multiple comparison ($P = 0.023$). This specific pattern in the vmPFC and angular gyrus may reflect the consistently reduced activity in those areas during "old updated" trials, as seen in the univariate analyses. In the hippocampus, neural similarity did not differ between trial types and was generally relatively low (Fig. 3B), in line with the idea that hippocampal activity patterns are fairly item specific, thus allowing pattern separation (Bakker et al. 2008; Yassa and Stark 2011).

Distinct Updating Representations in the Left dlPFC

In order to explicitly test which brain areas represent updated information distinctly from information that did not require updating, we next constructed 4 model RDMs: 1) a "not correctly updated" model, in which the activity patterns for "old updated" and "old identical" face–city pairs are comparable, 2) a "correctly updated" model, which assumes a certain overlap between activity patterns for "old updated" and "old identical" pairs (as a previously learned face is involved in both) but still expects the patterns to be more similar within each group of face–city pairs, 3) an "old identical distinct" model reflecting a clearly distinct activity pattern for the "old identical" face–city pairs but no specific pattern for "old updated" and "new" pairs, and 4) an "old updated distinct" model which assumes a highly distinct representation of "old updated" face–city pairs, as it purely reflects the processes active when information is updated without taking into account a specific overlap or similarity between "old identical" and "old updated" representations (Fig. 4A). To test whether the proposed models are significantly related to the brain RDMs in the respective ROIs we first considered results from the Wilcoxon signed rank test. These tests indicated a significant relatedness between the correctly updated, not correctly updated and "old updated" distinct models and the observed brain RDMs, respectively, within all ROIs (all $P < 0.01$), except for the left hippocampus (all $P > 0.119$). For the model old identical distinct, on the other hand, the results did not indicate any relatedness with the brain RDMs in none of the ROIs (all $P > 0.14$). We additionally performed the fixed effects condition label randomization test. Here, results indicated a non-significant trend (P (uncorrected) = 0.125) for the "old updated" distinct model in the dlPFC, while there were no other significant results in the other ROIs for any of the models. However, since the Wilcoxon signed rank test indicated a significant relatedness of the brain RDMs with most of the proposed models in most ROIs, we next compared the model fits of the 4 models in the ROIs statistically.

As the "old updated" distinct model reflects the unique representation of the updating process, we were primarily interested to specifically compare the fit of the "old updated" distinct model to the remaining models. Figure 4B suggests that the model fits were highest for the "old updated" distinct model in the left dlPFC, while all other regions showed a similar fit for the remaining models with specifically low fit values in the hippocampus (we obtained a strong trend for an ROI × model interaction: $F(5.856, 275.236) = 2.019$, $P = 0.065$, $\eta^2 = 0.041$). In the left dlPFC, post hoc t-tests indicated that the "old updated" distinct model may hold a superior fit compared with the "old identical" distinct model ($t(47) = -3.77$, $P < 0.001$) and a marginally significant better fit compared with the not correctly updated model ($t(47) = -2.003$, $P = 0.051$). The comparison between the "old updated" distinct and correctly updated model was, however, not significant ($t(47) = -1.513$, $P = 0.137$), which is not surprising since both models assume to a certain degree a differential representation of "old updated" and "old identical" trials, while the old updated distinct model purely reflects the processes active during updating of established memories with new information and does not take into account the similarity that may persist when an old face is shown. A different pattern than in the left dlPFC was observed in the angular gyrus, bilateral vmPFC, and right hippocampus. In these regions, the "old updated" distinct model was not significantly different from the "not correctly updated" model (all $P \geq 0.492$) and to a lesser extent than the left dlPFC from the "old identical" distinct model (all $P \leq 0.031$), indicating that there may be no clear distinction between the "old updated" and "not correctly updated" models in these regions (Fig. 5).

