Published in final edited form as:
Neuron. 2009 September 10; 63(5): 697–708. doi:10.1016/j.neuron.2009.08.011.

Recollection, familiarity, and cortical reinstatement: A multi-voxel pattern analysis

Jeffrey D. Johnson[1,*], **Susan G. R. McDuff**[2,*], **Michael D. Rugg**[1], and **Kenneth A. Norman**[2,3]

[1]Center for the Neurobiology of Learning and Memory, and Department of Neurobiology and Behavior, University of California, Irvine

[2]Department of Psychology, Princeton University

[3]Princeton Neuroscience Institute, Princeton University

Summary

Episodic memory retrieval is thought to involve reinstatement of the neurocognitive processes engaged when an episode was encoded. Prior fMRI studies and computational models have suggested that reinstatement is limited to instances in which specific episodic details are recollected. We used multi-voxel pattern-classification analyses of fMRI data to investigate how reinstatement is associated with different memory judgments, particularly those accompanied by recollection versus a feeling of familiarity (when recollection is absent). Classifiers were trained to distinguish between brain activity patterns associated with different encoding tasks, and were subsequently applied to recognition-related fMRI data to determine the degree to which patterns were reinstated. Reinstatement was evident during both recollection- and familiarity-based judgments, providing clear evidence that reinstatement is not sufficient for eliciting a recollective experience. The findings are interpreted as support for a continuous, recollection-related neural signal that has been central to recent debate over the nature of recognition memory processes.

Keywords

reactivation; remember; know; strength; recognition

Introduction

Findings from psychological and neurobiological studies of memory have led to general agreement that many of the neurocognitive processes engaged when an event is encoded are re-engaged when the event is retrieved (Damasio, 1989; Rugg et al., 2008). This *reinstatement* of encoding-related processing during retrieval is a major component of several neurally-inspired models of episodic memory (e.g., Alvarez & Squire, 1994; McClelland et al., 1995; Norman & O'Reilly, 2003; Rolls, 2000; Shastri, 2002). According to the model of Norman and O'Reilly (2003), for example, the neural architecture of the hippocampus allows it to store non-overlapping representations of the patterns of cortical activity elicited when different events are encoded. When an effective retrieval cue for an event is presented, the appropriate hippocampal representation is reactivated, leading to reinstatement of the original pattern of cortical activity. Crucially, in the context of 'dual-process' theories of recognition memory (Mandler, 1980; for review, see Yonelinas, 2002), hippocampally-mediated

Address correspondence to: Jeffrey D. Johnson, Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, CA 92697-3800, Phone: 949-824-8861, Fax: 949-824-4807, jeff.johnson@uci.edu.
[*]These authors contributed equally to this work.

reinstatement is thought to support the *recollection* (or recall) of specific details associated with an episode. In contrast to the hippocampal memory system, a cortical system involving extra-hippocampal regions of the medial temporal lobe is capable of giving rise only to an acontextual (non-recollective) *familiarity* signal that corresponds to the scalar match between the cue and episode (Norman & O'Reilly, 2003).

With the exception of indirect neuropsychological evidence (e.g., Rubin & Greenberg, 1998), empirical support for recollection-related cortical reinstatement in humans comes largely from the use of functional neuroimaging. Studies employing event-related functional magnetic resonance imaging (fMRI) have been particularly useful in this regard by enabling the neural correlates of recollection to be contrasted according to the nature of the recollected content (e.g., Johnson & Rugg, 2007; Kahn et al., 2004; Wheeler & Buckner, 2004; Woodruff et al., 2005). In combination with the neural measure, these studies adopted behavioral procedures —such as the *remember/know* (Tulving, 1985) or source memory procedures (Johnson et al., 1993)—to identify trials where subjects retrieved specific episodic details.[1] Arguably the most convincing evidence from these studies in favor of reinstatement comes in the form of regionally-specific double dissociations in the cortical patterns associated with the remembering differential content. For instance, Kahn et al. (2004) reported that remembering words studied in a visual imagery task activated left parahippocampal cortex to a greater extent than did remembering words studied in a phonological task, whereas the reverse contrast was associated with activation of left premotor cortex. Similarly, Woodruff et al. (2005) reported that two regions of left fusiform cortex—shown previously to be functionally specialized for the processing of visually-presented words and pictures—exhibited dissociable activity with respect to remembering words vs. pictures.

In a direct test of the reinstatement hypothesis, Johnson and Rugg (2007) investigated the extent to which content-specific neural correlates of remembering overlapped with regions that were selectively active when the relevant content was encoded. Subjects were first presented with a series of words and required to use the words either in a sentence or in a visual imagery task, and then undertook a remember/know test. Brain regions where greater activity was associated with remember compared to know responses exhibited specificity according to the class of study episode: Words studied with the sentence task elicited greater activity in medial prefrontal cortex, whereas words studied with the imagery task elicited greater activity in occipital and fusiform cortex. Importantly, the regions demonstrating these dissociations were a subset of regions exhibiting differential activity when the two classes of words were initially studied (see Kahn et al., 2004, for a similar, across-experiment comparison). Thus, these findings established a direct link between the neural correlates of study processing and the phenomenological experience of remembering, consistent with the notion that processing selectively engaged during encoding is reinstated during retrieval.

Although the aforementioned findings convincingly demonstrate a relationship between cortical reinstatement and recollection, they do not address two important questions about the status of reinstatement effects when items are reportedly judged on the basis of familiarity (known). First, are such judgments at all associated with reinstatement? In the three fMRI studies described above, although behavioral methods designed to separate recollection and familiarity were employed, in two of the studies there were insufficient numbers of trials to evaluate reinstatement when know responses were given (Johnson & Rugg, 2007; Woodruff et al., 2005), while the relevant contrasts were not reported in the remaining study (i.e., for incorrect source judgments; Kahn et al., 2004). If it transpires that reinstatement effects are

[1] To minimize confusion, we hereafter use the term *remembering* to refer to the experience of retrieving specific episodic details, and we use the term *knowing* to refer to the experience of recognizing an item without retrieving specific episodic details. We reserve the terms *recollection* and *familiarity* for describing the processes and neural signals often thought to respectively underlie those experiences.

evident during know responses, a second question arises: Do these effects differ from those associated with remembering, either in magnitude or localization?

Resolution of these two questions has important implications for the ongoing theoretical debate about the nature of processes contributing to recognition memory. According to some dual-process theorists (see Yonelinas, 2002), remember and know responses reflect the influence of qualitatively distinct processes. Remember responses are thought to reflect recollection of specific details, whereas know responses are thought to be based on familiarity (in the absence of recollection). Accepting this assertion at face value leads to the prediction that cortical reinstatement will be present for remember responses but absent for know responses. Contrary to this view, another class of theories posits that different recognition judgments are not based on a clear-cut distinction between two memory processes or signals. Rather, the judgments result from assessing a single, continuous 'memory strength' signal (e.g., Donaldson, 1996; Dunn, 2004) or an amalgam of continuous signals (Wixted & Stretch, 2004; Wixted, 2007) present across all of the different judgments. That is, each test item is associated with a particular level of the strength signal, and a detection process is used to decide whether the strength exceeds a criterion, thus determining the response. If it is assumed that the degree of retrieval-related reinstatement co-varies with memory strength, then reinstatement effects should follow a graded profile: largest when subjects report remembering episodic details yet also present for know responses that the subject attributes to familiarity.

The present fMRI study was designed to explore the relationship between cortical reinstatement and distinct phenomenological bases of recognition memory, as evidenced by different behavioral correlates (i.e., remember vs. know). Subjects first completed a study phase where they viewed a series of words and undertook three different encoding tasks that elicited distinct patterns of cortical activity (*Artist*, *Function*, and *Read*; McDuff et al., 2009). During a later test phase, recognition memory for the studied words was assessed using a modified remember/know procedure, in which one of five responses was required to each test item (Yonelinas et al., 2005). One response was used to indicate that details associated with studying an item were remembered. The remaining four responses were used to rate the confidence with which an item was known to be studied or not studied, presumably on the basis of item familiarity in the absence of recollection. fMRI data acquired during both the study and test phases allowed for direct comparison between encoding- and retrieval-related activity (Johnson & Rugg, 2007). According to the view that remember (but not know) judgments veridically index retrieval of specific episodic details, reinstatement effects should be confined to items endorsed as remembered. Alternatively, according to theories positing a memory strength continuum, reinstatement effects should be evident, albeit in weaker form, for test items associated with know judgments.

In contrast to the previous studies of reinstatement described above, we employed multi-voxel pattern analyses (MVPA) of the fMRI data (for reviews, see Haynes & Rees, 2006; Norman et al., 2006). MVPA is well-suited for characterizing reinstatement because it quantifies the relationship between patterns of brain activity acquired during one experimental phase (the study phase in our case) and any 'reactivated' patterns from another phase (our test phase). Moreover, because MVPA involves classifying correlated patterns of activity across multiple voxels, it is often considered to be more sensitive than 'mass-univariate' fMRI analyses, which might fail to detect differences in signals that are weak at the single-voxel level or even when spatially smoothed across voxels (see Haynes & Rees, 2006; Norman et al., 2006). In the present study, two types of MVPA were implemented. The first type was designed to maximize the sensitivity of detecting reinstatement across different recognition memory judgments, by making use of a subset of voxels that best distinguished between the encoding tasks (for similar implementations, see McDuff et al., 2009; Polyn et al., 2005). The second type of MVPA involved classifying data from 'searchlights' (spheres) of voxels (Kriegeskorte et al., 2006;

Mur et al., 2009) and provided information about whether the spatial distribution of reinstatement effects throughout the brain differed according to the type of memory judgment.

