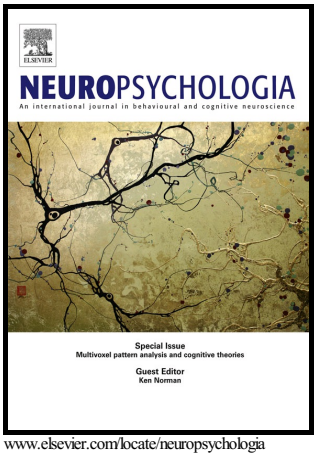
Knowledge Supports Memory Retrieval through Familiarity, Not Recollection

Wei-Chun Wang, Nadia M. Brashier, Erik A. Wing, Elizabeth J. Marsh, Roberto Cabeza



www.elsevier.com/locate/neuropsychologia

PII: S0028-3932(18)30019-8
DOI: https://doi.org/10.1016/j.neuropsychologia.2018.01.019
Reference: NSY6648

To appear in: *Neuropsychologia*

Received date: 31 August 2017
Revised date: 18 December 2017
Accepted date: 15 January 2018

Cite this article as: Wei-Chun Wang, Nadia M. Brashier, Erik A. Wing, Elizabeth J. Marsh and Roberto Cabeza, Knowledge Supports Memory Retrieval through Familiarity, Not Recollection, *Neuropsychologia,* https://doi.org/10.1016/j.neuropsychologia.2018.01.019

Knowledge Supports Memory Retrieval through Familiarity, Not Recollection

Wei-Chun Wang, Nadia M. Brashier, Erik A. Wing, Elizabeth J. Marsh, & Roberto Cabeza

Duke University

Author Note

Wei-Chun Wang, Center for Cognitive Neuroscience, Duke University; Nadia M. Brashier, Center for Cognitive Neuroscience, Duke University; Roberto Cabeza, Center for Cognitive Neuroscience, Duke University; Elizabeth J. Marsh, Department of Psychology & Neuroscience, Duke University.

Correspondence concerning this article should be addressed to Wei-Chun Wang, Center for Cognitive Neuroscience, Duke University, Durham, NC, 27708. E-mail: ww83@duke.edu

Conflicts of interest: none

Abstract

Semantic memory, or general knowledge of the world, guides learning and supports the formation and retrieval of new episodic memories. Behavioral evidence suggests that this knowledge effect is supported by recollection—a more controlled form of memory retrieval generally accompanied by contextual details—to a greater degree than familiarity—a more automatic form of memory retrieval generally absent of contextual details. In the current study, we used functional magnetic resonance imaging (fMRI) to investigate the role that regions associated with recollection and familiarity play in retrieving recent instances of known (e.g., *The Summer Olympic Games are held four years apart*) and unknown (e.g., *A flaky deposit found in port bottles is beeswing*) statements. Our results revealed a surprising pattern: Episodic retrieval of known statements recruited regions associated with familiarity, but not recollection. Instead, retrieval of unknown statements recruited regions associated with recollection. These

data, in combination with quicker reaction times for the retrieval of known than unknown statements, suggest that known statements can be successfully retrieved on the basis of familiarity, whereas unknown statements were retrieved on the basis of recollection. Our results provide insight into how knowledge influences episodic retrieval and demonstrate the role of neuroimaging in providing insights into cognitive processes in the absence of explicit behavioral responses.

Keywords: knowledge, episodic memory, semantic memory, recollection, familiarity

1. Introduction

One fundamental idea in memory research is the distinction between memory for personally-experienced events, or *episodic memory*, and general knowledge of the world, or *semantic memory* (Tulving, 1972, 1984). While the field traditionally emphasizes how episodic memories and knowledge differ (in phenomenology, development, and vulnerability to change), more recent work highlights the complex relationship between the two. For example, knowledge about sports (e.g., Rawson & Van Overschelde, 2008), people (e.g., Van Overschelde & Healy, 2001), and aviation (e.g., Meade, Nokes, & Morrow, 2009) support new learning (episodic memories) in those domains. Conversely, episodic memory supports the retrieval of knowledge, such as when drawing upon a personal memory of "my auntie's fruit bowl" to generate exemplars of fruits (Vallee-Tourangeau, Anthony, & Austin, 1998, p. 562).

