Testing Promotes Long-Term Learning via Stabilizing Activation Patterns in a Large Network of

Brain Areas

Attila Keresztes[1], Daniel Kaiser[2], Gyula Kovács[1,2,3,4], Mihály Racsmány[1,4]

[1] Department of Cognitive Science, Budapest University of Technology and Economics,

Budapest, Hungary

[2] Institute of Psychology, University of Regensburg, Regensburg, Germany

[3] DFG Research Unit Person Perception, Friedrich-Schiller-University of Jena, Jena, Germany

[4] These authors contributed equally to this study

Corresponding Authors: Gyula Kovács, Mihály Racsmány

Gyula Kovács: Inst. Psychol. Univ. Jena, Leutragraben 1 31, 07743, Jena, Germany, Tel: 0049

3641945936, Fax: 00493641945182, email: gkovacs@cogsci.bme.hu

Mihály Racsmány: Budapest University of Technology and Economics, Egry József u.1.,

Budapest, Hungary, Tel: 00364631269, 00364631072, email: racsmany@cogsci.bme.hu

1

Abstract

The testing effect refers to the phenomenon that repeated retrieval of memories promotes better long-term retention than repeated study. To investigate the neural correlates of the testing effect we used event-related functional magnetic resonance imaging methods while participants performed a cued recall task. Prior to the neuroimaging experiment, participants learned Swahili-German word pairs, then half of the word pairs were repeatedly studied, whereas the other half were repeatedly tested. For half of the participants, the neuroimaging experiment was performed immediately after the learning phase; a one-week retention interval was inserted for the other half of the participants. We found that a large network of areas identified in a separate 2-back functional localizer scan were active during the final recall of the word pair associations. Importantly, the learning strategy (retest or restudy) of the word pairs determined the manner in which the retention interval affected the activations within this network. Recall of previously restudied memories was accompanied by reduced activation within this network at long retention intervals, but no reduction was observed for previously retested memories. We suggest that retrieval promotes learning via stabilizing cue related activation patterns in a network of areas usually associated with cognitive and attentional control functions.

Understanding the neural basis of how we lose access to previously encoded knowledge is a fundamental question of cognitive science as well as the psychology of learning and education. Since the seminal work of Ebbinghaus (Ebbinghaus 1885/1964) the effect of the retention interval on forgetting has been one of the central topics of memory research. Several factors have been identified that could potentially explain aspects of the strong connection between retention interval and forgetting. Two such factors are the negative effect of acquiring new information after encoding the target event and the effect of sleep on memory consolidation (Roediger et al. 2010). Although some core processes of forgetting - such as the failure of memory consolidation and the consequences of interference resolution from competing irrelevant memories during retrieval - have already been identified (Uncapher and Wagner 2009; Wimber et al. 2009; Levy et al. 2010), our knowledge of the neural mechanisms of long-term forgetting is far from comprehensive. Hence, it is not surprising that some of the most remarkable experimental results regarding forgetting are those that demonstrated that even a single factor (an additional retrieval after memory encoding) can significantly reduce the negative influence of retention interval on recall performance (Spitzer 1939; Tulving 1967; Carrier and Pashler 1992; Roediger and Karpicke 2006a).

The finding that additional retrieval practice promotes better long-term retention and a slower forgetting rate than the simple restudy of the same information has been termed the *testing effect,* an effect that is currently attracting considerable attention (Roediger and Butler 2011). This phenomenon contradicts what is typically thought about successful learning and is also in conflict with general educational practice, in which testing is only the checkpoint of consecutive study phases (Roediger and Karpicke 2006b).

3

Furthermore, recent experiments have demonstrated that the rate of forgetting is influenced by learning strategy. Although retesting had no mnemonic advantage over restudying at short retention intervals, it produced significantly higher learning performance than an equal amount of restudying when the retention interval was longer than one day (Wheeler et al. 2003; Karpicke and Roediger 2008; Toppino and Cohen 2009). These results suggest that the efficiency of testing over restudying has a positive correlation with the length of retention interval. Although this interaction between learning strategy and retention interval seems to be an important aspect of human learning, the responsible functional neural networks have not yet been identified.

As a first step in seeking for the neural correlates of the *testing effect*, we investigated areas of the human brain that are known to be involved in cue-driven episodic retrieval (ER) processes. In previous experiments, ER was typically studied with associative cued-recall and recognition tasks (Rugg and Henson 2002). These experiments demonstrated that successful memory retrievals are associated with activations in a large cortical network, including the prefrontal (PFC), posterior parietal (PPL) and medial temporal cortices (MTL), and hippocampus (Fletcher and Henson 2001; Rugg 2004; Spaniol et al. 2009; Kim 2011). Importantly, this retrieval-related network has a striking overlap with the network activated by working memory (WM) tasks (Cabeza et al. 2002). This result corresponds to WM theories that assume that WM activation is crucial for enhancing the efficiency of retrieval cues in guiding memory search (Bunting 2006; Unsworth and Engle 2006, 2007). Interestingly, two recent neuroimaging studies (Kuhl et al. 2007; Eriksson et al. 2011) demonstrated that when compared to a single retrieval, repeated retrieval practice leads to a reduced activation of a large portion of these regions, including the bilateral ventrolateral PFC, inferior frontal cortices (BA 9/44), the right DLPFC

4

(BA 45/46), the left precuneus (BA 39) and the bilateral superior parietal lobule (BA 7). These results were considered to be evidence that repeated testing reduces cognitive control demands during future episodic retrieval by making the cue-target link easier to process (Kuhl et al. 2007). Furthermore, as Karpicke (2012) pointed out, each time a person retrieves a piece of information from memory, the future accessibility of this information improves because retrieving enhances the effectiveness of the specific retrieval cue in reconstructing all associated memories. According the account of Karpicke and colleagues (2012), this effectiveness is driven by a mechanism that by each retrieval act refines the search set and renders it smaller. This in turn may reduce the demand on WM to accomplish successful retrieval (Karpicke and Blunt 2011; Karpicke 2012). Altogether, these findings indicate the possible role of a network of areas related to WM in producing the long-term advantage of testing.

The aim of the current study was to investigate the role of cortical areas related to updating informations in working memory, attentional control and controlled retrieval in the testing effect. We predicted that retrieval during the test phase promotes long-term memory advantages via efficient retrieval cue processing. Furthermore, we assumed that following repeated successful retrieval attempts, a given retrieval cue can efficiently activate WM and cognitive control related networks even after long retention intervals. This would be beneficial for all future associative search processes, leading to the positive effect of retrieval (i.e. the testing effect). In contrast, without an initial retrieval attempt during learning, processing of retrieval cues may load heavily on control processes during tests following short retention intervals, and might not be effective following longer retention intervals. Thus, we compared the neural correlates of the associative recall of memories learned with two different learning strategies (retesting vs. restudying) after either a short or a long retention interval.

5

Materials and methods

Participants

Twenty-nine healthy participants (2 left handed, 20 females, mean ± SD age: 22.93 ± 2.26 years) were recruited at the University of Regensburg. All participants were native German speakers and gave informed written consent to participate in the study, which was approved by the ethics committee of the University of Regensburg. None of the participants had any history of neurological diseases, and all had normal or corrected-to-normal visual acuity. We excluded three participants from the final analysis: for one person, fMRI data acquisition failed, and the other two participants did not follow instructions.

