10 – May– 2016

Knowledge Acquisition During Exam Preparation Improves Memory and Modulates Memory Formation

Garvin Brod[a], Ulman Lindenberger[a,b], Anthony D. Wagner[c], & Yee Lee Shing[a,d]

[a]Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany; [b]European University Institute, San Domenico di Fiesole, Italy; [c]Department of Psychology, Stanford University, Palo Alto, CA, USA; [d]Division of Psychology, University of Stirling, Stirling, UK

Forthcoming in *Journal of Neuroscience*

Address correspondence to Dr. Yee Lee Shing, Division of Psychology, University of Stirling, Stirling, UK. Phone: +44 1786 466841, Email: yee.shing@stir.ac.uk

Abstract

According to the schema-relatedness hypothesis, new experiences that make contact with existing schematic knowledge are more easily encoded and remembered than new experiences that do not. Here we investigate how real-life gains in schematic knowledge affect the neural correlates of episodic encoding, assessing medical students three months before and immediately after their final exams. Human participants were scanned with functional magnetic resonance imaging (fMRI) while encoding associative information that varied in relatedness to medical knowledge (face–diagnosis versus face–name pairs). As predicted, improvements in memory performance over time were greater for face–diagnosis (high knowledge-relevance) than for face–name (low knowledge-relevance) pairs. Improved memory for face–diagnosis pairs was associated with smaller subsequent memory effects in the anterior hippocampus, along with increased functional connectivity between the anterior hippocampus and left middle temporal gyrus, a region important for the retrieval of stored conceptual knowledge. The decrease in the anterior hippocampus subsequent memory effect correlated with knowledge accumulation, as independently assessed by a web-based learning platform with which participants studied for their final exam. These findings suggest that knowledge accumulation sculpts the neural networks associated with successful memory formation, and highlight close links between knowledge acquired during studying and basic neurocognitive processes that establish durable memories.

Significance Statement

In a medical students sample, we tracked knowledge accumulation via a web-based learning platform and investigated its effects on memory formation before and after participants' final medical exam. Knowledge accumulation led to significant gains in memory for knowledge-related events and predicted a selective decrease in hippocampal activation for successful memory formation. Furthermore, enhanced functional connectivity was found between hippocampus and semantic processing regions. These findings (i) demonstrate that knowledge facilitates binding in the hippocampus by enhancing its communication with the association cortices, (ii) highlight close links between knowledge induced in the real world and basic neurocognitive processes that establish durable memories, and (iii) exemplify the utility of combining laboratory-based cognitive neuroscience research with real-world educational technology for the study of memory.

Introduction

Information contained in new experiences is learned more easily when learners can relate the information to their prior knowledge (Bransford and Johnson, 1972). This long-standing observation has commonly been attributed to individuals' ability to assimilate new experiences with existing schemas. Schemas are commonly portrayed as structured associative information, representationally distributed in neocortex, that allows for more elaborative encoding and provides a search frame during retrieval (Bartlett, 1932; Alba and Hasher, 1983; Ghosh and Gilboa, 2014). The integration of new experiences with existing schemas during encoding is associated with activity in anterior parts of the hippocampus (Ranganath and Ritchey, 2012; Poppenk et al., 2013; Schlichting et al., 2015). Connections between newly formed associations and existing schemas presumably facilitate binding in the HC and the integration of these associations into existing knowledge structures (van Kesteren et al., 2012; McClelland, 2013; Preston and Eichenbaum, 2013). Consistent with this framework, studies with rodents have shown that pre-existing neocortical schemas allow rapid integration of new information, as measured by accelerated HC-independence for consolidation of schema-related flavor–place associations, which is absent when connected neocortical regions are blocked (Tse et al., 2007, 2011). However, evidence from human subjects is limited (see van Kesteren et al., 2014, for findings in the parahippocampal gyrus).

The human anterior HC receives cross-domain inputs mainly from entorhinal cortex and binds these inputs into integrated memory representations (Prince et al., 2005; Davachi, 2006; Zimmer et al., 2006; Liang et al., 2013). HC has long-range connections to a number of neocortical areas, in particular medial prefrontal cortex (mPFC) and lateral temporal regions (Poppenk et al., 2013). Both mPFC and lateral temporal regions, especially the middle temporal gyrus (MTG), are involved in the retrieval of stored conceptual knowledge (Badre and Wagner, 2007; Lau et al., 2008; Binder et al., 2009; Turken and Dronkers, 2011).

An underlying assumption of most work investigating schema effects on memory is that knowledge accumulation in a domain, for example via formalized instruction, should lead to enhanced memorability of new information in the domain, in the sense of a "Matthew effect" (Stanovich, 1986) or cumulative advantage. However, systematic longitudinal studies are needed to strengthen a causal interpretation of the relationship between increasing knowledge and enhanced memory for episodes related to that knowledge, which have been lacking thus far.

Here, we utilized a real-world educational setting that involved extensive knowledge acquisition to fill this gap. Specifically, we tested a sample of medical students who used a web-based learning platform to prepare for a state-regulated final medical exam. New episodic learning was tested with a face–word associative memory task in two fMRI sessions, one three months before (T1) and one right after the final exam (T2). Critically, half of the faces were paired with medical diagnosis words (*high* relevance of schematic knowledge), while the other half were paired with first names (*low* relevance of schematic knowledge).

We predicted that, first, knowledge-related facilitation of episodic memory would be greater for high-relevance episodes (face–diagnosis pairs) than for low-relevance episodes (face–name pairs). Second, we predicted that as knowledge was gained (i.e., from T1 to T2), anterior HC activation associated with successful memory formation would decrease. This decrease was predicted to occur in the high-relevance condition only, putatively reflecting facilitated binding due to greater schematic support. Third, we predicted an activation increase in regions associated with the representation of schematic knowledge (Lau et al., 2008; Binder et al., 2009), such as lateral temporal regions, as well as an increase in functional connectivity between anterior HC and lateral temporal regions across time (T1 to T2), particularly for the high-relevance condition. Finally, we examined whether any of these

predicted changes would be related to individual differences in knowledge increase from T1 to T2, as independently assessed by the web-based learning platform.

Materials and Methods

Participants

Thirty-five medical students (20 women, age range = 23–29 years, mean age = 25.9 years) who gave written informed consent participated in the study. All participants were right-handed and had no history of psychiatric or neurological disorders. Recruitment took place via e-mails sent out to Berlin-based users of the commercial web-based learning platform AMBOSS, which prepares medical students for their final exam. Participants were paid 76 Euros. Two participants did not return for the second (T2) measurement; data from two further participants were excluded because they did not make proper use of the memory confidence scale (see below), leaving too few remembered trials ($n \leq 10$) for analysis. We thus analyzed data from 31 participants. The study also included a control group of 16 medical students (mean age 25.0 years) who were 0.5–1 year prior to taking the final exam to assess changes in brain structure (not reported in this paper, but see behavioral results of the memory task on p. 14). The control group participants were also tested twice in the course of three months, but did not study intensively during this time (which was confirmed via questionnaires), and did not use the web-based learning platform. Ethics approval was obtained from the ethics committee of the German Psychological Society (DGPs).