Successful Memory Updating Linked to Updating-Specific dlPFC Representation

Finally, we asked whether the above differences in the representational activity patterns between “old updated” versus “old identical” face–city pairs were predictive of subjects' behavioral memory updating performance. Our above findings suggest that the left dlPFC might show especially consistent activity patterns during the memory updating process. We therefore assumed that subjects highly proficient in updating memories were also the ones in which the neural pattern representation in the left dlPFC was most consistent during the updating process and therefore best characterized by the “old updated” distinct model. In contrast, in subjects performing poorly at updating their memories, we expected the angular gyrus to come to the fore, an area implicated, for example, in the representation of prior knowledge (van Kesteren et al. 2012; Wagner et al. 2015). To test this hypothesis, we used a median split to subdivide our participants into “successful updaters” (mean: 54% correct) and “poor updaters” (mean 15% correct), harnessing the large interindividual variance in updating performance (between 0% and 96 % correct). We then compared the RSA model fits for successful and poor updaters in our brain areas of interest, focusing on the left dlPFC and right angular gyrus (please see Fig. 5, for an overview of model fits in the remaining regions, see Supplemental Fig. S1) and again on the fit of the “old updated” distinct model compared with the remaining models. We again considered the Wilcoxon signed rank test in a first analysis and results showed a similar pattern for the poor updaters, similar to the analyses across all participants (with no significant relatedness between the “old identical distinct” model and the brain RDMs in any of the ROIs (all $P >$ 0.67), and a significant relatedness between the remaining 3 models and the brain RDMs in all ROIs (all $P < 0.022$) except in the left hippocampus (all $P > 0.10$)), while in the successful updaters, the pattern was slightly different, indicating a significant relatedness between the brain RDMs and the “old identical” distinct model in the angular gyrus ($P = 0.025$), while this was not significant in any other ROI (all $P > 0.17$). The remaining models were significantly related to the brain RDMs in all ROIs ($P < 0.03$) except the left hippocampus in the correctly updated, not correctly updated and “old updated” distinct models ($P > 0.33$) and the right hippocampus in the “old updated” distinct model ($P = 0.13$). We again performed a fixed effects condition label randomization test. While results from the randomization test for each group separately indicated a trend in the dlPFC for the “old updated” distinct model (P (uncorrected) = 0.0665) for successful updaters, there was no other indication of a significant relatedness of any other model to the data in any other ROI for either group using the randomization test (all $P > 0.12$). However, since the Wilcoxon signed rank test indicated a significant relatedness between the brain RDMs and 3 of the proposed models we again compared the Spearman's rank correlation coefficient of the different models in the different ROIs using a repeated measures ANOVA, although results need to be interpreted with caution. Results indicated a significant interaction between ROI, model, and group (successful vs. poor updaters; $F(1.799, 82.737) = 6.663$, $P = 0.003$, $\eta^2 = 0.127$). In successful updaters, the “old updated” distinct model showed a significantly better fit in the left dlPFC compared with the “old identical” distinct model, indicated by post hoc t-tests ($t(23)$ =−3.131, $P = 0.005$), the not correctly updated model ($t(23)$ = −2.537, $P = 0.018$), and even the correctly updated model ($t(23)$ = −2.112, $P = 0.046$; Figure 5). Thus, successful updaters appear to be characterized by showing an especially consistent activity pattern during the updating process in the left dlPFC. In the angular gyrus, however, there was no better fit of the “old updated” distinct model compared to any other model (all $P \geq$ 0.468). Specifically, in successful updaters, the fit for the “old identical” distinct model was not significantly different from the fit for the “old updated” distinct model ($t(23) = -0.630$, $P = 0.535$). This may be due to a better representation of “old identical” information in the angular gyrus, allowing participants to better recognize information that did not change. In poor updaters, there was no evidence for such a specifically consistent activity pattern during the updating process in the left dlPFC. Specifically, whereas there was a better fit for the “old updated” distinct model compared with the “old identical” distinct model both in the left dlPFC and angular gyrus (both $P \leq 0.041$), none of the other model comparisons were significant (all $P \geq 0.447$). Together, these results suggest that successful updating of existing memories is driven by a consistent and distinct activity pattern during the updating process in the left dlPFC, and that a failure to show this distinct pattern may be linked to an increased inability to update already established memories with new information.