Stimuli and Design

The stimuli were 60 Swahili-German word pairs translated from the Swahili-English normalized data published by Nelson and Dunlosky (1994). We used word pairs with moderate recall probabilities according to the Nelson and Dunlosky (1994) normalized data. Thirty word pairs were randomly assigned to both the retest and the restudy conditions (see below).

Procedure

The full experiment was run in two parts. In the first part participants completed an initial learning phase (learning Swahili-German word pairs). In the second part, participants were scanned in three sessions: First, they completed a final test for the material studied during the initial learning phase, second, they were asked to lay still and relax while a structural scan was performed, and third, they performed an n-back task. After these scanning sessions, the second part of the experiment ended with an off-scan test for all the material studied during the initial learning phase.

In the initial learning phase, participants learned the Swahili-German word pairs alone in a quiet room, seated in front of a computer screen (80 Hz, 1280 x 1024 resolution, viewing distance: 65 cm). First, participants were presented with all 60 Swahili-German word pairs subsequently. Each pair was presented randomly for 5000 ms in the center of the screen with the Swahili word on the left and its German meaning on the right. Participants were instructed to memorize all of the pairs for the later test-phase. They were also told that they would see the Swahili word during later testing and be asked to recall its German meaning. Next, participants learned the 60 word pairs through six learning cycles; each cycle included one retest, one restudy, and one feedback block. Unknown to the participants, half of the word pairs were assigned to the retest strategy condition and half to the restudy strategy condition. The retest-restudy words varied randomly across participants. In the retest blocks, all 30 word pairs assigned to the retest condition were tested once, in random order. During a trial, the Swahili member of the word pair appeared on the left side of the screen, and participants were instructed to recall and type in the German meaning in a box that appeared on the right side of the screen. Participants had 8000 ms to accomplish the task. In the restudy blocks, all 30 word pairs assigned to the restudy condition were presented randomly, each for 5000 ms, with the Swahili word on the left and its German meaning on the right. In each feedback block, all 60 word pairs were presented again, each for 1500 ms. These feedback blocks served to enhance the effect of testing (Roediger and Butler 2011). In each learning cycle, the order of the retest and restudy blocks was random, and each cycle ended with a feedback block.

Next, half of the participants (n=15) were assigned to the short retention interval group, while the other half (n=14) to the long retention interval group. As noted above, three participants' data were excluded from the analyses, leaving n = 13 in both groups. In the short

7

retention interval group, the second part of the experiment (final test of the Swahili-German words in the fMRI scanner) was performed right after the learning phase (on average, there was a 20 min interval between the end of the learning phase and the beginning of the scanning). In the long retention interval group, this final test and the scanning were performed exactly one week after the learning phase. In order to avoid self-testing during the retention interval, all participants were told that the fMRI part of the experiment would examine social cognition and that it would be unrelated to the 'memory experiment' they had just performed. In both cases, participants were informed about the security issues of the scanning procedure prior to the final test. In the scanner, stimuli were back-projected via an LCD video projector (JVC, DLA-G20, Yokohama, Japan, 72 Hz, 800 x 600 resolution) onto a translucent circular screen (diameter= 30 degree), placed inside the scanner bore 63 cm from the observer. Stimulus presentation was controlled via Presentation (Version 14.1 Build 09.21.09).

The final test phase consisted of cued recall trials (which were similar to the trials of the retest block during the learning phase) intermixed with fixation trials. Each of the 60 word pairs was tested once. In each trial, the Swahili word appeared in the middle of the screen, and participants were instructed to silently recall its German meaning. Participants were told to press a response button if they knew the answer, but to refrain from saying the word out loud. Each trial lasted 10 sec, irrespective of whether the participant responded. Each cued recall trial was preceded by fixation trials (1000, 3000 or 5000 ms) that were used to jitter the cue onset during the test phase. The three types of fixation trials appeared equally often and were randomized in order. Participants were told to press the response button as quickly as possible because we were interested in observing how fast they could remember the word. To measure their correct recollection rate, we specifically instructed them that they should press the response button only

if they would be able to report the German word at a follow-up test immediately after scanning in the laboratory. Participants had a 30 sec rest period after the thirtieth cued recall trial. During the follow-up test right after the scanning sessions, participants were asked to recall the remembered words. In all further analyses, we considered a word pair to be remembered only if the participant signaled during scanning that (s)he remembered it and if (s)he could report the answer correctly in the follow-up test. Incorrect trials (i.e. trials in which the participant had responded that they had known the response, but could not report the correct target at the follow-up test) were dismissed from further analyses.

Scanning parameters and data acquisition

Imaging was performed using a 3-Tesla MR head scanner (Siemens Allegra, Erlangen, Germany). For the functional series, we continuously acquired images (34 slices, 20 deg tilted relative to axial; T2* weighted EPI sequence, TR =2000 ms, TE = 30 ms; flip angle = 90 deg; 64 x 64 matrices; in-plane resolution: 3x3 mm; slice thickness: 3mm, 10% gap). High-resolution sagittal T1-weighted images were acquired using a magnetization-prepared rapid gradient-echo sequence (MP-RAGE; TR = 2250 ms; TE = 2.6 ms; 1 mm isotropic voxel size) to obtain a 3D structural scan. Details of preprocessing and statistical analysis are given elsewhere (Kovács et al. 2008; Cziraki et al. 2010; Kovács et al. 2012). Briefly, the functional images were corrected for acquisition delay, realigned, normalized to the MNI-152 space, resampled to 2 x 2 x 2 mm resolution and spatially smoothed with a Gaussian kernel of 8 mm FWHM (SPM8, Welcome Department of Imaging Neuroscience, London, UK).

Region of interest (ROI) analysis was based on the results of separate functional localizer runs which were 5 runs of the following 2 x 30 sec blocks: a 30 sec epoch of letters (700 ms exposition time + 300 ms blank for each letter) preceded by an instruction to "respond if the

current letter is the same as the one presented two letters previously (2-back)", followed by a 4 sec blank period and another 30 sec period of letters (700 ms exposition time + 300 ms blank for each letter) preceded by the instruction screen "respond if the current letter is a 'D' (detect a D)". This functional localizer was similar to the one used for localizing the cortical network activated by a 2-back task in Drobyshevsky, Baumann, and Schneider (Drobyshevsky et al. 2006). The data were analyzed using the MARSBAR 0.42 toolbox for SPM (Brett et al. 2002).

----------------Fig 1 approximately here--------------

The locations of ROI areas were determined individually as areas responding more strongly during the 2-back task than during the detection task in the functional localizer scans ($p_{uncorrected} < 10^{-6}$; T=4.86, df=273). The coordinates of the areas are presented in Table 1. The ROIs were selected individually on the single subject level from the thresholded T-maps. Areas lying closest to the corresponding reference cluster (based on the results of the previous literature and the results of the random-effects analysis for differential contrasts; $p_{uncorrected} < 10^{-3}$; T=3.12, df = 241) were considered as their appropriate equivalents at the single subject level. A time series of the mean voxel values within an 8 mm radius sphere around the local maximum of the areas of interest was calculated and extracted from our event-related sessions using finite impulse response (FIR) models (Ollinger et al. 2001). The convolution of a reference hemodynamic response function (HRF) with boxcars (which represented the onsets and durations of the experimental conditions) was used to define the regressors for a general linear model.