General Design and Procedure

Participants were tested twice, once three months prior to their final state-regulated written medical exam (T1), and again shortly after (mean: 11.7 days, range: 1–22 days) their written exam (T2). In between, they intensively prepared for the exam using the web-based

learning platform. The exam takes three days (five hours per day) and consists of a total of 320 multiple-choice questions, covering all clinical subjects taught during the final three years of medical school in Germany. On the learning platform, they followed a structured 100-day learning plan that guided them through all of the topics relevant for the final exam. Each learning day consisted of solving exam questions from previous years and reading relevant information (e.g., on symptoms, etiology, epidemiology, pathophysiology) about the medical syndromes and diagnoses covered in the questions. Exam questions were multiple-choice questions with five response options, of which only one was correct. Participants received detailed feedback on their responses. We wish to stress that, although the web-based learning platform was critical in our study design because it allowed close monitoring of knowledge accumulation in our sample, our design does not permit any conclusions about the effectiveness of studying with the platform in comparison to other methods of studying.

Measuring Learning Performance and Success

We measured participants' learning performance using data provided by the learning platform. We focused on daily measures of the number of questions answered on the platform, and the correctness of the answers (% correct) as a measure of their knowledge. To measure the increase in medical knowledge from T1 to T2, we first calculated, for each individual, the average accuracy (percent correct) for answered questions during the first week of studying on the learning platform (T1 measurement). This score was subtracted from the average percent correct score achieved during the week before the real exam (T2 measurement), in which participants solved mock exams consisting of previously non-studied questions.

Because the medical students differed in their initial level of knowledge, with some students having high knowledge already at T1 and others lower knowledge, and because there is a finite number of old exam questions, participants' change in accuracy (T2-T1) can be

expected to negatively correlate with accuracy at T1. This was indeed the case, such that participants who started with high level of knowledge showed less change in accuracy ($r = –.82$, $p < .001$). To control for this negative correlation between change and initial performance on the platform, we used a residualized change score (Cohen et al., 2003). To obtain a validity check of our learning measures, we asked the participants to report their final exam scores.

Memory Stimuli, Task, and Behavioral Analyses

The encoding phase took place in the MRI scanner and participants were instructed to memorize 140 face–word pairs, in which half of the words were diagnoses and the other half were first names (Figure 1). We predicted that encoding of a face-diagnosis association would activate a medicine-related network of schematic knowledge in the participants, and that this network would increase in strength and connection over the period of intensive learning. Over time, this should then lead to a differential encoding advantage for face–diagnosis over face–name pairs. Two parallel stimulus lists were created to allow counterbalancing across participants and the two study–test fMRI sessions. A total of 140 medical diagnoses and 140 common German first names were used together with 140 neutral face pictures. Each face was pseudorandomly combined with one diagnosis and one name. Thus, while participants saw each face twice, once at T1 and once at T2, they saw each diagnosis and name only once during the whole study. To ensure that our diagnosis stimuli were sensitive to change in knowledge, they were chosen from a wide array of diagnoses relevant for the final exam based on a rating by four recent medical graduates regarding their difficulty and prevalence. Based on these ratings, highly frequent diagnoses (e.g., hypertension) as well as highly similar diagnoses (e.g., Type 1 Diabetes vs. Type 2 Diabetes) were discarded. The face stimuli consisted of pictures of Caucasian young adults taken from the Center for Vital Longevity Face Database (Minear and Park, 2004). Faces and names were matched for

gender. To avoid highly implausible face–diagnosis pairs, these pairs were matched for gender (e.g., pre-eclampsia was used only for female faces) as well as age specificity (only diagnoses were chosen that can affect young adults). Face–diagnosis and face–name pairs were presented for 5 seconds each in an interleaved fashion (in pseudorandom order). Trials were separated by a variable fixation cross period of 2–5 seconds (mean: 3.5 seconds). During each session (T1 or T2), there were two experimental blocks, each consisting of 70 trials.

Before entering the scanner, participants were instructed to try to memorize both the face–diagnosis and face–name pairs equally well and were told that there would be a memory test outside of the scanner. To ensure that the participants were paying attention to the task and to promote elaborative encoding, they were asked to indicate for each face-word pair whether or not the name / diagnosis fit with the face, responding with their left / right index finger. Left / right response options were counterbalanced across participants. The encoding phase took 20 minutes in total and was performed after the structural scans.

The retrieval phase took place outside of the scanner, about 10 minutes after the end of the encoding session. Participants were presented with all 140 faces again in a pseudorandom order. For each face, they were given 4 first names or 4 diagnoses, of which one had been presented with the face during the encoding phase (target), whereas the other three were names / diagnoses seen with other faces during encoding (lures). Participants indicated their choice via button press. Afterwards, they were asked to indicate their decision confidence on a scale of 1 (guess) to 4 (very sure). They were given no time limit for their responses, but were told to answer as quickly and as correctly as possible.

Data were analyzed using R (R Core Team, 2014). A repeated-measures ANOVA was performed with condition (diagnoses / names) and time (T1, T2) as within-subjects factors to test for differences in memory (% correctly retrieved associations, independent of decision confidence).

fMRI Data Acquisition and Preprocessing

T2*-weighted echo-planar images were acquired using a 3T Siemens TIM Trio MRI scanner (direction = transverse (interleaved ascending), FOV = 216 mm, TR = 2500 ms, TE = 30 ms, number of slices = 45, slice thickness = 2.5 mm, matrix = 72 x 72, voxel size = 3 x 3 x 2.5 mm, distance factor = 20%, 2 runs with 232 volumes each, including 4 dummy volumes each). To attenuate signal dropout in orbitofrontal regions, the slice orientation was tilted upwards vertically by 15 degrees after alignment to the anterior commissure–posterior commissure plane (Weiskopf et al., 2006). To estimate geometric distortion and signal loss in the EPI, an additional 53-seconds fieldmap was acquired. Structural data was acquired using a T1-weighted 3D magnetization-prepared rapid gradient echo sequence (TR 2500 ms, TE 2500 ms, sagittal orientation, spatial resolution 1 x 1 x 1 mm).