Knowledge supports both the elaboration and organization of incoming information (Ericsson & Kintsch, 1995), with consequent benefits for later memory. A diehard baseball fan—with intimate knowledge of the rules, players' tendencies, and strategies—experiences a baseball game vastly differently than a cricket game. Our question involves understanding why the baseball game will later be remembered better than the cricket game. From a dual-process view of episodic memory retrieval (Yonelinas, 2002), this benefit could result from an increase in

recollection, familiarity, or both (but see, Wixted, 2007). *Recollection* involves relatively more conscious effort to think back to a particular time and place (i.e., contextual details), such as vividly remembering a game winning walk-off home run. In contrast, *familiarity* is relatively more automatic and does not invoke reliving; instead one simply knows the information (e.g., the fact that a home run won the game, without a sense of reliving that moment in time).

Separating the contributions of recollection and familiarity to episodic retrieval is challenging, especially since both are likely involved when discriminating old from new events (Yonelinas, 2002). For example, a face may be correctly identified as "old" because one remembers seeing the person in a particular place and time, or because the person feels familiar. One strategy is to measure properties that suggest recollection, such as memory for context, as opposed to simply collecting old/new judgments. Notably, it is easier to remember the background context associated with a famous face than an unfamiliar one (Reder et al., 2013), suggesting a role for recollection in the benefits of knowledge.

A second strategy involves measuring the subjective phenomenology associated with retrieval by asking people to label retrieved events as "remembered" or "known." While mapping processes onto remember-know judgments is imperfect (Donaldson, 1996; Wixted & Stretch, 2004), "remember" responses do capture at least part of the recollective process. For example, "remember" responses drive the memory benefits of meaning-based over perceptual-based encoding (Gardiner, 1988) and the advantage of pictures over words (Rajaram, 1996). Critically, knowledge appears to benefit episodic memory through "remembering" rather than "knowing." For example, people with more Star Trek knowledge (defined by the ability to discriminate Star Trek lifeforms from lures) were more likely to say that they "remembered" reading information from a Star Trek text than were novices, while "remember" responses were similar for the two groups when they were tested on a (control) psychology text (Long & Prat,

2002). Similar effects occur with vocabulary terms; after studying a list of technical terms, students were better able to discriminate old and new terms from their academic major, and that benefit reflected remembering rather than knowing (Brandt, Cooper, & Dewhurst, 2005). While these studies suggest that knowledge involves recollection-based retrieval, this finding may be unique to expert domains, as experts process stimuli in their domain of expertise more deeply (Kawamura, Suzuki, & Morikawa, 2007) and encode and retrieve information based on interitem associations within the domain of expertise (Kalakoski & Saariluoma, 2001). Therefore, expertise effects may not extend to general knowledge in a non-expert domain.

We took a different approach to disentangling recollection and familiarity by examining activation of brain areas associated with these two retrieval processes. Recollection-based retrieval has been associated with posterior midline, ventral parietal, anterior prefrontal, and hippocampal regions (Kim, 2013; Rugg & Vilberg, 2013; Skinner & Fernandes, 2007; Spaniol et al., 2009), particularly posterior hippocampus (Poppenk, Evensmoen, Moscovitch, & Nadel, 2013; Ranganath & Ritchey, 2012), whereas familiarity-based retrieval has most commonly been associated with perirhinal cortex (Henson, Cansino, Herron, Robb, & Rugg, 2003; Skinner & Fernandes, 2007), but also lateral prefrontal, including inferior frontal gyrus (IFG), and temporal regions (for a review, see Skinner & Fernandes, 2007). The present study (1) investigated the brain areas involved when knowledge supports episodic retrieval, using a recognition memory test, with a direct manipulation of knowledge (as opposed to comparing two different tasks), and (2) examined the mechanism underlying the benefits of knowledge, by linking them to brain areas associated with recollection and familiarity. We used three converging approaches to answer these questions.