Discussion

The present findings provide novel insights into the neural mechanisms underlying the flexible updating of established memories. In particular, our data indicate a key role of the left dlPFC in this updating process. First, activity in left dlPFC and functional coupling of left dlPFC with the hippocampus were increased when learned associations had to be updated. Importantly, this increased (left) dlPFC activity during updating trials was seen both in comparison to “old identical” trials and “new” trials, showing that this activity was not due to simply retrieving old information or encoding new information but specifically to the updating of existing associations. Both, left dlPFC activity and prefrontal–hippocampal coupling were directly related to behavioral updating performance. Second, results of neural pattern similarity analysis in left dlPFC showed a distinct pattern during the updating process and the distinctiveness of these activity patterns separated individuals that were successful in memory updating from those that were not. Notably, these results hint to the fact that left dlPFC and hippocampus may act in concert to support memory updating. Unlike the left dlPFC, the hippocampus showed no increased activity during the presentation of trials that required updating in our univariate analyses. Similarly, hippocampal pattern similarity was relatively low in general and may lend support to the idea that hippocampal neural patterns are specific to particular associations (Leutgeb et al. 2007; Yassa and Stark 2011). Finally, activity in areas supporting prior knowledge, such as the angular gyrus, appeared to interfere with successful memory updating.

The dlPFC has been classically related to cognitive control and working memory (Curtis and D'Esposito 2003; Egner and Hirsch 2005) as well as attentional processes, novelty detection and cognitive load (Yamaguchi and Knight 1991; Johnson et al. 2007; Szczepanski and Knight 2014). All of these processes are highly relevant in the context of memory updating, which is a complex process that builds heavily on attention as well as cognitive control capacities and requires the individual to hold both the old and the new information in mind, thus posing a cognitive load. Without these basic cognitive processes memory updating would be unthinkable. Beyond these classical

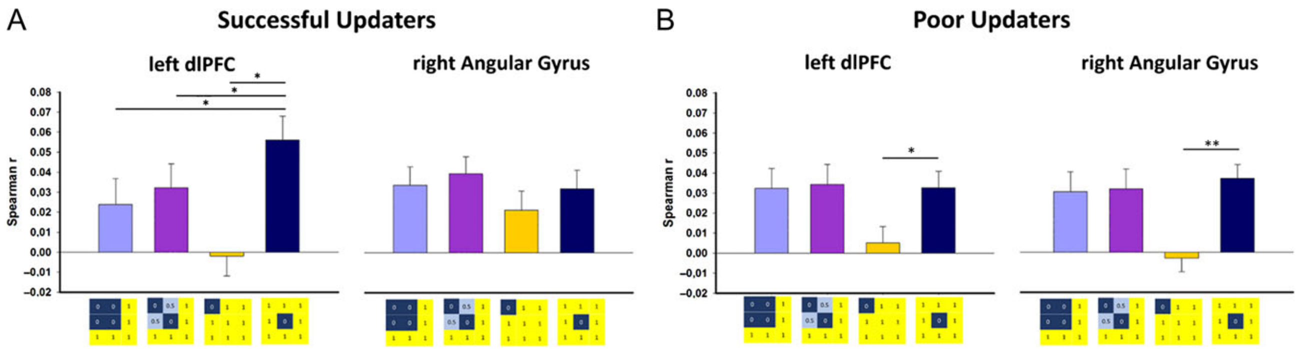


Figure 5. Model representation distinguishes between successful and poor updaters. We distinguished between (*A*) successful and (*B*) poor updaters based on a median split of updating success (successful updaters: mean performance of 54% correct, poor updaters: mean performance of 15% correct). (*A*) In the successful updaters, the "old updated distinct" model was represented with the highest fit value in the left dlPFC (Spearmans *r* = 0.056) and (*B*) while this was not the case in poor updaters (Spearmans *r* = 0.033). Error bars indicate standard error of the mean. *P < 0.05, **P < 0.001.