Results

Behavioral performance

Figure 1 displays the mean proportions of responses (*Remember*, *Sure Old*, *Unsure Old*, *Unsure New*, and *Sure New*) and corresponding RTs to items presented during the test phase. As is apparent in the figure, items previously studied with the Artist and Function tasks primarily elicited Remember responses, followed by Sure Old (know) responses; items studied with the Read task were associated mostly with Sure Old, Unsure Old, and Unsure New responses; and new items primarily elicited New responses. Because the pattern-classification analyses (see below) are restricted to old items, we also focused the behavioral analyses on those items. Additionally, given the low proportions of Unsure Old, Unsure New, and Sure New responses for Artist and Function items, the corresponding trials were collapsed into an *Other* category for each task. ANOVA of the response proportions, incorporating factors of task (Artist, Function, and Read) and response (Remember, Sure Old, and Other), revealed a significant interaction ($F_{2.3,34.3} = 64.05$, $p < .001$; degrees of freedom corrected according to Greenhouse & Geisser, 1959). The interaction indicated a trade-off between response categories, such that Artist and Function items elicited more Remember responses than did Read items (min. $t_{15} = 7.73$, $p < .001$), whereas the opposite was true for Other responses (min. $t_{15} = 13.15$, $p < .001$). ANOVA of the RT data (including the same factors as above) gave rise to a significant main effect of response ($F_{1.4,21.7} = 26.51$, $p < .001$) and its interaction with task ($F_{3.2,47.5} = 7.38$, $p < .001$). Subsidiary ANOVAs revealed a task effect only for Remember responses ($F_{1.4,20.4} = 12.25$, $p < .005$), whereby RTs were shorter for Artist and Function items than for Read items (min. $t_{15} = 3.45$, $p < .005$). Notably, the response proportions and RTs for Artist and Function items were statistically equivalent.

Whole-brain MVPA

As described in the Introduction, our primary aim was to determine the relationship between patterns of brain activity elicited at study and those elicited at test. MVPA was employed to provide a sensitive index of the strength of the study-test relationship, but this index is blind to the loci of voxels expressing any such relationship. Although the specific brain regions discriminating between the three study tasks (as determined through classifier training) were largely inconsequential, it was important to ensure that the voxels were biologically meaningful —that is, they constituted sizable clusters rather than dispersed individual voxels, and encompassed cortical areas expected to be active in cognitive tasks such as those employed here. Accordingly, we created a group mean *importance map* for each study task condition, by combining the voxel-wise input values and the trained classifier's weights (see Experimental Procedures; McDuff et al., 2009). As shown in Figure 2, it is apparent that the patterns of important voxels—clustering largely in bilateral occipital, superior parietal, and left inferior frontal cortex—meet our aforementioned criteria for meaningfulness.

The accuracy of the classifier in determining the prior encoding history of old test items was operationalized as the probability that the classifier's output for the correct study task was greater than the output for each of the other two tasks. Classifier accuracy for a given test item was assessed beginning with the time point (TR) in which the test item onset (hereafter, TR 1) and continuing for six additional time points (up to TR 7). Thus the accuracy measure provided information about classification performance as a function of time. Overall classifier accuracy for all of the old items, regardless of the response given, is shown in Figure 3. As is clear from the figure, classifier accuracy was maximal at TR 4, coinciding with the expected peak of the hemodynamic response for a transient stimulus (as estimated by convolving hemodynamic and impulse response functions). To test whether classifier accuracy exceeded chance (.33) at any of the TRs, a series of one-sample t-tests was conducted. Accuracy was significantly above chance for TRs 3 through 7 (min. $t_{15} = 3.75$, $p < .005$), and each of these effects remained significant following correction for multiple comparisons across the seven TRs (using the Holm-Bonferroni procedure with an overall $p < .05$; Holm, 1979). These findings demonstrate that the patterns of fMRI data associated with old test items differed systematically, so as to allow the classifier to accurately determine an item's prior encoding history.

Of more relevance to our primary aim was whether the classifier's ability to correctly identify the encoding condition of test items was limited to items eliciting a recollection response, or whether above-chance accuracy also extended to items accorded non-recollective responses. To address this issue, we investigated classifier accuracy according to response category (Remember, Sure Old, and Other). Classifier accuracy was first assessed separately for the three tasks and then averaged across tasks. The resulting accuracy values for each response category are shown in Figure 3, in which it is apparent that accuracy once again followed the expected (hemodynamically-corrected) time course. One-sample t-tests (corrected for multiple comparisons as before) revealed that accuracy was above chance when test items were endorsed with either a Remember or Sure Old response. Significant effects were observed at TRs 3 through 5 for Remember responses (min. $t_{15} = 3.26$, $p < .01$), and at TRs 4 and 5 for Sure Old responses (min. $t_{15} = 4.30$, $p < .001$). There were no significant effects for Other responses. Thus, the classifier was capable of assigning test items to the appropriate encoding condition only when the items elicited a phenomenological sense of recollection (remembering) or high-confidence familiarity (knowing).

Having demonstrated that reinstatement was above chance for both Remember and Sure Old responses, we set out to investigate the relative strength of reinstatement for Remember vs. Sure Old trials. To accomplish this goal, we switched from computing classifier accuracy to measuring classifier output strength: the real-valued output for the classifier node representing the actual (true) task condition. The key advantage of using classifier output strength is that it tracks the raw magnitude of reinstatement on each trial, whereas the accuracy measure computes a binary score for each trial (based on whether the actual task output is higher than the other task outputs) and discards information about the actual magnitude of reinstatement. We restricted these analyses of reinstatement strength to the data from TRs 3 through 5, based on our earlier findings that classifier accuracy was maximal during this time period.

As with our classifier accuracy analysis, the classifier output values for each response type were first averaged across trials within each study task and then averaged across tasks. This averaging procedure was especially important here because (as mentioned above) there were significant across-task differences in responding. That is, items from the Artist and Function tasks elicited more Remember responses and fewer Other responses than did items from the Read task. This discrepancy raises the possibility that effects of response type (e.g., Remember vs. Other) on classifier output strength will be confounded with effects of task (Artist/Function vs. Read). Our averaging procedure eliminates this potential confound by ensuring that each task is equally represented within each response type.

Figure 4A shows the output values for each response, averaged over the Artist, Function, and Read conditions. As can be seen in the figure, the output for Remember responses is higher than that for Sure Old (except at TR 5), which is in turn higher than for Other responses. The output values were subjected to pair-wise comparisons between response categories. For the Remember vs. Sure Old comparison, t-tests revealed no significant differences. For the comparison of Sure Old vs. Other, there was a significant difference at TR 5 ($t_{15} = 2.71$, $p < .025$; corrected for multiple comparisons). Finally, the Remember vs. Other comparison gave rise to a significant difference at TR 4 ($t_{15} = 2.96$, $p < .01$). Thus, although the output values appeared to follow a graded profile across responses, the differences did not consistently reach significance.

We hypothesized that the weak results of the previous analyses might be attributable to the low number of Remember trials in the Read condition. For reasons described above, our analysis weighted Artist, Function, and Read trials equally when estimating reinstatement for each response type. However, the actual number of Read-task Remember trials was extremely small. (Each subject had at least one Remember response in the Read condition, but most subjects had fewer than five such responses.) Classifier estimates based on very small numbers of trials can be highly volatile. To address this issue, we conducted a further analysis that was identical to the foregoing one, but was restricted to test items associated with the Artist and Function tasks (since these tasks, unlike the Read task, had adequate numbers of trials in each response bin). As in the previous analysis, for each response type, we first averaged classifier estimates within each task and then averaged these estimates across tasks. This procedure ensures that the Artist and Function tasks are equally represented within each response type, so effects of response type on reinstatement can not be attributed to task differences.[2]

The output values averaged over the Artist and Function conditions are shown in Figure 4B, segregated according to response. For the Remember vs. Sure Old comparison, pair-wise t-tests revealed a significant difference for each of the three TRs (min. $t_{15} = 2.13$, $p < .05$; corrected for multiple comparisons). The Remember vs. Other comparison also gave rise to significant differences for all TRs (min. $t_{15} = 2.85$, $p < .05$). In addition, the comparison of Sure Old vs. Other responses revealed a significant effect at TR 5 ($t_{15} = 3.18$, $p < .001$). In contrast to the results based on all three study tasks, these results clearly demonstrate a graded profile of classifier output across response categories. Specifically, output was highest for items eliciting Remember responses, intermediate for Sure Old responses, and lowest for Other responses.

Searchlight MVPA

The foregoing results provided evidence that the magnitude of reinstatement differs across recognition memory judgments. In the Introduction, a further question was posed about whether the spatial patterns of reinstatement effects associated with recollection- vs. familiarity-based memory judgments also differ. This question was addressed with searchlight-based classification analyses.[3] Based on our previous results demonstrating reinstatement for both Remember and Sure Old responses, we tested several possibilities regarding how the patterns of brain regions exhibiting reinstatement effects might differ according to response. First, regions might exhibit equivalent levels of above-chance reinstatement for both response types. Second, reinstatement might differ quantitatively for recollection- and familiarity-based responses, whereby both are associated with above-chance reinstatement, but the effects occur at a greater magnitude for one of the responses. Finally, the patterns of reinstatement could differ qualitatively, such that the reinstatement exhibited in some regions is evident selectively for one of the responses but absent for the other response.

Given that reinstatement was previously shown to be most prominent at TRs 3 through 5, the searchlight results were simplified by averaging the classifier output values over these time points (rather than creating separate maps for each TR). For reasons outlined earlier, only the data from Artist and Function test trials were used for these analyses (first averaged separately, and then across tasks). Two types of maps—output and accuracy—were created from these results. Output maps corresponded to the real-valued output from the actual (true) task node

[2] We also re-analyzed the accuracy data based on only the Artist and Function conditions, which produced qualitatively similar results to those reported here (see Supplemental Material).