Data were preprocessed and analyzed using FEAT in FSL (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl; Smith et al., 2004). Functional data were corrected for motion (MCFLIRT), slice acquisition times (interleaved), and local field inhomogeneities (BBR / FUGUE), then high-pass filtered (80 Hz), and spatially smoothed using a 5-mm full-width half-maximum Gaussian filter, resulting in a final estimated spatial smoothness of 6.9 x 6.8 x 6.6 mm^3. Data were first coregistered with the structural image and then spatially normalized into a common space (Montreal Neurological Institute (MNI) 152 standard-space 2 mm^3).

fMRI Analyses

Brain Activation

First-level analyses were conducted for individual participants, separately for the two runs at T1 and at T2. Using general linear modeling (GLM), regressors were generated by convolving the impulse function related to the onset and length of encoding events with a

Gamma hemodynamic response function (5 seconds boxcar function). Using behavioral data from retrieval, we sorted encoding trials according to their later memory outcome to investigate subsequent memory effects (SMEs, remembered > forgotten contrast, see Brewer et al., 1998; Wagner et al., 1998; Paller and Wagner, 2002). Five types of events were modeled with separate regressors in the GLM. Trials that received a correct retrieval response with a confidence rating of above 1 (i.e. non-guessing trials) were classified as remembered diagnosis or remembered name events; those that received an incorrect response were classified as forgotten diagnosis or forgotten name events. A fifth regressor of no interest was included for all remembered events that received "guessing" ratings on the confidence scale. Overall, the number of correct guesses was low (mean_T1/T2 = 8.9/8.6 out of 140 trials across the two runs), and was higher for the name (mean_T1/T2 = 13.3/12.7) than for the diagnosis (mean_T1/T2 = 4.5/4.4) condition (T1: t(30)=5.50, p<.001; T2: t(30)=5.49, p < .001). This condition difference was due to the stronger tendency to give a "guess" rating for the name (mean_T1/T2 = 38.7/34) than for the diagnosis (mean_T1/T2 = 12.6/9.8) condition (T1: t(30)=7.09, p < .001; T2: t(30)=6.58, p < .001). SMEs, defined by the remembered > forgotten contrast, were computed for the face–diagnosis and face–name condition separately.

In a next step, the two runs were combined using a within-subjects fixed-effects analysis and normalized into MNI space. To test for changes in brain activation from T1 to T2 that differed by condition (diagnosis, name), a within-subjects fixed-effects analysis was performed that tested for differences in SME between time points (T1, T2) that were larger for the diagnoses than for the names, and vice versa (memory x time x condition interaction). To better understand any significant pattern observed in the memory x time x condition interaction, additional within-subjects fixed-effects analyses were performed to test for differences in the SME between T1 and T2 for each of the conditions separately. Across-subjects analyses were carried out using a mixed-effects model in the FLAME framework in

FSL. Z-statistic images were thresholded voxel-wise at a threshold of $z > 2.3$. Multiple comparison correction was performed using the 3DClustSim program of the AFNI software package (http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html), which computes minimum cluster-extent thresholds for specific regions of interest (ROI's) using Monte Carlo simulation analysis. In addition to a whole-brain gray-matter mask, based on our a priori hypothesis about changes in the anterior HC, we created an anatomical mask of the bilateral anterior HC using the probabilistic Harvard-Oxford Subcortical Structural Atlas, including voxels located at the anterior 35% of the long axis of the HC and with at least 25% probability of being inside the HC. Smoothness of our group-level data was estimated on the residual time series image using AFNI's 3dFWHMx (http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dFWHMx.html). A simulation with 10,000 runs yielded minimum cluster extents of 17 (HC) and 143 (whole-brain) voxels to maintain a family-wise error rate of $p < .05$.

To determine the source of the observed change in SME in the anterior HC for the diagnosis condition, and to find out whether this change was related to individual differences in knowledge increase (as measured by behavioral performance during the first and last week of studying on the learning platform), percent signal change was extracted from the memory x time interaction contrast and correlated with the residualized change score.

Brain Connectivity

We tested for changes in coupling between the anterior HC and lateral temporal regions using psycho-physiological interaction (PPI) analysis separately for the left and right HC. The time courses of left and right anterior HC (using an anatomical mask as described in the previous section) served as the physiological regressor in the model. Psychological regressors were defined as representing the remembered > forgotten (SME) contrast (i.e., remembered – forgotten, and in addition remembered + forgotten to model the shared

variance), which were convolved with a Gamma hemodynamic response function. Finally, a PPI regressor representing the interaction of the psychological and the physiological regressors was created. The three regressor types were then added to the existing brain activation GLM, replacing the corresponding remembered and forgotten events. This was done separately for the two conditions, runs, and time points, which were then combined at higher levels using within-subjects fixed-effects analyses and between-subjects mixed-effects models in the same way as the activation analyses. In an additional analysis, to test whether the observed connectivity changes were related to changes in knowledge, the residualized gain score was entered as a covariate into the between-subjects mixed effects model.

Results

Knowledge Accumulation Predicts Final Exam Score

During the 100 days of intensive studying, participants answered on average 7460 (range: 3702–10605; SD = 1818.19) questions from earlier exams, thus, on average 75 questions per day (range: 0–440; SD=17.15). During the first week of studying at our T1 measurement, 69% (range: 48–83%, SD = 9.6%) of the earlier exam questions were answered correctly. During the wrap-up period in the last week before the real final exam, 83% (range: 69–92%; SD = 5.4%) of the questions were answered correctly (change in performance on answering questions: $t(30)=10.02$, $p < .001$). Mean performance on the actual written final exam was 80% (range: 65–90, SD = 6.4). The total number of old exam questions studied during the 100 days was positively related to the final exam score ($r = .41$, $p = .01$). Importantly, the residualized gain in knowledge (first to last week) correlated with the final exam score ($r = .47$, $p = .005$; Figure 2, left). In sum, during the 100 days, participants used the learning platform extensively and substantially increased their medical knowledge, which

was measured by the change in performance between the first and last week of studying. Furthermore, performance on the learning platform was highly predictive of their exam success later on.

Enhanced Memory Improvement for Knowledge-Related Information

A repeated-measures ANOVA revealed (a) a main effect of condition, indicating better memory for diagnoses as compared to names, $F(1,30) = 75.24$, $p < .001$; (b) a main effect of time, indicating improved memory performance at T2, $F(1,30) = 8.24$, $p = .01$; and (c) a marginally significant interaction, $F(1,30) = 3.83$, $p = .059$, suggesting a greater increase in memory for face–diagnosis than for face–name pairs (Figure 2, right). In addition, to confirm that this enhanced memory improvement for face–diagnosis pairs in the exam candidates (EC) was specific to increased medical knowledge, we compared it to a control group (CG) of medical students who did not study intensively during this period. At T1, memory performance was similar between the two groups for the high knowledge-relevance condition (EC: 67.8%, SD = 11.5; CG: 67.1%, SD = 9.2; $t(45) = .21$, $p = .83$), and for the low knowledge-relevance condition (EC: 54.6%, SD = 13.6; CG: 49.1%, SD = 10.7; $t(45) = 1.4$, $p = .17$). At T2, memory performance for the high knowledge-relevance condition was better in the exam candidates (EC: 75.2%, SD = 11.9; CG: 68.1%, SD = 13.9; $t(45)=1.83$, $p = .037$), but again was similar for the low knowledge-relevance condition (EC: 58.1%, SD = 15.5; CG: 58.2%, SD = 12.4; $t(45) = .02$, $p = .98$). A three-way mixed ANOVA revealed a reliable group x condition x time interaction ($F(1,45) = 11.44$, $p = .001$). In particular, the EC ($\Delta = +7.2\%$, SE = .017) showed a greater increase in memory for the high knowledge-relevance condition than the CG ($\Delta = +1.1\%$, SE = .028) from T1 to T2 ($t(45) = 2.0$, $p = .02$, one-tailed). For the low knowledge-relevance condition, the CG ($\Delta = +9.1\%$, SE = .027) showed a numerically greater increase in memory than the EC ($\Delta = +3.5\%$, SE = .024) from T1 to T2, which did not reach significance ($t(45) = 1.4$, $p = .16$, two-tailed).