To better understand how general knowledge influences memory for events, we manipulated whether or not stimuli were known to participants. We used facts drawn from

different domains (history, geography, science, etc.) that extensive piloting demonstrated were *known* (e.g., *The composer who worked in deafness was Beethoven*) or *unknown* (e.g., *The stick used in the game of shinty is a caman*). To examine episodic retrieval during a recognition memory test, we identified, for both known and unknown statements, regions showing greater activity for hits than correct rejections (*repetition enhancement*—RE)—as RE reflects the formation and retrieval of new episodic representations (Henson, 2003; Henson, Shallice, Gorno-Tempini, & Dolan, 2002). We also identified, for both known and unknown statements, regions showing less activity for hits than correct rejections (*repetition suppression*—RS)—as RS reflects the fluent processing of pre-existing semantic representations (Henson, 2003; Henson et al., 2002).

Second, we examined whether regions showing RE or RS effects related to individual differences in recognition memory performance for unknown versus known statements. Without a clear relationship with behavior, the putative correlates of episodic retrieval may be conflated with other processes such as priming (Dew & Cabeza, 2011; Paller, Voss, & Boehm, 2007). Therefore, significant correlations between neural repetition effects and behavior would support the idea that different regions support memory retrieval for information with or without pre-existing semantic representations.

Finally, using multi-voxel pattern analysis (MVPA), we examined whether unknown and known stimuli involved different memory representations. Previous studies demonstrated that activity patterns can distinguish old and new recognition memory trials (Rissman, Greely, & Wagner, 2010). Using a searchlight procedure, we investigated whether representational differences could also classify hits and correct rejections during the recognition test for unknown and known statements. Such a finding would support the idea that episodic retrieval of unknown and known statements differ not only in terms of memory processes, but also in their memory representations.

In sum, we investigated how knowledge supports episodic retrieval. After studying known and unknown statements, participants made old/new recognition decisions about these statements intermixed with new known and unknown statements. We investigated (1) the effects of this knowledge manipulation on regions showing RE or RS effects, (2) the correlation between RE or RS effects in these regions and individual differences in recognition memory performance for unknown and known statements, and (3) the difference in memory representations for unknown and known statements as detected by MVPA.

2. Material and method

2.1 Participants

The Duke University Institutional Review Board approved all procedures. Thirty-one native English speakers from Duke University and the surrounding communities participated for monetary compensation. Seven participants were excluded (three due to technical malfunctions with the scanner or testing computer, and four due to poor performance: two fell asleep, one was at chance, and the fourth failed to use the full scale). The final sample included 24 participants (age M = 23.17, SEM = 0.68; education M = 15.33, SEM = 0.41; 10 female), which allowed for an equal distribution of counterbalance orders.

2.2 Materials

Materials consisted of 360 trivia statements collected from the Internet that referred to known or unknown facts. Pilot participants (N = 47) rated these items from 1 (*definitely false*) to 6 (*definitely true*). For unknown facts, we created a true framing (e.g., *The inhabitable part of the world is the ecumene*) and a matching false framing (e.g., *The inhabitable part of the world is the*

toponym) that referred to a plausible, but incorrect, alternative.[1] Two thirds of the statements were unknown; of these items, half were true and the other half were false (counterbalanced across participants). Pilot participants responded similarly to unknown true and unknown false items (i.e., *guess false* or *guess true* responses from > 75% of participants, regardless of framing). As a result, we collapsed across framing in our analysis. The remaining one third of the items were known facts (e.g., *The capital of Spain is called Madrid*). These statements all appeared in a true framing, and pilot participants reliably and confidently rated them as true (i.e., *probably true* or *definitely true* responses from > 90% of participants).

2.3 Procedure

Following informed consent, participants incidentally encoded 180 statements (outside the scanner) by rating their interest on a 6-point scale from 1 (*very uninteresting*) to 6 (*very interesting*). The scale was reversed for half of the participants. Each statement appeared for 4 s, followed by a fixed 1 s interstimulus interval fixation. To maximize subsequent recognition performance, participants completed this task twice.

