

Neural correlates of actual and predicted memory formation

Yun-Ching Kao¹, Emily S Davis¹ & John D E Gabrieli^{1,2}

We aimed to discover the neural correlates of subjective judgments of learning—whereby participants judge whether current experiences will be subsequently remembered or forgotten—and to compare these correlates to the neural correlates of actual memory formation. During event-related functional magnetic resonance imaging, participants viewed 350 scenes and predicted whether they would remember each scene in a later recognition-memory test. Activations in the medial temporal lobe were associated with actual encoding success (greater activation for objectively remembered than forgotten scenes), but not with predicted encoding success (activations did not differ for scenes predicted to be remembered versus forgotten). Conversely, activations in the ventromedial prefrontal cortex were associated with predicted but not actual encoding success, and correlated with individual differences in the accuracy of judgments of learning. Activations in the lateral and dorsomedial prefrontal cortex were associated with both actual and predicted encoding success. These findings indicate specific dissociations and associations between the neural systems that mediate actual and predicted memory formation.

A critical aspect of learning is knowing how to learn. Knowing how to learn reflects an interaction between memory processes that encode experience into long-term memory and introspective (or metamemory) processes that evaluate whether information has been learned successfully. Such judgments of learning (JOLs) guide the allocation of cognitive and mnemonic resources so that information that has been sufficiently learned is no longer studied, whereas information that has not yet been successfully learned can be further encoded into long-term memory. Functional neuroimaging studies have delineated neural systems that seem to determine whether information is successfully or unsuccessfully learned during encoding^{1,2}, but there is no evidence as yet about the neural systems that, during study, mediate judgments about whether information has or has not been learned.

JOLs are known to influence the success of learning^{3–5}. Studies of JOL ask people to judge, during encoding, whether particular pieces of information or stimuli are successfully or unsuccessfully encoded (that is, whether they are likely to be remembered or forgotten in a later test of retention). Superior JOLs are associated with superior learning: students with higher scholastic performance evaluate their learning and predict their test performance better than students with lower scholastic performance scores^{3,4}. The effects of JOLs on successful learning are most potent in situations when people are given the most opportunity to select study strategies, such as self-paced learning, and least potent when there is minimal opportunity for this, as in experimenter-paced learning⁶. With training, people can improve their ability to accurately assess what they will remember or forget, especially under self-paced learning circumstances⁷.

The unknown neural circuitry that mediates JOLs may be the same as that known to mediate learning itself¹, or there may be some distinction between memory systems that encode experience and the metamemory systems that evaluate the success of that encoding. Indeed, psychopharmacological and neuropsychological studies indicate that these memory and metamemory processes depend upon at least partially distinct neural circuitry^{8–10}. Nitrous oxide and the benzodiazepine lorazepam both impair memory performance, but do not impair the ability to form accurate JOLs^{8,9}. Participants given lorazepam perform poorly in recalling previously learned word pairs, but actually perform as well as those given a placebo in forming accurate JOLs about the likelihood of later recall⁹.

Neuropsychological studies have also reported dissociations between JOLs and memory performance. Individuals with frontal lobe lesions are impaired at making JOLs, despite normal recognition-memory performance^{10,11}. In one study, individuals with unspecified frontal lesions and those with a variety of posterior lesions memorized a 4 × 4 matrix of faces¹⁰. The participants predicted the number of faces they would later be able to match to the original matrix location. Whereas subjects with right frontal lesions showed better memory performance than did those with right posterior lesions, they were less accurate in making memory predictions. These results suggest that JOLs may be dissociable from actual encoding success for experimenter-paced learning. Furthermore, these results point to the potential importance of prefrontal cortex (PFC) in forming accurate JOLs.

Currently, there is no evidence about the neural basis of JOLs in normal healthy adults. Although little is known about the neural systems supporting the prediction of encoding success, there is

¹Department of Psychology, Stanford University, 420 Jordan Hall, Stanford, California 94305, USA. ²Present address: Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building 46-4033, Cambridge, Massachusetts 02139, USA. Correspondence should be addressed to Y.-C.K. (ykao@psych.stanford.edu).

Received 13 September; accepted 12 October; published online 13 November 2005; doi:10.1038/nn1595

mounting evidence about the neural systems that mediate actual encoding success. The ability to successfully encode experiences into long-term memory has been associated with the functioning of medial temporal lobe (MTL) and PFC. MTL structures are essential for the formation of new declarative memories^{12,13}, and MTL activation is greater during the encoding of scenes¹ and words² that will later be remembered than for those that will later be forgotten. Activations in PFC during encoding also predict subsequent remembering^{1,2,14}. One interpretation is that PFC contributes to memory formation by supporting semantic elaborations and executive operations that monitor, regulate and facilitate memory processes¹⁵. Thus, convergent behavioral evidence from patients with frontal lesions and imaging evidence suggest that JOLs, which involve the self-monitoring of encoding processes, may also depend on PFC functioning.

Although neuropsychological studies have implicated PFC as important in JOL accuracy, it is unknown which specific subregions support JOLs. One clue comes from a neuropsychological study of a metamemory judgment made at retrieval, the feeling-of-knowing (FOK) judgment¹⁶. In FOK studies, people are asked whether they feel they would recognize information that they have failed to recall. The accuracy of the FOK can be gauged by then asking a person to actually make the recognition memory. Individuals with the most inaccurate FOK judgments had in common damage to the ventromedial prefrontal cortex (VMPFC)¹⁶. The finding that VMPFC is essential in FOK metamemory judgment at retrieval raises the possibility that the same brain region participates in JOL metamemory judgments at encoding.