[3] For comparison, a parallel GLM-based analysis is reported in the Supplemental Material.

for a given trial and were used to identify differences in the magnitude of reinstatement according to the designated test response. Accuracy maps were constructed for each response category by determining whether the output value for the actual task node was greater than that for the other two nodes (one of which was the Read node). The two map types were used together to ensure that any voxels exhibiting response-related differences in reinstatement magnitude also showed above-chance reinstatement.

Regions exhibiting equivalent reinstatement for Remember and Sure Old responses were identified by the intersection (identified by inclusive masking) of voxels showing above-chance classifier accuracy for the two response categories: $\text{Remember}_{acc} > .33$ and $\text{Sure Old}_{acc} > .33$ (each thresholded at $p < .01$). Further, any voxels where the correct classifier output differed according to the Remember vs. Sure Old contrast (thresholded liberally at $p < .1$) were removed via exclusive masking. Figure 4 shows the outcome of this contrast procedure, which identified regions of left lateral temporal cortex, superior frontal gyrus, and inferior frontal gyrus (each surviving a cluster-level correction of $p < .05$; Worsley et al., 1996). Thus, these regions showed reinstatement effects during both recollective- and familiarity-based memory judgments.

There were no supra-threshold clusters where reinstatement was at above-chance levels for both Remember vs. Sure Old responses, but where these effects also differed in magnitude. This pattern of results was tested by contrasting the classifier output for the two responses (Remember > Sure Old or Sure Old > Remember) in combination with verifying that reinstatement was evident for both ($\text{Remember}_{acc} > .33$ and $\text{Sure Old}_{acc} > .33$; each at $p < .01$).

Finally, qualitatively different patterns of reinstatement for Remember and Sure Old responses were identified by testing for selective effects associated with either response. Regions exhibiting selective Remember-related reinstatement were identified with the Remember > Sure Old contrast of raw classifier output, while ensuring that reinstatement for the former response category in these voxels also achieved above-chance accuracy ($\text{Remember}_{acc} > .33$, each thresholded at $p < .01$). Additionally, any voxels where reinstatement for Sure Old responses differed from chance (Sure Old_{acc} vs. .33, bi-directional; $p < .1$) were excluded. As shown in Figure 5, two clusters of voxels in medial posterior cortex were identified, one in the vicinity of retrosplenial cortex and the other in posterior cingulate. The analogous contrast procedure used to identify selective reinstatement for Sure Old responses (Sure Old > Remember and $\text{Sure Old}_{acc} > .33$, excluding $\text{Remember}_{acc} > .33$) revealed no significant effects. Thus, there was a single dissociation in the reinstatement effects associated with recollection- and familiarity-based memory, which took the form of regions showing selective reinstatement for recollection.

Discussion

The aim of the present study was to elucidate the relationship between reinstatement of encoding-related neural activity during retrieval and different phenomenological correlates of recognition memory judgments (specifically, judgments associated with remembering vs. knowing that a test item was previously encountered). Using two forms of multivariate pattern-classification analyses of fMRI data, we assessed the extent to which patterns of brain activity associated with retrieval can be used to correctly classify the prior encoding history of test items. These analyses demonstrated that MVPA is capable of detecting the relationship between brain patterns activated during encoding and those that are reactivated at test (also see McDuff et al., 2009; for analogous findings in free recall, see Polyn et al., 2005). Two novel and theoretically substantive findings emerged: one involved the different levels of reinstatement that were associated with recognition judgments having distinct subjective bases, and the other concerned the cortical regions that exhibited these reinstatement effects when the different types of judgments were made. We discuss these findings in turn below.

Using a whole-brain MVPA approach designed to be maximally sensitive to detecting reinstatement effects, we were able to classify with above-chance accuracy the prior encoding task that was undertaken for a given test item, regardless of whether the item was correctly judged as old in association with a Remember or a Sure Old response. The results for Remember responses are consistent with findings from a recent study where it was demonstrated that the neural correlates of Remember responses overlapped with regions that were selectively active when the test items were initially studied (Johnson & Rugg, 2007; also see Kahn et al., 2004). In keeping with our prior interpretation, the present findings are taken to indicate the reinstatement of study content at the time of retrieval. The reinstated content likely reflects a recapitulation of the cognitive operations that were engaged by the different tasks during the study phase, given that there were no physical differences between test items that correlated with their prior study task.

The present findings also constitute a theoretically important extension to our prior conceptualization of reinstatement. As was noted in the Introduction, reinstatement (or content-specificity) was previously evaluated for only those recognized test items that were accompanied by either a Remember response (Johnson & Rugg, 2007; Woodruff et al., 2005) or a correct source memory attribution (Kahn et al., 2004). Here, however, we have demonstrated that reinstatement is also evident for test items correctly recognized in the absence of any avowed retrieval of specific episodic details. More specifically, the magnitude of reinstatement, as measured by classifier output values, decreased in a graded manner across Remember, Sure Old, and Other responses. Therefore, although reinstatement has been shown here, as previously, to be correlated with the phenomenal experience of remembering, the current study provides compelling evidence that reinstatement is not uniquely associated with such an experience.

In a second set of analyses, we employed searchlight-based classifications to characterize the similarities and differences among reinstatement effects associated with Remember vs. Sure Old responses. These analyses yielded two results. First, multiple regions exhibited reinstatement for both response types, with common reinstatement effects evident in left-lateralized regions of inferior frontal gyrus, superior frontal gyrus, and lateral temporal cortex (Figure 5). Second, an additional set of regions was associated with reinstatement for Remember responses but not for Sure Old responses. These selective effects were in the vicinity of retrosplenial cortex and posterior cingulate (Figure 6). Thus, the searchlight analyses demonstrated that the reinstatement effects associated with Sure Old responses were in a subset of the cortical regions associated with Remember-related reinstatement.

The above findings, together with the graded reinstatement effects described earlier, suggest that the Remember and Sure Old responses relied on a common process or signal. An obvious account of the findings is that test items gave rise to different levels of a continuous memory signal, and that criteria placed along this continuum were used to assign items to the different response categories (Donaldson, 1996; Dunn, 2004). By this argument, Remember responses were made when test items evoked a signal that exceeded the strictest criterion, Sure Old responses resulted from a memory signal that fell between this criterion and one that was less strict, and Other responses were due to the signal falling short of both criteria. A similar account has been applied to results from a recent behavioral study that combined remember/know and source memory judgments (Wais et al., 2008). In that study, subjects' source memory performance was above chance even for items they reported to not remember, leading the authors to suggest that the retrieval of source information contributed to the memory strength of the resulting know responses. This is not to say that recognition memory judgments are guided solely by a single process; instead, one can assume the involvement of multiple continuous processes, whereby all types of judgments are influenced to some degree by each process (Wixted, 2007; Wixted & Stretch, 2004). Importantly, the reinstatement effects for both Remember and Sure Old responses in the present study are inconsistent with models in which such responses are thought to selectively tap into qualitatively different memory processes (i.e., recollection vs. familiarity, respectively; Yonelinas, 2002). Our findings provide crucial evidence that cortical reinstatement effects constitute a neural signature of previously-hypothesized instances of ‘sub-threshold’ recollection.

As described in the Introduction, our work here was largely inspired by a framework in which learning and memory rely on complementary systems: a hippocampal system capable of rapidly encoding non-overlapping conjunctions of the cortical patterns that represent specific episodes, and an extra-hippocampal (cortical) system that exploits overlapping representations of the general statistical structure evident across similar episodes (McClelland et al., 1995; O’Reilly & Norman, 2002; O’Reilly & Rudy, 2001). Models derived from this framework, along with related models of hippocampal function, have proposed that a hippocampally-stored cortical representation mediates the reinstatement of a corresponding cortical pattern, leading to recollection (Alvarez & Squire, 1994; Norman & O’Reilly, 2003; Rolls, 2000; Shastri, 2002). Consistent with this proposal, and with our findings of graded reinstatement effects, a GLM analysis (see Supplemental Material) gave rise to greater right hippocampal activity associated with Remember compared to Sure Old responses. By contrast, activity in hippocampus was not enhanced for Sure Old relative to Other responses (or correctly-rejected new items). Although this latter finding might suggest that the reinstatement observed for Sure Old responses is attributable to some mechanism other than hippocampally-mediated recollection, the result is likely due instead to the hippocampus being involved additionally in novelty-related encoding processes (also see Düzel et al., 2003; Stark & Okado, 2003; Stern et al., 1996). Such processing is elicited to a greater extent by relatively unfamiliar items (e.g., new items and those given Other responses) and would thereby counteract any reinstatement-related enhancement of hippocampal activity elicited by familiar items.

As we alluded to above, a parsimonious account of the present results, in relation to the Norman and O’Reilly (2003) model, supposes that the mapping between the neural correlates of hippocampally-mediated recollection and subjects’ behavioral responses is more continuous than sometimes conceptualized by dual-process models of memory. It is important, however, not to overlook a discontinuity between the reinstatement effects associated with items endorsed as recollected versus those related to confident old judgments. Although reinstatement-related neural signals associated with Sure Old responding were observed in multiple cortical regions, the effects were evidently insufficient to support the phenomenal experience of recollection. At the moment, it is not possible to discern between two explanations of this discontinuity. On the one hand, as evidenced by the effects in additional cortical regions for Remember compared to Sure Old responses, recollection might result from a quantitative increase in either the number of regions exhibiting reinstatement or the magnitude of reinstatement in those areas. Alternatively, it is possible that the specific loci of reinstatement effects associated with Remember responses, such as in medial parietal cortex, carry crucial qualitative information that drives the episodic evidence above the appropriate decision threshold. Importantly, while the interpretation of this response-related distinction in the neural signal is an important topic for follow-up research, its resolution does not detract from our main finding that reinstatement plays a role in phenomenologically-distinct forms of recognition memory.