Knowledge Accumulation is Associated with Decreased Hippocampal Activation During Successful Memory Formation

At the neural level, we found a significant memory x time x condition interaction. Specifically, a cluster in the right anterior HC (peak voxel: 26, -8, -20; Figure 3a) showed an across-time decrease in the SME that was larger for face–diagnosis pairs (high knowledge relevance) than for face–name pairs (low knowledge relevance). Follow-up analyses for the two conditions separately (memory x time) showed a decrease in SME from T1 to T2 for face–diagnosis pairs in a cluster in the right anterior HC (peak voxel: 26, -6, -22) overlapping with the cluster identified in the three-way interaction; no significant decrease in SME across time was observed for the face-name pairs. Thus, the memory x time x condition interaction was driven by a decrease in SME for the condition with high knowledge relevance (see extracted % signal change for illustration). Testing for increases in SME for either condition did not yield any significant effects. No decreases in SME were detected outside of the HC for either condition.

To examine whether the observed right HC T1 to T2 decrease in the SME for face–diagnosis pairs was related to individual differences in medical knowledge increase, percent signal change was extracted from the memory x time interaction cluster in the right anterior HC and correlated with the residualized gain between the first and last week of studying on the learning platform. This analysis revealed a significant positive correlation between knowledge increase and hippocampal SME decrease ($r = .32$, $p = .04$, Figure 3, right).

Knowledge Accumulation is Associated with Increased Hippocampus–Neocortical Connectivity During Successful Memory Formation

PPI analyses were performed to assess whether the anterior HC showed differential changes in connectivity with neocortical areas for the two conditions. For the right HC seed, testing for regions where across-time increases in functional connectivity for subsequently

remembered vs. forgotten episodes (i.e., SME in connectivity) are more pronounced for the face–diagnosis condition than for the face–name condition revealed no significant effects. However, we observed several sizable neocortical clusters with voxels just below the significance threshold ($z > 2.3$). Given that PPI contrasts have less statistical power than activation contrasts, which is due to multicollinearities between the interaction term and the psychological and physiological terms (O'Reilly et al., 2012), we performed an additional test at a lowered voxel threshold of $z > 1.96$. To account for the lower voxel threshold for multiple comparison correction, we performed another Monte Carlo simulation using 3DClustSim, which yielded a minimum cluster extents of 497 voxels (whole-brain) to maintain a family-wise error rate of $p < .05$. This analysis revealed a significant memory x time x condition interaction in the left posterior MTG (peak: -60, -48, 8; Figure 4). To follow up, we tested for across-time increases in functional SME connectivity separately for face-diagnosis and face-name condition (note: voxel threshold $z > 2.3$, cluster threshold $p < .05$). The follow-up analyses showed that the observed three-way interaction reflected a specific increase for the face–diagnosis condition, as indicated by an overlapping cluster in left posterior MTG (peak: -60, -48, 8), which was not present for the face–name condition. In addition, both conditions displayed an extensive network of neocortical regions with greater functional connectivity to the HC seed at T2 than T1 (for an overview of the results, see Table 2). Testing for regions where the time-related increase in SME was larger for the face–name condition revealed no significant effects. In addition, changes in connectivity were not related to gains in knowledge.

For the left HC seed, testing for regions where the time-related increase in SME was larger for face–diagnosis pairs than for face–name pairs (voxel threshold $z > 2.3$, cluster threshold $p < .05$) yielded two clusters in the left and right lingual gyrus / temporal-occipital fusiform cortex (peaks: 14, -64, 2; -34, -40, -6). Follow-up tests for PPI increases from T1 to

T2 yielded a highly similar network to the one observed for the right HC, including left MTG (-58, -48, 8) for the face–diagnosis condition only, as well as an extensive network of neocortical regions for both conditions (see Table 2). Again, no enhanced increases in SME were observed for the face–name condition and changes in connectivity were not related to gains in knowledge.

Discussion

By following a group of medical students who studied for their final medical exam, we were able to show that an increase in schematic knowledge induced by three months of intensive studying was associated with gains in memory for knowledge-related episodic events. These gains were further associated with a selective decrease in SME in the right anterior HC during encoding. This decrease was related to individual differences in the accumulation of knowledge, as measured by participants' performance on the learning platform. Furthermore, we observed an increase in connectivity SME between the anterior HC and the left posterior MTG, a brain area that is key to semantic processing (Badre and Wagner, 2007; Hickok and Poeppel, 2007; Binder et al., 2009; Turken and Dronkers, 2011). For the first time, our study demonstrates close links between changes in knowledge induced in a real-world educational setting and changes in encoding-related brain activation as observed with a laboratory memory paradigm.

Prior knowledge facilitates the acquisition of new, related information, presumably because it provides a pre-existing associative network that offers many links to which the new information can be bound and assimilated (Piaget, 1952; van Kesteren et al., 2012; Brod et al., 2013; Ghosh and Gilboa, 2014). Recently, the notion that binding in the HC is facilitated by the presence of a schema has received increased attention (Wang and Morris, 2010; van Kesteren et al., 2012; McClelland, 2013; Preston and Eichenbaum, 2013), mainly sparked by

a study that showed accelerated consolidation (HC-independence) of schema-related information in rodents (Tse et al., 2007). In humans, this facilitation due to schema is expected to manifest itself as a decrease in HC activation during successful memory formation, that is, a decrease in the difference in HC activation for later remembered versus forgotten events (van Kesteren et al., 2012; Zeithamova et al., 2012). To date, empirical evidence for this prediction with human subjects has been limited, except for some hints from a study by Zeithamova and colleagues (2012), showing that increased activation in mPFC, coupled with decreased activation in HC across study episodes within a session, predicted successful inference. This decrease in HC activity across study repetitions was taken to reflect either progressively more efficient coding in the HC or a decreased need for binding due to stronger overlaps with earlier events (Zeithamova et al., 2012). Indeed, comparing HC SME between the face–diagnosis pairs and face–name pairs during our T1 measurement did not reveal any HC differences: both conditions yielded strong anterior HC activity (see Table 1). Arguably, due to the HC being highly active during successful encoding of both schema-related and schema-unrelated information, condition differences in activation that are due to varying levels of schema knowledge are typically too subtle to be reliably measured by a one-occasion fMRI research design. In contrast, comparisons over time within individuals whose knowledge base is expanding may be more sensitive to capture the deceasing relationship between successful encoding and activation magnitude in the anterior HC. Hence, we conclude that extending one's knowledge base through the acquisition of schematic knowledge facilitates binding in the HC, presumably by increasing the number of potential associative links to the knowledge base.