In the present study, we used event-related functional magnetic resonance imaging (fMRI) to investigate the neural basis of JOLs. Specifically, we examined whether subjective predictions of encoding success (JOLs) depend on the same or different neural circuits underlying actual encoding success. Participants were scanned while predicting whether or not scenes were successfully encoded (that is, whether they would be later remembered or forgotten) (Fig. 1). Afterwards, outside the scanner, participants were given an old or new recognition-memory test. For analysis, items were sorted on the basis of whether they were given a “will remember” (R) or a “will forget” (F) JOL, and whether they were later actually remembered (r) or forgotten (f) during the recognition-memory test. This design allowed us to investigate the neural substrates underlying predicted encoding success (JOLs) compared to actual encoding success (learning itself). We found specific brain regions that were related exclusively to actual memory formation, or exclusively to predicted memory formation.

RESULTS

Task performance

Participants made an average of 144 (s.d. = 51) “will remember” (R) and 104 (s.d. = 43) “will forget” (F) predictions. There was no significant difference in either the propor-

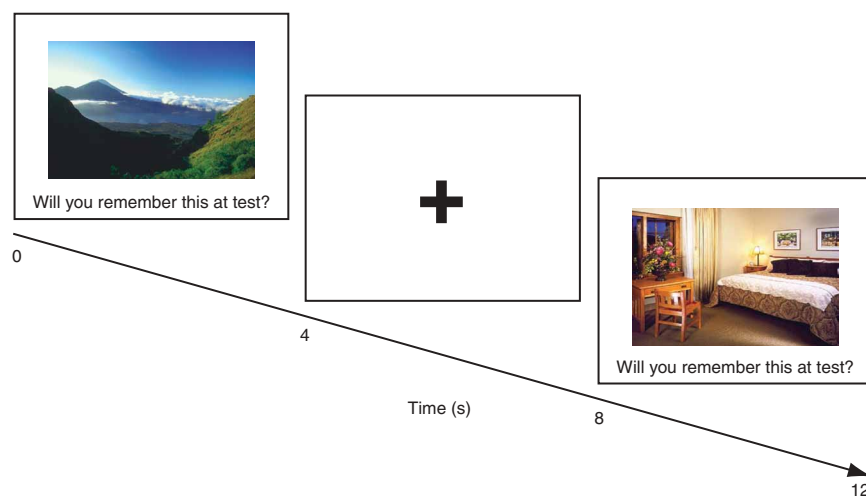


Figure 1 Task design. Scenes and fixations were presented for 4 s. For each scene, participants made judgments of learning by predicting whether or not they would remember the scene in a later recognition-memory test.

tion ($t_{15} = 1.90$, $P > 0.05$) or latency ($t_{15} = 0.91$, $P > 0.05$) of R versus F responses. In the post-scan recognition test, participants made accurate old or new judgments 71% of the time (s.d. = 7%) and had a mean d' -prime of 0.83 (s.d. = 0.33). (D-prime is a measure of recognition memory sensitivity, independent of decision criteria.) Using participants' responses on the post-scan memory test, we sorted trials at study based on predicted encoding success and actual encoding success. Items remembered with low confidence were excluded from all analyses to minimize the influence of guessing; however, items forgotten with low confidence were not excluded because such exclusion would have resulted in an insufficient number of misses for data analysis. During encoding, scenes were given either R or F predictions and were either subsequently remembered with high confidence (r) or subsequently forgotten (f). Thus, there were four possible trial outcomes (Fig. 2a,b): (i) scenes were given a “will remember” prediction and were later remembered (Rr), (ii) scenes were given a “will remember” prediction but were later forgotten (Rf), (iii) scenes were given a “will forget” prediction but were later remembered (Fr) and (iv) trials were given a “will forget” prediction and were later forgotten (Ff). A 2×2 repeated-measures analysis of variance (ANOVA) revealed no significant main effects for prediction or retrieval outcome, indicating suitable and unbiased measurement per trial type.

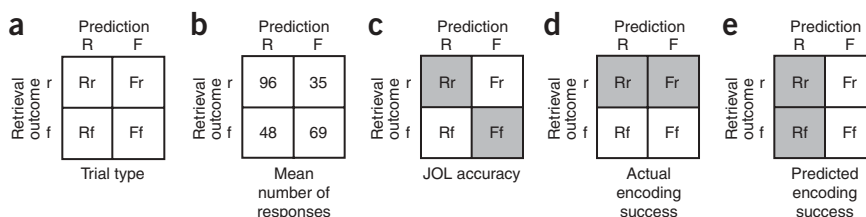


Figure 2 Schematics of the four trial types and trial combinations for statistical analyses. (a) At study, there were two possible JOL predictions: “will remember” (R) and “will forget” (F). At the recognition-memory test, scenes were either remembered (r) or forgotten (f). (b) Mean numbers of responses for the four trial types. (c) JOL accuracy was assessed by comparing correct JOL trials (gray) to incorrect JOL trials (white). (d) In the actual encoding success analysis, scenes that were later remembered (gray) were compared to scenes that were later forgotten (white). (e) In the predicted encoding success analysis, R predictions (gray) were compared to F predictions (white).

Table 1 Brain regions associated with actual and predicted encoding success

Study task	MNI coordinates			Peak Z-score	Cluster size	Region	BA	Peak Z-score
	x	y	z					
Actual encoding success	-22	-72	45	4.74	97	L	Parietal	7
	38	-60	-15	4.72	135	R	Fusiform	37
	45	6	22	4.34	11	R	Inferior frontal (lateral PFC)	44/6
	-45	-66	-4	4.27	142	L	Fusiform	37
	38	-84	15	4.07	78	R	Middle temporal	19
	-34	30	-22	3.98	12	L	Inferior frontal	47
	-41	6	26	3.87	5	L	Inferior frontal (lateral PFC)	44/6
	15	-54	15	3.47	5	R	Posterior cingulate	30
	34	-48	49	3.35	6	R	Parietal lobule	40
Predicted encoding success	-22	12	56	4.67	153	L	Superior frontal	6
	-41	-66	0	4.18	64	L	Middle occipital	37
	22	30	-15	4.07	58	R	Middle frontal	11
	-22	-72	34	4.23	53	L	Precuneus	7
	30	-54	34	4.02	49	R	Precuneus	7
	-11	42	-22	4.45	26	L	Inferior frontal (VMPFC)	11/47
	-45	36	19	4.08	17	L	Inferior frontal (lateral PFC)	44/6
	-15	0	-22	3.49	10	L	Amygdala	34
	49	-66	-4	4.79	8	R	Inferior temporal	37
	-41	24	-19	3.86	7	L	Inferior frontal	47
	52	-78	8	3.46	5	R	Middle temporal	39

Only clusters of five or more voxels and a significance of $P < 0.001$ uncorrected are reported. BA, Brodmann's area; PFC, prefrontal cortex; VMPFC, ventromedial prefrontal cortex; L, left; R, right.