----------------Table 1 approximately here--------------

Data analysis

10

We performed a two-way mixed design ANOVA on final recall accuracy and final recall RTs with strategy (2, retest, restudy) as the within-subject factor and retention interval (2, short, long) as the between-subject factor. As for the BOLD signal, trials were analyzed and separately modeled at the onset of the stimuli (duration=10 sec). The peak of the event-related averages at 6-8 sec post-stimulus onset was used as an estimate of the response magnitude and averaged across repetitions for each condition and participant separately. We performed three-way mixed design ANOVAs on the peaks with strategy (2, retests, restudy) and success (2, remembered, forgotten) as the within-subject factors and retention interval (2, short, long) as the between-subject factor.

Results

Behavior results

Participants learned on average 75% of Swahili-German associations until the end of the initial learning phase. Recall success (in percentages) for retest items increased from cycle one to cycle six, (M = .15, SE = .02 in cycle 1, M = .29, SE = .03 in cycle 2, M = .42, SE = .04 in cycle 3, M = .57, SE = .04 in cycle 4, M = .67, SE = .04 in cycle 5, and M = .75, SE = .04 in cycle 6). Performance in the short vs. long retention interval groups did not differ in any of the learning cycles (all $ps > .33$).

The upper panel of Fig. 2 presents the performances of the participants at the final test, expressed as the proportion of correctly recalled words for the retest and restudy strategy conditions and for the short and long retention interval groups, separately. As expected, retention interval had a significant main effect on the final recall accuracy ($F(1,24) = 14.26$, $p < 0.001$): participants' overall recall accuracies were lower after a one-week retention interval (M = 44.74%, SE = 4.56%) than after a 20-minute retention interval (M = 69.1%, SE = 4.56%).

Although strategy had no main effect on recall accuracy, we observed a significant interaction between strategy and retention interval (F(1,24) = 5.80, p = 0.024). Post-hoc tests demonstrated that this result arose because the recall accuracies of the retest condition were significantly higher (M = 50.26%, SE = 6.93%) than those of the restudy condition (M = 39.23%, SE = 3.25%) in the long retention interval condition (t(12) = 2.33, p = 0.038). However, there were no differences in the short retention interval condition, (t(12) = 0.92, ns, M = 67.44%, SE = 4.55% and M = 70.77%, SE = 4.66% for retest and restudy, respectively). This result confirms previous findings in which repeated retrieval lead to better long-term retention than additional study, even though the two conditions produce similar performances on short intervals (Roediger and Karpicke 2006b).

----------------Figure 2 approximately here--------------

Analysis of RTs (Fig 2, lower panel) revealed a significant main effect of strategy (F(1,24) = 8.93, p = 0.006), that was due to shorter recall RTs overall in the retest condition (M = 2411 ms, SE = 148 ms) compared to the restudy condition (M = 2859 ms, SE = 149 ms). Retention interval also had a main effect with shorter RTs in the short retention interval group (M = 2249 ms, SE = 182 ms) than the long retention interval group (M = 3021 ms, SE = 182 ms). In contrast to the ANOVA on final recall accuracy, the ANOVA on RTs did not reveal any significant interaction between strategy and retention interval.

----------------Table 2 approximately here--------------

fMRI results

Interaction of learning strategy and retention interval. The main aim of the current study was to determine whether there are cortical areas which show activation patterns that reflect the interaction of learning strategy and retention interval of the task, similarly to previous behavioral

results (Roediger and Karpicke 2006a; Karpicke and Roediger 2008). To this end, we performed a three-way mixed design ANOVA on the extracted BOLD signals. We reasoned that if an area is related to the superior performance observed after repeated retrieval and long retention periods, then the activity of that area should show a significant interaction of learning strategy and retention interval. Table 2 presents main effects and interactions for each area separately. A number of ROIs demonstrated this type of interaction. Fig. 3 presents the average (+-SE) BOLD signal as a function of time for four representative areas as well as the extracted peak activations for all areas with significant interactions. As can be observed in the HRFs, the basis of the interaction between learning strategy and retention interval was that activations in the restudy condition were higher when compared to those in the retest condition after short retention interval, but the opposite effect was observed after long retention interval: retest activations exceeded those of the restudy condition. Post-hoc t-tests (see Supplementary material, Table S1.) showed that from short to long retention interval, activation did not decrease significantly for retested items in any of the ROIs (all ps > .33), and only one region, the right thalamus showed a significant increase (p < .026, all other ps > .18). In contrast, for restudied items, all areas showed a nominal decrease of activation, that was significant in several areas, including the left insula, the anterior cingulate bilaterally, the left anterior prefrontal, the right middle orbitofrontal, and the right superior parietal cortex.

This finding suggests that, when compared to repeated study, repeated retrieval leads to higher activations in a network of areas activated during a WM task after long retention intervals, which, in turn, leads to superior memory performance. Thus, the activity of these areas could serve as the functional basis of the behaviorally observed testing effect.

----------------Figure 3 approximately here--------------

13

Interaction of learning strategy, retention interval and retrieval success. Interestingly, a subset of the areas (anterior PFC (BA10) bilaterally, the left insula, the left ACC, the right inferior parietal area (BA40), the right thalamus, and fusiform gyrus, bilaterally), the activities of which were modulated by strategy and retention interval also showed modulation according to the success of retrieval. This modulation was manifested in the significant three-way interaction between strategy, retention interval, and success. Fig. 4a shows examples of the HRF for two such areas, the left insula (upper panel) and the left ACC (lower panel), while Fig. 4b shows the extracted peak activations for all other areas with significant three-way interactions. At short retention intervals previously restudied items elicit larger activations than previously retested items in these areas, irrespective of final recall success. At long retention intervals, however, the result is different: previously restudied items elicit smaller activations than previously retested items, although this appears to be driven by the nearly complete absence of BOLD signal change for the previously restudied *and* forgotten items.

----------------Figure 4 approximately here--------------

Areas related to retrieval success. The main effect of retrieval success was significant in right frontal areas (right DLPFC, right posterior/dorsal PFC and the right anterior PFC), left ACC, left insula, inferior parietal ROIs bilaterally, thalamus bilaterally, a midbrain area, and the left fusiform gyrus. In addition, we found a tendency for a main effect of success in superior parietal ROIs bilaterally ($p=0.06$ and $p=0.05$ for the left and right hemispheres, respectively). Because retrieval success interacted with either strategy or strategy and retention interval in several areas, we ran post-hoc paired samples *t*-tests (see Supplementary material, Table S2.) separately for the retest and restudy items in both the short and long retention interval conditions. This analysis revealed that only a few areas showed an effect of success at short retention

interval: the left fusiform gyrus for retested items only, the right thalamus for both type of items, and the midbrain for restudied items only. At long retention intervals, however, the effect of success was significant in all but two of the above ROIs for restudied items (no significant effect was found in the left fusiform gyrus, and the midbrain ROI). In contrast, for retested items, only the right thalamus showed a significant effect. Briefly, most ROIs were activated differently during successful vs. unsuccessful retrieval attempts of the restudied items and mainly at long retention intervals. This effect contributed to the main effect of retrieval success.