Contrary to expectations, we did not find increased SME in lateral temporal regions across the two time points for the high knowledge-relevance condition. However, for this condition, anterior HC showed an increase in SME in functional connectivity with the left

MTG across the two time points. This increase was not present for the low knowledge-relevance condition (see Table 2). However, the memory x time x condition analysis did not identify a cluster exceeding our significance criterion ($z > 2.3$, cluster corrected). A follow-up analysis at a lowered voxel threshold ($z > 1.96$) that used an adjusted cluster correction to maintain the same cluster threshold ($p < .05$) did reveal the hypothesized three-way interaction, specifically in the left MTG cluster identified for the high knowledge-relevance condition. Increases in HC–neocortical connectivity were observed for both high and low knowledge-relevance conditions in frontal and parietal regions as well (see Table 2). Thus, in order to be more certain about the specificity of the connectivity increases between anterior HC and neocortical regions involved in semantic processing, further studies are needed. A potential next step could be to investigate whether the observed increased HC–neocortical connectivity persists during resting-state. Activity in left MTG is consistently observed during the retrieval of conceptual knowledge, especially lexical-semantic knowledge, and lesions to this region lead to severe deficits in the retrieval of word meaning (Lau et al., 2008; Binder et al., 2009; Turken and Dronkers, 2011). The accumulation of knowledge might have facilitated episodic memory formation in our task specifically by enhancing neural communication between HC and association cortex, including MTG. Connectivity increases are often observed in training studies and have been proposed to reflect improved communication between brain areas (e.g. Büchel et al., 1999; Kelly and Garavan, 2005). In the memory domain, increased connectivity between HC and lateral temporal lobe regions is associated with successful memory formation (Gagnepain et al., 2010). Gagnepain et al. (2010) demonstrated that a decreased HC SME and increased hippocampal–neocortical connectivity together were associated with the memory benefit due to priming. These findings underscore the importance of cortical inputs to the HC, which may alter memory formation processes in the HC. Therefore, the observed increase in connectivity between the HC and the left MTG in

our study may reflect increased connections to nodes within the more extensive medical knowledge network, which in turn facilitated hippocampal binding of faces to diagnoses.

In contrast to previous studies pointing to the importance of mPFC for schema-related memory processing (van Kesteren et al., 2010, 2014; Warren et al., 2014; Brod et al., 2015), the mPFC did not show reliable condition differences in our study. However, we found tentative evidence for an increasing mPFC involvement for the high knowledge-relevance condition in the PPI analysis. While this finding is in line with claims about the mPFC biasing HC processing when prior knowledge is highly relevant, a bias that can be assumed to increase with increasing schema strength, it has to be treated with caution because of the lack of a significant memory x time x condition interaction, In general, the lack of a strong mPFC engagement in our task, which is apparently inconsistent with earlier studies, may reflect crucial differences among the tasks. In memory tasks, mPFC activation has been assumed to reflect the evaluation of new experiences based on their fit, or congruence, to schematic knowledge. When congruency is high, the mPFC is furthermore assumed to inhibit the HC (van Kesteren et al., 2012; Brod et al., 2013; Ghosh et al., 2014; Warren et al., 2014). The present task conditions, however, did not vary in congruency, but in relevance, as prior medical knowledge was highly relevant for diagnoses, but less so for names. Future studies should vary both congruency and relevance to obtain a more complete picture of mPFC contributions to schema-related modulations of episodic memory.

A potential concern in our design is the use of the same faces at the two testing sessions, which might have induced proactive interference at T2. Proactive interference is typically observed in paired-associate cued recall when the cue has been previously associated with a different response. We used a forced choice recognition task in which the participants had to choose from either four names or four diagnoses (never mixed). Thus, even though the

participants saw the T1 faces again three months later at T2, interference from the initial association was unlikely to be strong..

Outlook

A strength of the present study is its high external validity. We examined how neural activation during the encoding of knowledge-relevant associative episodes changed as a function of increasing domain knowledge. Both the domain knowledge and its increase were not engineered in the laboratory, but resulted from future doctors preparing for their final medical exams. Thus, learning intensity and knowledge acquisition was beyond the scope of a laboratory experiment, but could be monitored closely via a web-based learning platform. In particular, performance on the learning platform was highly predictive of later exam success. The associative memory task consisted of encoding and later remembering face–name and face–diagnosis associations, and bore some resemblance to real-world memory situations that doctors encounter in their professional lives. The diagnoses were selected by medical professionals and carefully piloted to capture the purported knowledge gains while preparing for the final medical exam. We observed that participants with greater knowledge gains showed greater decreases in right HC SME for knowledge-relevant pairs. Thus, we observed, for the first time, an association between changes in knowledge due to intensive learning in a real-world educational setting and changes in brain activation in a laboratory episodic memory paradigm (see van Kesteren et al., 2014, for related cross-sectional findings). This finding extends the literature on the effects of intensive, real-world studying on brain plasticity (Draganski et al., 2006) and connectivity at rest (Mackey et al., 2013) in that it uncovers changes in brain activation and connectivity in a transfer task and relates those to gains in content knowledge. Increments in content knowledge across three months of intensive study correlated with HC SME decrements for knowledge-relevant material. The knowledge gains were assessed using a web-based learning platform on which the participants

studied for their final exam. We believe that combining laboratory-based cognitive neuroscience research with real-world educational settings and relating the two, in our case via educational technology, holds great promise for the study of memory in itself as well as for bridging the proclaimed gap between cognitive neuroscience and education (Ansari & Coch, 2006; Blakemore & Bunge, 2012). We suggest that future neuroscience research would profit from making greater use of knowledge acquisition in real-world contexts, such as schooling, vocational training, and the workplace, to better understand the neural pathways, areas, and mechanisms through which knowledge affects memory for new information in individuals of different ages.

Funding

This work was funded by the Max Planck Society, a Heinz Maier Leibnitz prize awarded to YLS by the German Research Foundation, and a Gottfried Wilhelm Leibniz prize to UL by the German Research Foundation. Garvin Brod was supported by a doctoral fellowship of the International Max Planck Research School on the Life Course (LIFE; www.imprs-life.mpg.de), and a scholarship of the German Academic Exchange Service (DAAD).

Acknowledgments

This study was conducted within the “Cognitive and Neural Dynamics of Memory across the Lifespan” (CONMEM) project at the Center for Lifespan Psychology, Max Planck Institute for Human Development, in partial fulfilment of the doctoral dissertation of GB. GB was supported by a fellowship of the International Max Planck Research School on the Life Course (LIFE; **www.imprs-life.mpg.de**). Parts of the analyses were performed during a research visit of GB at the Stanford Memory Laboratory, Department of Psychology, Stanford University. We thank Torben Brod, Friederike Fenski, Thomas Grandy, Andreas Horn, Daniel Kutzer, and Rebekka Rust for their help in choosing the medical diagnoses, and Silvia Bunge, Thackery Brown, Attila Keresztes, Markus Werkle-Bergner, and Marlieke van Kesteren for valuable discussions.

References

Alba JW, Hasher L (1983) Is memory schematic? Psychol Bull 93:203–231.

Badre D, Wagner AD (2007) Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 45:2883–2901.