Participants then entered the scanner to perform a recognition memory task to examine episodic retrieval (i.e., is this statement old or new) and a truth rating task (i.e., is this statement true or false) in four separate counterbalanced ABBA runs (i.e., episodic-semantic-semantic-episodic for half of the participants, and semantic-episodic-episodic-semantic for the other half). The imaging data in the semantic task are not discussed further as they were collected to answer a different research question (Wang, Brashier, Wing, Marsh, & Cabeza, 2016). In the task of

[1] We created true and false framings of unknown facts to examine another research question unrelated to the relationship between knowledge and episodic retrieval. Response distributions for true and false unknown items completely overlapped.

interest, participants indicated whether statements were old or new on a 6-point scale from 1 (*definitely new*) to 6 (*definitely old*), with the scale being reversed for half of the participants. Overall, participants rated 60 old unknown, 60 new unknown, 30 old known, and 30 new known statements, divided equally across the two recognition runs. The different item types were intermixed, and each statement appeared for 5 s with a jittered (mean = 3 s; range: 1-8 s) interstimulus interval fixation.

2.4 Image Acquisition & Analysis

Images were collected on a 3T General Electric scanner with an 8-channel head coil at the Duke University Brain Imaging and Analysis Center. Functional images were acquired using a SENSE spiral sequence (64 x 64 matrix, repetition time = 2000 ms, echo time = 27 ms, field of view = 24cm, flip angle = 60°) and consisted of 34 axial slices acquired in an interleaved fashion. Slice thickness was 3.8 mm, resulting in 3.75 x 3.75 x 3.8 mm voxels. Additionally, high-resolution structural images were collected using a 3D, T1-weighted FSPGR sequence (256 x 256 matrix, 166 slices, 1 mm isotropic voxels).

Data were preprocessed with SPM12 (Wellcome Trust Centre for Neuroimaging). After discarding the first three scans of each run, the functional data for each participant were slice-time corrected, realigned, and coregistered to their respective anatomical images. The anatomical images were then segmented into separate grey and white matter images that were used to normalize the functional and anatomical images into MNI (Montreal Neurological Institute) space. Lastly, the normalized functional data were denoised using the DRIFTER toolbox (Särkkä et al., 2012) and spatially smoothed with an 8mm isotropic full-width at half-maximum Gaussian filter.

To assess the contributions of knowledge to episodic retrieval, we focused our analyses on high-confidence hits (i.e., *probably old* and *definitely old* responses) and correct rejections (i.e., *probably new* and *definitely new* responses). We modeled these responses for both known and unknown statements, resulting in four conditions of interest (mean trials per condition ranged from 28 to 52). Statistical analyses were performed in SPM12 using the general linear model (GLM). A high-pass filter of 128 s and grand mean scaling were applied to the data and serial correlations in the time series were accounted for using the autoregressive model (AR[1]). The functional data were modeled using a canonical hemodynamic response function with temporal derivatives with the stimulus onsets serving as event onsets. Guesses, incorrect responses (i.e., misses, false alarms), and missed trials were modeled separately. Note that guesses were excluded (Furman, Mendelsohn, & Dudai, 2012; Wagner et al., 1998), but the overall patterns were the same when (1) guesses were included, (2) the analyses were limited to the highest-confidence responses, and (3) false statements were removed from the analyses (thus balancing the number of unknown and known trials). Additional covariates of no interest included the six motion parameters estimated during realignment, baseline and session effects, global mean and motion outliers obtained from the Artifact Detection Toolbox (http://www.nitrc.org/projects/artifact_detect), and white matter and CSF signal time courses.

The MVPA searchlight analysis was conducted using The Decoding Toolbox (Hebart, Görgen, & Haynes, 2014) on unsmoothed single trial betas calculated using the least squares single (LSS) approach (Mumford, Turner, Ashby, & Poldrack, 2012), and subsequently smoothed with an 8mm isotropic full-width at half-maximum Gaussian filter for group analyses. A leave-one-run-out cross-validation scheme was utilized within a 3 voxel searchlight sphere.