To test whether any additional areas showed differential activation for the retest and restudy strategies, we performed a whole-brain analysis as well. This analysis revealed no significant activations at the p_{FWE}<0.05 level at an extended threshold of 50 voxels in the short or long retention interval groups, neither for the retest > restudy nor for the restudy > retest contrasts. Similarly we observed no significant activations for the interaction of retention interval and strategy. To further explore our data, we ran the same analyses at the more liberal $p_{uncorrected}$<0.0001 level (with an extended threshold of 50 voxels) as well. At the short retention interval, again, no significant activations were found for either the retest > restudy or the restudy > retest contrasts. In contrast, as shown in Figure 5A, at the long retention interval, the retest vs. restudy contrast revealed significant activations in a medial frontal/anterior cingulate area (8,38,10) and in an area in the occipital lobe at around the early visual cortex (2,-92,2). Importantly, the interaction of retention interval and strategy, as shown in Fig 5B revealed two clusters of activations bilaterally over the inferior frontal gyrus (30,28,-2; k=61; Z=4.89 and -32,26,-2; k=124; Z=4.23), corresponding to the bilateral insular cortices in our ROI analyses. The restudy vs. retest contrast did not reveal any significant activation at the long retention interval either.

As the comparison of data presented in Table 2 and in Figure 5 indicates, both the voxelwise and the region of interest based approach provide evidence showing that activations in the insular and the cingulate cortices are modulated by the interaction of strategy and retention interval. In addition, the results of the whole-brain analysis revealed only one additional area, the early visual cortex, which has been previously suggested to play a role in both working-memory and episodic memory retrieval related tasks (Cabeza et al., 2003; Kim, 2011).

----------------Figure 5 approximately here--------------

Finally, to check the specificity of the results to the previously described areas we applied the same ROI analyses as above to additional areas, using ROIs defined by the complementary contrast of the functional localizer scan (detection > 2-back). This contrast, showing areas that are more active during the cognitively less loaded task, activated a network of areas, very similar to the recently described default network (Buckner, Andrews-Hanna, Schacter, 2008; Gusnard, Raichle, 2001; Mazoyer, Zago, Mellet, Bricogne, Etard, Houdé, Crivello et al., 2001; Shulman, Fiez, Corbetta, Buckner, Miezin, Raichle, Petersen 1997) and included the medial posterior cingulate (BA 30/31, x:-3, y:-51, z:28), the orbito-frontal (BA 10/11; -1, 55, -9) and the superior frontal gyrus (BA,9/10; -5,62, 14). The ROI analyses of these three areas, did not show any significant main effect of strategy or delay, nor any interactions (all ps > .15), supporting further the specificity of our results to areas related to cognitive and attentional control functions.

Discussion

The major findings of our study are the following. [1] Parietal and frontal areas, as well as the thalamus, the left fusiform gyrus, and a midbrain area were activated when participants had to recall previously learned memory items. The same areas were also activated during active updating and manipulating of information in WM during a 2-back task. [2] In most ROIs

identified by the functional localizer 2-back task, the learning strategy of participants determined how the retention interval affected activations during the final test: repeated study and repeated retrieval of the learning material led to different BOLD signals during final recall after short and long retention intervals. In addition, the effect of learning strategy was different for participants who had to retain the memories for a few minutes vs. for a week. [3] For several ROIs identified by the functional localizer 2-back task, the interaction of learning strategy and retention interval was also influenced by retrieval success. Our results show, for the first time, that the long-term behavioral advantage of repeated retrieval over repeated study is due to the differential activation of a large network involving parietal, frontal and insular cortical areas, as well as the thalamus and the fusiform gyrus.

Memory retrieval activates a network of areas activated during updating and manipulating information in working memory. The anterior and dorsolateral part of the prefrontal cortex, the superior and inferior parietal cortex, the anterior cingulum, the thalami bilaterally, an area in the midbrain, the left fusiform gyrus, and the left insula were activated both during the 2-back localizer task and episodic recall of words. This result supports earlier findings of Cabeza et al. (2002) and Ranganath et al. (2003) who showed that these regions, together with the cerebellum, were activated in both ER and WM tasks. The WM task used in our study involves on-line monitoring, updating, and manipulation of remembered information (Owen et al. 2005), and is therefore assumed to place great demands on a number of key processes within WM. Our findings suggest that participants may have leaned on these cortical areas to effectively process retrieval cues during associative recall. Indeed, theories of ER suggest that WM is necessary for several steps of the recall process, such as the initiation of a search process for a specific target memory or the monitoring of the accessed responses (Fletcher et al. 1998; Henson et al. 1999;

17

Cabeza et al. 2002; Ranganath et al. 2003). Determining whether the currently found activations of areas identified by a 2-back task during the cued recall task are due to any of these steps was beyond the scope of the current study (designed to evaluate the possible effects of repeated retrieval vs. that of repeated study) and requires further neuroimaging studies.

Neural correlates of testing effect: learning strategy affects long-term stability of activations during recall in a network of areas activated during updating and manipulating information in working memory. Secondly and more importantly, our behavioral results confirm the existence of testing effect in an fMRI scanner; a long retention interval produced a lower memory performance for previously restudied items compared to the performance on previously retested items. In addition, the analysis of RTs during final recall revealed that repeated retrieval of memories generally increased the effectiveness of retrieval cues; participants could recall the items faster in the retest condition than items in the restudy condition, irrespective of the length of the retention interval.

Furthermore, the imaging data obtained during final cued recall suggests that repeated retrieval of memories might contribute to the long-term stability of memory traces via the activation of retrieval-related areas whereas repeated study does not modulate these activations. In other words, during ER the activation of a network activated by a WM task is largely influenced by the learning strategy of the participants, which is a possible neural correlate of the *testing effect*. At short retention intervals, there is a significant activation of this network, irrespective of the learning strategy. At long retention intervals, this activation is more pronounced for memories that have been encoded through repeated retrieval compared to memories encoded through repeated study.

The effect of learning strategy depends on retrieval success. Our results indicate that at short retention intervals, retrieval cues activate areas in a network also activated by a WM task, irrespective of retrieval strategy, and more importantly, irrespective of retrieval success. In other words, the BOLD activations, associated with successfully recalled and forgotten words, were similar for both retested and restudied items. Similarly, after a week-long retention interval, these areas were activated for the previously retested memories, irrespective of recall success. However, for the previously restudied items, activation at final recall after a week-long retention interval depended largely on recall success, with virtually no BOLD signal change during retrieval attempts of restudied but forgotten items. This result suggests that at short retention intervals, cues related to the restudied memories activate areas of this network (and to a larger extent than cues related to retested memories). At long retention intervals, however, lower activation of the same areas suggests that the cue processing is not initiated in many trials, which might lead to lower recall accuracy for previously restudied items compared to previously retested items, i.e. the emergence of the testing effect.

Our results show that when a target memory of a cue-target association has been repeatedly retrieved during learning, cue processing will activate an overlapping network related to ER and WM tasks, even after a long retention interval. In contrast, for target memories that have been repeatedly studied, the cues might only activate these overlapping networks when the retention interval is short. Our neuroimaging results suggest that some of the restudied memories cannot be recalled after a week-long retention interval, most likely because of the failures of retrieval-related cue processing.

In interpreting our findings two relevant neuroimaging studies should be mentioned. Erikson and colleagues (2011), investigating the effect of repeated successful retrieval on

changes in brain activity, found that the more times an item had been successfully retrieved during a prescan learning phase the higher the activity level in the ACC and the lower the activity level in the superior parietal and mid-ventrolateral cortex was during a final retrieval phase. According to Erikson and colleagues' (2011) interpretation, decreased activation in the fronto-parietal network reflected reduced demands on cognitive control mechanisms necessary for successful retrieval. In a more recent study, Wiklund-Hörnquist and colleagues (2012) showed that repeated and successful retrieval during scanning was paralleled by decreases in the activity level of brain areas in orbitofrontal, insular as well as medial frontal regions, and the ACC (BA 47, 45, 6, 32, respectively). These results are in line with our present finding showing that, in the short retention interval condition, activity level of fronto-parietal networks was lower following repeated retrieval than following repeated study cycles.