We calculated JOL accuracy to examine whether participants could reliably assess their learning to predict future retrieval performance. Correct JOL trials consisted of trials in which JOL predictions matched retrieval outcomes (Fig. 2c). Participants made JOL predictions with above-chance accuracy (mean \pm s.d. $64 \pm 5\%$). To investigate whether participants could discriminate between well-learned and

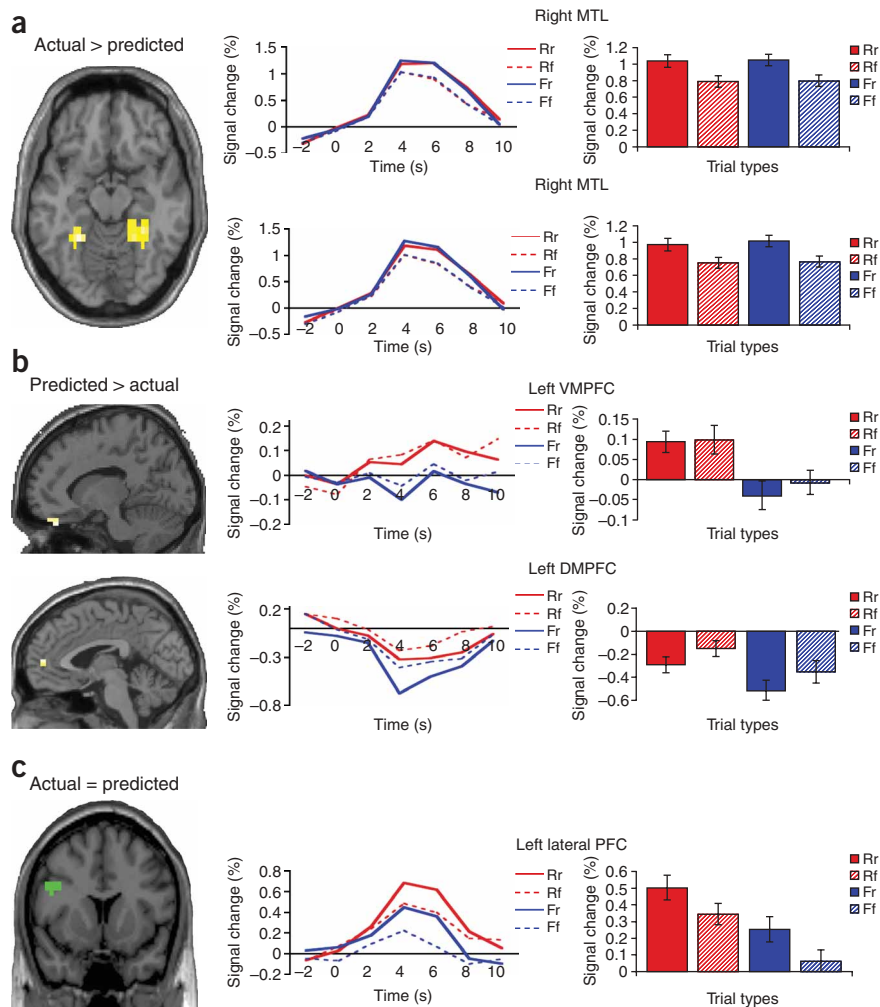
poorly-learned items, the Goodman-Kruskal gamma was calculated for each participant. The gamma statistic indicates the strength of association between two ordinal variables and is the preferred measure in metamemory research¹⁷. The mean gamma score was 0.53 (s.d. = 0.04) and was statistically greater than zero ($t_{15} = 12.64$, $P < 0.001$), indicating that participants were able to discriminate between items

Table 2 Distinct and overlapping regions between actual and predicted encoding success

Study task	MNI coordinates			Peak Z-score	Cluster size	Region	BA	
	x	y	z					
Actual > predicted	-26	-48	-11	4.62	35	L	Fusiform/parahippocampal	37
	-26	-66	-11	4.17	5	L	Posterior cingulate	31
	30	-42	-19	4.16	44	R	Fusiform/parahippocampal	37
Predicted > actual	-52	6	15	3.73	5	L	Precentral	44
	56	-24	-4	3.71	11	R	Superior temporal	21
	30	12	-19	3.68	17	R	Inferior frontal	47
	-11	42	-26	3.61	5	L	Orbital frontal (VMPFC)	11
	0	54	8	3.61	7	L	Medial frontal (DMPFC)	10
	4	42	-4	3.40	7	R	Anterior cingulate	32
	-26	12	64	3.32	5	L	Middle frontal	6
Actual = predicted	-41	6	26	3.87	18	L	Inferior frontal (lateral PFC)	44/6
	-22	-72	41	4.23	32	L	Superior parietal lobule	7
	-45	-66	4	4.27	44	L	Inferior temporal gyrus	37
	26	-72	34	3.49	28	R	Precuneus	19
	-22	6	49	3.18	14	L	Cingulate gyrus	32

Only clusters of five or more voxels and a significance of $P < 0.005$ uncorrected are reported. BA, Brodmann's area; VMPFC, ventromedial prefrontal cortex; DMPFC, dorsomedial prefrontal cortex; PFC, prefrontal cortex; L, left; R, right.