roles of the dlPFC, there is also evidence for its role in long-term memory processes (Rossi et al. 2001; Simons and Spiers 2003). Specifically, the dlPFC has been implicated in memory control processes, supporting strategic search, retrieval and evaluation of stored representations (Fletcher and Henson 2001). Moreover, the dlPFC has been linked to the verification and monitoring of recollected information (Burgess and Shallice 1996). These latter functions of the dlPFC dovetail with its key role in memory updating that we propose here, requiring the evaluation of and comparison between stored and currently presented associations in order to detect the new information. However, as participants identified the updated item (i.e., city) with high precision on experimental day 2, the mere recognition that associations have been modified does not appear to be sufficient for successful memory updating, nor does the mere attention to the updated information or novelty detection that are both reflected in participants judgments, and we presume that the role of the dlPFC goes beyond that of a "discrepancy detector." Instead, the dlPFC may orchestrate the retrieval of stored representations and the encoding of modified associations, facilitating actual memory updating, most likely in interaction with MTL areas. It may be this interaction and integration of stored representations and incoming information that distinguishes the updating of established memories from working memory updating.

The dlPFC has recently also been assigned a critical role in working memory updating (D'Ardenne et al. 2012). In working memory, however, both the original and the new information is maintained in the dlPFC (Curtis and D'Esposito 2003; Barbey et al. 2013) without the need to interact with long-term storage sites, whereas this interaction is essential for the updating of consolidated memories. We observed a significant increase in the crosstalk between the dlPFC and the hippocampus, when learned associations were updated and this functional connectivity was directly linked to updating success. Although PPI data do not allow conclusions about the direction of the interaction, the fact that we did not find a significant increase in hippocampal activity for updated items, nor an updating-specific activity pattern in the hippocampus, lets us assume that this crosstalk was mainly driven by the dlPFC.

Interactions between the dlPFC and hippocampus may facilitate memory updating in several ways. The dlPFC, when detecting modified associations, may suppress the activation of the original memory representations in the MTL, as shown in intentional forgetting paradigms (Anderson et al. 2004). Moreover, the dlPFC activation may foster the storage of the modified information in MTL areas or its incorporation into the existing trace. Our findings of a positive interaction between dlPFC and hippocampus, which was directly associated with updating success, speaks in favor of the latter alternative and renders a suppression effect, which should be reflected in a negative interaction, rather unlikely. It is to be noted, however, that the contrasts "old updated" versus "old identical" and "old updated" versus "new" may include a number of different processes, such as noticing that an item was updated, reactivating the existing memory trace, or the attempt to update memories. Although we cannot distinguish between these processes during the presentation of updated information in the present study, all of these processes may be relevant for successful memory updating. Interestingly, an earlier study that used an AB/AC interference paradigm, resembling our task design to test the relation between memory reactivation and integration of competing memories provided evidence for an involvement of the dlPFC in the processing of competing memories (Kuhl et al. 2011). Although there are important differences between this previous study and the present study, for instance, with respect to the degree to which participants were instructed to update their memories and to the age of the memories that were required to update, both studies agree in that they point to an important role of the dlPFC in memory updating.

While we consider a suppression-like mechanism as main source of successful updating rather unlikely, our data still suggest that representations of prior knowledge might hamper updating processes. Prior knowledge is thought to be represented in parietal areas, including the angular gyrus, and its relevance for the ongoing task is assumed to be detected by the vmPFC (van Kesteren et al. 2012). In line with these ideas, the angular gyrus and vmPFC were less active when modified information, compared with original information, was shown. Recruitment of the angular gyrus during the presentation of updated information was associated with impaired updating performance. Moreover, the specific increase in neural similarity in the angular gyrus for originally encoded item pairs distinguished successful from unsuccessful updaters. Together, these findings suggest that the recruitment of prior knowledge representations during the presentation of the original information is beneficial for updating performance, presumably as it allows a sharp discrimination between original and modified information, while the recruitment of those areas during the presentation of modified information impedes updating success. Albeit we did not find updating-related changes in dlPFC–angular gyrus connectivity, the dlPFC might also coordinate angular gyrus activity through its connection to the vmPFC, which was increased for updated items in our study. An additional account describes the function of the medial prefrontal cortex (mPFC) in the process of integrating new memories into existing structures and storing these memories (Schlichting and Preston 2016) as seen in schema-based learning. This would also describe the current observations of heightened vmPFC activity when "old identical" information is presented. While an interaction between the mPFC and hippocampus is thought to facilitate this integration process (Kroes and Fernandez 2012; Schlichting and Preston 2016), our results did not show this pattern, which may however be due to differences in study design and task instructions.