Presently, there is no widely accepted theoretical account of the testing effect. We discuss two possible theories that have been raised in recent discussions. According to the elaborative encoding hypothesis (Carpenter, 2009, 2011) attempts to reconstruct target memories during repeated retrieval produce extra information related to the cues which might mediate retrieval during later tests (Pyc and Rawson, 2010). At long retention intervals, when target memories become harder to be reconstructed from single cues, it is the use of extra cues that would produce the long-term advantage of repeated retrieval over repeated study. In contrast, the search set constraining theoretical framework (Karpicke and Smith, 2012; Karpicke and Zaromb, 2010; Karpicke, 2012) suggests that retrieval prompts a process, probably through effective temporal context reinstatement, which narrows the cue-related search set, and even a single retrieval can decrease the number of potentially retrievable items in response to a specific retrieval cue (Karpicke & Blunt, 2011; Karpicke & Zaromb, 2010; Karpicke, 2012). In this account retrieval

20

is a discrimination process, where the effectiveness of a given cue will be determined by its ability to specify a given memory fragment in the context of many similar and interfering memory features.

The aim of the present study was not to contrast experimentally these two theoretical frameworks. However, the observed interaction between learning strategy and retention interval (with activations in areas activated during a working memory task being higher in the restudy than in the retest condition after short retention interval, and lower after long retention interval) in our study, and results of earlier studies showing that each retrieval act leads to a decrease in fronto-parietal activations that is correlated with memory efficiency (Kuhl et al., 2009, Erikkson et al., 2011; Wiklund-Hörnquist, 2012) provide indirect support to the search set constraining framework.

In addition, the fact that retested memories were recalled with shorter RTs than restudied memories during final recall at both short and long retention intervals, also suggests that repeatedly retrieving memories increased the effectiveness of retrieval cues. One possible interpretation of the fMRI results, together with the pattern of RT findings is the following. At short retention intervals, repeated retrieval of associative memories leads to reduced demands on WM compared to restudying the same memories. This may be due to the fact that the search set and potentially activated features are significantly constrained during repeated testing cycles. According to this idea, a network of areas also related to WM, and cognitive and and attentional control in general (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011), is responsible for calibrating the processing of cues to search long-term memories and delimit the search set to the target items. This result suggests that at short retention intervals, cues related to the restudied memories activate areas of this network (and to a larger extent than cues related to retested

memories), as a direct consequence of the extended search set and larger amount of activated semantic elements following repeated study. At long retention intervals, however, lower activation of the same areas suggests that the cue processing is not initiated in many trials, which might lead to lower recall accuracy for previously restudied items compared to previously retested items, i.e. the emergence of the testing effect. In sum, we suggest that the average RT advantage of the retest condition is the consequence of a smaller search set at short retention intervals, while it is due to the effective and more successful target reconstruction following long retention interval. This interpretation is supported by the fact that the RT advantage was accompanied by higher recall performance only following long retention interval.

In sum, these findings suggest that the retention interval of the first retrieval of a target memory, after learning, will determine the activation of overlapping areas in networks activated in ER and WM tasks. The first retrieval attempt of a cue-target association may trigger cue processing only when the retention interval between initial learning and retrieval is short. In contrast, when the retention interval is long, participants cannot effectively process the cue and a large percentage of retrieval attempts fail. Thus, the *testing effect* may be a consequence of processes that, through each additional retrieval act, conserve the effectiveness of the retrieval cue to access a specific memory. Based on our findings, we suggest that this strengthening arises from an effective and stable response for specific episodic cues in a network of brain areas related to cognitive control functions.

Acknowledgments

23

References

Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using an SPM toolbox. In: Paper presented at: 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan. Neuroimage. 16(2):372-373.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008). The Brain's Default Network: Anatomy, Function, and Relevance to Disease. Ann N Y Acad Sci. 1124: 1–38.

Bunting M. 2006. The role of processing difficulty in the predictive utility of working memory span. Psychon Bull Rev. 13(6):998–1004.

Cabeza R, Dolcos F, Graham R, Nyberg L. 2002. Similarities and differences in the neural correlates of episodic memory retrieval and working memory. Neuroimage. 16(2):317–330.

Cabeza R, Dolcos F, Prince SE, Rice HJ, Weissman DH, & Nyberg L (2003). Attention-related activity during episodic memory retrieval: a cross-function fMRI study. Neuropsychologia, 41(3), 390-399.

Carrier M, Pashler H. 1992. The influence of retrieval on retention. Mem Cognit. 20(6):633–642.

Cziraki C, Greenlee MW, Kovács G. 2010. Neural correlates of high-level adaptation-related aftereffects. J Neurophysiol. 103(3):1410–1417.

Drobyshevsky A, Baumann SB, Schneider W. 2006. A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function. Neuroimage. 31(2):732–744.

Ebbinghaus H. 1885/1964. Memory: A contribution to experimental psychology. New York (NY): Dover.

Eriksson J, Kalpouzos G, Nyberg L. 2011. Rewiring the brain with repeated retrieval: a parametric fMRI study of the testing effect. Neurosci Lett. 505(1):36–40.

24

Fletcher PC, Henson RNA. 2001. Frontal lobes and human memory. Brain. 124(5):849–881.

Fletcher PC, Shallice T, Frith CD, Frackowiak RS, Dolan RJ. (1998). The functional roles of prefrontal cortex in episodic memory. II. Retrieval. Brain. 121(7):1249–1256.

Gusnard DA, Raichle ME (2001). Searching for a baseline: functional imaging and the resting human brain. Nat Rev Neuroscience. 2(10): 685-694.

Henson RNA, Shallice T, Dolan RJ. 1999. Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. Brain. 122(7):1367–1381.

Karpicke JD. 2012. Retrieval-based learning: active retrieval promotes meaningful learning. Curr Dir Psychol Sci. 21(3):157–163.

Karpicke JD, Blunt JR. 2011. Retrieval practice produces more learning than elaborative studying with concept mapping. Science. 331(6018):772-775.

Karpicke JD, Roediger HL. 2008. The critical importance of retrieval for learning. Science. 319(5865):966-968.

Kim H. 2011. Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. Neuroimage. 54(3):2446–2461.

Kovács G, Iffland L, Vidnyánszky Z, Greenlee MW. 2012. Stimulus repetition probability effects on repetition suppression are position invariant for faces. Neuroimage. 60(4):2128–2135.

Kovács G, Raabe M, Greenlee MW. 2008. Neural correlates of visually induced self-motion illusion in depth. Cereb Cortex. 18(8):1779–1787.

Kuhl BA, Dudukovic NM, Kahn I, Wagner AD. 2007. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat Neurosci. 10(7):908–914.

Levy BJ, Kuhl BA, Wagner AD. 2010. The functional neuroimaging of forgetting. In: Della Sala
    S. Forgetting (Current issues in memory). New York (NY): Psychology Press. p 135–
    164.