Two caveats to the interpretation of the present findings deserve further discussion. First, the precise time course of the neural events driving reinstatement effects cannot be determined by fMRI data alone (for similar discussion, see Johnson et al., 2008; Johnson & Rugg, 2007; Kahn et al., 2004; Maratos et al., 2001; Woodruff et al., 2005). On the one hand, the effects might occur shortly after test item onset, as would be expected if they were a key determinant of the recognition memory judgment. Alternatively, the effects could be a consequence of the memory judgment, possibly reflecting the deployment of attention toward particular types of retrieved content, or the maintenance of that content in working memory in service of further evaluation. These two accounts can only be adjudicated by employing a neural measure with much higher temporal resolution than fMRI, such as event-related potentials (ERPs; for an example of content-specific ERP effects during retrieval, see Johnson et al., 2008). A second caveat is that the classifier was trained solely to detect patterns of activity that discriminated between the three encoding tasks. This training procedure gives the classifier the ability to detect when test items are accompanied by activity related to the recollection of task-specific details. However, the procedure does not enable the classifier to detect activity related to the recollection of ‘non-diagnostic’ details (i.e., details shared by all three tasks) or activity associated to familiarity-based processing—both of which likely contribute to some degree to subjects’ responses in this task.

To conclude, the fMRI findings reported here are consistent with the idea that the retrieval of episodic memories involves reinstating patterns of cortical activity that were engaged during encoding. The present findings extend previous results by demonstrating that reinstatement is not restricted to instances in which subjects reportedly retrieve specific episodic information, emphasizing that the presence of a content-dependent neural signal is not sufficient for eliciting a phenomenological sense of remembering (and the ensuing response). Rather, in situations where recognition is indicated as being guided solely by a strong feeling of familiarity or knowing, reinstatement is also evident (albeit at a lower magnitude) and recruits largely the same pattern of brain regions that were associated with remember-related effects. Finally, the current study adds to a growing body of evidence demonstrating the benefits of using multivariate classification analyses to detect subtle, yet informative patterns in fMRI data.

Experimental Procedures

Subjects

Sixteen volunteers (11 females) between 18 and 31 years of age (M = 22) were recruited from the undergraduate and graduate student community of Princeton University and remunerated for their participation. All subjects reported being right-handed, native-English speakers with normal or corrected-to-normal vision, no history of neurological disease, and no other contraindications for MRI. Informed consent was obtained in accordance with the Princeton University Institutional Review Board guidelines.

Stimuli

The stimuli were 306 words drawn from the MRC database (Coltheart, 1981; Wilson, 1988; http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). The words were between four and nine letters long (M = 5.5, SD = 1.3), had written frequencies between one and 50 per million (M = 16.9, SD = 13.2; Kucera & Francis, 1967), and had scores of at least 500 on scales of familiarity (M = 580.9, SD = 34.6), concreteness (M = 539.2, SD = 27.5), and imagability (M = 581.7, SD = 31.3). Words with emotional connotations or referring to alcoholic beverages were not used. Twenty-seven additional words with similar characteristics served as practice stimuli. All words were displayed visually in white lowercase 30-point Helvetica font.

For each subject, 162 words were randomly selected from the pool to serve as study items. Study items were randomly assigned to three study blocks and three encoding task conditions (see below), resulting in 18 items per task per block. A subset of 144 study items (16 drawn from each task/block combination) were re-presented as old items during the test phase and intermixed with the remaining 144 non-studied (new) words. The test items were randomly divided into four test blocks (72 items per block), with the constraint that each block had equal numbers of old and new items.

Behavioral tasks

Subjects were instructed and completed a practice version of the experiment outside the scanner. In the scanner, the experiment consisted of three study blocks followed by four test blocks, with each block corresponding to a separate run of fMRI acquisition. Blocks were separated by breaks of around 1–2 minutes. A longer (~10 min) break occurred between the study and test phases, during which anatomical data were acquired while a nature video was shown. All experimental stimuli were displayed on a screen positioned at the head of the magnet bore, which was viewed through a mirror placed in front of the subject's eyes.

For the study phase, subjects were presented with a series of words and had to complete an *Artist*, *Function*, or *Read* task for each word (Davachi et al., 2003; Dzulkifli & Wilding, 2005; Johnson et al., 1997; see McDuff et al., 2009, for use of the same tasks). For the Artist task, subjects were to imagine how an Artist would draw the object denoted by the word and then rate the difficulty of drawing (1 = easy to 5 = hard). For the Function task, subjects had to think of different functions for the object and then respond according to how many were generated (1 to 5). For the Read task, subjects were to silently pronounce the word backwards and rate the difficulty (1 = easy to 5 = hard). The study phase was subdivided into miniblocks of three consecutive trials, during which a single encoding task was performed. The miniblocks were employed to allow for efficient segregation of the hemodynamic responses according to task (also see McDuff et al., 2009), while not requiring long lags between individual study items (which would have significantly increased scanning time). Miniblocks began with a 4-sec display of a task instruction (e.g., *Do ARTIST task*) and the response options, which remained on the screen throughout the miniblock. Each word appeared in the center of the screen for two seconds, and subjects were instructed to withhold their response until a response cue (*) appeared. Both the word and response cue remained on the screen for two seconds. Responses were made by pressing one of five keys mapped to the right hand. The second and third words of the miniblock were presented similarly and followed immediately by another miniblock. Each study block comprised 18 miniblocks (six per task) which were randomly ordered such that no task was completed twice consecutively. (An analysis of behavioral performance during the study phase is reported in the Supplemental Material.)

For the test phase, subjects were shown a series of intermixed old and new words and required to make one of five responses to each word (following Yonelinas et al., 2005). Subjects were to respond with their right thumb when they could remember specific details surrounding the word's presentation during the study phase (Remember). It was emphasized that subjects should give a Remember response if they remembered *any* details, regardless of whether the details were directly related to the study tasks or unrelated. The instructions also included a description of some examples of task-unrelated details, such as a personal thought elicited by a study item (e.g., something about your own dog in response to seeing the word *dog*) and an environmental stimulus co-occurring with an item (e.g., an unexpected background noise). If no study details were remembered, subjects used a four-point scale to rate their confidence that the word was either old or new. The right index through little fingers were mapped respectively to Sure Old, Unsure Old, Unsure New, and Sure New responses. Each test word was displayed centrally for three seconds, during which subjects were instructed to make their response. Responses outside the 3-sec period were infrequent and not analyzed. Test words were followed by relatively long inter-item lags, during which a plus sign was centrally displayed, which helped to segregate the hemodynamic responses elicited by individual items. Each test block contained 48 trials with 5-sec lags, 18 with 7-sec lags, and 6 with 9-sec lags, divided equally between old and new items.

Data acquisition and preparation

MRI data were acquired with a Siemens Allegra 3T scanner at the Center for the Study of Brain, Mind, and Behavior at Princeton University. A T_1-weighted anatomical volume (176 sagittal slices, 2-sec TR, 4.38-msec TE, 1-mm^3 voxels, 78° flip angle, and 256-mm^2 FOV) was acquired with an MP-RAGE sequence. Functional volumes consisted of T_2*-weighted echoplanar images with blood oxygenation level dependent (BOLD) contrast and the following parameters: 2-sec TR, 30-msec TE, 34 slices, 3.9-mm slice thickness, 3-mm^2 in-plane resolution, 75° flip angle, and 192-mm^2 FOV. The fMRI data were acquired in 7 separate blocks, with 152 volumes for each of 3 study blocks and 326 volumes for each of 4 test blocks. Five additional fMRI volumes collected at the beginning of each block permitted T_1 equilibration and were discarded before analysis. The onset of each study and test item coincided with the acquisition onset of an fMRI volume. The fMRI data were pre-processed using the AFNI software package (Cox, 1996; http://afni.nimh.nih.gov/afni). All volumes were spatially realigned to the first volume of the first study block, and the data in each volume were temporally shifted to the onset of the middle slice. Voxels exhibiting signal spikes were replaced via a temporal smoothing algorithm. Linear and quadratic trends were removed from each run to minimize the influence of scanner drift. The fMRI data were z-scored separately for each voxel and block. Notably, the fMRI data were neither spatially normalized nor smoothed prior to being used for the classification analyses.

For the classification analyses (see below), only a subset of the fMRI data were used—those volumes (TRs) determined as being associated with the study and test items, based on the lag in timing between stimulus events and the assumed resulting hemodynamic response. For the study phase, the hemodynamic lag was accounted for by convolving the onset of each study word with a synthetic hemodynamic response function (HRF; the gamma variant of AFNI's *waver*). Due to our use of miniblocks of study items assigned to a single task, this convolution produced a relatively dispersed (boxcar-like) HRF for each miniblock rather than three distinct item-specific HRFs. The convolved values at each time point (TR) were then normalized (from 0 to 1) across time. TRs with normalized values ≥ 0.5 were assigned to the corresponding (immediately presented) study task, whereas all other study phase TRs were excluded from the classification. With the first TR (hereafter TR 1) marking the onset of a miniblock's first word, the binarization procedure resulted in TRs 4 through 9 being assigned to the task completed during that miniblock. Given the relatively slow cycling through miniblocks, no study phase TR was assigned to more than one task. For the test phase, the fMRI volumes used in the classification corresponded to the TRs during which items onset (TR 1) followed by six subsequent TRs (2 through 7). Using a series of consecutive test phase TRs allowed us to assess classifier performance as a function of time.

