



# Neural Correlates of Metamemory: A Comparison of Feeling-of-Knowing and Retrospective Confidence Judgments

## Citation

Chua, Elizabeth F., Daniel L. Schacter, and Reisa A. Sperling. 2009. Neural correlates of metamemory: A comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience* 21(9): 1751-1765.

## Published Version

doi:10.1162/jocn.2009.21123

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3627271>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Neural Correlates of Metamemory: A Comparison of Feeling-of-Knowing and Retrospective Confidence Judgments

Elizabeth F. Chua<sup>1,2,3</sup>, Daniel L. Schacter<sup>2,3</sup>, and Reisa A. Sperling<sup>1,3</sup>

## Abstract

■ Metamemory refers to knowledge and monitoring of one's own memory. Metamemory monitoring can be done *prospectively* with respect to subsequent memory retrieval or *retrospectively* with respect to previous memory retrieval. In this study, we used fMRI to compare neural activity during prospective feeling-of-knowing and retrospective confidence tasks in order to examine common and distinct mechanisms supporting multiple forms of metamemory monitoring. Both metamemory tasks, compared to non-metamemory tasks, were associated with greater activity in medial prefrontal, medial parietal, and lateral parietal regions, which have previously been implicated in internally directed cognition. Furthermore, compared to non-metamemory tasks, metamemory tasks were associated with less activity in occipital regions, and in lateral inferior frontal and dorsal medial prefrontal regions, which

have previously shown involvement in visual processing and stimulus-oriented attention, respectively. Thus, neural activity related to metamemory is characterized by both a shift toward internally directed cognition and away from externally directed cognition. Several regions demonstrated differences in neural activity between feeling-of-knowing and confidence tasks, including fusiform, medial temporal lobe, and medial parietal regions; furthermore, these regions also showed interaction effects between task and the subjective metamemory rating, suggesting that they are sensitive to the information monitored in each particular task. These findings demonstrate both common and distinct neural mechanisms supporting metamemory processes and also serve to elucidate the functional roles of previously characterized brain networks. ■

## INTRODUCTION

Metamemory, broadly defined as knowledge about one's own memory function, requires the monitoring of memory processes, and can be thought of as the on-line ability to gather information about the current state of the memory system (Nelson & Narens, 1990). In metamemory tasks, subjects are explicitly asked to make judgments about the state of their own memories. These tasks may be *prospective*, and ask subjects to judge their future memory performance, or they may be *retrospective*, and ask subjects to judge their prior memory performance. Neuroimaging studies have begun to elucidate the neural correlates of various metamemory processes. For example, recent research has shown lateral prefrontal and parietal involvement in a prospective metamemory task (Maril, Simons, Weaver, & Schacter, 2005; Kikyo & Miyashita, 2004; Maril, Simons, Mitchell, Schwartz, & Schacter, 2003; Kikyo, Ohki, & Miyashita, 2002). In contrast, other studies have implicated medial prefrontal, medial temporal, medial parietal, and lateral parietal regions in a retrospective metamemory

task (Chua, Schacter, Rand-Giovannetti, & Sperling, 2006; Moritz, Glascher, Sommer, Buchel, & Braus, 2006). In this study, we examine whether these regions are similarly engaged during prospective and retrospective metamemory monitoring tasks in order to gain further insight into the specific roles of parietal, prefrontal, and medial temporal regions in specific aspects of metamemory.

Theoretical conceptions of metamemory, as formulated by Nelson and Narens (1990), propose that memory processes can be split into two levels: an *object* level and a *meta* level. In the Nelson and Narens model of metamemory, the meta level contains an imperfect model of the object level (i.e., a simulation). Information from the object level is available to the meta level via monitoring mechanisms. The meta level then can modify the object-level processes or change the state of the object-level processes via control processes. In the case of memory retrieval, the object-level information being monitored is the content of retrieval (e.g., the target and related information). Monitoring processes are then required to evaluate the relevance and validity of the retrieved information in terms of the task goals. Thus, when referring to metamemory monitoring in this study, we are referring to monitoring for accuracy and relevance of the retrieved information.

<sup>1</sup>Brigham and Women's Hospital, Boston, MA, <sup>2</sup>Harvard University, <sup>3</sup>Massachusetts General Hospital

Metamemory monitoring can occur at different stages of retrieval, as revealed by performance on different tasks (Nelson & Narens, 1990). The two metamemory tasks we investigate in this functional magnetic resonance imaging (fMRI) study are the prospective feeling-of-knowing (FOK) and the retrospective confidence judgment (CONF) tasks. The FOK paradigm requires subjects to make predictions about their future ability to remember previously learned information that is currently inaccessible, and FOK ratings have been predictive of recognition accuracy (Nelson, 1984). FOK has been classically studied using a recall–judgment–recognition paradigm (Hart, 1965). Subjects first perform a cued recall test, and if they fail to recall the target, they are then asked to give an FOK rating. Subjects then perform a recognition test. In contrast, during retrospective confidence judgment tasks, subjects are given the recognition test, which is then followed by rating their confidence in the accuracy of the previously made recognition decision.

Both FOK and CONF involve memory monitoring (Koriat & Goldsmith, 1996), but the judgments are thought to be based on different sources of information. FOK judgments are believed to be based on partial access to the semantic, perceptual, or affective attributes of target (e.g., Koriat, 1993), familiarity of the cue (e.g., Metcalfe, Schwartz, & Joaquim, 1993), or a combination of the two processes (Koriat & Levy-Sadot, 2001). In contrast, CONF judgments are thought to be based on the strength of the underlying memory trace, ease of retrieval, and also on heuristics that are applied to the specific study and test conditions, and to the subject's own memory (Shaw & Zerr, 2003; Bradfield, Wells, & Olson, 2002; Busey, Tunnicliff, Loftus, & Loftus, 2000; Belli, Lindsay, Gales, & McCarthy, 1994; Yonelinas, 1994).

Previous neuroimaging studies have investigated the neural basis of FOK and CONF using two different approaches. The first approach is to examine task-related neural activity by comparing the metamemory task to a control task. Comparisons of CONF to recognition showed increased activity in lateral parietal, medial parietal, and right orbito-frontal regions (Chua et al., 2006). A second approach is to compare different levels of FOK or CONF within each task (e.g., comparing high and low CONF), which has been done for both FOK (Maril et al., 2003, 2005; Kikyo & Miyashita, 2004; Kikyo et al., 2002) and CONF (Chua et al., 2006; Moritz et al., 2006). Higher levels of FOK have been associated with greater activity in several prefrontal and parietal regions (Maril et al., 2003, 2005; Kikyo & Miyashita, 2004; Kikyo et al., 2002). In contrast, higher levels of CONF have shown greater activity in the medial temporal lobe (MTL), and several regions along the cingulate gyrus, both anteriorly and posteriorly (Chua et al., 2006; Moritz et al., 2006). In this study, we use both approaches to examine how brain regions modulate based on (1) monitoring task demands regardless of behavioral response, and (2) subjective metamemory judgments.

We investigated neural activity during CONF and FOK using a face–name associative memory task. Prior to scanning, subjects were familiarized with novel faces and then encoded names associated with those faces. During scanning, subjects performed FOK judgments, forced-choice recognition, and CONF judgments, as well as an additional task of subjectively rating the attractiveness of the faces. During FOK, subjects retrieved information, monitored the outputs from retrieval, and made a subjective decision about their future ability to recognize the name. During recognition, subjects chose which name was associated with the face based on the information they retrieved. During CONF, subjects monitored their recognition decision and made a subjective judgment about their previous memory performance. These three tasks differed in the degree to which they required retrieval, monitoring, and subjective decisions, but all three probed memory. During a nonmemory task, attractiveness judgments, subjects were required to evaluate the pleasantness of a face and make a subjective decision about the face. Unlike FOK and CONF, which necessarily preceded and followed recognition trials, respectively, attractiveness judgments could be performed prior to or after recognition. Importantly, comparing the two metamemory tasks to attractiveness judgments allowed us to control for timing of the judgment (pre- or postrecognition), and it also allowed us to compare tasks that both required subjective decisions, even though one probed memory and the other did not.

This design allowed us to determine the specificity of activity related to metamemory. We examined (1) the neural correlates of metamemory by comparing metamemory tasks (i.e., FOK and CONF) to non-metamemory tasks (i.e., recognition and attractiveness ratings); (2) the brain regions that differentiate between two metamemory tasks by directly comparing FOK and CONF; and (3) the neural correlates of the subjective metamemory rating by examining MR signal modulation based on the level of FOK and MR signal modulation based on the level of CONF. Based on previous research, we predicted that both metamemory tasks would show greater activity in medial and lateral parietal regions compared to the non-metamemory tasks, which would demonstrate common neural mechanisms underlying metamemory. In contrast, we predicted that FOK and CONF would show modulation based on the level of the judgment in different brain regions, with level of FOK modulating prefrontal regions and level of CONF modulating MTL, medial prefrontal, and medial parietal regions.