Bartlett FC (1932) Remembering: An experimental and social study. Cambridge, UK: Cambridge University Press.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex 19:2767–2796.

Bransford J, Johnson M (1972) Contextual prerequisites for understanding: Some investigators of comprehension and recall. J Verbal Learning Verbal Behav 11:717–726.

Brewer JB, Zhao Z, Desmond JE, Glover GH, Gabrieli JDE (1998) Making memories: brain activity that predicts how well visual experience will be remembered. Science (80-) 281:1185–1187.

Brod G, Lindenberger U, Werkle-Bergner M, Shing YL (2015) Differences in the neural signature of remembering schema-congruent and schema-incongruent events. Neuroimage 117:358–366.

Brod G, Werkle-Bergner M, Shing Y (2013) The influence of prior knowledge on memory: a developmental cognitive neuroscience perspective. Front Behav Neurosci 7:139.

Büchel C, Coull JT, Friston KJ (1999) The predictive value of changes in effective connectivity for human learning. Science (80-) 283:1538–1541.

Cohen J, Cohen P, West SG, Aiken LS (2003) Applied multiple correlation/regression analysis for the behavioral sciences. Routledge.

Davachi L (2006) Item, context and relational episodic encoding in humans. Curr Opin Neurobiol 16:693–700.

Draganski B, Gaser C, Kempermann G, Kuhn HG, Winkler J, Büchel C, May A (2006) Temporal and spatial dynamics of brain structure changes during extensive learning. J Neuroscience 26:6314–6317.

Gagnepain P, Henson R, Chételat G, Desgranges B, Lebreton K, Eustache F (2011) Is neocortical–hippocampal connectivity a better predictor of subsequent recollection than local increases in hippocampal activity? New insights on the role of priming. J Cogn Neurosci *23*:391-403.

Ghosh VE, Gilboa A (2014) What is a memory schema? A historical perspective on current neuroscience literature. Neuropsychologia 53:104–114.

Ghosh VE, Moscovitch M, Melo Colella B, Gilboa A (2014) Schema Representation in Patients with Ventromedial PFC Lesions. J Neurosci 34:12057–12070.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Kelly AMC, Garavan H (2005) Human functional neuroimaging of brain changes associated with practice. Cereb Cortex 15:1089–1102.

Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics:(de) constructing the N400. Nat Rev Neurosci 9:920–933.

Liang JC, Wagner AD, Preston AR (2013) Content representation in the human medial temporal lobe. Cereb Cortex 23:80–96.

Mackey AP, Singley ATM, Bunge SA (2013). Intensive reasoning training alters patterns of brain connectivity at rest. J Neurosci 33:4796-4803.

McClelland JL (2013) Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. J Exp Psychol Gen 142:1190–1210.

Minear M, Park DC (2004) A lifespan database of adult facial stimuli. Behav Res Methods, Instruments, Comput 36:630–633.

O'Reilly JX, Woolrich MW, BehrensTE, Smith SM, Johansen-Berg H. (2012) Tools of the trade: psychophysiological interactions and functional connectivity. Social cognitive and affective neuroscience 7:604–609.

Paller KA, Wagner AD (2002) Observing the transformation of experience into memory. Trends Cogn Sci 6:93–102.

Piaget J (1952) The origins of intelligence in children. 8.

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. Trends Cogn Sci 17:230–240.

Preston A, Eichenbaum H (2013) Interplay of hippocampus and prefrontal cortex in memory. Curr Biol 23:R764–R773.

Prince SE, Daselaar SM, Cabeza R (2005) Neural correlates of relational memory: Successful encoding and retrieval of semantic and perceptual associations. J Neurosci 25:1203–1210.

R Core Team (2014) R: A Language and Environment for statitical computing.

Ranganath C, Ritchey M (2012) Two cortical systems for memory-guided behaviour. Nat Rev Neurosci 13:713–726.

Schlichting ML, Mumford JA, Preston AR (2015) Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat Commun 6:8151.

Smith S, Jenkinson M, Woolrich M (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23:208–219.

Stanovich KE (1986) Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Read Res Q:360–407.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner P a, Wood ER, Witter MP, Morris RGM (2007) Schemas and memory consolidation. Science 316:76–82.

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris RGM (2011) Schema-dependent gene activation and memory encoding in neocortex. Science 333:891–895.

Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. Front Syst

586 Neurosci 5:1.

587 van Kesteren MTR, Rijpkema M, Ruiter DJ, Fernández G (2010) Retrieval of associative
588 information congruent with prior knowledge is related to increased medial prefrontal
589 activity and connectivity. J Neurosci 30:15888–15894.

590 van Kesteren MTR, Rijpkema M, Ruiter DJ, Morris RGM, Fernández G (2014) Building on
591 prior knowledge: Schema-dependent encoding processes relate to academic performance.
592 J Cogn Neurosci 26:2250–2261.

593 van Kesteren MTR, Ruiter DJ, Fernández G, Henson RN (2012) How schema and novelty
594 augment memory formation. Trends Neurosci 35:211–219.

595 Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale a M, Rosen BR, Buckner
596 RL (1998) Building memories: remembering and forgetting of verbal experiences as
597 predicted by brain activity. Science 281:1188–1191.

598 Wang S-H, Morris RGM (2010) Hippocampal-neocortical interactions in memory formation,
599 consolidation, and reconsolidation. Annu Rev Psychol 61:49–79, C1–C4.

600 Warren DE, Jones SH, Duff MC, Tranel D (2014) False Recall Is Reduced by Damage to the
601 Ventromedial Prefrontal Cortex: Implications for Understanding the Neural Correlates of
602 Schematic Memory. J Neurosci 34:7677–7682.

603 Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for
604 reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3
605 T and 1.5 T. Neuroimage 33:493–504.

606 Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and ventral medial prefrontal
607 activation during retrieval-mediated learning supports novel inference. Neuron 75:168–
608 179.

609 Zimmer H, Mecklinger A, Lindenberger U (2006) Handbook of binding and memory:
610 Perspectives from cognitive neuroscience.

611

Figure Captions

Figure 1 Memory Task. In the MR scanner, participants memorized face–word pairs, which were presented for 5 seconds each. Half of the words were diagnoses (left example) and half were first names (right example). Outside of the scanner, participants were presented with all of the studied faces, together with four first name or four diagnosis options, of which only one had been presented with the face during the encoding phase. The other three were familiar names or diagnoses that had been paired with other faces. The participants' task was to select the option that had been presented with the face during encoding.

Figure 2 Correlation between medical knowledge gains and final exam score; memory performance. Gains in medical knowledge, assessed via the web-based learning platform, correlated with the final exam score (r = .53, p < .001). Gains in associative memory performance were more pronounced for face–diagnosis pairs than for face–name pairs. Standard errors reflect the pooled error term of the within-subjects F statistic.