Figure 3 Statistical activation maps and percent signal change. Activation maps are rendered onto the MNI normalized canonical single-subject brain. Line graphs represent the percent signal change in brain activation as a function of time, for each of the four trial types. Bar graphs represent the mean percent signal change from 4 s to 8 s after stimulus presentation, for each of the four trial types. Error bars indicate s.e.m. R, "will remember" predictions; r, later remembered; f, later forgotten. (a) Regions of interest (ROIs) defined from actual > predicted encoding success contrast. In bilateral medial temporal lobe (MTL), only the main effect of actual encoding success was significant (solid lines above dotted lines). (b) ROIs defined from predicted > actual encoding success contrast. Ventromedial prefrontal cortex (VMPFC) and dorsomedial prefrontal cortex (DMPFC) showed a significant main effect for predicted encoding success (red lines above blue lines) but not actual encoding success. However, the main effect for actual encoding success showed a trend toward significance in DMPFC. (c) ROIs defined from regions in which predicted encoding success matched actual encoding success. Left lateral PFC showed significant main effects for both actual and predicted encoding success. Coordinates in **Table 2**.



that would be later remembered and items that would be later forgotten with above-chance accuracy. Response times for correct and incorrect JOLs differed such that correct JOL predictions were made faster than incorrect JOL predictions (mean response times of 1,509 and 1,574 ms, respectively; $t_{15} = 2.73$, $P < 0.05$). Pearson's correlation revealed a significant correlation between participants' JOL accuracy (gamma score) and their recognition-memory performance (d -prime) ($r_{15} = 0.52$, $P < 0.05$).

Neural correlates of actual encoding success

To assess the neural correlates of actual encoding success, we identified regions in which activation was significantly greater during encoding for scenes subsequently remembered with high confidence than for scenes subsequently forgotten (Rr and Fr trials > Rf and Ff trials) (**Fig. 2d**). Activation was significantly greater for subsequently remembered scenes than for subsequently forgotten scenes in bilateral MTL regions including posterior parahippocampal and fusiform gyri, but not in the hippocampus proper. In addition, there was greater activation for remembered than for forgotten scenes in bilateral inferior frontal gyrus (lateral PFC) corresponding to Brodmann's area 44/6 and in right posterior cingulate gyrus (**Table 1**). There were no significant activations for the opposite contrast (subsequently forgotten scenes compared to subsequently remembered scenes).

Neural correlates of predicted encoding success

We assessed predicted encoding success by comparing R predictions to F predictions across actual retrieval outcomes (**Fig. 2e**). This contrast yielded significant activations ($P < 0.001$) in left lateral PFC, left VMPFC corresponding to Brodmann's area 11/47, left amygdala, right

middle temporal and right inferior temporal regions, bilateral precuneus and left middle occipital regions (**Table 1**). The opposite contrast, comparing F prediction trials to R predictions trials, did not reveal any significant brain activations.

Comparing actual and predicted encoding success

To identify brain regions more specifically related to either actual encoding success or predicted encoding success, we carried out paired t -tests using the two contrasts discussed above. Performing a paired t -test is analogous to performing a linear contrast comparing Fr trials (which are associated with actual encoding success but not predicted encoding success) to Rf trials (which are associated with predicted encoding success but not actual encoding success).

The brain regions with greater activation for actual encoding success than predicted encoding success included bilateral MTL and left posterior cingulate (**Table 2**). The bilateral MTL regions included parahippocampal and fusiform gyri (**Fig. 3a**). The bilateral MTL regions were submitted to regions-of-interest (ROI) analyses. Peak percent signal change for all four trial types were entered into 2×2 repeated-measures ANOVA to assess the relationship between actual encoding success (retrieval outcome) and predicted encoding success (JOL predictions). The ANOVA revealed a significant main effect for actual encoding success in both the right and left MTL regions (right MTL, $F_{1,15} = 25.37$, $P < 0.001$; left MTL,

$F_{1,15} = 15.58, P < 0.001$). Subsequently remembered trials showed greater activation than did subsequently forgotten trials (Fig. 3a). However, neither the main effect for predicted encoding success nor the interaction effect reached significance. This suggests that these MTL regions are sensitive to actual encoding success and not to predicted encoding success.

Conversely, several brain regions were more activated for predicted than actual encoding success. These regions included prefrontal cortices, specifically left VMPFC, left dorsomedial prefrontal cortex (DMPFC) corresponding to Brodmann's area 10, right inferior frontal gyrus, left precentral gyrus and right anterior cingulate gyrus (Table 2). In the ROI analysis for VMPFC and DMPFC (Fig. 3b), there was a significant main effect for prediction, such that R predictions resulted in more activation than did F predictions (VMPFC, $F_{1,15} = 28.54, P < 0.001$; DMPFC, $F_{1,15} = 11.57, P < 0.001$). The main effect of actual encoding success and the interaction effect did not reach significance, although the main effect of actual encoding success showed a trend toward significance in DMPFC ($F_{1,15} = 3.71, P = 0.07$). Significant effects in VMPFC reflected increases in activation relative to baseline, whereas effects in DMPFC reflected decreases in activation relative to baseline¹⁸. VMPFC seemed to have a role in predicted encoding success, but not in actual encoding success. In contrast, DMPFC seemed to respond to both predicted and actual encoding success.

In the above analyses, no region showed greater activation for correct predictions than for incorrect predictions (Rr and Ff trials > Rf and Fr trials). We performed a voxelwise analysis to search for any such interactions, and neither this nor the reverse interaction yielded significant activations (even at a lowered threshold of $P < 0.01$).

We used a masking procedure to identify brain regions associated with both actual and predicted encoding success. Both the actual and predicted encoding success contrasts showed significant activations in left lateral PFC, left superior parietal lobule, bilateral inferior temporal gyrus and bilateral precuneus (Table 2), suggesting that these regions may mediate both actual and predicted encoding success. In the ROI analysis of left lateral PFC (Fig. 3c), there were significant main effects for both actual and predicted encoding success ($F_{1,15} = 27.22, P < 0.001$; and $F_{1,15} = 13.38, P < 0.001$, respectively; the interaction was not significant, $F_{1,15} = 0.109, P = 0.75$). The pattern of activation from the left lateral PFC region was representative of the other significant clusters.