## METHODS

### Participants

Twenty right-handed, healthy, young, native English speakers completed this study (11 women/9 men; ages = 20–30 years), but only the 13 subjects (6 women/7 men;

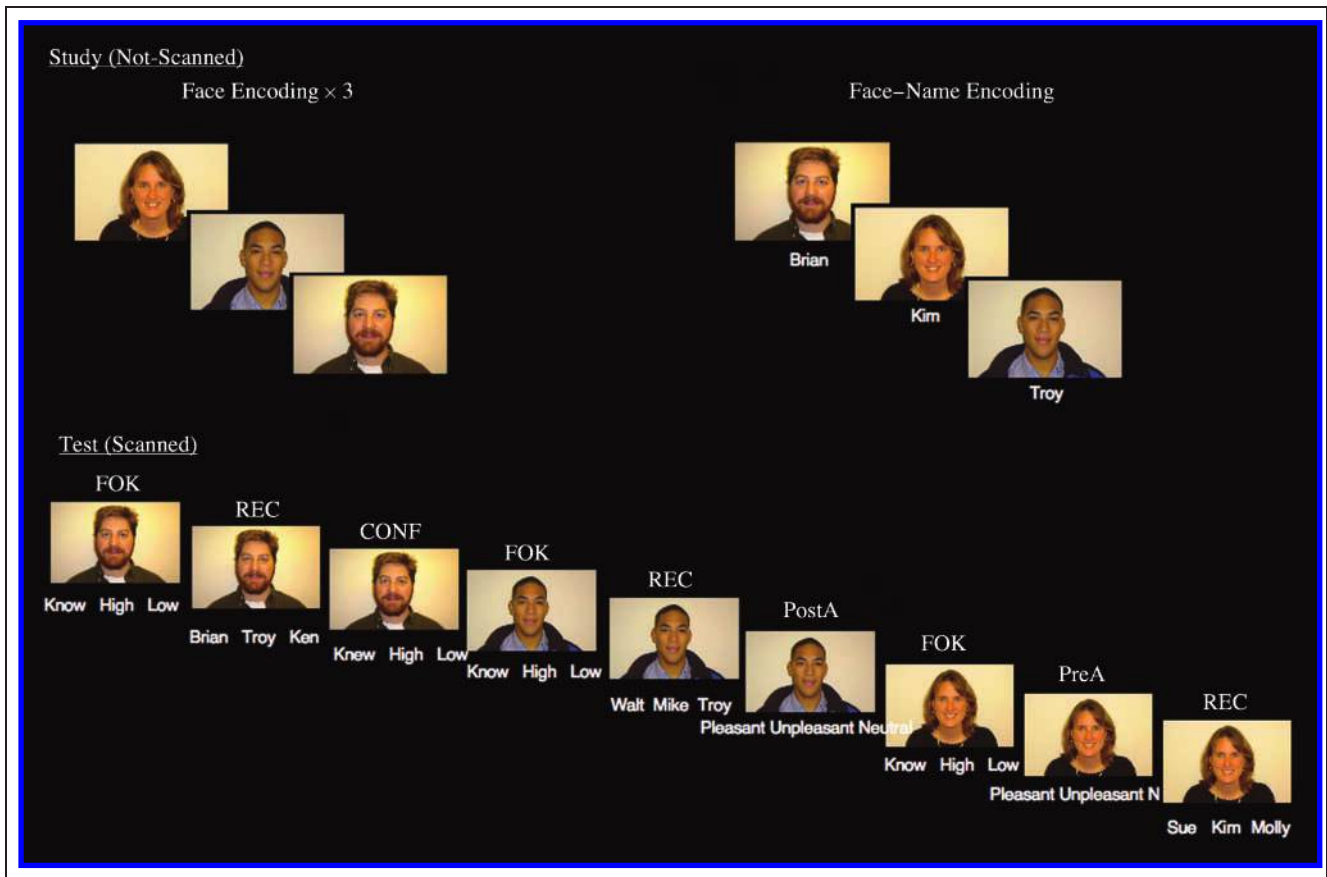
ages = 21–30 years) who had sufficient trials in each behavioral response category to be included in fMRI analyses were analyzed. All subjects were free from psychiatric and neurologic illness, and none were taking medications with known central nervous system effects. All subjects were screened for contraindications to MRI. Each subject provided written informed consent in a manner approved by the Human Research Committee at Brigham and Women’s Hospital.

**Procedure**

Participants completed face–name associative memory and metamemory tasks that included a prescan encoding phase and a scanned test phase. During the prescan study phase, subjects encoded faces alone and then face–name pairs (Figure 1) presented on a Macintosh PowerBook G4 using MacStim (WhiteAnt Occasional Publishing, West Melbourne, Australia). Subjects viewed 270 digital photographs of faces for 1.75 sec presented against a black background followed by a white fixation cross for 0.25 sec. Equal numbers of male and female faces, and equal numbers of young, middle-aged, and older faces, were presented. One-third of the faces presented were

nonwhite faces. For each face, subjects indicated via button press whether the face was female or male. Each face was viewed thrice in a gender decision task; presenting the face thrice helped control for potential differences in familiarity and novelty during the scanned test phase. After viewing the three face-alone runs, subjects saw a single presentation of each face that had been seen in the face-alone runs paired with a fictional first name. First names were assigned based on census lists obtained from the internet that list popular first names by decade. Faces were presented on a black background with the name printed in white underneath the face. Each face–name pair was presented for 1.75 sec each followed by 0.25 sec of visual fixation. Subjects were instructed to try to remember the name associated with the face for later testing and also to make a purely subjective decision about whether the name “fits” the face. Thus, by the end of the study phase, subjects had viewed each face a total of four times, which off-line pilot testing showed was sufficient for subjects to perform at ceiling on a face recognition test.

During scanning, subjects completed the test phase approximately 20 min after the study phase, which involved four different cognitive tasks: FOK, recognition



**Figure 1.** At study (not-scanned), subjects viewed novel faces three times in a face encoding task, and face–name encoding task. At test (scanned), subjects performed feeling-of-knowing (FOK), recognition (REC), confidence (CONF), and attractiveness (A) tasks (right). Attractiveness ratings were given either prerecognition (PreA) or postrecognition (PostA). The tasks were presented in three different randomized orders: FOK–REC–CONF, FOK–REC–PostA, FOK–PreA–REC in an event-related design.

(REC), CONF, and an attractiveness rating (A) (Figure 1). During the FOK task, subjects were shown a face seen during the study phase presented against a black background with the words “Know,” “High,” and “Low” printed underneath. Subjects were instructed to indicate via button response whether they “know” the name (i.e., they have free recall for the name), have “high” confidence that they will recognize the name later even though they do not currently recall the name, or have “low” confidence that they will recognize the name later. For the FOK task, upon failing to recall the name associated with the face (“know” response), subjects indicated their level of confidence that they would be able to correctly subsequently recognize the name associated with the face. During REC trials, subjects completed a three-alternative forced-choice task. Subjects saw the same face with three different names printed underneath (one correct, one name that was paired with a different face during the encoding trial, and one unique name) and indicated which name was correct with a button press. During CONF trials (which by definition followed recognition trials), subjects indicated their subjective confidence regarding whether they had chosen the correct name. Subjects saw the face with the words “Knew,” “High,” or “Low” printed in white underneath and were asked to indicate whether they had high or low confidence that they had chosen the correct name during the recognition task; they were instructed to choose the “Knew” option only for names that they had recalled earlier. During the attractiveness rating task, subjects performed an attractiveness judgment task and indicated whether the face is “pleasant,” “unpleasant,” or “neutral.”

As noted in the Introduction, the attractiveness rating was included for methodological reasons. Task order for FOK, REC, and CONF is constrained in that FOK must precede REC and REC must precede CONF. At a behavioral level, if CONF always follows REC, then subjects may make CONF decisions during REC. To help minimize this potential problem, we included “catch” trials in which subjects did not rate their confidence after recognition, by requiring subjects to perform an attractiveness rating (instead of a confidence rating) after REC. CONF trials occurred after 65% of the recognition trials. In addition to having behavioral effects, the constraints on task ordering also makes it difficult to deconvolve the hemodynamic response for each task. The attractiveness rating task was used in order to vary stimulus ordering. Attractiveness ratings occurred either prerecognition (PreA) or postrecognition (PostA). Thus, there were three different task orders: (1) FOK, REC, CONF; (2) FOK, REC, PostA; (3) FOK, PreA, REC (Figure 1). This additional task also had benefits for subsequent analyses in that it allowed contrasts that controlled for stimulus order and number of repetitions; for example, contrasting CONF and PostA compared tasks that both occurred postrecognition and were the third

presentation of the face in the scanner. The attractiveness rating was chosen because it is a non-memory-based subjective decision (not because of a particular interest in attractiveness).