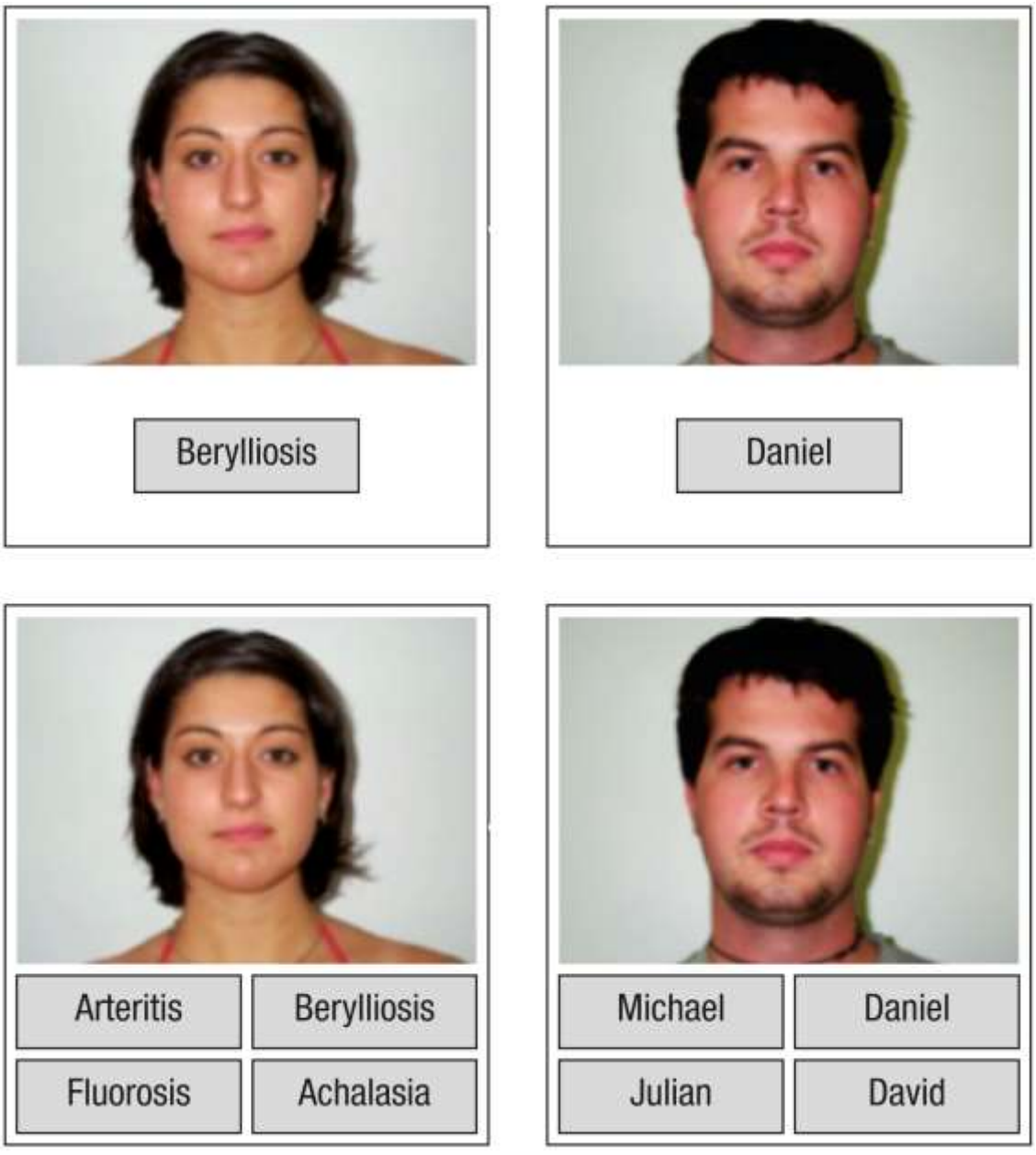
Figure 3 Time x condition interaction for SME in the right anterior HC. a) Reductions in the SME from pretest to posttest were larger for face–diagnosis pairs than for face–word pairs in a cluster of voxels in the right anterior HC (in red, peak voxel: 26, -8, -20; overlaid on the memory x time interaction for face–diagnosis pairs in yellow). b) Percent Signal Change of the time x condition interaction for SME in the right anterior HC. While the difference between subsequently remembered and forgotten events remained comparable across the two time points for the names condition, for the diagnosis condition, there was a significant memory x time interaction. This interaction was driven by the remembered trials that displayed decreased anterior HC activation. Rem = Remembered. Forg = Forgotten. c) Decrements of the SME for face–diagnosis pairs in this cluster of voxels correlated with individual differences in medical knowledge gains (r = .32, p = .04).

Figure 4 Time x condition interaction for SME in functional connectivity (PPI) between the right anterior hippocampus and the left middle temporal gyrus. Using the right anterior hippocampus as a seed, we found a significant time x condition

643 interaction with the left MTG, which was driven by a stronger increase for face–
644 diagnosis pairs as compared to face–name pairs.