For display purposes, each subject's anatomical data were normalized to a standard T_1-weighted template (ICBM452; http://www.loni.ucla.edu) in Talairach space (Talairach & Tournoux, 1988). The resulting normalization parameters were also applied to the results of the individual-subject classification analyses (i.e., the importance and searchlight maps, as described below), which were resampled into 3-mm^3 voxels, in order to perform additional group-based analyses. The importance maps were smoothed with an 8-mm FWHM Gaussian kernel in order to create group-wise maps. The searchlight maps were left unsmoothed.

Pattern-classification analyses

Analyses of the fMRI data were performed with the Multi-Voxel Pattern Analysis toolbox (MVPA; Computational Memory Laboratory, Princeton, NJ;

http://www.csbmb.princeton.edu/mvpa) and SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm) in MATLAB (The MathWorks, Natick, MA). MVPA involves using neural network classifiers to determine how patterns expressed in multiple voxels of fMRI data relate to different experimental conditions. In the present implementation, classifiers were trained on study phase data and then validated on test phase data. The ability of a classifier to determine the prior encoding condition of a test item was used as the putative measure of reinstatement. The present study employed two types of classification (Whole-brain MVPA and Searchlight MVPA). Both types were conducted on an individual subject basis, with the reported results reflecting group-wise descriptors or further group analyses of the individual results.

Whole-brain MVPA—The whole-brain MVPA procedure was similar to that used previously (McDuff et al., 2009; Polyn et al., 2005). The classifier consisted of a two layer (input and output; no hidden layer) feed-forward neural network, with full connections between input and output nodes. The input layer represented the fMRI data (one node per voxel) and the output layer corresponded to the task conditions (three nodes representing Artist, Function, and Read). A feature selection procedure (see below and Supplemental Material) was used to select the voxels to be included in the classification, regardless of how these voxels were distributed throughout the brain.

Training of the classifier began by initializing the input-output connection weights to random values between 0 and 1. Each training pattern of study phase fMRI data was then submitted to the classifier in random order. Classifier output for a given training pattern was determined by a sigmoid transfer function, producing values between 0 and 1. A cross-entropy function was used to calculate the classifier's prediction error following each training pattern, based on a comparison of the actual (true) and computed outputs. For example, the actual output values for a TR corresponding to the Artist task would be 1/0/0 for Artist/Function/Read. The classifier's weights were updated with a conjugate gradient descent version of the backpropagation algorithm (for further discussion, see Bishop, 1995; Duda et al., 2001; Rumelhart et al., 1996). Training continued until either the mean prediction error across the three output nodes fell below .001 or there were 500 passes through all of the training TRs. Subsequently, classifier validation involved submitting each test phase TR of fMRI data to the trained network and noting the resulting (computed) values of the output nodes. To reduce the prediction error associated with randomly initializing the network weights, the classification was repeated 50 times for each subject, with a fresh randomization for each repetition. Results reflect the average across the 50 repetitions.

Classifier performance can be hindered by the inclusion of input data that exhibit excessive noise or are uninformative of the experimental conditions (for further discussion, see Mitchell et al., 2004, and Norman et al., 2006). To maximize performance we implemented a feature (voxel) selection procedure that restricted classifier input to only those voxels showing the largest differences among the three study tasks. Using an additional independent classifier based solely on the study phase data, the optimal number of voxels was determined to be 1000 (consistent with McDuff et al., 2009; see Supplemental Material). For each subject, voxel selection began by setting up a GLM (implemented in AFNI) that included a regressor for the convolved time course of each study task and nuisance regressors generated from spatial realignment. The F-values from an ANOVA of the parameter estimates for the three tasks were then sorted, with the 1000 voxels exhibiting the largest values selected as input data.

To identify the voxels that were most influential in determining classifier output across subjects, we created importance maps for each subject by multiplying the average value of each input node by the three weights (post-training) connecting that node to the output layer. Voxels with positive values for both activity and weight were assigned positive importance values, voxels with negative activity and weight were assigned negative importance values, and voxels for which the activity and weight had opposite signs were assigned importance values of zero (McDuff et al., 2009; cf. Polyn et al., 2005). An across-subjects average map was created for each task, following spatial normalization and smoothing (see above).

Searchlight MVPA—The second type of classification followed an information-based searchlight approach (e.g., Kriegeskorte et al., 2006). For these analyses, the fMRI data were first divided into searchlights, consisting of all voxels falling within a sphere with a radius of 2 voxels. Each searchlight thus contained 33 voxels. A searchlight was centered on each voxel in a subject's brain, truncating those searchlights at the edge of the brain so as to exclude non-brain voxels.

A separate classification was conducted for each searchlight. As in our previous classifications, the input layer of the classifier consisted of the fMRI data (one node for each of the 33 voxels) while the output layer corresponded to the encoding tasks. We found that the searchlight analysis ran too slowly when we used our standard backpropagation classification procedure, so we switched to using Gaussian Naïve Bayes (GNB) classifiers for the searchlight analysis (Mitchell et al., 2004). GNB classification runs faster than backpropagation because it computes the input-output weights analytically (in contrast to backpropagation, which sets weights via an iterative error-correction procedure).

The task outputs generated for a given searchlight were assigned to its center voxel. Thus, the searchlight results constituted whole-brain maps of outputs for each of the three study tasks and for each TR during the test phase. To simplify the results, maps corresponding to TRs 3 through 5 for a given test item were averaged into a single map, and were then averaged according to experimental conditions. After spatial normalization of the searchlight maps, they were imported into SPM5 for further group-wise analysis. All of the effects reported as significant survived a cluster-wise threshold of $p < .05$ (Worsley et al., 1996).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant sponsor: National Institutes of Health

Grant numbers: R01-MH072966 (M.D.R.) & P50-MH062196 (K.A.N.)

References

Alvarez P, Squire LR. Memory consolidation and the medial temporal lobe: a simple network model. Proc. Natl. Acad. Sci. USA 1994;91:7041–7045. [PubMed: 8041742]

Bishop, C. Neural Networks for Pattern Recognition. New York: Oxford University Press; 1995.

Coltheart M. The MRC psycholinguistic database. Q. J. Exp. Psychol. A 1981;33:497–505.

Cox R. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res 1996;29:162–173. [PubMed: 8812068]

Damasio AR. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. Cognition 1989;33:25–62. [PubMed: 2691184]

Davachi L, Mitchell JP, Wagner AD. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. Proc. Natl. Acad. Sci. USA 2003;100:2157–2162. [PubMed: 12578977]

Donaldson W. The role of decision processes in remembering and knowing. Mem. Cognit 1996;24:523–533.

Duda, R.; Hart, P.; Stork, D. Pattern Classification. Vol. 2nd Edition. New York: Wiley; 2001.

Dunn JC. Remember-know: A matter of confidence. Psychol. Rev 2004;111:524–542. [PubMed: 15065921]

Düzel E, Habib R, Rotte M, Guderian S, Tulving E, Heinze HJ. Human hippocampal and parahippocampal activity during visual associative recognition memory for spatial and nonspatial stimulus configurations. J. Neurosci 2003;23:9439–9444. [PubMed: 14561873]

Dzulkifli MA, Wilding EL. Electrophysiological indices of strategic episodic retrieval processing. Neuropsychologia 2005;43:1152–1162. [PubMed: 15817173]

Greenhouse GW, Geisser S. On methods in the analysis of repeated measures designs. Psychometrika 1959;49:95–112.

Haynes JD, Rees G. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci 2006;7:523–534. [PubMed: 16791142]

Holm S. A simple sequentially rejective multiple test procedure. Scand. J. Stat 1979;6:65–70.

Johnson JD, Minton BR, Rugg MD. Content dependence of the electrophysiological correlates of recollection. NeuroImage 2008;39:406–416. [PubMed: 17933555]

Johnson JD, Rugg MD. Recollection and the reinstatement of encoding-related cortical activity. Cereb. Cortex 2007;17:2507–2515. [PubMed: 17204822]

Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychol. Bull 1993;114:3–28. [PubMed: 8346328]

Johnson MK, Kounios J, Nolde SF. Electrophysiological brain activity and memory source monitoring. Neuroreport 1997;8:1317–1320. [PubMed: 9175136]

Kahn I, Davachi L, Wagner AD. Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. J. Neurosci 2004;28:4172–4180. [PubMed: 15115812]

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. Proc. Natl. Acad. Sci. USA 2006;103:3863–3868. [PubMed: 16537458]

Kucera, H.; Francis, WN. Computational Analysis of Present-Day American English. Providence, RI: Brown University Press; 1967.

Mandler G. Recognizing: The judgment of previous occurrence. Psychol. Rev 1980;87:252–271.

Maratos EJ, Dolan RJ, Morris JS, Henson RNA, Rugg MD. Neural activity associated with episodic memory for emotional context. Neuropsychologia 2001;39:910–920. [PubMed: 11516444]

McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev 1995;102:419–457. [PubMed: 7624455]

McDuff SG, Frankel HC, Norman KA. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. J. Neurosci 2009;29:508–516. [PubMed: 19144851]

Mitchell T, Hutchinson R, Niculescu S, Pereira F, Wang X, Just M, Newman S. Learning to decode cognitive states from brain images. Mach. Learn 2004;57:145–175.