The stimuli were constructed to hold visual complexity constant in order to avoid fMRI activation patterns that varied according to stimulus complexity. Only the three words printed underneath the face changed based on task. Because the temporal ordering of the stimuli was constrained, the prescan study phase included multiple exposures to the faces in order to help minimize effects of stimulus familiarity and novelty in the scanner. During scanning, the trials were presented through Magnetic Resonance Technology goggles in an event-related design with varying fixed interstimulus intervals from 0.25 to 8 sec with self-paced stimulus offsets, which has shown to be feasible in rapid event-related fMRI (Maccotta, Zacks, & Buckner, 2001). The duration of the stimulus presentation was self-paced, and stimulus offsets occurred once subjects had pressed a button to make a behavioral response. Subjects had a maximum of 3.75 sec to respond. There were 270 FOK, 270 recognition, 174 CONF, 48 PreA, and 48 PostA trials. The tasks were presented across four runs with varying lengths because of the self-paced nature of the design and ranged from 6.33 to 10.36 min (mean = 8.426 min,  $SD = 0.831$ ). Subjects practiced this task with 18 faces prior to entering the scanner.

### Imaging Parameters

Whole-brain fMRI scans were collected on a 3-Tesla GE scanner using a gradient-echo sequence at Brigham and Women’s Hospital (TR = 2000 msec, TE = 30 msec, flip angle = 90) in an oblique coronal orientation perpendicular to the anterior commissure–posterior commissure line (28 slices, 5 mm, skip 1 mm,  $3.125 \times 3.125 \times 6$  mm voxels).

### Imaging Preprocessing

The fMRI data were preprocessed and analyzed using SPM2 (Wellcome Department on Cognitive Neurology) for Matlab (Mathworks, Natick, MA). Images were motion corrected using INRIAlign, a motion correction algorithm unbiased by local signal changes. No slice-timing correction was applied. The data were then spatially normalized to an EPI template based on the MNI1305 stereotactic space (resampled voxel size  $3 \times 3 \times 3$  mm) and then spatially smoothed using an 8-mm full-width half-maximum isotropic Gaussian kernel.

### fMRI Modeling

Data were analyzed according to a random effects general linear model in SPM2. Two different GLMs were generated: one to examine monitoring processes and

another to examine the level of FOK or CONF expressed. In both models, trials were modeled as events (i.e., modeled as a stick function) using the canonical hemodynamic response function alone, and both models included a high-pass filter of 70 sec. Motion parameters were not included in either model.

To examine monitoring processes, the conditions modeled in the GLM were based on the cognitive task (FOK, REC, CONF, PreA, and PostA) only and were collapsed across behavioral responses. This procedure allowed for greater power to detect differences than modeling based on behavioral response to each task, and this model focuses on the act of performing the task rather than the outcome of the task. First, data were analyzed at the subject level, with each run treated as a time series and modeled with the canonical hemodynamic response function. At the second step, data were averaged together treating each subject as a random effect.

A separate GLM was defined that incorporated the behavioral responses within each process (FOK, REC, CONF, PreA, PostA) for each subject. For FOK judgments, trials were modeled based on the level of FOK given and on the accuracy of the recognition judgment, making six different trial types: know-hit, high FOK-hit, low FOK-hit, know-miss, high FOK-miss, and low FOK-miss. CONF trials were also categorized based on the level of confidence expressed and the accuracy of the recognition judgment, making a similar six trial types: knew-hit, high CONF-hit, low CONF-hit, knew-miss, high CONF-miss, and low CONF-miss. Recognition trials were modeled based on memory accuracy, the level of FOK, and the level of CONF. Attractiveness judgments were modeled based on whether or not the occurred pre- or postrecognition and whether the judgment was pleasant, unpleasant, or neutral. The runs were concatenated in time and treated as a single time series in this analysis. Additional regressors were included in the model to account for run. The high-pass filter of 70 sec was done using regressors in order to make sure that the filtering was appropriate at the run level.

### Whole-brain Analyses

Whole-brain statistical maps were thresholded at  $p < .001$ , uncorrected at the voxel level, and then corrected for multiple comparisons at the cluster level of  $p < .05$ . The number of contiguous voxels required for significant clusters was 20 resampled voxels, and was defined based on Monte Carlo simulations using our imaging and analysis parameters (Slotnick, Moo, Segal, & Hart, 2003). Voxel coordinates are presented in MNI space.

### Region-of-Interest Analyses

Specific functionally defined regions of interest (ROI) were generated from significantly activated clusters from

comparisons of interest and subjected to further analyses. Percent signal change data were extracted using MarsBar (Brett, Anton, Valabregue, & Poline, 2002) from the significantly activated cluster, or in the case of clusters that spanned multiple regions, a 4-mm sphere around the local peak voxel. Post hoc repeated measures ANOVAs and paired  $t$  tests were calculated using SPSS and were considered significant at  $p < .05$ , two-tailed, unless otherwise noted.

### Specific Analyses

#### *Common Activity for Both Metamemory Tasks*

In order to determine which brain regions played a role in metamemory function, weighted contrasts comparing both of the metamemory tasks to the non-metamemory tasks were generated (i.e., CONF + FOK compared to REC + PreA + Post A). Because fMRI contrasts are relative to one another, we included both recognition and attractiveness judgments as non-metamemory tasks. This procedure allowed us to control for task order and number of repetitions, and also to rule out the explanation that the differences between metamemory and recognition tasks were solely driven by REC (Chua et al., 2006). Regions that showed common activity for both metamemory tasks were examined post hoc to determine whether or not the mapwise comparisons were driven by specific tasks. Paired  $t$  tests on percent signal change data were used to compare CONF to REC, CONF to PostA, FOK to REC, and FOK to PreA; to correct for multiple comparisons within the regions at  $p < .05$ ,  $t$  tests were considered significant at  $p < .0125$ , one-tailed.

Regions that showed common activity for both metamemory tasks were also examined to determine whether (1) these regions showed similar effects for all behavioral response types, regardless of recognition accuracy or the subjective level of FOK or CONF expressed; and/or (2) the regions modulated within tasks based on behavioral response. Percent signal change from these regions were entered in a  $2 \times 2 \times 2$  Task (FOK, REC)  $\times$  Level (high, low)  $\times$  Accuracy (hits, misses) repeated measures ANOVA. Then a second  $2 \times 2 \times 2$  Task (REC, CONF)  $\times$  Level (high, low)  $\times$  Accuracy (hits, misses) repeated measures ANOVA was performed. Separate ANOVAs comparing FOK to REC and CONF to REC were performed in order to sort recognition trials based on the metamemory judgment of interest (i.e., REC trials were sorted based on level of FOK for the ANOVA comparing FOK to REC, whereas REC trials were sorted based on level of CONF for the ANOVA comparing CONF to REC).

#### *Differential Activity within Metamemory Tasks*

We generated bidirectional contrasts that directly compared FOK and CONF to determine which brain regions

showed differential activity between the two metamemory tasks. Similar to the analyses for regions showing common metamemory effects, regions that showed differences between FOK and CONF were subjected to a  $2 \times 2 \times 2$  Task (FOK, CONF)  $\times$  Level (high, low)  $\times$  Accuracy (hits, misses) repeated measures ANOVA to determine whether this pattern was consistent across different behavioral responses and if these regions modulated based on the subjective metamemory rating given or recognition accuracy.

### Subjective Level of Feeling-of-Knowing or Confidence Expressed

The next set of whole-brain analyses aimed to examine, during FOK and CONF, the neural activity associated with the subjective level expressed. Our primary interest was in comparing high and low responses for FOK and CONF, and therefore, we generated contrasts for high FOK-hit and high FOK-miss versus low FOK-hit and low FOK-miss, and high CONF-hit and high CONF-Miss versus low CONF-hit and Low CONF-miss. We were less interested in the know/knew trials because there were relatively few of them, and they were mainly included as a response option to eliminate freely recalled responses. For ROI that distinguished high and low ratings for either FOK or CONF,  $2 \times 2 \times 2$  repeated measures ANOVAs for task (FOK and CONF), level (high and low), accuracy (hits and misses) were performed to examine whether these regions showed main effects of level or any Task  $\times$  Level interactions.

## RESULTS

### Behavioral Results

Not all subjects used the full response scale for either FOK or CONF, and some subjects had less than 10 trials in specific conditions, and thus, were not included in fMRI analyses. The average number of trials per condition is presented in Table 1. Thirteen subjects had sufficient trials for high and low FOK and CONF responses for both hits and misses. Know and knew trials were not included because there were few subjects with enough misses in these conditions. Not all “knew” trials were

**Table 1.** Average Number of Trials ( $\pm$ SEM) in Each Condition

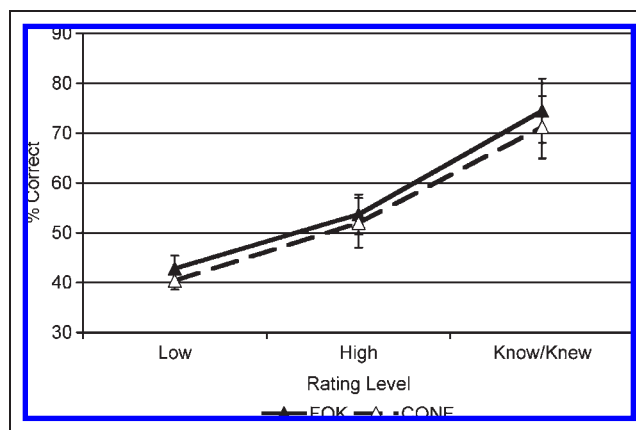
	FOK			CONF		
	Know	High	Low	Knew	High	Low
<i>n</i> = 13						
Hits	18 $\pm$ 5	49 $\pm$ 7	57 $\pm$ 10	19 $\pm$ 3	29 $\pm$ 4	27 $\pm$ 2
Misses	19 $\pm$ 8	48 $\pm$ 10	50 $\pm$ 9	14 $\pm$ 4	25 $\pm$ 3	41 $\pm$ 3

previously given a “know” rating (mean  $\pm$  SEM proportion of trials:  $0.21 \pm 0.07$ ; median proportion: 0.08).