Multiple comparison corrections ($p < .05$) were calculated with Monte Carlo simulations obtained in 3dClustSim (AFNI) using an uncorrected threshold of $p < .001$. Based on these calculations, whole brain univariate and MVPA analyses were conducted using a cluster size (cs) of no less than 27 contiguous voxels. Additionally, given the importance of the medial temporal lobes (MTL) in episodic retrieval, a region of interest (ROI) analysis of hippocampus, PRC, and parahippocampal cortex was also conducted using a cs of no less than 8 contiguous voxels. For conjunctions, a conjoint probability of $p < .001$ was achieved with uncorrected thresholds of $.010^2$ (Fisher, 1990). Brain-behavior correlation differences were calculated using the cocor package in R (Diedenhofen & Musch, 2015).

3. Results

3.1 Behavioral Results

We conducted a 2 (repetition: old, new) x 2 (knowledge: known, unknown) ANOVA on recognition memory ratings to examine interactions between knowledge and episodic memory (Figure 1A). There was a main effect of repetition ($F(1, 23) = 873.10$, $p < .001$, $\eta_p^2 = .97$), as participants unsurprisingly perceived old statements to be older than new statements. There was no main effect of knowledge ($F(1, 23) = 1.37$, $p > .05$, $\eta_p^2 = .06$), but consistent with evidence that knowledge benefits episodic memory (e.g., Poppenk, Köhler, & Moscovitch, 2010), there was a significant interaction between knowledge and repetition ($F(1, 23) = 48.76$, $p < .001$, $\eta_p^2 = .68$). Specifically, the mean oldness rating (on a 1-6 scale where 1 refers to "definitely new" and 6 refers to "definitely old") was higher for old items (i.e., rated more confidently as "old") if they referred to known ($M = 5.87$; $SEM = 0.05$) compared to unknown ($M = 5.43$; $SEM = 0.10$)

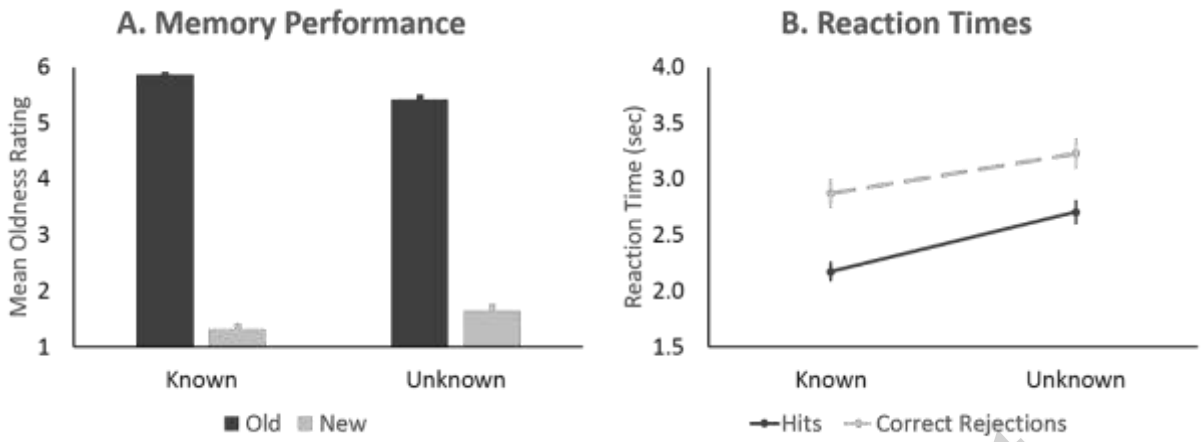


Figure 1. Mean oldness ratings for old and new statements (A) and reaction times for high-confidence hits and correct rejections (B) plotted as a function of knowledge. Error bars reflect standard error of the mean.

statements ($t(23) = 5.78$, $p < .001$, Cohen's $d = 1.02$), whereas the mean oldness rating was lower for new items (i.e., rated more confidently as "new") if they referred to known ($M = 1.32$; $SEM = 0.08$) compared to unknown ($M = 1.64$; $SEM = 0.13$) statements ($t(23) = -4.56$, $p < .001$, $d = -0.53$). An analysis of recognition sensitivity (d'), collapsing across confidence, yielded converging results; d' was greater for known ($M = 3.79$; $SEM = 0.14$) than unknown ($M = 2.94$; $SEM = 0.19$) statements ($t(23) = 6.14$, $p < .001$, $d = 1.01$).