Mazoyer B, Zago L, Mellet E, Bricogne S, Etard O, Houdé O, Crivello F, Joliot M, Petit L,
    Tzourio-Mazoyer N (2001). Cortical networks for working memory and executive
    functions sustain the conscious resting state in man. Brain Res Bull. 54: 287–298.

Nelson TO, Dunlosky J. 1994. Norms of paired-associate recall during multitrial learning of
    Swahili-English translation equivalents. Memory. 2(3):325–335.

Ollinger JM, Corbetta M, Shulman GL. 2001. Separating Processes within a Trial in Event-
    Related Functional MRI: II. Analysis. Neuroimage. 13(1):218–229.

Owen AM, McMillan KM, Laird AR, Bullmore E. 2005. N-back working memory paradigm: A
    meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp.
    25(1):46–59.

Ranganath C, Johnson MK, D'Esposito M. 2003. Prefrontal activity associated with working
    memory and episodic long-term memory. Neuropsychologia. 41(3):378–389.

Roediger HL, Butler AC. 2011. The critical role of retrieval practice in long-term retention.
    Trends Cogn Sci. 15(1):20–27.

Roediger HL, Karpicke JD. 2006a. Test-enhanced learning. Psychol Sci. 17(3): 249-255.

Roediger HL, Karpicke JD. 2006b. The power of testing memory: Basic research and
    implications for educational practice. Perspect Psychol Sci. 1(3):181-210.

Roediger HL, Weinstein Y, Agarwal PK. 2010. Forgetting: preliminary considerations. In: Della
    Sala S. Forgetting (Current issues in memory). New York (NY): Psychology Press. p 1–
    22.

Rugg MD, Henson RNA. 2002. Episodic memory retrieval: An (event-related) functional

    neuroimaging perspective. In: Parker AE, Wilding EL, Bussey T. The cognitive

    neuroscience of memory encoding and retrieval. Hove (UK): Psychology Press. p 3-37.

Rugg MD. 2004. Retrieval processing in human memory: electrophysiological and fMRI

    evidence. In: Gazzaniga MS. The cognitive neurosciences. 3rd ed. Cambridge (MA):

    MIT Press. p 727-738.

Shulman GL, Fiez JA, Corbetta M, Buckner RL, Miezin FM, Raichle ME, Petersen SE (1997).

    Common blood flow changes across visual tasks: II.: decreases in cerebral cortex. J Cogn

    Neurosci. 9(5):648–963.

Spaniol J, Davidson PSR, Kim ASN, Han H, Moscovitch M, Grady CL. 2009. Event-related

    fMRI studies of episodic encoding and retrieval: meta-analyses using activation

    likelihood estimation. Neuropsychologia. 47(8-9):1765–1779.

Spitzer HF. 1939. Studies in retention. J Educ Psychol. 30(9):641-656.

Toppino TC, Cohen MS. 2009. The testing effect and the retention interval. Exp Psychol

    (formerly Z Exp Psychol). 56(4):252–257.

Tulving E. 1967. The effects of presentation and recall of material in free-recall learning. Journal

    of Verbal Learning and Verbal Behavior. 6(2):175–184.

Uncapher MR, Wagner AD. 2009. Posterior parietal cortex and episodic encoding: insights from

    fMRI subsequent memory effects and dual-attention theory. Neurobiol Learn Mem.

    91(2):139–154.

Unsworth N, Engle RW. 2006. A temporal–contextual retrieval account of complex span: An

    analysis of errors. J Mem Lang. 54(3):346–362.

Unsworth N, Engle RW. 2007. Individual differences in working memory capacity and retrieval: A Cue-dependent search approach. In: Roediger HL. The Foundations of Remembering: Essays in Honor of Henry L. Roediger, III. New York (NY): Psychology Press. p 241–258.

Wheeler M, Ewers M, Buonanno J. 2003. Different rates of forgetting following study versus test trials. Memory. 11(6):571–580.

Wiklund-Hörnqvist C, Karlsson L, Eriksson J, Jonsson B & Nyberg L (2012, May). The neural mechanisms underlying test-enhanced learning: an event-related fMRI study. Talk presented at the EARLI SIG 22: Neuroscience and Education, London, England.

Wimber M, Rutschmann RM, Greenlee MW, Bäuml KH. 2009. Retrieval from episodic memory: Neural mechanisms of interference resolution. J Cogn Neurosci. 21(3):538–549.

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, & Wager TD (2011). Large-scale automated synthesis of human functional neuroimaging data. Nat Methods. 8(8): 665-670.

Table 1.

Average MNI Coordinates of regions of interest identified during the localizer run (2-back task).

	Mean cluster center		
	x	*y*	*z*
DLPFC, left (~BA9/45)	-47 (1)	28 (2)	32 (1)
DLPFC, right (~BA9/45)	42 (2)	37 (2)	32 (2)
Anterior prefrontal, left (~BA10)	-36 (1)	55 (1)	14 (2)
Anterior prefrontal, right (~BA10)	37 (1)	56 (1)	14 (2)
Posterior/dorsal prefrontal, left (~BA6)	-28 (1)	3 (1)	58 (1)
Posterior/dorsal prefrontal, right (~BA6)	34 (1)	5 (1)	60 (1)
Anterior cingulate, left (~BA32)	-11(1)	28(2)	31 (1)
Anterior cingulate, right (~BA9/32)	8 (1)	34 (1)	31 (1)
Middle orbitofrontal, right (~BA11)	30 (1)	54 (2)	-11 (1)
Inferior Parietal, left (~BA7)	-35 (1)	-54 (2)	48 (2)
Inferior Parietal, right (~BA40)	43 (1)	-46 (1)	49 (1)
Superior Parietal, left (~BA7)	-20 (1)	-71 (1)	55 (1)
Superior Parietal, right (~BA7)	28 (1)	-66 (1)	57 (1)
Insula, left	-36 (1)	23 (1)	3 (1)
Insula, right	36 (1)	23 (1)	-1 (1)
Midbrain	2 (1)	-24 (1)	-11 (1)
Thalamus, left	-9 (1)	-13 (1)	11 (1)
Thalamus, right	12 (1)	-11 (1)	10 (1)
Fusiform gyrus, left	-42 (1)	-59 (3)	-17 (1)
Fusiform gyrus, right	39 (2)	-60 (3)	-17 (1)

Note: Standard errors are shown in brackets.

Table 2.

Significant main effects and interactions affecting percent signal change in regions of interest.