Subjects performed above the chance rate (33%) for a three-alternative forced-choice recognition task, correctly recognizing the name associated with the face on  $50.1 \pm 3\%$  of the recognition trials. There were significant differences in recognition accuracy based on subjective rating for both FOK and CONF. For “know,” “high,” and “low” responses during FOK, subjects chose the correct name  $72.7 \pm 5.9\%$ ,  $53.7 \pm 4.0\%$ , and  $42.8 \pm 2.6\%$  of the time, respectively, which exhibited a significant linear effect [ $F(1, 12) = 48.74, p < .0001$ ]. For “knew,” “high,” and “low” responses during CONF, subjects chose the correct name for  $69.2 \pm 5.8\%$ ,  $52.0 \pm 4.9\%$ , and  $40.4 \pm 1.8\%$ , respectively, which showed a significant linear effect [ $F(1, 12) = 26.79, p < .0004$ ]. The pattern of results for FOK and CONF with greater percentage of correct trials for “know/knew” than “high,” and “high” greater percentage correct than “low” (Figure 2) shows that the ratings were meaningful and related to accuracy.

There were significant differences in reaction time (RT) between the different cognitive tasks performed in the scanner. Because the trials were self-paced with the stimulus offsets being a function of the RT, the duration of the stimulus presentation also differed. The mean RT ( $\pm$ SEM) for each task was: FOK =  $1.61 \pm 0.07$  sec, REC =  $2.10 \pm 0.5$  sec, CONF =  $1.11 \pm 0.05$  sec, PreA =  $1.66 \pm 0.06$  sec, and PostA =  $1.60 \pm 0.07$  sec. Paired *t* tests showed that the RT for all tasks differed from each other ( $p < .001$ ), with the exception of FOK and PostA.

There were no within-task differences in RT during the two metamemory tasks. During FOK, there were no significant differences in RTs based on subjective rating, accuracy, or their interaction (high FOK-hit =  $1.71 \pm 0.09$ , low FOK-hit =  $1.64 \pm 0.09$ , high FOK-miss =  $1.72 \pm 0.08$ , and low FOK-miss =  $1.65 \pm 0.09$ ). During



**Figure 2.** Calibration curves depicting the proportion of face-name pairs correctly recognized in each rating category (know/knew, high, and low) for both feeling-of-knowing (FOK) and confidence judgments (CONF).

CONF, subjects showed no significant differences in RTs based on subjective rating, accuracy, or their interaction (high CONF–hit =  $1.19 \pm 0.06$ , low CONF–hit =  $1.17 \pm 0.08$ , high CONF–miss =  $1.18 \pm 0.05$ , and low CONF–miss =  $1.14 \pm 0.06$ ).

## Imaging Results

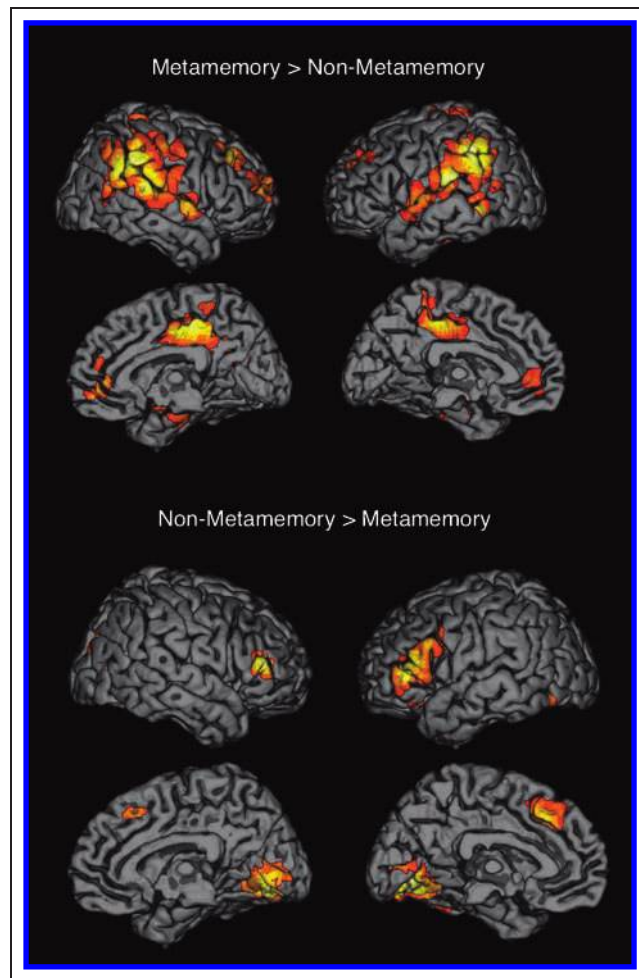
### *Common Activity in Feeling-of-Knowing and Confidence Judgments*

The contrast of metamemory > non-metamemory (i.e., FOK + CONF > REC + PreA + PostA; Figure 3) revealed differences in the right MTL (BA 35/36), bilateral superior frontal (BA 9/10), ventral medial prefrontal (BA 32), mid and posterior cingulate (BA 31), and a large lateral parietal/temporal area that included the inferior parietal lobule (IPL; BA 40), the tempo-parietal junction (TPJ; BA 40/42), and the superior temporal gyrus (STG; BA 42/22). The contrast of non-metamemory > metamemory (i.e., REC + PreA + PostA > FOK + CONF) revealed dif-

ferences in bilateral inferior prefrontal (BA 44/45), dorsal medial prefrontal (BA 8/32), and occipital regions (BA 18/19), and also in the right cuneus (BA 19).

ROI analyses comparing each metamemory judgment to recognition and attractiveness judgments (i.e., CONF vs. REC, CONF vs. PostA, FOK vs. REC, and FOK vs. PreA) revealed that some regions showed consistent differences between metamemory and non-metamemory tasks, whereas other regions only showed differences between metamemory tasks and recognition, and still others showed differences only for either FOK or CONF compared to the non-metamemory tasks (Table 2). The TPJ (BA 40/42), STG (BA 42/22), right IPL (BA 40), and mid/posterior cingulate (BA 31) regions showed consistently greater activity for metamemory compared to non-metamemory tasks (i.e., consistent differences for CONF vs. REC, CONF vs. PostA, FOK vs. REC, and FOK vs. PreA), and right inferior prefrontal (BA 44/45), dorsal medial prefrontal (BA 8/32), and bilateral occipital (BA 18/19) regions showed consistently less activity for metamemory compared to non-metamemory tasks. In contrast, bilateral superior frontal (BA 9/10) and ventral medial prefrontal (BA 32) regions showed greater activity for both metamemory tasks compared to recognition, but not to attractiveness ratings.

The next set of ROI analyses examined the regions that showed differential activity in metamemory and non-metamemory tasks based on behavioral responses. All the regions that had previously shown greater activity for metamemory compared to recognition [i.e., IPL (BA 40), TPJ (BA 40/42), STG (BA 42/22), mid/posterior cingulate (BA 31), superior frontal (BA 9/10), and ventral medial prefrontal (BA 32) regions] showed significant main effects of task, but no other main effects or interactions with the exception of the left IPL (BA 40). The left IPL (BA 40) showed a significant Task  $\times$  Accuracy interaction for FOK and REC, with greater activity for misses than hits during FOK and greater activity for hits than misses during REC. Similarly, the right MTL (BA 35/36), which had shown greater activity during FOK compared to non-metamemory tasks, showed a significant main effect of FOK > REC, but no other main effects or interactions. Some of the regions that had shown less activity during metamemory tasks compared to non-metamemory tasks, such as the right inferior prefrontal (BA 44/45) and dorsal medial prefrontal (BA 8/32) regions, similarly showed main effects of task, but no other main effects or interactions. Occipital (BA 18/19) and the left inferior prefrontal (BA 44/45) regions showed significant main effects of task, and also showed evidence of modulation based on the subjective metamemory rating given and recognition accuracy. The left occipital region (BA 18/19) showed greater activity during FOK ratings for high than low responses for hits, but not misses [ $F(1, 12) = 9.42, p < .01$ ]; the right occipital region (BA 18/19) showed a marginal main effect of accuracy for FOK and REC with hits greater than



**Figure 3.** Whole-brain analyses comparing metamemory tasks (feeling-of-knowing and confidence judgments) to non-metamemory tasks (recognition and attractiveness ratings).