Individual differences in JOL accuracy

In addition to identifying separable networks for actual and predicted encoding success, we were also interested in assessing which brain regions support the ability to make accurate predictions. The contrast between correct and incorrect predictions did not yield any significant clusters. Therefore, to investigate whether differences in brain activations might reflect individual differences in JOL accuracy, we submitted the four ROIs identified in the previous analyses to correlational analyses: MTL regions identified in the actual > predicted encoding success analysis, VMPFC and DMPFC regions identified in the predicted > actual encoding success analysis and the left lateral PFC region identified in the predicted = actual encoding success analysis. Of the four ROIs, only VMPFC activation was significantly correlated with variation in JOL accuracy (Fig. 4). Specifically, gamma scores were positively correlated with VMPFC activation on trials given accurate predictions (Fig. 4a; Rr trials, $r_{15} = 0.58, P < 0.01$; Ff trials, $r_{15} = 0.43, P < 0.05$). However, gamma scores were not correlated with activation on trials given inaccurate predictions (Rf and Fr trials) (Fig. 4b). Thus, individuals who made more accurate JOL predictions showed greater VMPFC activations. Additional correlation analyses between brain

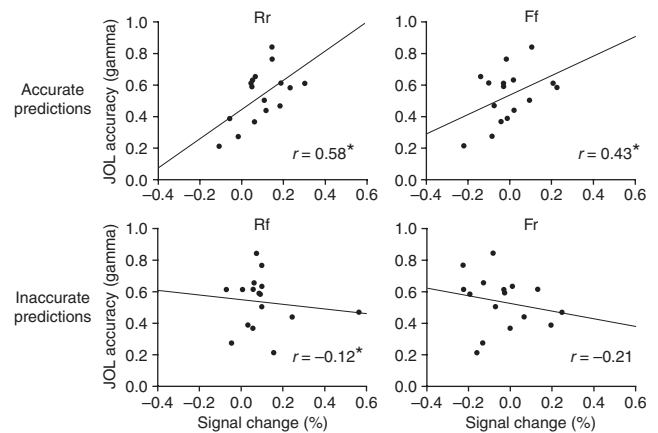


Figure 4 Ventromedial prefrontal cortex (VMPFC) and individual differences in JOL accuracy. The scatter diagrams are plots of individual gamma scores of JOL accuracy as a function of percent signal change in VMPFC, for each of the four trial types. R, “will remember” predictions; F, “will forget” predictions; r, later remembered; f, later forgotten. * $P < 0.05$.

activations and recognition-memory performance, as indexed by d-prime, did not reach significance for any of the four ROIs.

DISCUSSION

The present study investigated the neural basis of judgments of learning (JOLs) by asking participants to study scenes and predict whether or not each scene would be later remembered. The goals of this study were to examine the brain regions supporting the predictions of encoding success (participants’ judgments that they had successfully learned a scene) and compare those regions to brain regions supporting actual encoding success (whether participants actually remembered that scene when tested). We identified four patterns of brain-behavior relationships in brain regions known to be important for memory, thought and introspection.

First, actual encoding success engaged MTL regions (that is, activations were greater during the encoding of subsequently remembered scenes than that of subsequently forgotten scenes). MTL activations were unrelated to the predicted encoding success. This suggests that the MTL processes engaged during encoding reflect memory but not JOL processes. Second, predicted encoding success (that is, greater activation for the scenes that subjects predicted they would remember than for the scenes subjects predicted they would forget) engaged a network of PFC regions including medial regions such as VMPFC and DMPFC. This suggests that medial PFC processes engaged during encoding may support JOL processes. Third, JOL accuracy—the strength of the relation between predicted and actual encoding success—may involve processes mediated by VMPFC: participants who made more accurate JOL predictions showed greater VMPFC activations than participants who made less accurate predictions. Fourth, both actual and predicted encoding success engaged left lateral PFC. Thus, there were both dissociations and associations between the neural systems that seemed to be engaged by objective memory encoding and subjective appraisals of that encoding.

MTL and actual encoding success

MTL regions seemed to support actual encoding success, but not the subjective evaluation of encoding success. There is substantial evidence that the hippocampus and surrounding cortices are crucial for declarative memory formation^{12,13}. Individuals such as H.M. who have

extensive MTL injury show a severe anterograde amnesia that includes an apparent inability to successfully encode new experiences¹³. In normal adults, neuroimaging studies have frequently shown MTL activation during episodic encoding tasks, with greater activation for items successfully than for those unsuccessfully encoded^{1,2}. For complex scenes, MTL activations associated with successful encoding have sometimes included the hippocampus¹⁹, but have typically been most robust in the parahippocampal and fusiform gyri^{1,14}. This may reflect parahippocampal specialization for visual-spatial memory²⁰. Thus, the current finding—that activations in the parahippocampal and anterior fusiform cortices are associated with successful scene encoding—is consistent with prior imaging findings.

Unexpected and new, however, is the discovery that MTL regions seemed not to be involved in predicting encoding success (that is, there was no difference in activation for items that had a “will be remembered” prediction compared to items that had a “will be forgotten” prediction). The current imaging study examined the neural correlates of subjective metamemory judgments during encoding, but a number of studies have found MTL activation associated with subjective judgments during retrieval. In one study, hippocampal activation was found for stimuli that were consciously remembered, relative to stimuli that were accurately judged as having been studied but seemed merely familiar²¹. In another study, hippocampal activation at retrieval occurred equally for items thought to have been studied whether they were actually studied or not, whereas parahippocampal activations did differentiate between whether items were actually studied or not²². The present findings, however, suggest that MTL regions, despite their essential involvement in actual encoding, may not be involved in the subjective evaluation of encoding success. MTL, and especially hippocampal, activations associated with the subjective evaluations of retrieval may reflect a different role for MTL structures at encoding versus retrieval. Alternatively, other regions associated with retrieval evaluation, such as prefrontal^{23,24} or parietal cortices²⁵, may actually mediate subjective evaluation at retrieval.