For imaging analyses reported below, we focused on high confidence hits and correct rejections, and reaction times for these trials are illustrated in Figure 1B. We identified main effects of repetition ($F(1, 23) = 139.46$, $p < .001$, $\eta_p^2 = .86$) and knowledge ($F(1, 23) = 50.74$, $p < .001$, $\eta_p^2 = .69$). These main effects, however, should be interpreted in the context of a marginally significant two-way interaction ($F(1, 23) = 3.57$, $p = .07$, $\eta_p^2 = .13$). Specifically, the difference in reaction times between hits and correct rejections (i.e., hits faster than correct rejections) was greater for known than unknown statements.

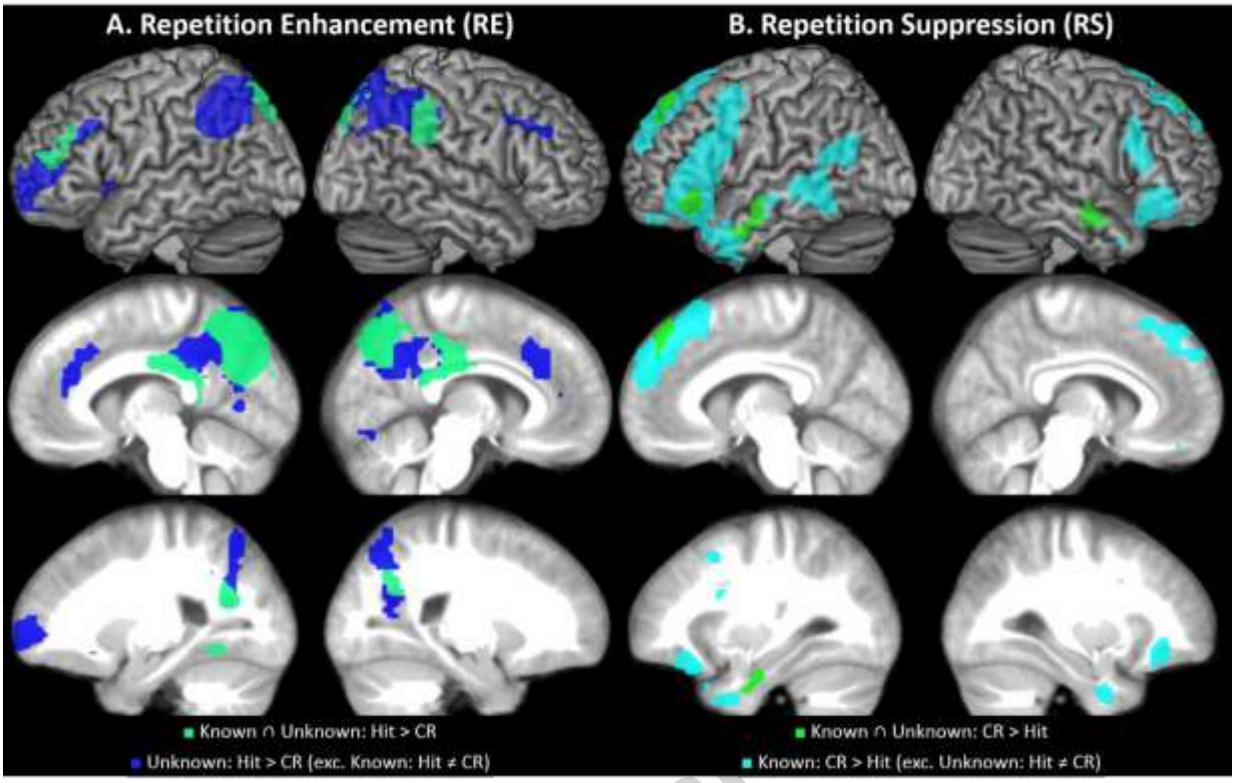


Figure 2. Regions exhibiting repetition enhancement (RE) effects for both known and unknown statements (green) or for unknown statements only (blue) (A). Regions exhibiting repetition suppression (RS) effects for both known and unknown statements (green) or for known statements only (cyan) (B). Smoothed for visualization purposes.

3.2 fMRI Results

We used three converging approaches to investigate the contributions of knowledge to episodic retrieval: (1) the effects of the knowledge manipulation on regions showing repetition effects, (2) the correlation between RE or RS in these