**Table 2.** Regions of Interest were Generated Based on Significant Whole-brain Level Contrasts Comparing All Metamemory Tasks (FOK, CONF) to All Non-metamemory Tasks (REC, PreA, PostA)

<i>Region</i>	<i>MNI Coordinates</i>	<i>BA</i>	<i>CONF &gt; REC</i>	<i>CONF &gt; PostA</i>	<i>FOK &gt; REC</i>	<i>FOK &gt; PreA</i>
<i>Metamemory &gt; Non-metamemory</i>						
Tempo-parietal junction						
Left	-60 -30 21	40/42	$p < .002$	$p < .013^*$	$p < .00001$	$p < .002$
Right	63 -30 18	40/42	$p < .0004$	$p < .007$	$p < .00001$	$p < .002$
Superior temporal gyrus						
Left	-42 -18 -9	42/22	$p < .0006$	$p < .08^*$	$p < .00001$	$p < .01$
Right	45 -15 -6	42/22	$p < .00004$	$p < .013^*$	$p < .00005$	$p < .003$
Cingulate gyrus	3 -27 48	31	$p < .03^*$	$p < .006$	$p < .00001$	$p < .004$
R. Inferior parietal lobule	60 -45 39	40	$p < .00003$	$p < .011$	$p < .00001$	$p < .011$
<i>Metamemory &gt; Recognition</i>						
Superior frontal						
Left	-36 42 33	9	$p < .003$	<i>ns</i>	$p < .00003$	<i>ns</i>
Right	30 60 21	9/10	$p < .0003$	<i>ns</i>	$p < .00002$	<i>ns</i>
Ventral medial prefrontal	6 42 3	32	$p < .0001$	<i>ns</i>	$p < .00001$	<i>ns</i>
L. Inferior parietal lobule	-57 -48 42	40	$p < .00001$	$p < .00008$	$p < .00002$	<i>ns</i>
<i>FOK &gt; Non-metamemory</i>						
R. Medial temporal lobe	33 -30 -12	36	<i>ns</i>	<i>ns</i>	$p < .0001$	$p < .004$
<i>Region</i>	<i>MNI Coordinates</i>		<i>REC &gt; CONF</i>	<i>PostA &gt; CONF</i>	<i>REC &gt; FOK</i>	<i>PreA &gt; FOK</i>
<i>Non-metamemory &gt; Metamemory</i>						
Occipital						
Left	18 -69 -6	18/19	$p < .00001$	$p < .001$	$p < .0004$	$p < .005$
Right	-15 -75 -6	18/19	$p < .0001$	$p < .006$	$p < .0001$	$p < .02^*$
R. inferior frontal	54 39 12	44/45	$p < .00001$	$p < .001$	$p < .001$	$p < .002$
Dorsal medial prefrontal	-6 27 42	8/32	$p < .001$	$p < .007$	$p < .0001$	$p < .03^*$
<i>Non-metamemory &gt; CONF</i>						
R. Cuneus	33 -81 30	19	$p < .001$	$p < .01$	<i>ns</i>	<i>ns</i>
L. Inferior frontal	-45 27 15	44/45	$p < .02^*$	$p < .01$	<i>ns</i>	$p < .02^*$

Paired *t* tests compared each metamemory task, feeling-of-knowing (FOK), and confidence (CONF), to recognition (REC) and the relevant attractiveness rating (PreA, PostA). ROIs were categorized based on the results of the post hoc *t* tests as: (1) showing differences for metamemory compared to non-metamemory if all four comparisons in the header row were significant; (2) showing differences for metamemory compared to recognition if the metamemory tasks were different from recognition and not different from attractiveness ratings; or (3) showing differences for one metamemory task only (FOK or CONF) if only one metamemory task differed from recognition and attractiveness ratings. Reported *p* values are one-tailed and were considered significant at  $p < .0125$  in order to correct for multiple comparisons.

\*Indicates *p* value that does not survive correction for multiple comparisons.

misses [ $F(1, 12) = 4.61, p < .053$ ], and during CONF, but not REC, showed greater activity for low than high responses [ $F(1, 12) = 5.07, p < .044$ ]. The left inferior frontal region (BA 44/45) also showed a main effect of

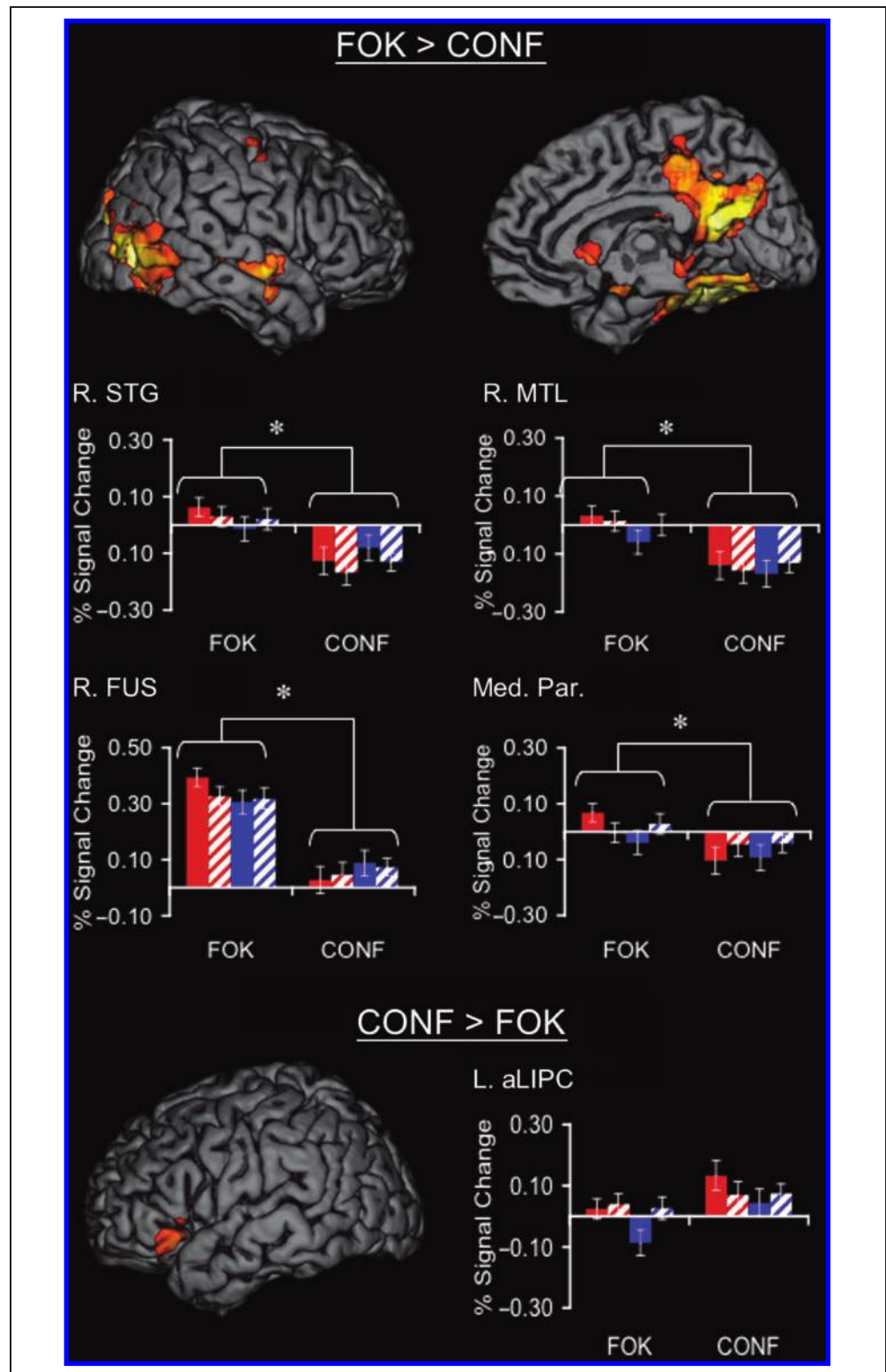
level for FOK and REC sorted by FOK [ $F(1, 12) = 7.4, p < .019$ ], and CONF and REC sorted by CONF [ $F(1, 12) = 6.09, p < .03$ ], with greater activity for high than low responses.

*Differential Activity in Feeling-of-Knowing and Confidence Judgments*

Regions that showed differential activity for the two metamemory tasks were assessed by directly contrasting CONF and FOK (Figure 4). An anterior left infe-

rior prefrontal region (BA 47) showed greater activity during CONF than FOK. Several regions showed greater activity during FOK compared to CONF, including bilateral occipital (BA 18/19), fusiform (BA 37), hippocampal formation, medial parietal (including the parts of the posterior cingulate, retrosplenial cortex, and

**Figure 4.** Regions that showed differences in activity for feeling-of-knowing (FOK) and confidence judgments (CONF). Regions for FOK > CONF are shown on the right hemisphere only, but a similar pattern was observed on the left. There was greater activity during FOK than CONF medial parietal (Med. Par; BA 31/7), medial temporal lobe (MTL), fusiform (Fus; BA 37), and superior temporal gyrus (STG; BA 21). An anterior left inferior prefrontal region (aLIPC; BA 47) was significant for CONF > FOK. Graphs depict percent signal change in these regions sorted by task (FOK or CONF), level (high or low FOK or CONF), and recognition accuracy (hits or misses) with significant effects of tasks and task interactions (also see Table 3). Post hoc analyses with  $2 \times 2 \times 2$  Task  $\times$  Level  $\times$  Accuracy ANOVAs on percent signal change data indicated that FOK > CONF regions showed consistent task differences across behavioral response, whereas CONF > FOK regions may have been driven by specific behavioral effects.