Medial PFC and predicted encoding success

Several regions in PFC, such as VMPFC and DMPFC, were engaged during the prediction of encoding success. JOLs involve introspective processes about the mental states associated with successful or unsuccessful memory formation, and it has been proposed that medial PFC supports the ability to represent mental states of the self and others²⁶. Accordingly, neuroimaging studies have found medial PFC activations when people process information about themselves relative to other people^{27,28}. Thus, the medial PFC activations in the present study that were associated with predictions of encoding success may reflect processes involved in the introspective evaluation of internal mental states.

In addition to differentiating between JOL predictions, activations in DMPFC showed a trend toward a significant main effect for actual encoding success. The ROI analyses revealed that DMPFC activations decreased compared to baseline, consistent with other studies that find deactivations in medial PFC¹⁸. Greater deactivations were associated with subsequently remembered items than with subsequently forgotten items. This is consistent with the interpretation that deactivations for items later remembered reflect the successful disengagement of ongoing baseline mental processes that are task irrelevant and the allocation of neurocognitive resources to the task at hand²⁹. For predicted encoding success, items given “will forget” predictions showed greater deactivation than did items given “will remember” predictions, perhaps reflecting participants’ reallocation of resources in attempts to remember items they deemed harder to remember.

VMPFC and JOL accuracy

Only VMPFC activation correlated with individual differences in JOL accuracy. Ever since the case of Phineas Gage, who sustained bilateral VMPFC damage³⁰, the VMPFC has been thought to play an important role in judgment and decision-making by integrating somatic markers or emotions with goals and expectations³¹. VMPFC is also involved in determining whether information is contextually appropriate, the noetic feeling of ‘rightness’³². For judgments about memory, VMPFC may support accurate JOL predictions by forming an internal model of what constitutes successful learning and comparing information available from the current state of learning to this internal model³³.

Cognitive theories have proposed that accurate JOLs result from monitoring information that is predictive of later remembering, such as encoding strategy, list position, word-pair relatedness or length of study time³⁴. However, people can also use information that is not predictive of future remembering when forming their JOLs. This can lead to a false sense of successful encoding and thereby lead to inaccurate predictions. For instance, speed of encoding is a salient cue that increases people’s confidence that an item will be later remembered (JOL magnitude), but the speed of encoding does not predict differences in future memory retrieval^{35,36}. One possible role of the VMPFC is to flexibly evaluate reliable indicators of encoding success to form accurate JOLs. Future research is needed to determine whether individual differences in JOL accuracy reflect transient changes during task performance (such as effort, alertness or JOL strategy)³⁷ or stable differences in metamemory abilities^{3,4}.

The VMPFC has also been shown to support other metamemory judgments, such as the FOK judgments made at retrieval. Among individuals with PFC damage, those with the most inaccurate FOKs had damage to VMPFC¹⁶. In an fMRI study with normal adults, greater FOK accuracy was associated with greater VMPFC activation³⁸. The VMPFC is also implicated in predictions about outcomes other than memory performance. For instance, individuals with orbitofrontal lesions are unable to weigh advice and make predictions about economic outcome³⁹. These convergent lesion and imaging findings support the interpretation that the quality of the internal monitoring of memory success or failure depends upon processes mediated by VMPFC.

Lateral PFC, and actual and predicted encoding success

Whereas MTL and VMPFC were differentially involved in either actual or predicted encoding success, lateral PFC activation occurred for both actual and predicted encoding success. In broad terms, this is consistent with evidence that lateral PFC contributes to successful memory encoding^{2,14}, perhaps through semantic elaboration and organizational strategies that enhance successful encoding⁴⁰. It might, however, have been expected that a brain region activated by both actual and predicted encoding success would also support the accuracy of JOL predictions. Our results identified several brain regions, such as lateral PFC, that tracked both actual and predicted encoding success; however, none of these regions correlated with JOL accuracy.

The finding that lateral PFC tracked predicted encoding success is consistent with results from other studies comparing objective and subjective memory processes^{23,24,41,42}. In studies of subjective judgments made during retrieval, lateral PFC activations varied with the strength of FOK judgments (analogous to the “will remember” versus “will forget” JOL judgments), but not with FOK accuracy across participants^{23,24}. During FOKs, participants are bringing online the information associated with the target item⁴³. Graded activations in lateral PFC may be associated with the amount of partial retrieval of the target item brought into working memory²⁴. Lateral PFC may have a

similar role in JOLs, such that the degree of activation reflects how powerfully the information to be learned is held in working memory and contributes to the subjective experience of how successfully an item is learned. Lateral PFC activations may also reflect an increased encoding effort associated with “will remember” predictions; this is consistent with evidence that the magnitude of lateral PFC activation during encoding varies with both encoding effort and subsequent retrieval outcome, whereas MTL activations vary only with subsequent retrieval outcome⁴².

A fundamental issue in human memory research is the relationship between objective (veridical) and subjective (experiential) dimensions of memory. The relationship between objective and subjective memory processes can be studied during learning (memory encoding at study) or retrieval (recall or recognition at test). A number of functional neuroimaging studies have examined the relationship between objective and subjective memory processes at retrieval^{16,23,24,38,41}. The present study reports initial evidence about the neural substrates of subjective memory during learning. Such subjective or metamemory processes during learning are of particular interest, because these processes can actually enhance the effectiveness of learning by guiding the allocation of resources at a time when information remains available for learning. Indeed, in the present study, greater accuracy in JOLs (including greater accuracy in predicting both future remembering and future forgetting) was correlated with greater accuracy in recognition memory and greater VMPFC activation. Thus, these processes, and the neural circuits that mediate them, constitute a critical component of the way in which knowing how to learn empowers learning itself.