It has been argued that the updating of consolidated memories is based on memory reconsolidation processes (for a recent review see Lee et al. 2017). Specifically, it is assumed that the reactivation of a consolidated memory renders the reactivated trace labile again so that it needs to be stabilized anew during a period of reconsolidation (Nadel and Land 2000; Hupbach et al. 2008; Nader and Einarsson 2010; Dudai 2012). During the reconsolidation window, memories can be weakened, strengthened, or updated (Dudai 2006; Alberini 2011). Reactivation-dependent memory modifications have by now been shown across tasks and species (Hupbach et al. 2007; Lee 2008; Stollhoff et al. 2008; Schwabe et al. 2012). Although we did not aim to probe reconsolidation processes, a reconsolidation-like mechanism might still have been active. In the updating phase, we presented the face first, which may have served as a reminder for the respective face–city pair, and the updated (or original) city name shortly thereafter, which may be considered a contextual reinstatement, possibly aiding an inability to successfully update memories (Gershman et al. 2013). Previous evidence from a study using TMS pointed also to a critical role of the dlPFC in the postreactivation strengthening of the episodic memories (Sandrini et al. 2013). In contrast to the present study, however, this study did not address the actual updating of memories by new information. In fact, we are not aware of any reconsolidation study that tested the neural underpinnings of the updating of long-term memories in humans. Closely related to the reconsolidation concept and to the current findings, however, is the fundamental and highly controversial issue of the fate of the original memory trace after memory updating (Hardt et al. 2009; Bermúdez-Rattoni and McGaugh 2017). Is the original trace overwritten by the representation of the updated information? Or are there 2 competing traces and successful updating reflects mainly the predominance of the trace representing the modified association? Experimental studies in humans, such as the present, can hardly solve this issue. Animal studies employing state-of-the art molecular techniques, however, could make a significant contribution to this long-standing debate. Moreover, as our data suggested updating-related activity mainly in the left dlPFC and there is some evidence for distinct roles of the right and left dlPFC in retrieval processes (Rossi et al. 2001; Javadi and Walsh 2012), future studies may use brain stimulation techniques to test whether memory updating processes are indeed lateralized.

In sum, our data provide novel insights into the neural mechanisms underlying a fundamental feature of memory, its ability to update in light of new information. In particular, we show that the dlPFC, most likely through its interaction with the hippocampus, is essential for keeping memories up to date, enabling them to effectively guide choice or simulate upcoming events (Schacter et al. 2007) and thus to prepare the organism for the future.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Authors' Contributions

L.S. conceived and designed the experiment, L.M.K. performed research, L.M.K. and L.C.D. analyzed the data, L.S. and G.J. supervised research and analysis, L.M.K. and L.S. drafted the manuscript, all authors contributed to the manuscript.

Funding

This study was funded by the German Research Foundation (DFG; SCHW1357/16-1).

Notes

The authors thank Hendrik Heinbockel, Jan Patrick Kalweit, Svetlana Wiese, Vladimir Demchenko, Lea Seule, Simone Roik, and Felix Kalbe for their help during data collection and Lukas Mönch for programming the task. The authors further greatly acknowledge the help of Daniel Kutzner, Wiebke Schmidt, and Christoph Schaede during fMRI scanning. *Conflict of Interest*: All authors declare no conflict of interest.