Effect	Area	F	df	p
Main effect of success	Inferior Parietal, right (~BA40)	16.82	1,22	0.0001
	Thalamus, right	34.41	1,10	0.0001
	DLPFC, right (~BA9/45)	10.53	1,22	0.004
	Posterior/dorsal prefrontal, right (~BA6)	9.52	1,22	0.005
	Inferior Parietal, left (~BA7)	8.05	1,23	0.009
	Anterior cingulate, left (~BA32)	8.57	1,20	0.008
	Thalamus, left	10.21	1,10	0.01
	Midbrain	7.95	1,9	0.02
	Anterior prefrontal, right (~BA10)	5.55	1,21	0.028
	Fusiform, left	5.5	1,16	0.032
	Insula, left	4.69	1,24	0.041
	Superior Parietal, left (~BA7)	3.97	1,24	0.058
	Superior Parietal, right (~BA7)	4.23	1,23	0.051
Main effect of retention interval	Middle orbitofrontal, right (~BA11)	6.3	1,22	0.02
Interaction: success X strategy	DLPFC, right (~BA9/45)	5.24	1,22	0.032
	Anterior prefrontal, right (~BA10)	5.28	1,21	0.032
Interaction: strategy X retention interval	Insula, left (~BA7)	19.86	1,24	0.0001
	Fusiform, left	13.65	1,16	0.002
	DLPFC, right (~BA9/45)	9.39	1,22	0.006
	Anterior cingulate, right (~BA9/32)	7.34	1,19	0.014
	Thalamus, right	7.63	1,10	0.02
	Insula, right	6.05	1,22	0.022
	Inferior Parietal, left (~BA7)	5.62	1,23	0.026
	Superior Parietal, right (~BA7)	5.14	1,23	0.033
	DLPFC, left (~BA9/45)	4.85	1,23	0.038
	Inferior Parietal, right (~BA40)	4.32	1,22	0.05
	Middle orbitofrontal, right (~BA11)	4.21	1,22	0.052
	Posterior/dorsal prefrontal, right (~BA6)	4.12	1,22	0.055
Interaction: success X strategy X retention interval	Insula, left	9.14	1,24	0.006
	Anterior cingulate, left (~BA32)	9.08	1,20	0.007
	Thalamus, right	9.84	1,10	.011
	Fusiform, left	7.77	1,16	.013
	Fusiform, right	7.29	1,12	.019
	Inferior Parietal, right (~BA40)	5.35	1,22	0.03
	Anterior prefrontal, right (~BA10)	4.73	1,21	0.041
	Anterior prefrontal, left (~BA10)	4.27	1,23	0.05

Note: A 2x2x2 ANOVA was performed on percent signal changes in regions of interest (see Table 1.). Retrieval success (remembered vs. forgotten) and learning strategy (retest vs. restudy) were varied within participants, retention interval (20 minutes vs. one week) was varied between participants.