METHODS

Participants. Six female and 14 male native English speakers between the ages of 19 and 25 (mean \pm s.d. = 21 \pm 2) participated in the study. Participants were right-handed, with normal or corrected vision, and were without any neurological or psychiatric conditions or structural brain abnormalities. We discarded data from four participants because they made excessive head movements during data acquisition. We obtained informed consent from all participants according to the requirements of the Panel on Human Participants in Medical Research at Stanford University.

Stimulus materials. Materials consisted of 700 pictures of indoor and outdoor scenes. Pictures were randomly distributed into ten lists of 70 pictures. Five lists with a total of 350 scenes were presented during study, and the remaining five lists were presented as foils during the recognition-memory test. Lists were counterbalanced across participants such that all lists were presented as old and new items.

Task procedure. In a rapid event-related design, we scanned participants while they studied 350 scenes randomly intermixed with 350 fixations (Fig. 1). Participants were explicitly instructed to memorize the scenes for an upcoming memory test. Scenes were presented for 3 s with an inter-stimulus interval of 1 s. For each scene, participants made JOL predictions about the likelihood of remembering the scene during the memory test. The instructions (“Will you remember this at test?”) appeared on the bottom of the screen, prompting participants to make “will remember” or “will forget” JOLs. Participants were instructed to press a button with their right index finger to indicate a “will remember” prediction and with their middle finger to indicate a “will forget” prediction.

After scanning, participants were given a recognition test consisting of 350 old and 350 new pictures. For each trial, participants made two judgments: (i) whether the item was old or new and (ii) whether their judgment was made with high or low confidence. In this self-paced recognition test, each trial lasted a maximum of 8 s.

Imaging procedure. Magnetic resonance imaging was performed using a 3-T GE Signa scanner. Before functional imaging, a spin-echo T1-weighted

anatomical image was acquired (30 coronal slices; slice thickness = 6 mm; TE = 30 ms; TR = 2,000 ms; field of view = 24 \times 24 cm²). A shim procedure was used to improve B_0 magnetic field homogeneity. Functional images were then obtained in the same slice location as the anatomical images using a T2*-sensitive two-dimensional gradient-echo spiral-in/out sequence. Scenes were presented over five scanning sessions lasting approximately 9 min each. Bite bars made out of dental compress were used to restrict head movement.

Imaging analyses. Imaging data were preprocessed and analyzed using SPM99. We corrected for differences in the acquisition time of the functional images and then performed a motion correction using sinc interpolation. The T2* anatomical image was co-registered to the mean functional image that was created during motion correction. The anatomical image was then segmented into gray matter, white matter and cerebral spinal fluid. Anatomical and functional images were spatially normalized based on parameters determined by normalizing the segmented gray matter image to a gray matter template from the MNI series using a 12-parameter affine transformation. Finally, images were resampled into 3.75 \times 6 \times 3.75 mm voxels and spatially smoothed with an isotropic Gaussian kernel of 7 mm full-width at half maximum (FWHM).

Statistical models were constructed for individual participants using a general linear model. Regressor functions were constructed for each of the four trial types (Rr, Rf, Fr and Ff). Trials were modeled as events assuming a canonical hemodynamic response function. Subject-specific effects were estimated using a fixed-effects model. Linear contrasts were computed to generate subject-specific SPM(t) contrasts representing statistical differences in brain activation between conditions. Contrasts constructed at the single participant level were then input into a second-level group analysis using a random-effects model. Group contrasts were constructed by using a one-sample t -test. All reported clusters survived a P -threshold of 0.001 (uncorrected for multiple comparisons) and consisted of five or more significant voxels.

To identify voxels that differed between group-level contrasts for actual and predicted encoding success, we conducted one-tailed paired t -tests of the contrasts. Reported clusters for the paired t -test survived a P -threshold of 0.005 and consisted of five or more significant voxels. In addition, the contrast for actual encoding success was masked by the contrast for predicted encoding success to identify voxels that were significant in both contrasts.

ROI analyses were employed to characterize the statistical effects of each of the four trial types. ROIs were functionally defined and included all significant voxels in the cluster. Data were extracted by selective averaging with respect to peristimulus time out to 10 s after stimulus onset. ROI data are expressed as percent signal change calculated by taking the average signal from 4 s to 8 s after stimulus onset. These data were then subjected to a repeated-measures ANOVA and Pearson's r correlations.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

The authors thank S. Gabrieli, P. Mazaika, J. Cooper, A.R. Preston and P. Sokol-Hessner for their assistance or comments. This research was sponsored by grants from the US National Institute of Mental Health to Y.-C.K. (MH073234) and J.D.E.G. (MH59940).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/natureneuroscience/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Brewer, J.B., Zhao, Z., Desmond, J.E., Glover, G.H. & Gabrieli, J.D. Making memories: brain activity that predicts how well visual experience will be remembered. *Science* **281**, 1185–1187 (1998).
2. Wagner, A.D. *et al.* Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* **281**, 1188–1191 (1998).
3. King, J.F., Zechmeister, E.B. & Shaughnessy, J.J. Judgments of knowing: the influence of retrieval practice. *Am. J. Psychol.* **93**, 329–343 (1980).
4. Maki, R.H. & Berry, S.L. Metacomprehension of text material. *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 663–679 (1984).