References

Alberini CM. 2011. The role of reconsolidation and the dynamic process of long-term memory formation and storage. Front Behav Neurosci. 5:1–10.

Alvarez P, Squire LR. 1994. Memory consolidation and the medial temporal lobe: a simple network model. Proc Natl Acad Sci USA. 91:7041–7045.

Aly M, Turk-Browne NB. 2016. Attention stabilizes representations in the human hippocampus. Cereb Cortex. 26:783–796.

Anderson MC, Ochsner KN, Kuhl B, Cooper J, Robertson E, Gabrieli SW, Glover GH, Gabrieli JD. 2004. Neural systems underlying the suppression of unwanted memories. Science. 303:232–235.

Baddeley AD, Dale HCA. 1966. The effect of semantic similarty on retroactive inteference in long- and short-term memory. J Verbal Learning Verbal Behav. 5:417–420.

Bakker A, Kirwan CB, Miller M, Stark CE. 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science. 319:1640–1642.

Barbey AK, Koenigs M, Grafman J. 2013. Dorsolateral prefrontal contributions to human working memory. Cortex. 49: 1195–1205.

Bermúdez-Rattoni F, McGaugh JL. 2017. Memory reconsolidation and memory updating: two sides of the same coin? Neurobiol Learn Mem. 142:1–3.

Bilek E, Schafer A, Ochs E, Esslinger C, Zangl M, Plichta MM, Braun U, Kirsch P, Schulze TG, Rietschel M, et al. 2013. Application of high-frequency repetitive transcranial magnetic stimulation to the DLPFC alters human prefrontal-hippocampal functional interaction. J Neurosci. 33: 7050–7056.

Blumenfeld RS, Parks CM, Yonelinas AP, Ranganath C. 2011. Putting the pieces together: the role of dorsolateral prefrontal cortex in relational memory encoding. J Cogn Neurosci. 23:257–265.

Burgess N, Maguire EA, O'Keefe J. 2002. The human hippocampus and spatial and episodic memory. Neuron. 35:625–641.

Burgess PW, Shallice T. 1996. Confabulation and the control of recollection. Memory. 4:359–411.

Chun MM, Turk-Browne NB. 2007. Interactions between attention and memory. Curr Opin Neurobiol. 17:177–184.

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE. 1997. Temporal dynamics of brain activation during a working memory task. Nature. 386:604–608.

Cole MW, Schneider W. 2007. The cognitive control network: integrated cortical regions with dissociable functions. Neuroimage. 37:343–360.

Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G. 1995. Automated multi-modality image registration based on information theory. In: Bizais Y, Barillot C, Di Paola R, editors. Proc. information processing in medical imaging. Dordrecht, the Netherlands: Kluwer Academic Publishers. p. 263–274.

Curtis CE, D'Esposito M. 2003. Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci. 7:415–423.

Depue BE, Curran T, Banich MT. 2007. Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. Science. 317:215–219.

Dudai Y. 2006. Reconsolidation: the advantage of being refocused. Curr Opin Neurobiol. 16:174–178.

Dudai Y. 2012. The restless engram: consolidations never end. Annu Rev Neurosci. 35:227–247.

D'Ardenne K, Eshel N, Luka J, Lenartowicz A, Nystrom LE, Cohen JD. 2012. Role of prefrontal cortex and the midbrain dopamine system in working memory updating. Proc Natl Acad Sci USA. 109:19900–19909.

Egner T, Hirsch J. 2005. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nat Neurosci. 8:1784–1790.

Eichenbaum H. 1999. The hippocampus and mechanisms of declarative memory. Behav Brain Res. 103:123–133.

Fletcher PC, Henson R. 2001. Frontal lobes and human memory. Brain. 124:849–881.

Fuster JM, Alexander GE. 1971. Neuron activity related to short-term memory. Science. 173:652–654.