Mur M, Bandettini PA, Kriegeskorte N. Revealing representational content with pattern-information fMRI—an introductory guide. Soc. Cogn. Affect. Neurosci 2009;4:101–109. [PubMed: 19151374]

Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev 2003;110:611–646. [PubMed: 14599236]

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci 2006;10:424–430. [PubMed: 16899397]

O'Reilly RC, Norman KA. Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. Trends Cogn. Sci 2002;6:505–510. [PubMed: 12475710]

O'Reilly RC, Rudy JW. Conjunctive representations in learning and memory: principles of cortical and hippocampal function. Psychol. Rev 2001;108:311–345. [PubMed: 11381832]

Polyn SM, Natu VS, Cohen JD, Norman KA. Category-specific cortical activity precedes retrieval during memory search. Science 2005;310:1963–1966. [PubMed: 16373577]

Rolls ET. Hippocampo-cortical and cortico-cortical backprojections. Hippocampus 2000;10:380–388. [PubMed: 10985277]

Rubin DC, Greenberg DL. Visual memory-deficit amnesia: a distinct amnesic presentation and etiology. Proc. Natl. Acad. Sci. USA 1998;95:5413–5416. [PubMed: 9560290]

Rugg MD, Johnson JD, Park H, Uncapher MR. Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspective. Prog. Brain Res 2008;169:339–352. [PubMed: 18394485]

Rumelhart, D.; Durbin, R.; Golden, R.; Chauvin, Y. Backpropagation: the basic theory. In: Chauvin, Y.; Rumelhart, D., editors. Backpropagation: Theory, Architectures, and Applications. Mahwah, NJ: Erlbaum; 1996. p. 1-34.

Shastri L. Episodic memory and cortico-hippocampal interactions. Trends Cogn. Sci 2002;6:162–168. [PubMed: 11912039]

Stark CEL, Okado Y. Making memories without trying: Medial temporal lobe activity associated with incidental memory formation during recognition. J. Neurosci 2003;23:6748–6753. [PubMed: 12890767]

Stern CE, Corkin S, Gonzalez RG, Guimaraes AR, Baker JR, Jennings PJ, Carr CA, Sugiura RM, Vedantham V, Rosen BR. The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. Proc. Natl. Acad. Sci. USA 1996;93:8660–8665. [PubMed: 8710927]

Talairach, J.; Tournoux, P. Co-planar Stereotaxic Atlas of the Human Brain. 3-dimensional Proportional System: An Approach to Cerebral Imaging. New York: Thieme Medical Publishers; 1988.

Tulving E. Memory and consciousness. Can. Psychol 1985;26:1–12.

Wais PE, Mickes L, Wixted JT. Remember/know judgments probe degrees of recollection. J. Cogn. Neurosci 2008;20:400–405. [PubMed: 18004949]

Wheeler ME, Buckner RL. Functional-anatomic correlates of remembering and knowing. NeuroImage 2004;21:1337–1349. [PubMed: 15050559]

Wilson M. The MRC psycholinguistic database: machine readable dictionary. Behav. Res. Methods Instrum. Comput 1988;20:6–11.

Wixted JT. Dual-process theory and signal-detection theory of recognition memory. Psychol. Rev 2007;114:152–176. [PubMed: 17227185]

Wixted JT, Stretch V. In defense of the signal detection interpretation of remember/know judgments. Psychon. Bull. Rev 2004;11:616–641. [PubMed: 15581116]

Woodruff CC, Johnson JD, Uncapher MR, Rugg MD. Content-specificity of the neural correlates of recollection. Neuropsychologia 2005;43:1022–1032. [PubMed: 15769488]

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp 1996;4:58–73.

Yonelinas AP. The nature of recollection and familiarity: a review of 30 years of research. J. Mem. Lang 2002;46:441–517.

Yonelinas AP, Otten LJ, Shaw KN, Rugg MD. Separating the brain regions involved in recollection and familiarity in recognition memory. J. Neurosci 2005;25:3002–3008. [PubMed: 15772360]

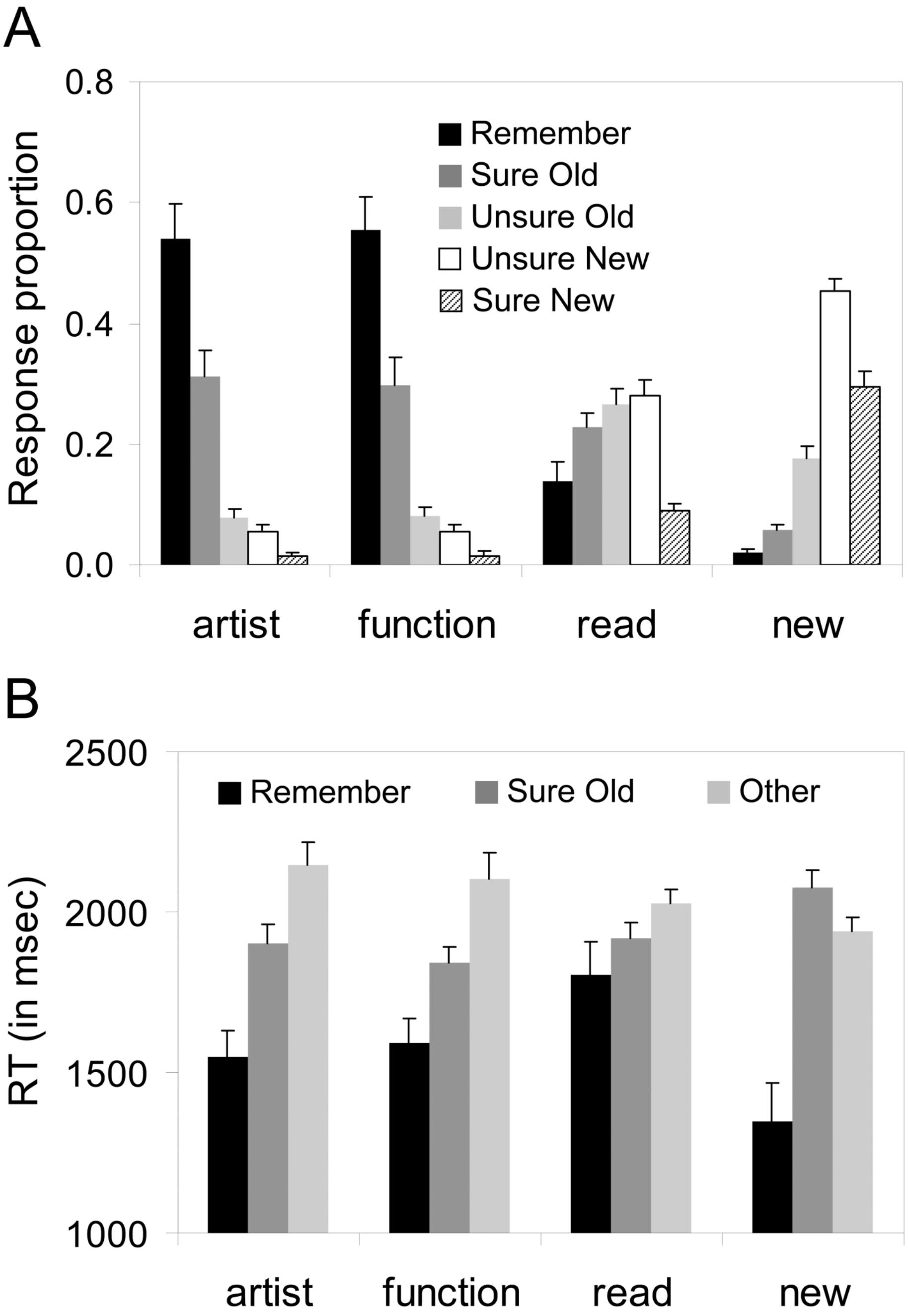

Figure 1. Behavioral Performance

(A) Mean (+SEM) proportions of responses according to the test item condition. (B) Mean (+SEM) response time (RT) data. The Other category reflects collapsed Unsure Old, Unsure New, and Sure New responses (due to low individual trial numbers). The RT data for Remember responses to new items are based on only 12 subjects contributing such responses.