**Table 3.** *p* Values from Post Hoc Repeated Measures ANOVAs (Task × Level × Accuracy) for Clusters Showing Greater Activity during Feeling-of-Knowing Compared to Confidence Judgments for Task and Task × Level Interactions

<i>Region</i>	<i>BA</i>	<i>FOK &gt; CONF</i>	<i>Task × Level</i>	<i>High FOK &gt; High CONF</i>	<i>Low FOK &gt; Low CONF</i>	<i>High FOK &gt; Low FOK</i>	<i>Low CONF &gt; High CONF</i>
R. Fusiform	37	$p < .00001$	$p < .032$	$p < .00001$	$p < .00003$	$p < .021$	<i>ns</i>
L. Fusiform	37	$p < .0001$	$p < .05$	$p < .0003$	$p < .0003$	<i>ns</i>	<i>ns</i>
Med. Parietal	31/7	$p < .005$	<i>ns</i>	$p < .005$	$p < .009$	$p < .08$	<i>ns</i>
R. Hippocampal formation		$p < .00005$	$p < .005$	$p < .00001$	$p < .005$	$p < .012$	<i>ns</i>
L. Hippocampal formation		$p < .054$	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>	<i>ns</i>
R. Superior temporal	21	$p < .0003$	$p < .012$	$p < .0002$	$p < .003$	$p < .048$	<i>ns</i>

Regions that showed Task × Level interactions showed consistent differences between tasks (significant high FOK > high CONF and low FOK > low CONF), and the interaction was driven by differences between high and low responses within tasks.

precuneus; BA 31/7), and right superior temporal (BA 21) regions.

The regions that showed significant differences in activity comparing FOK and CONF were entered into post hoc ROI analyses using  $2 \times 2 \times 2$  repeated measures ANOVA testing for effects of task (FOK and CONF), subjective level of FOK or CONF expressed (high and low), recognition accuracy (hits and misses), and their interactions (Figure 4; Table 3). As expected, occipital (BA 18/19), fusiform (BA 37), hippocampal formation, medial parietal region (BA 31/7), and right superior temporal (BA 21) regions showed main effects of task at the ROI level with FOK > CONF ( $p < .005$ ), and the left hippocampal formation ROI was marginally significant ( $p < .054$ ). In addition to main effects of task, the right fusiform (BA 37), left fusiform (BA 37), right hippocampal formation, and right STG (BA 21) showed significant Task × Level interaction effects ( $p < .05$ ) that were driven by within-task differences, specifically that during FOK there tended to be greater activity during high than low responses, whereas during CONF there tended to be greater activity during low than high responses. There were no other significant main effects or interactions in these regions. The left anterior inferior prefrontal region (BA 47) showed a significant main effect of level [ $F(1, 12) = 8.61, p < .012$ ], with greater activity during high compared to low responses.

#### *Subjective Level of Feeling-of-Knowing or Confidence Expressed*

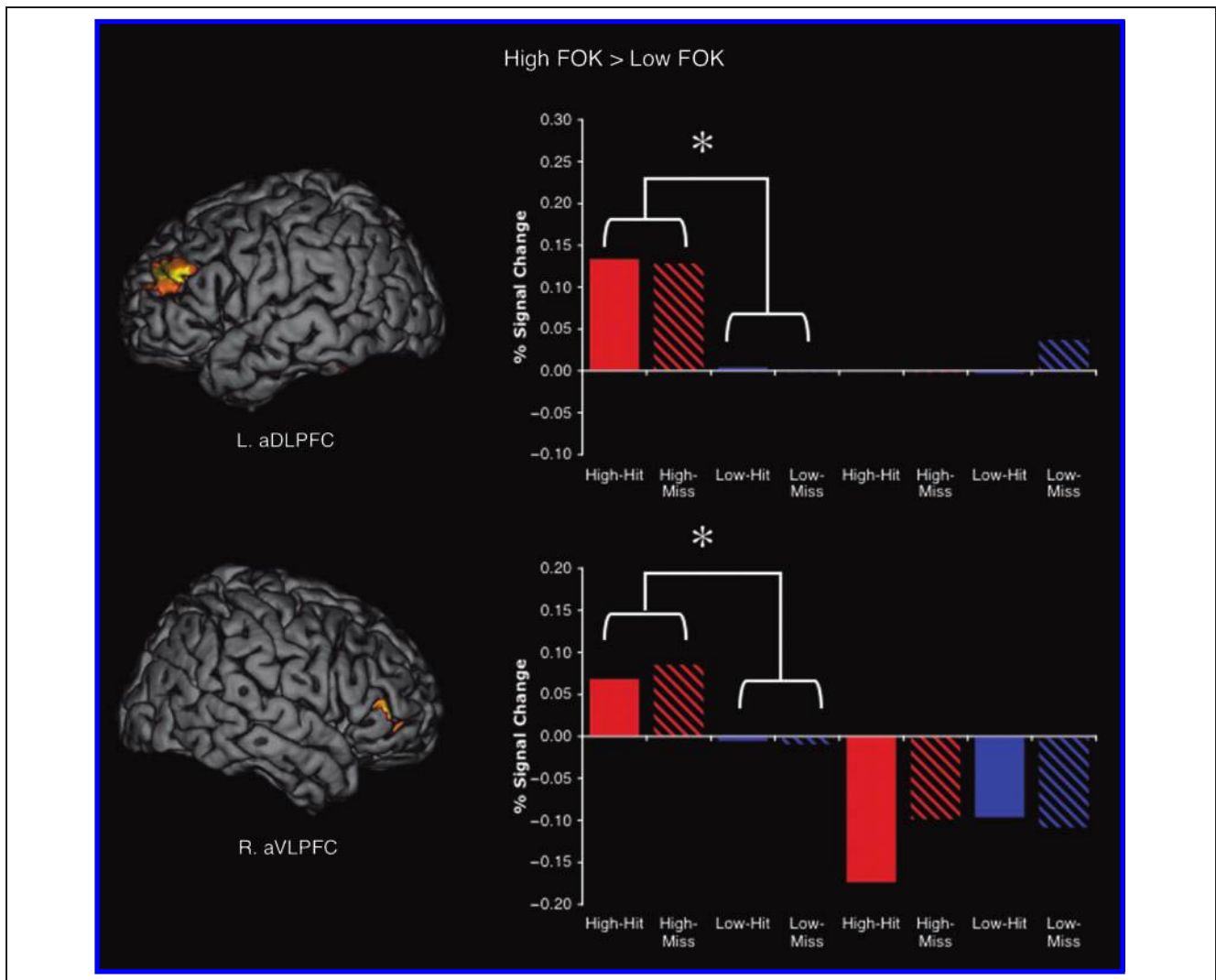
Whole-brain analyses for FOK judgments showed two prefrontal regions with greater activity for high FOK compared to low FOK (Figure 5), including the anterior right ventrolateral prefrontal cortex (aVLPFC; 45, 39, 3; BA 45) and the anterior left dorsolateral prefrontal cortex (aDLPFC; -39, 42, 24; BA 10/9). We then performed an ANOVA that examined effects of metamemory task (FOK or CONF), level of the rating (high or low CONF/FOK), accuracy (hit or miss), and their interaction, and showed

a Task × Level interaction [ $F(1, 12) = 12.04, p < .005$ ] in the right aVLPFC (BA 45) and a marginally significant Task × Level interaction [ $F(1, 12) = 3.96, p < .07$ ] in the left aDLPFC (BA 10/9) region. The right aVLPFC (BA 45) also showed a main effect of task [ $F(1, 12) = 6.46, p < .026$ ], with greater activity during FOK than CONF being driven by signal increases or baseline activity during FOK and signal decreases during CONF. Thus, the regions that modulated by level of FOK did not modulate by level of CONF. For CONF judgments, there were no regions that showed significantly greater activity for high confidence responses compared to low confidence responses at the whole-brain level.

## DISCUSSION

The goal of this study was to identify brain regions that are specifically involved in metamemory monitoring, in terms of the particular task performed and the subjective metamemory judgment that was made. Both metamemory tasks were characterized by greater activity in medial prefrontal, mid/posterior cingulate, and lateral parietal and temporal regions, and less activity in occipital, lateral inferior frontal, and dorsal medial prefrontal regions, compared with non-metamemory tasks. Based on previous findings that suggest medial prefrontal, medial parietal, and lateral parietal regions are involved in internally directed cognition (Gusnard & Raichle, 2001; Raichle et al., 2001), and occipital, lateral inferior frontal, and dorsal medial prefrontal regions are involved in stimulus directed cognition (e.g., Corbetta & Shulman, 2002), common metamemory monitoring mechanisms appear to be characterized by both a shift toward internally directed cognition and away from externally directed cognition.