Gershman SJ, Schapiro AC, Hupbach A, Norman KA. 2013. Neural context reinstatement predicts memory misattribution. J Neurosci. 33:8590–8595.

Gilboa A, Marlatte H. 2017. Neurobiology of schemas and schema-mediated memory. Trends Cogn Sci. 21:618–631.

Hardt O, Wang SH, Nader K. 2009. Storage or retrieval deficit: the yin and yang of amnesia. Learn Mem. 16:224–230.

Hupbach A, Gomez R, Hardt O, Nadel L. 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. Learn Mem. 14:47–53.

Hupbach A, Hardt O, Gomez R, Nadel L. 2008. The dynamics of memory: context-dependent updating. Learn Mem. 15:574–579.

Javadi A.H., Walsh, V. 2012. Transcranial direct current stimulation (tDCS) of the left dorsolateral prefrontal cortex modulates declarative memory. Brain Stimul. 5:231–241

Jing HG, Madore KP, Schacter DL. 2017. Preparing for what might happen: an episodic specificity induction impacts the generation of alternative future events. Cognition. 169:118–128.

Johnson JA, Strafella AP, Zatorre RJ. 2007. The role of the dorsolateral prefrontal cortex in bimodal divided attention: two transcranial magnetic stimulation. Studies. J Cogn Neurosci. 19:907–920.

Karpicke JD, Roediger HLI. 2008. The critical importance of retrieval for learning. Science. 319:966–968.

Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci. 2:1–28.

Kroes MC, Fernandez G. 2012. Dynamic neural systems enable adaptive, flexible memories. Neurosci Biobehav Rev. 36: 1646–1666.

Kuhl BA, Bainbridge WA, Chun MM. 2012. Neural reactivation reveals mechanisms for updating memory. J Neurosci. 32: 3453–3461.

Kuhl BA, Rissman J, Chung MM, Wagner AD. 2011. Fidelity of neural reactivation reveals competition between memories. PNAS. 108:5903–5908.

Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. 2010. Presentation and validation of the Radboud Faces Database. Cogn Emot. 24:1377–1388.

Lee JL. 2008. Memory reconsolidation mediates the strengthening of memories by additional learning. Nat Neurosci. 11: 1264–1266.

Lee JLC, Nader K, Schiller D. 2017. An update on memory reconsolidation updating. Trends Cogn Sci. 21:531–545.

Lenartowicz A, McIntosh AR. 2005. The role of anterior cingulate cortex in working memory is shaped by functional connectivity. J Cogn Neurosci. 17:1026–1042.

Leutgeb JK, Leutgeb S, Moser M-B, Moser EI. 2007. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science. 315:961–966.

Loftus EF. 1975. Leading questions and the eyewitness report. Cogn Psychol. 7:560–572.

Lundqvist D, Flykt A, Öhman A 1998. The Karolinska Directed Emotional Faces—KDEF, CD ROM from Department of Clinical Neuroscience.

Manenti R, Cotelli M, Calabria M, Maioli C, Miniussi C. 2010. The role of the dorsolateral prefrontal cortex in retrieval from long-term memory depends on strategies: a repetitive transcranial magnetic stimulation study. Neuroscience. 166: 501–507.

McLaren DG, Ries ML, Xu G, Johnson SC. 2012. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. Neuroimage. 61:1277–1286.

Murray LJ, Ranganath C. 2007. The dorsolateral prefrontal cortex contributes to successful relational memory encoding. J Neurosci. 27:5515–5522.

Nadel L, Hupbach A, Gomez R, Newman-Smith K. 2012. Memory formation, consolidation and transformation. Neurosci Biobehav Rev. 36:1640–1645.

Nadel L, Land C. 2000. Memory traces revisited. Nat Rev Neurosci. 1:209–212.

Nader K, Einarsson EO. 2010. Memory reconsolidation: an update. Ann N Y Acad Sci. 1191:27–41.

Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. 2014. A toolbox for representational similarity analysis. PLoS Comput Biol. 10:1–11.