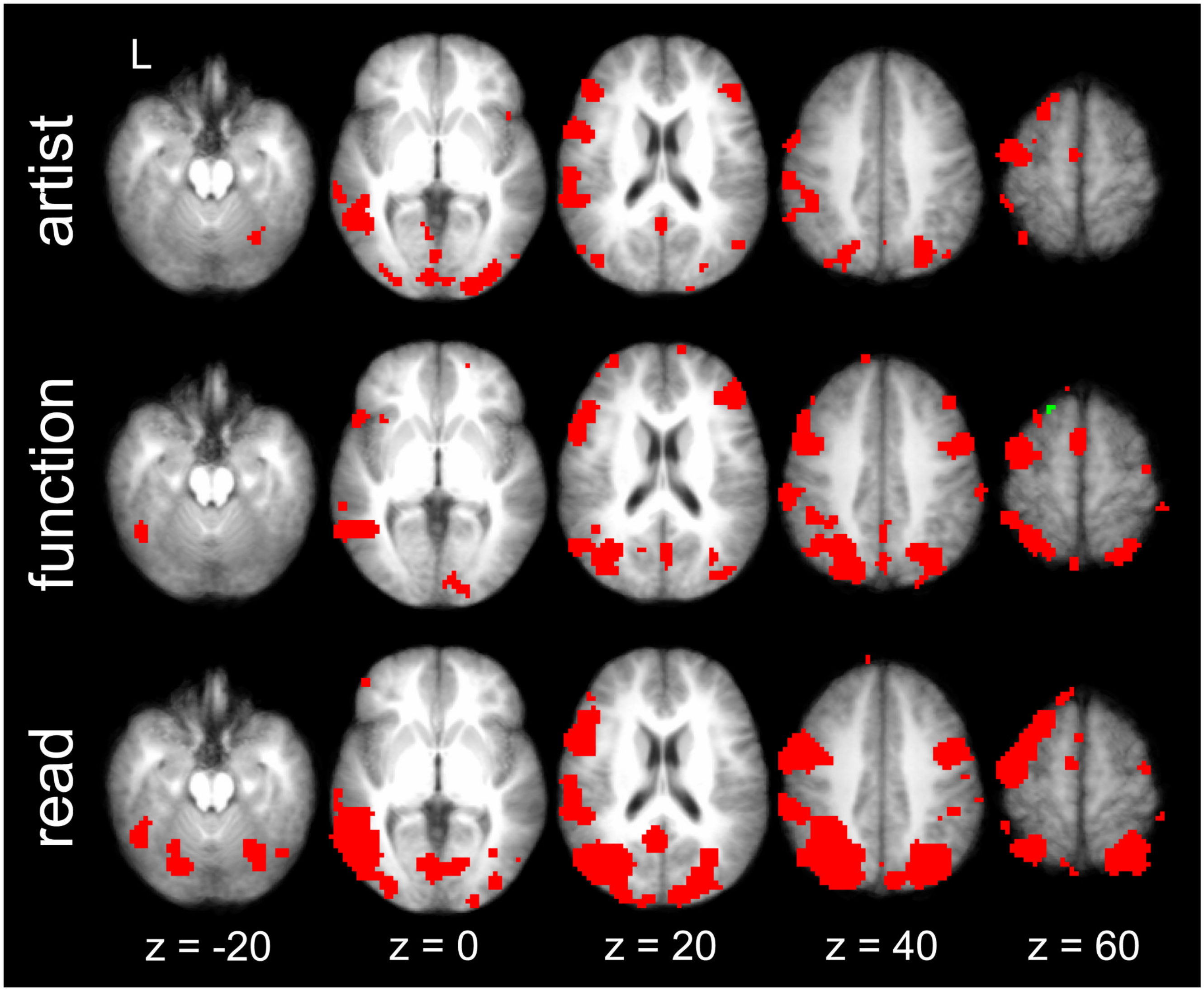

Figure 2. Importance Maps

Group mean importance maps for the three study tasks, overlaid on axial slices of the mean normalized anatomical data (coordinates in Talairach space). The colored areas depict voxels where importance values exceeded arbitrary thresholds of .001 positively (red) and -.001 negatively (green; see middle row, right-most column). L = left.

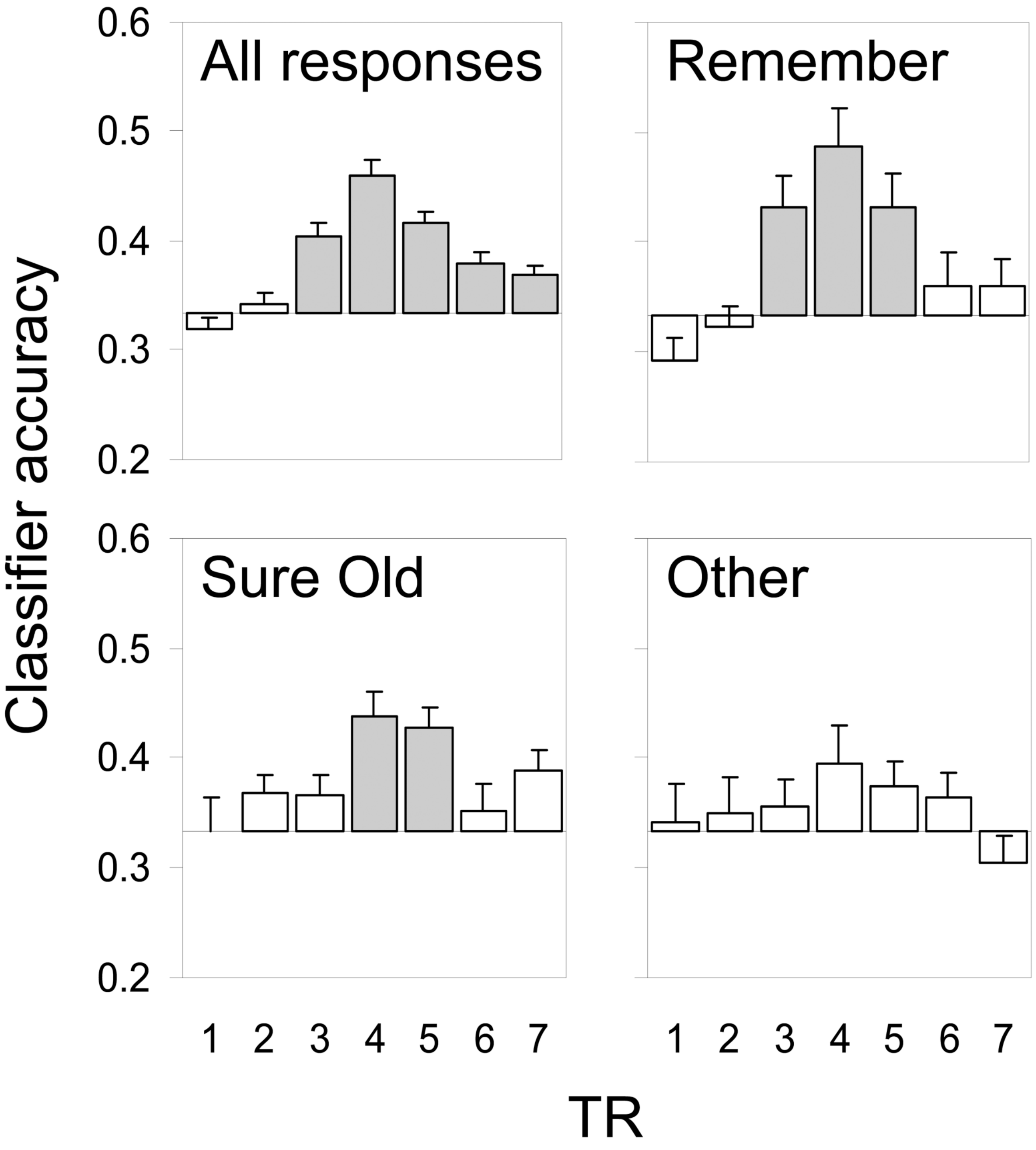

Figure 3. Classifier Accuracy

Mean classifier accuracy (+SEM) collapsed across all response categories and separated by response category. Time point (TR) 1 corresponds to test item onset. Shaded bars indicate the TRs during which classifier accuracy was significantly above chance (.33; correcting for multiple comparisons).

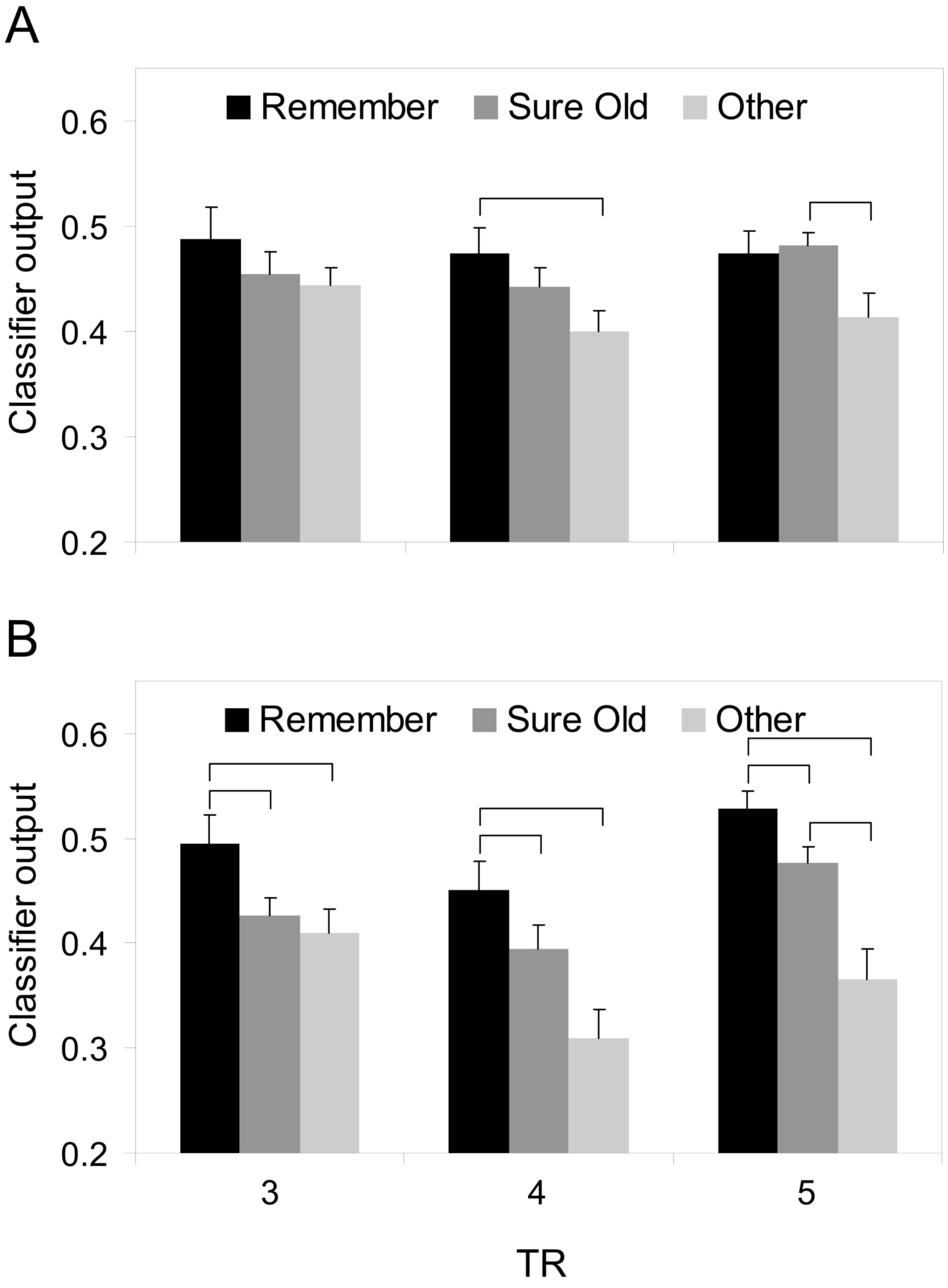

Figure 4. Classifier Output

Mean values (+SEM) of the classifier's correct output node, (A) averaged over all three study tasks, and (B) over only the Artist and Function tasks. Each bar reflects classifier output for a given response category and time point (TR). Brackets indicate significant differences between responses (correcting for multiple comparisons).