Many regions demonstrated differences in neural activity between FOK and CONF, including fusiform, hippocampal formation, and medial parietal regions; furthermore, these regions also showed interaction effects between task and the subjective metamemory rating,



**Figure 5.** Regions that modulated based on high or low feeling-of-knowing (FOK), regardless of accuracy. Mapwise comparisons revealed greater activity during high FOK compared to low FOK in the left anterior dorsolateral prefrontal (BA 10/9; top) and the right anterior ventrolateral prefrontal (BA 45; bottom) regions. Repeated measures ANOVAs on percent signal change in these regions showed Task  $\times$  Level interactions and indicated that these regions modulated based on level of FOK expressed but not on level of confidence (CONF) expressed.

suggesting that they are sensitive to the information being monitored in each particular task. Other lateral prefrontal regions modulated based on the subjective level of FOK and/or level of CONF expressed. These findings demonstrate both common neural mechanisms supporting metamemory monitoring demands and distinct mechanisms relating to the specific metamemory task, and may also serve to elucidate the functional roles of previously characterized brain networks.

### Common Neural Correlates of Performing Metamemory Tasks

We compared metamemory to non-metamemory tasks in order to determine which brain regions were associated with performing metamemory tasks, regardless of behavioral response. Several regions showed differential neural activity for both metamemory tasks (FOK and

CONF) compared to non-metamemory tasks (REC and attractiveness), and this finding was consistent across subjective metamemory ratings and recognition accuracy. Because metamemory tasks explicitly require subjects to monitor the products of memory retrieval, common modulation of activity in brain regions during both metamemory tasks may reflect shared aspects of metamemory monitoring. However, metamemory monitoring is associated with a number of other processes, and modulation in these regions may be related to these processes. For example, memory monitoring requires a shift from the external stimulus to the internally generated products of retrieval. Less activity during metamemory tasks in the occipital cortex (BA 18/19), which is known to be involved in processing visual information, is consistent with a shift to internally directed representations compared to external stimulus representations during metamemory monitoring. Along similar lines, the right inferior prefrontal

(BA 44/45) and dorsal medial prefrontal (BA 8/32) regions, which showed less activity in metamemory tasks compared to non-metamemory tasks, have been implicated in stimulus-driven attention (e.g., Corbetta & Shulman, 2002), and decreased activity in these regions may be related to a shift away from externally directed attention.

Previous research has shown greater activity in the dorsal medial prefrontal cortex (BA 32) associated with tip-of-the-tongue states (Maril et al., 2005; Maril, Wagner, & Schacter, 2001), which involves a subjective metamemory rating, so it may be surprising that this region showed less activity during metamemory tasks compared to non-metamemory tasks. However, the types of analyses done in our study and the Maril et al. (2001, 2005) studies are different; we compared metamemory to non-metamemory tasks and they compared different subjective metamemory judgments. The activation in the dorsal medial prefrontal cortex (BA 32) associated with tip-of-the-tongue states is typically interpreted in terms of conflict monitoring (Maril et al., 2001, 2005). We showed greater activity in the dorsal medial prefrontal cortex during a three-alternative forced-choice recognition paradigm, and this task may elicit increased conflict monitoring because subjects may experience conflict when they attempt to choose between target and distracter names.

There was greater activity in regions associated with internally directed cognition during metamemory tasks compared to non-metamemory tasks. The posterior cingulate (BA 31) and the right IPL (BA 40), which showed greater activity during metamemory compared to non-metamemory tasks, and the ventral medial prefrontal region (BA 32) and the left IPL (BA 40), which showed greater activity during metamemory tasks compared to recognition, have previously been characterized as being involved in “default” mode processing (Gusnard & Raichle, 2001; Raichle et al., 2001). One reason these regions were characterized as part of the “default network” was that they show consistent task-induced deactivations across a wide variety of tasks (Shulman et al., 1997). It has been hypothesized that these task-induced deactivations reflect a shift from internal to external processing (Gusnard & Raichle, 2001; Raichle et al., 2001). Accordingly, the greater activity in the ventral medial prefrontal cortex (BA 32), the posterior cingulate (BA 31), and the IPL (BA 40) for metamemory compared to non-metamemory tasks is likely due to a shift toward internally directed cognition.

Although comparisons of metamemory to non-metamemory tasks showed differences in regions similar to “default” network regions, it is worth noting that many of the clusters of activation extended into neighboring regions that are likely to subservise different functions and represent engagement of additional functional networks. The posterior cingulate (BA 31/23), the IPL (BA 40), and the ventral medial prefrontal (BA 32/10/12) cortex have been shown to be core regions of the “default” network, but these regions also correlated with

other subsystems, including the MTL and dorsal medial prefrontal subsystems (Buckner, Andrews-Hanna, & Schacter, 2008). These other subsystems have shown correlations with lateral temporal and parietal regions. It is therefore likely that the large cluster of activation that included the IPL (BA 40), the TPJ (BA 40/42), and the STG (BA 42/22) reflects the engagement of these different subsystems. A recent study compared thinking about others’ thoughts to thinking about others’ appearance or bodily sensations, and showed greater MR signal in the posterior cingulate (BA 31) and the TPJ (BA 42/22) for thinking about others’ thoughts (Saxe & Powell, 2006). Based on these findings, in conjunction with our own findings regarding the involvement of these regions in metamemory, we suggest that the posterior cingulate (BA 31) and TPJ (BA 42/22) regions may play a role in monitoring or thinking about cognition.

The differences between metamemory and non-metamemory tasks in ventral medial prefrontal (BA 32), posterior cingulate (BA 31), and lateral temporal/parietal (BA 40/42/22) regions were driven by deactivations during non-metamemory tasks and activity nearer to baseline during metamemory tasks. Deactivations are quite common with passive baseline tasks, such as our own, and would likely change to activations with an active baseline task (Stark & Squire, 2001). fMRI comparisons always involve contrasting one condition to another condition, even if a task is labeled as a “baseline” task; therefore, the relative difference between conditions is important. We previously showed that the posterior cingulate (BA 31) and lateral parietal regions (BA 40) showed greater activity for CONF compared to REC, and in this study, we replicated those findings. These regions often show greater signal decreases during more difficult tasks (McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003), and one interpretation of the data could be that the differential activity during metamemory tasks and recognition is because recognition is more cognitively demanding. Importantly, in this study, we also showed that these regions showed greater activity for CONF compared to attractiveness, which demonstrates that these differences are not solely driven by recognition demands. Furthermore, we also observed greater activity during FOK compared to REC and attractiveness, which also suggests that these regions modulate based on metamemory task demands.

The ventral medial prefrontal cortex (BA 32), unlike the posterior cingulate (BA 31), TPJ (BA 40/42), and STG (BA 42/22) regions, showed greater activity for metamemory tasks compared to recognition, but not attractiveness ratings. Previous studies have implicated the medial prefrontal cortex in self-related processing (e.g., Mitchell, Banaji, & Macrae, 2005; Johnson et al., 2002; Frith & Frith, 1999). Thus, it may be that metamemory tasks and attractiveness judgments both require self-related processing, but recognition tasks require less self-related processing.

## Differences between Feeling-of-Knowing and Confidence Judgments

Although FOK and CONF share some common mechanisms, both cognitively and neurally, they also differ in the information being monitored and the basis for the judgment (Shaw & Zerr, 2003; Bradfield et al., 2002; Koriat & Levy-Sadot, 2001; Busey et al., 2000; Belli et al., 1994; Yonelinas, 1994; Koriat, 1993; Metcalfe et al., 1993). Thus, we expect differential neural activity during the two metamemory tasks to be related to the differences in cognitive bases for FOK and CONF. FOK showed greater activity in several regions at the map level compared to CONF, including medial parietal (BA 31/7), fusiform (BA 37), right superior temporal (BA 21), and hippocampal formation regions. A subset of these regions (right fusiform [BA 31/7], left fusiform [BA 37], right hippocampal formation, and right STG [BA 21]) also showed differences in activity related to the subjective level expressed during the tasks; there was greater activity for high compared to low FOK responses, whereas during confidence there was the opposite effect, with low responses showing greater activity than high responses. This pattern shows that brain regions may modulate based on both performing the task and the behavioral response. It is likely that the fusiform activity is related to processing of the cue (in this case, a face) that occurs during FOK judgments (e.g., Metcalfe et al., 1993) and that hippocampal activity is related to partial access to the semantic, perceptual, or affective attributes of the target (e.g., Koriat, 1993), which serve for the basis of FOK, but not confidence judgments. Furthermore, the findings that the fusiform (BA 37) and the right hippocampal formation show greater activity during high FOK compared to low FOK judgments, together with the opposite effect during confidence judgments, also suggest that activity in these regions is sensitive to the information being monitored and the subjective outcome judgment.

A medial parietal region (BA 31/7) showed greater activity during FOK compared to CONF, in addition to showing greater activity during metamemory tasks compared to non-metamemory tasks. Although both tasks require metamemory monitoring, they may not be equivalent; during the FOK task, subjects attempt to recall the name and must monitor the retrieved information in order to complete the task, whereas in CONF, subjects have already retrieved and monitored information and are required to come up with a final judgment. One possible explanation for the observed differences in medial parietal activity is that this region is sensitive to the amount of information being internally generated and monitored.