5. Thiede, K.W., Anderson, M.C. & Theriault, D. Accuracy of metacognitive monitoring affects learning of texts. *J. Educ. Psychol.* **95**, 66–73 (2003).
6. Mazzoni, G. & Cornoldi, C. Strategies in study time allocation: why is study time sometimes not effective? *J. Exp. Psychol. Gen.* **122**, 47–60 (1993).
7. Dunlosky, J., Kubat-Silman, A.K. & Hertzog, C. Training monitoring skills improves older adults' self-paced associative learning. *Psychol. Aging* **18**, 340–345 (2003).
8. Dunlosky, J. *et al.* Inhalation of 30% nitrous oxide impairs people's learning without impairing people's judgments of what will be remembered. *Exp. Clin. Psychopharmacol.* **6**, 77–86 (1998).
9. Izaute, M. & Bacon, E. Specific effects of an amnesic drug: effect of lorazepam on study time allocation and on judgment of learning. *Neuropsychopharmacology* **30**, 196–204 (2005).
10. Vilkki, J., Surma-aho, O. & Servo, A. Inaccurate prediction of retrieval in a face matrix learning task after right frontal lobe lesions. *Neuropsychology* **13**, 298–305 (1999).
11. Kennedy, M.R. & Yorkston, K.M. Accuracy of metamemory after traumatic brain injury: predictions during verbal learning. *J. Speech Lang. Hear. Res.* **43**, 1072–1086 (2000).
12. Scoville, W.B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
13. Corkin, S. What's new with the amnesic patient H.M.? *Nat. Rev. Neurosci.* **3**, 153–160 (2002).
14. Kirchhoff, B.A., Wagner, A.D., Maril, A. & Stern, C.E. Prefrontal circuitry for episodic encoding and subsequent memory. *J. Neurosci.* **20**, 6173–6180 (2000).
15. Fletcher, P.C. & Henson, R.N. Frontal lobes and human memory: insights from functional neuroimaging. *Brain* **124**, 849–881 (2001).
16. Schnyer, D.M. *et al.* A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: evidence from patients with lesions to frontal cortex. *Neuropsychologia* **42**, 957–966 (2004).
17. Nelson, T.O. A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychol. Bull.* **95**, 109–133 (1984).
18. Gusnard, D.A., Akbudak, E., Shulman, G.L. & Raichle, M.E. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. USA* **98**, 4259–4264 (2001).
19. Stern, C.E. *et al.* The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93**, 8600–8665 (1996).
20. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
21. Eldridge, L.L., Knowlton, B.J., Furmanski, C.S., Bookheimer, S.Y. & Engel, S.A. Remembering episodes: a selective role for the hippocampus during retrieval. *Nat. Neurosci.* **3**, 1149–1152 (2000).
22. Cabeza, R., Rao, S.M., Wagner, A.D., Mayer, A.R. & Schacter, D.L. Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proc. Natl. Acad. Sci. USA* **98**, 4805–4810 (2001).
23. Kikyo, H., Ohki, K. & Miyashita, Y. Neural correlates for feeling-of-knowing: an fMRI parametric analysis. *Neuron* **36**, 177–186 (2002).
24. Maril, A., Simons, J.S., Mitchell, J.P., Schwartz, B.L. & Schacter, D.L. Feeling-of-knowing in episodic memory: an event-related fMRI study. *Neuroimage* **18**, 827–836 (2003).
25. Wheeler, M.E. & Buckner, R.L. Functional dissociation among components of remembering: control, perceived oldness, and content. *J. Neurosci.* **23**, 3869–3880 (2003).
26. Frith, C.D. & Frith, U. Interacting minds—a biological basis. *Science* **286**, 1692–1695 (1999).
27. Kelley, W.M. *et al.* Finding the self? An event-related fMRI study. *J. Cogn. Neurosci.* **14**, 785–794 (2002).
28. Schmitz, T.W., Kawahara-Baccus, T.N. & Johnson, S.C. Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *Neuroimage* **22**, 941–947 (2004).
29. Daselaar, S.M., Prince, S.E. & Cabeza, R. When less means more: deactivations during encoding that predict subsequent memory. *Neuroimage* **23**, 921–927 (2004).
30. Damasio, H., Grabowski, T., Frank, R., Galaburda, A.M. & Damasio, A.R. The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science* **264**, 1102–1105 (1994).
31. Tranel, D. Emotion, decision making, and the ventromedial prefrontal cortex. in *Principles of Frontal Lobe Function* (eds. Stuss, D.T. & Knight, R.T.) Ch. 22, 338–353 (Oxford University Press, London, 2002).
32. Moscovitch, M. & Winocur, G. The frontal cortex and working with memory. in *Principles of Frontal Lobe Function* (eds. Stuss, D.T. & Knight, R.T.) Ch. 12, 188–209 (Oxford University Press, London, 2002).
33. Nelson, T.O. & Narens, L. Metamemory: a theoretical framework and new findings. *Psychol. Learn. Motiv.* **26**, 125–141 (1990).
34. Koriat, A. Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *J. Exp. Psychol. Gen.* **126**, 349–370 (1997).
35. Hertzog, C., Dunlosky, J., Robinson, A.E. & Kidder, D.P. Encoding fluency is a cue used for judgments about learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **29**, 22–34 (2003).
36. Benjamin, A.S., Bjork, R.A. & Schwartz, B.L. The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *J. Exp. Psychol. Gen.* **127**, 55–68 (1998).
37. Kelemen, W.L., Frost, P.J. & Weaver, C.A., III. Individual differences in metacognition: evidence against a general metacognitive ability. *Mem. Cognit.* **28**, 92–107 (2000).
38. Schnyer, D.M., Nicholls, L. & Verfaellie, M. The role of VMPC in metamemorial judgments of content retrievability. *J. Cogn. Neurosci.* **17**, 832–846 (2005).
39. Gomez-Beldarrain, M., Harries, C., Garcia-Monco, J.C., Ballus, E. & Grafman, J. Patients with right frontal lesions are unable to assess and use advice to make predictive judgments. *J. Cogn. Neurosci.* **16**, 74–89 (2004).
40. Petrides, M. Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Phil. Trans. R. Soc. Lond. B* **351**, 1455–1461 (1996).
41. Chua, E.F., Rand-Giovannetti, E., Schacter, D.L., Albert, M.S. & Sperling, R.A. Dissociating confidence and accuracy: functional magnetic resonance imaging shows origins of the subjective memory experience. *J. Cogn. Neurosci.* **16**, 1131–1142 (2004).
42. Reber, P.J. *et al.* Neural correlates of successful encoding identified using functional magnetic resonance imaging. *J. Neurosci.* **22**, 9541–9548 (2002).
43. Schacter, D.L. & Worling, J.R. Attribute information and the feeling of knowing. *Can. J. Psychol.* **39**, 467–475 (1985).