regions and individual differences in recognition memory performance of unknown and known statements, and (3) the difference in memory representations for unknown and known statements as detected by MVPA.

3.2.1 Effects of knowledge on regions showing repetition effects.

We first identified regions showing repetition effects for both known and unknown statements by calculating the conjunction (i.e., hit > CR for known statements ∩ hit > CR for

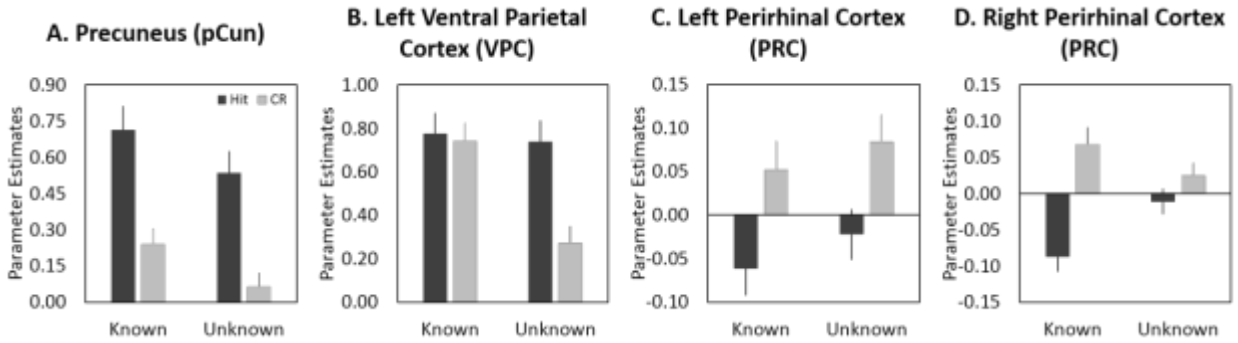


Figure 3. Precuneus exhibited repetition enhancement (RE) for both known and unknown statements (A). Left ventral parietal cortex exhibited RE for unknown statements only (B). Left anterior hippocampus/perirhinal cortex exhibited repetition suppression (RS) for both known and unknown statements (C). Right temporal pole/perirhinal cortex exhibited RS for known statements only (D).

unknown statements). RE effects shared by unknown and known statements (green regions in Figure 2A; top panel of Table 1) were found mainly in the precuneus (see Figure 3A), PCC, and anterior PFC regions. RS effects (i.e., CR > hit) shared by unknown and known statements (green regions in Figure 2B; top panel of Table 2) were found mainly in the left PRC (see Figure 3C), temporal pole, and lateral temporal cortex.