645

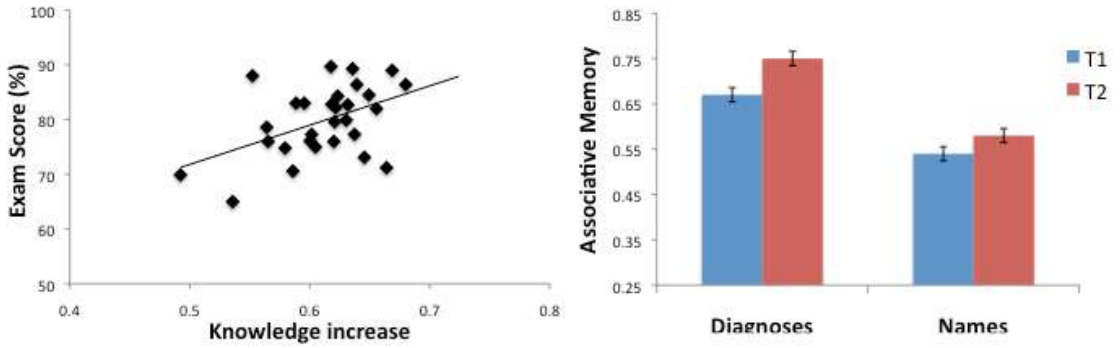
646 Figure 1



647

648

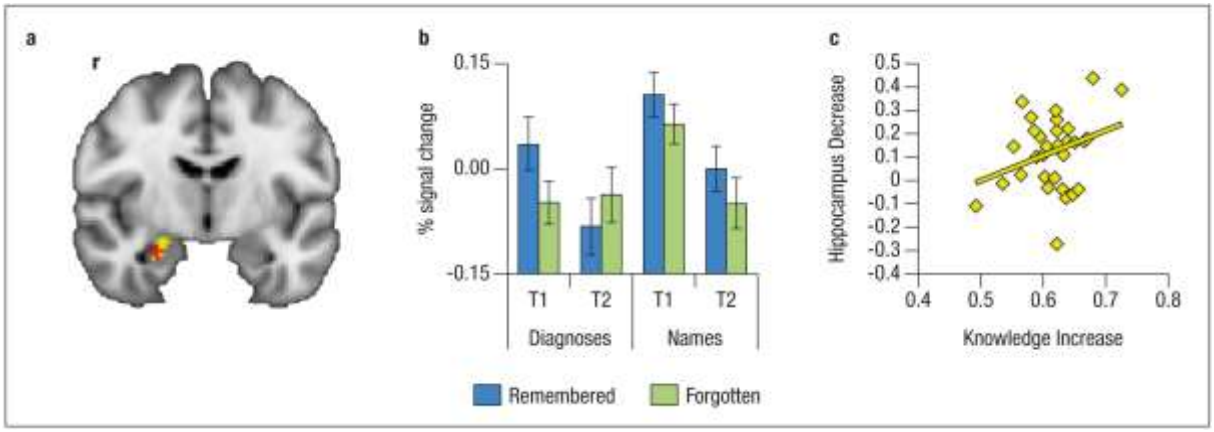
649 Figure 2
650



651

652

653 Figure 3

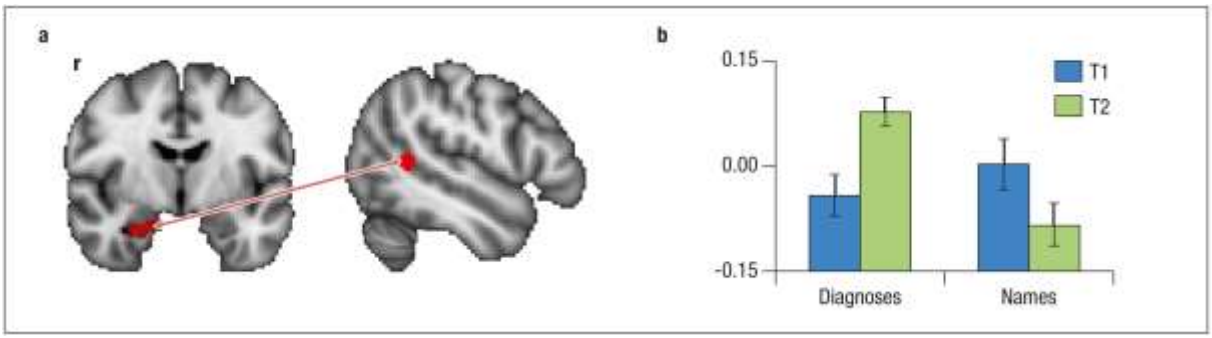


654

655

656

657 Figure 4



658

659

Table 1. Regions exhibiting stronger activation at T1 for later remembered as compared to later forgotten diagnoses or names (Subsequent memory effect, upper part). Regions exhibiting a stronger subsequent memory effect for diagnoses than for names, and vice versa, at T1 (Condition x Memory interaction). Voxel threshold: $z > 2.3$, cluster threshold: $p < .05$. rem = remembered, forg = forgotten.

Region	x	y	z	# voxels	Z-Max
Subsequent Memory Effect (Rem > Forg) at T1					
Right Lateral Occipital Cortex	34	-88	18	3030	4.25
Left Inferior Temporal Gyrus	-40	-54	-18	2116	4.69
Left Inferior Frontal Gyrus / Frontal Pole	-42	42	-6	1662	3.91
Bilateral Superior Frontal Gyrus	-2	52	44	1381	4.69
Right Hippocampus / Amygdala	20	-6	-14	1010	4.27
Left Hippocampus / Amygdala	-16	-4	-12	509	3.88
Bilateral ventromedial PFC	-4	48	-14	401	3.73
Left Lateral Occipital Cortex	-48	-70	36	310	3.55
Right Inferior Frontal Gyrus	56	34	12	256	3.60
Right Precentral Gyrus	48	8	30	221	3.79
Left Superior Frontal Gyrus	-6	20	54	187	3.16
Diagnosis (Rem > Forg) > Name (Rem > Forg) at T1					
Right Angular Gyrus	54	-48	28	246	3.37
Frontal Pole	32	64	6	220	3.31
Right Paracingulate Gyrus	8	48	26	169	2.90
Bilateral Precuneus	-2	-58	62	158	3.13
Right Middle Temporal Gyrus	70	-28	-8	148	3.36
Name (Rem > Forg) > Diagnosis (Rem > Forg) at T1					
Left Lateral Occipital Cortex	-42	-78	-6	382	3.28
Left Precentral Gyrus	-46	-4	46	227	3.63
Right Lateral Occipital Cortex	38	-58	-6	224	3.78

Table 2. Regions exhibiting increases in functional connectivity with the left and right anterior hippocampus, respectively, for subsequently remembered vs. forgotten episodes (i.e., SME in connectivity). Voxel threshold: $z > 2.3$, cluster threshold: $p < .05$. rem = remembered, forg = forgotten.

Region	x	y	z	# voxels	Z-Max
Seed: Right hippocampus					
Face–diagnosis condition: memory (rem>forg) x time (T2>T1)					
Left Middle / Inferior Frontal Gyrus	-38	40	30	1498	4.36
Left Insular Cortex	-34	-16	14	1004	3.83
Right Superior Frontal Gyrus	2	54	40	772	3.56
Right Frontal Pole / Medial Prefrontal Cortex	34	38	-6	695	4.16
Right Lateral Occipital Cortex	58	-62	6	478	3.7
Left Middle Temporal Gyrus	-60	-48	8	440	4.09
Right Lingual Gyrus	14	-66	2	352	3.51
Right Postcentral Gyrus	16	-34	74	181	3.35
Right Temporal Pole	50	12	-12	160	3.52
Right Putamen	28	10	-6	158	3.45
Left Postcentral Gyrus	-10	-36	78	149	3.42
Face–name condition: memory (rem>forg) x time (T2>T1)					
Left Frontal Pole / Medial Prefrontal Cortex	-6	56	6	1999	4.13
Left Postcentral Gyrus	-60	-22	20	215	3.8
Left Postcentral Gyrus	-50	-12	28	191	3.58
Right Central Opercular Cortex	60	4	2	166	3.45
Left Precuneus	-6	-58	42	150	3.41
Right Precentral Gyrus	64	6	26	148	3.67
Seed: Left Hippocampus					
Face–diagnosis condition: memory (rem>forg) x time (T2>T1)					
Left Frontal Pole / Inferior Frontal Gyrus	-38	44	-4	1975	4.13
Bilateral Superior / Medial Prefrontal Cortex	-2	58	26	357	3.28
Medial Prefrontal Cortex	-10	38	-10	354	3.25
Left Middle Temporal Gyrus	-58	-48	8	317	3.53

Right Frontal Pole	24	56	28	179	3.19
Left Posterior Hippocampus	-32	-32	-6	147	3.66
Face–name condition: memory (rem>forg) x time (T2>T1)					
Left Frontal Pole / Medial Prefrontal Cortex	-8	50	-6	1190	3.82
Left Precuneus	-8	-60	42	299	3.56
Right Central Opercular Cortex	58	-8	12	286	3.43
Right Frontal Pole	8	54	32	203	3.14
Left Frontal Pole	-36	-48	18	150	3.28

674