Preston AR, Eichenbaum H. 2013. Interplay of hippocampus and prefrontal cortex in memory. Curr Biol. 23:1–21.

Ritchey M, Montchal ME, Yonelinas AP, Ranganath C. 2015. Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. Elife. 4:1–19.

Rossi S, Cappa SF, Babiloni C, Pasqualetti P, Miniussi C, Carducci F, Babiloni F, Rossini PM. 2001. Prefrontal cortex in long-term memory: an "interference" approach using magnetic stimulation. Nat Neurosci. 4:948–952.

Sandrini M, Brambilla M, Manenti R, Rosini S, Cohen LG, Cotelli M. 2014. Noninvasive stimulation of prefrontal cortex strengthens existing episodic memories and reduces forgetting in the elderly. Front Aging Neurosci. 6: 1–9.

Sandrini M, Censor N, Mishoe J, Cohen LG. 2013. Causal role of prefrontal cortex in strengthening of episodic memories through reconsolidation. Curr Biol. 23:2181–2184.

Schacter DL, Addis DR, Buckner RL. 2007. Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci. 8:657–661.

Schacter DL, Alpert NM, Savage CR, Rauch SL, Albert MS. 1996. Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. Proc Natl Acad Sci USA. 93:321–325.

Schiller D, Monfils MH, Raio CM, Johnson DC, Ledoux JE, Phelps EA. 2010. Preventing the return of fear in humans using reconsolidation update mechanisms. Nature. 463:49–53.

Schlichting ML, Preston AR. 2016. Hippocampal-medial prefrontal circuit supports memory updating during learning and post-encoding rest. Neurobiol Learn Mem. 134:91–106.

Schwabe L, Nader K, Pruessner JC. 2014. Reconsolidation of human memory: brain mechanisms and clinical relevance. Biol Psychiat. 76:274–280.

Schwabe L, Nader K, Wolf OT, Beaudry T, Pruessner JC. 2012. Neural signature of reconsolidation impairments by propranolol in humans. Biol Psychiat. 71:380–386.

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

Squire LR, Zola-Morgan S. 1991. The medial temporal lobe network. Science. 253:1380–1386.

Stollhoff N, Menzel R, Eisenhardt D. 2008. One retrieval trial induces reconsolidation in an appetitive learning paradigm in honeybees (*Apis mellifera*). Neurobiol Learn Mem. 89: 419–425.

Szczepanski SM, Knight RT. 2014. Insights into human behavior from lesions to the prefrontal cortex. Neuron. 83:1002–1018.

Tulving E. 2002. Episodic Memory: From mind to brain. Annu Rev Psychol. 53:1–25.

van Kesteren MT, Ruiter DJ, Fernandez G, Henson RN. 2012. How schema and novelty augment memory formation. Trends Neurosci. 35:211–219.

van Veen V, Carter CS. 2002. The anterior cingulate as a conflict monitor: fMRI and ERP studies. Physiol Behav. 77:477–482.

Wagner IC, van Buuren M, Kroes MC, Gutteling TP, van der Linden M, Morris RG, Fernandez G. 2015. Schematic memory components converge within angular gyrus during retrieval. Elife. 4:1–28.

Weible AP. 2013. Remembering to attend: the anterior cingulate cortex and remote memory. Behav Brain Res. 245:63–75.

Wimmer EG, Shohamy D. 2012. Preference by association: how memory mechanisms in the hippocampus bias decisions. Science. 338:270–273.

Wolosin SM, Zeithamova D, Preston AR. 2013. Distributed hippocampal patterns that discriminate reward context are associated with enhanced associative binding. J Exp Psychol Gen. 142:1264–1276.

Yamaguchi S, Knight RT. 1991. Anterior and posterior association cortex contributions to the somatosensory P300. J Neurosci. 11:2039–2054.

Yassa MA, Stark CE. 2011. Pattern separation in the hippocampus. TrendsNeurosci. 34:515–525.

Zeithamova D, Dominick AL, Preston AR. 2012. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron. 75: 168–179.