Figure 1. Activation maps for the functional localizer task (2-back vs. detection). Regions consistently activated across subjects are color-coded according to P_{FDR} <0.001. The z-coordinate in Talairach space is indicated above each section. For anatomical details of the activations see Table 1.
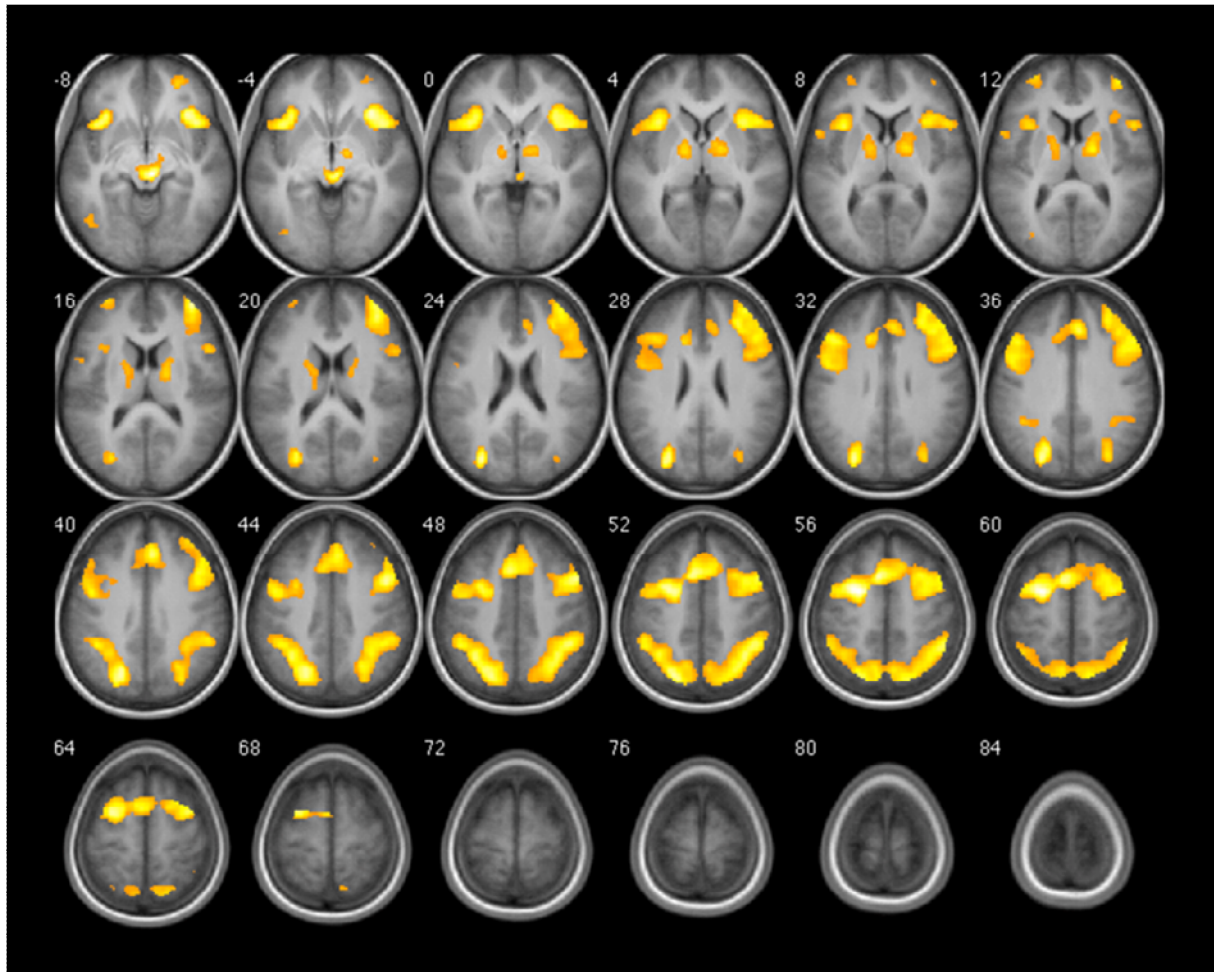
Figure 2. Accuracy (Upper panel) and RTs (Lower panel) of recall during the final test phase performed in the scanner as a function of retention interval and learning strategy. Error bars represent standard error of the mean.
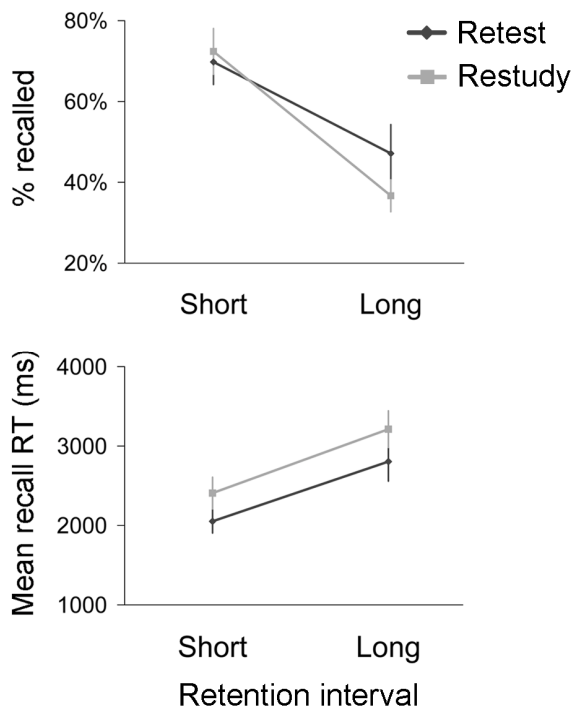
Figure 3. Activations during final recall in regions of interest where significant two-way interactions between learning strategy (retest vs. restudy) X retention interval (short vs. long) were observed. A) HRF functions (left column) and peak percent signal changes (right column) are shown for five representative areas. B) Peak percent signal changes in eight additional areas showing interaction of strategy and retention interval (see Table 2). Retested and restudied word pairs are shown separately for the short retention interval and long retention interval groups. Error bars represent standard error of the mean.
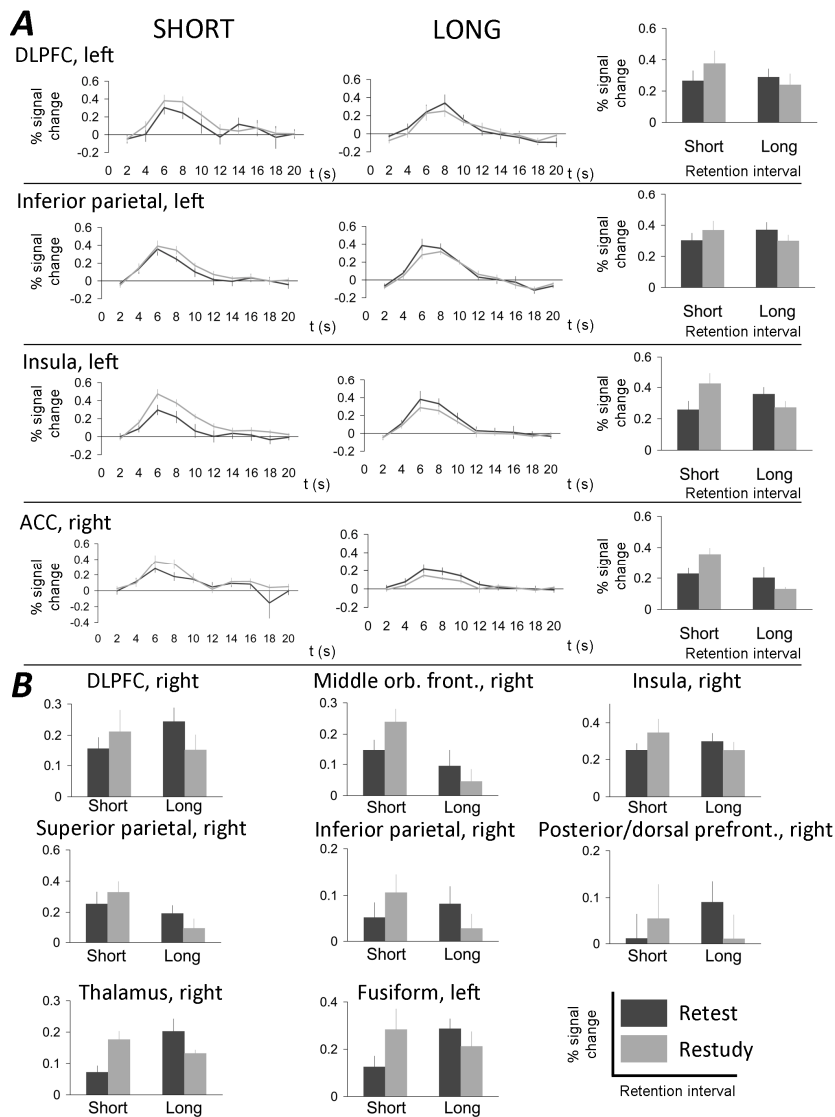
Figure 4. Activations during final recall in regions of interest where significant three-way interactions (learning strategy X retention interval X final recall success) were observed. A) HRF functions for the left insula (upper panel) and left ACC (lower panel). B) Peak percent signal changes in six additional areas showing the three-way interaction. Data are shown separately for the short retention interval and long retention interval groups. Error bars represent standard error of the mean.
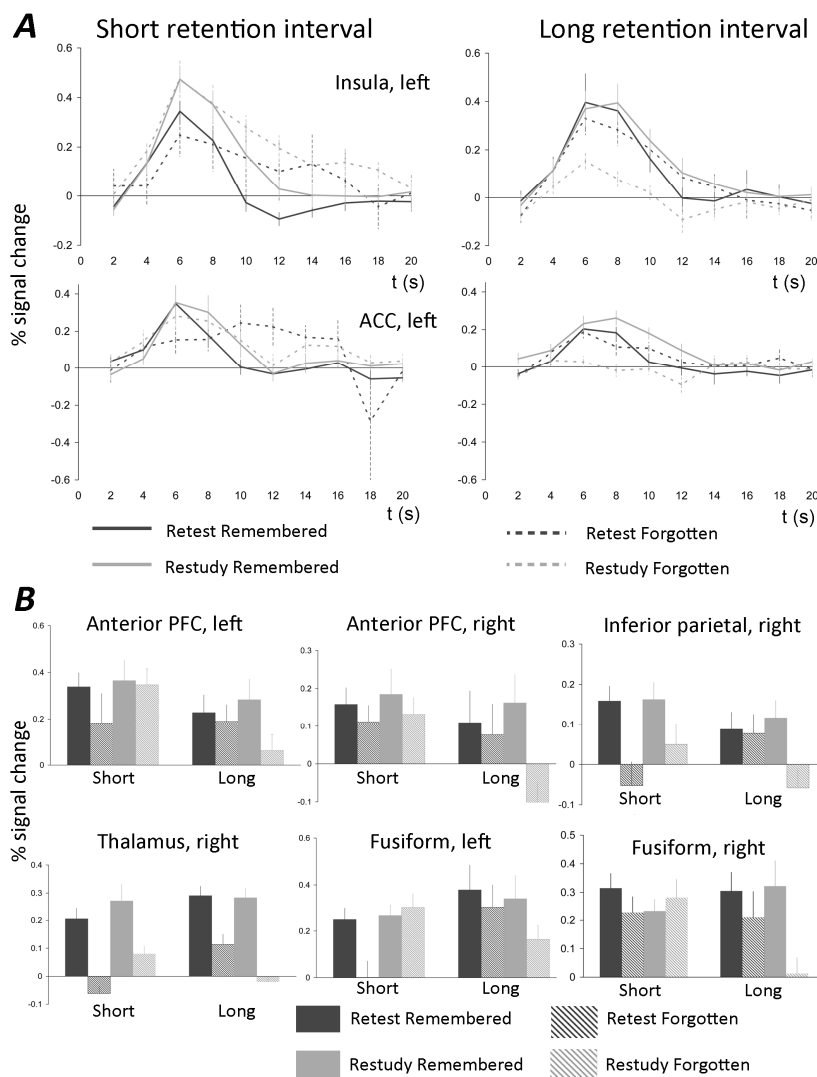
Figure 5. Results of the whole-brain analyses for the two contrasts with significant activations. Both panels show contrast images with activations (in yellow) significant at a threshold of p < 0.0001 (uncorrected) and an extended threshold of 50 voxels. A) The effect of strategy at long retention interval (retest vs. restudy contrast) revealed significant activations in the medial frontal/anterior cingulate cortex and in an area in the occipital lobe at around the early visual cortex. B) The interaction of strategy and retention interval ([(retest vs. restudy) long] vs. [(retest vs. restudy) short] contrast) revealed significant activations in the insula, bilaterally.