There are a few caveats worth considering related to our findings. First, there were significant differences in RT for all comparisons of interest. RT was not entered as a covariate in our model because RT was highly

collinear with specific task functions. Another issue is that the numbers of trials varied between tasks, resulting in unequal power to detect differences between tasks; however, given that there were changes in both directions (e.g., metamemory > non-metamemory, and non-metamemory > metamemory), this inequality cannot fully account for our findings. Furthermore, there may be task order effects, especially comparing FOK to CONF; at the time of the CONF trials, subjects had seen the face two more times than they had during FOK trials. Thus, repetition suppression effects may have influenced our effects (Schacter, Wig, & Stevens, 2007; Henson & Rugg, 2003). Some regions that showed greater activity for FOK compared to confidence judgments, however, also showed effects of level of FOK, indicating that their activity is not entirely driven by repetition suppression. A final caveat relates to the design of the FOK task. In our design, subjects were asked to attempt to recall the name associated with the face and to make the FOK judgment in a single step. This observation raises the possibility that FOK-related activity is confounded with recall-related neural activity. Although this possibility cannot be rejected unequivocally, if the activity were related to recall, we would expect that all of the regions that showed greater activity for FOK compared to REC would also show greater activity compared to CONF, but this effect was not observed.

## Subjective Level of Feeling-of-Knowing or Confidence Expressed

Another component in understanding the neural basis of metamemory involves determining which brain regions modulate based on the subjective level of FOK or level of CONF expressed. Mapwise comparisons of high and low FOK and CONF revealed that the right VLPFC (BA 45) and the left DLPFC (BA 10/9) modulated based on the level of FOK, but not based on the level of CONF. This finding is consistent with reports in the literature that indicate different neural representation of level of FOK and CONF in separate studies; it has been shown that prefrontal and parietal regions tend to modulate based on the level of FOK (Maril et al., 2003, 2005; Kikyo & Miyashita, 2004; Kikyo et al., 2002), whereas the MTL, medial parietal, and medial prefrontal regions tend to modulate based on the level of CONF (Chua et al., 2006; Moritz et al., 2006). However, previous studies of FOK had not analyzed both correct and incorrect memory responses separately, and our study confirms that activity in these regions is indicative of subjective experience and not objective accuracy. However, to our knowledge, FOK and CONF have not been performed previously within the same study. By comparing them directly, we confirmed that the right VLPFC (BA 45) and the left DLPFC (BA 10/9) modulated based on the level of FOK, but not on the level of CONF. Although the right VLPFC (BA 45) and the left DLPFC (BA 10/9) did not modulate

based on the confidence level, a more anterior and inferior region of the left inferior prefrontal gyrus (BA 47) did show a main effect of level, indicating there may be regions that consistently modulate based on the subjective level of metamemory rating.

### **A Broader Understanding of the Neural Basis of Metamemory**

The literature on the cognitive neuroscience of metamemory, particularly for FOK, currently focuses on the importance of the prefrontal cortex in metamemory judgments (for a review, see Schwartz & Bacon, 2008). Evidence that the prefrontal cortex is critical for metamemory judgments typically comes from analyses of monitoring accuracy and/or the subjective level of the judgment. The findings that highlight the importance of the frontal lobes in metacognition suggest a relationship between metacognition and executive control processes (Botvinick, 2007; Shimamura, 2000). Consistent with the metamemory literature, we also showed modulation in the lateral prefrontal cortex based on the subjective level of FOK (BA 47, 10/9, 45) and level of CONF (BA 47) expressed. However, a fuller understanding of the neural basis of metamemory would also include which brain regions are involved in performing metamemory tasks, regardless of behavioral performance. In this study, we documented brain regions that are involved in performing multiple metamemory tasks, only one metamemory task, and that modulate based on the subjective metamemory rating given. Both metamemory tasks were characterized by greater activity in regions subserving internal attention and less activity in regions subserving external attention, which we suggest is related to monitoring one's own memory. Comparing FOK and CONF revealed brain regions that have roles in metamemory that are related to the specific metamemory task, which included the hippocampal formation, the fusiform (BA 37), the medial parietal cortex (BA 31/7), and the left inferior prefrontal cortex (BA 47). These findings do not undermine ideas about the role of the frontal cortices in metamemory, but instead show other ways that metamemory processes are represented in the brain, including neural activity that modulates based on the task and also based on behavioral response.

The brain regions that we have discussed with respect to metamemory have been previously implicated in studies that have focused more on factors related to accurate memory retrieval. Many of these studies used combined measures of subjective and objective factors (e.g., high confidence correct trials) to examine memory (Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000; Henson, Rugg, Shallice, & Dolan, 2000; Henson, Rugg, Shallice, Josephs, & Dolan, 1999). The finding that many of these regions show modulation based on metamemory task and the subjective judgment suggests that future research may benefit from attempting to distin-

guish between the memory and metamemory components. It is, of course, possible—and even likely—that high levels of FOK and confidence are related to retrieved content (Koriat & Goldsmith, 1996), making metamemory judgment and objective accuracy difficult to disentangle. Nonetheless, it may be useful to dissociate them when trying to understand the functional contributions of specific brain regions to memory function.

### **Acknowledgments**

We thank Kristina Depeau, Kim Celone, and Saul Miller for help with data acquisition. This work was supported by NINDS: K23-NS02189 (R. S.); NIA: P01-AG-04953 (R. S.) and R01-AG027435 (R. S.); the AFAR Beeson Scholars in Aging Program (R. S.); NIMH: MH60941 (D. S.); a National Science Foundation Graduate Research Fellowship (E. C.); NCRR: P41RR14075; the MIND Institute; the Athinoula A. Martinos Center for Biomedical Imaging; and the Harvard Center for NeuroDiscovery.

Reprint requests should be sent to Elizabeth F. Chua, Harvard Psychology, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, or via e-mail: echua@nmr.mgh.harvard.edu.

### **REFERENCES**

- Belli, R. F., Lindsay, D. S., Gales, M. S., & McCarthy, T. T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Memory & Cognition*, *22*, 40–54.
- Botvinick, M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, *7*, 356–366.
- Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology*, *87*, 112–120.
- Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002). *Region of interest analysis using an SPM toolbox*. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, Sendai, Japan.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.
- Busey, T. A., Tunnickliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*, 26–48.
- Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: Neural correlates of the cognitive process and subjective level of confidence in recognition memory. *Neuroimage*, *29*, 1150–1160.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215.
- Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, *3*, 1149–1152.
- Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science*, *286*, 1692–1695.
- Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*, 685–694.

- Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208–216.
- Henson, R. N., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia, 41*, 263–270.
- Henson, R. N., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. *Journal of Cognitive Neuroscience, 12*, 913–923.
- Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience, 19*, 3962–3972.
- Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain, 125*, 1808–1814.
- Kikyo, H., & Miyashita, Y. (2004). Temporal lobe activations of “feeling-of-knowing” induced by face–name associations. *Neuroimage, 23*, 1348–1357.
- Kikyo, H., Ohki, K., & Miyashita, Y. (2002). Neural correlates for feeling-of-knowing. An fMRI parametric analysis. *Neuron, 36*, 177.
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.
- Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 34–53.
- Maccotta, L., Zacks, J. M., & Buckner, R. L. (2001). Rapid self-paced event-related functional MRI: Feasibility and implications of stimulus- versus response-locked timing. *Neuroimage, 14*, 1105–1121.
- Maril, A., Simons, J. S., Mitchell, J. P., Schwartz, B. L., & Schacter, D. L. (2003). Feeling-of-knowing in episodic memory: An event-related fMRI study. *Neuroimage, 18*, 827–836.
- Maril, A., Simons, J. S., Weaver, J. J., & Schacter, D. L. (2005). Graded recall success: An event-related fMRI comparison of tip of the tongue and feeling of knowing. *Neuroimage, 24*, 1130–1138.
- Maril, A., Wagner, A. D., & Schacter, D. L. (2001). On the tip of the tongue. An event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron, 31*, 653–660.
- McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of Cognitive Neuroscience, 15*, 394–408.
- Metcalf, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 851–861.
- Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17*, 1306–1315.
- Moritz, S., Glascher, J., Sommer, T., Buchel, C., & Braus, D. F. (2006). Neural correlates of memory confidence. *Neuroimage, 33*, 1188–1193.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26*, 125–173.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A., 98*, 676–682.
- Saxe, R., & Powell, L. J. (2006). It’s the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science, 17*, 692–699.
- Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology, 17*, 171–176.
- Schwartz, B. L., & Bacon, E. (2008). Metacognitive neuroscience. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory*. New York: Taylor & Francis.
- Shaw, J. S., III, & Zerr, T. K. (2003). Extra effort during memory retrieval may be associated with increases in eyewitness confidence. *Law and Human Behavior, 27*, 315–329.
- Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition, 9*, 313–323; discussion 324–316.
- Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Meizin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience, 9*, 648–663.
- Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J., Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Research, Cognitive Brain Research, 17*, 75–82.
- Stark, C. E., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences, U.S.A., 98*, 12760–12766.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1341–1354.