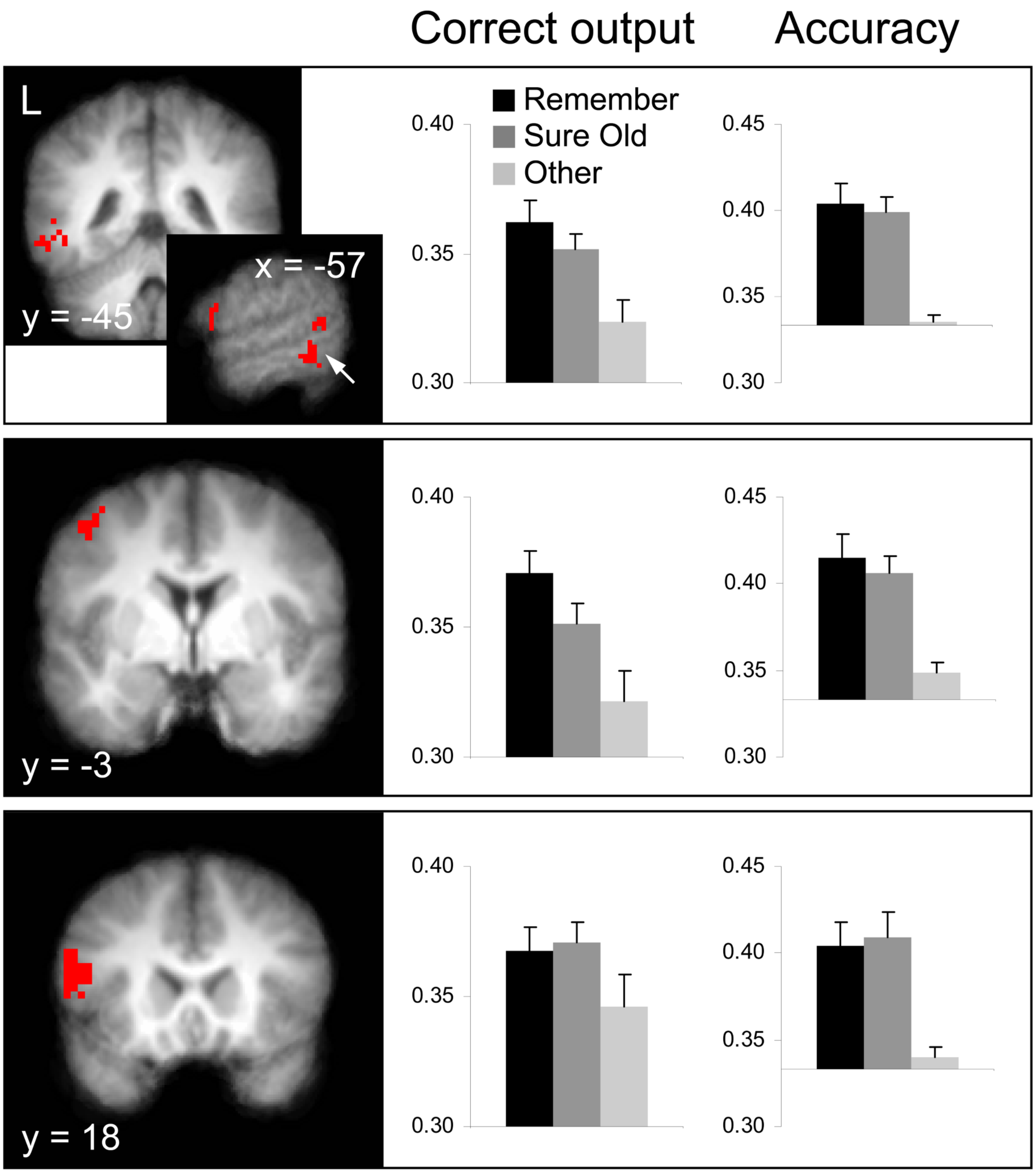

Figure 5. Equivalent Reinstatement Effects

Results of searchlight analyses where reinstatement was equivalent for test items designated with Remember and Sure Old responses (see main text for details of the contrast procedure). Histograms reflect the mean (+SEM) output values at the correct classifier node (left column) and classifier accuracy (right column; chance = .33) within the depicted clusters in lateral temporal cortex, superior frontal gyrus, and inferior frontal gyrus. All effects depicted here survived a cluster-wise threshold of p < .05 and are overlaid on the mean anatomical image (coordinates in Talairach space). L = left.

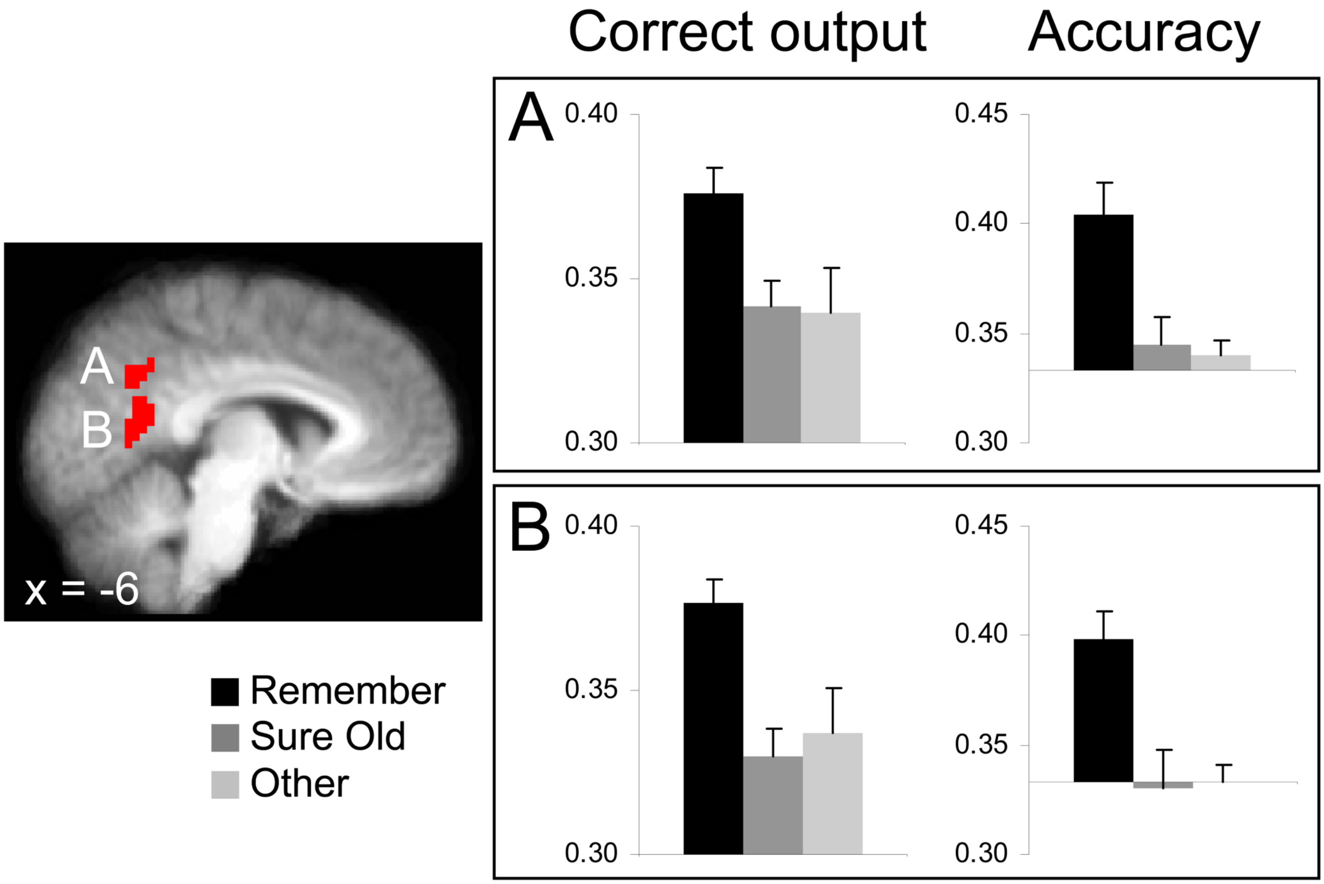

Figure 6. Selective Reinstatement Effects

Results of searchlight analyses showing selective reinstatement for test items designated with Remember responses (compared to Sure Old responses; see main text for contrast procedure). The histograms provide the mean (+SEM) output value at the correct classifier node and the mean classifier accuracy within the depicted clusters of (A) posterior cingulate and (B) retrosplenial cortex. Both effects survived a cluster-wise threshold of p < .05. See Figure 5 caption for further display details.