We then turned to repetition effects specific to retrieving unknown versus known statements, which we identified by exclusively masking one of these conditions (at $p < .01$). Regarding RE effects, which reflect retrieval of new episodic representations, we did not find any regions that were exclusive to known statements, but several regions were exclusive to unknown statements (blue regions in Figure 2A; bottom panel of Table 1), including ventral parietal cortex (see Figure 3B), PCC, and anterior PFC. Regarding RS effects, which reflect processing of pre-existing semantic representations, we did not find any regions that were exclusive to unknown statements, but several regions were exclusive to known statements (cyan regions in Figure 2B; bottom panel of Table 2), including a large swath in the left IFG, left

Table 1

Regions Exhibiting Repetition Enhancement (RE) Effects

Region	Hemi	MNI x	MNI y	MNI z	Z	k
Known ∩ Unknown: Hit > Correct Rejection						
Precuneus	M	-4	-74	45	7.00	674
Posterior Cingulate	M	-4	-37	22	5.78	175
Anterior PFC	L	-38	46	15	3.64	36
Supramarginal Gyrus	R	40	-37	37	3.25	47
Unknown: Hit > Correct Rejection						
Ventral Parietal Cortex	L	-38	-52	37	6.66	375
	R	40	-48	45	4.88	234
Posterior Cingulate	M	-1	-30	34	4.69	149
Precuneus	M	10	-56	30	4.23	
Anterior PFC	L	-23	60	3	4.46	164
Anterior Cingulate	M	7	23	41	3.98	97

Table 2

Regions Exhibiting Repetition Suppression (RS) Effects

Region	Hemi	MNI x	MNI y	MNI z	Z	k
Known ∩ Unknown: Correct Rejection > Hit						
Anterior Hippocampus	L	-27	-11	-23	3.17	27
Perirhinal Cortex	L	-27	0	-31	2.36	
Lateral Temporal Cortex	R	63	-3	-20	3.06	27
Temporal Pole	L	-46	12	-31	3.00	31
Known: Correct Rejection > Hit						
Inferior Frontal Gyrus	R	40	38	-12	7.01	340
Middle Frontal Gyrus	R	52	23	34	4.52	
Inferior Frontal Gyrus	L	-46	30	-1	6.29	1502
Middle Frontal Gyrus	L	-38	4	49	4.80	
Superior Frontal Gyrus	L	-4	42	41	6.07	
Medial PFC	M	-12	57	37	5.37	
Superior Frontal Gyrus	R	7	34	49	4.53	
Temporal Pole	L	-31	8	-42	5.19	
Lateral Temporal Cortex	L	-50	-44	-1	4.77	
Perirhinal Cortex	L	-34	-14	-31	3.88	
	R	29	0	-35	6.25	71
Temporal Fusiform	R	40	-7	-27	4.21	
Temporal Pole	R	33	16	-39	3.72	

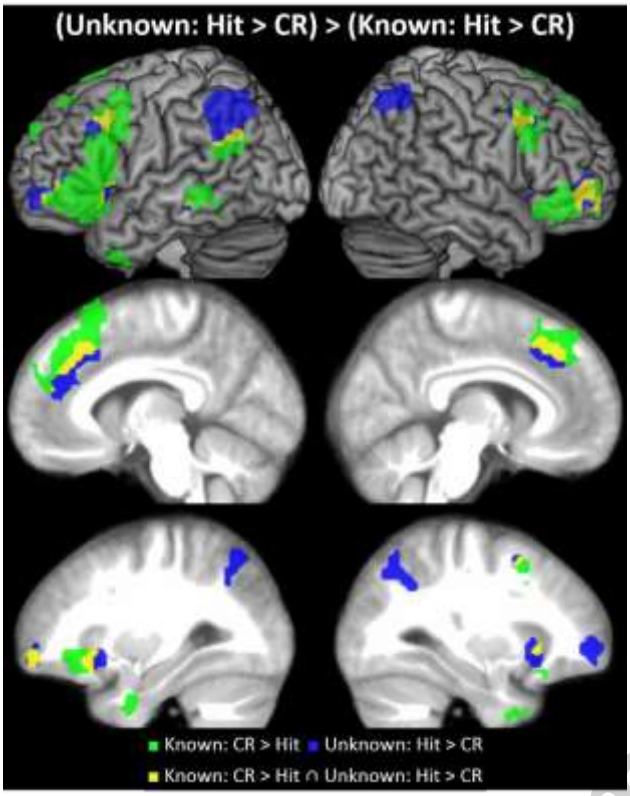


Figure 4. Regions exhibiting a significant interaction [(unknown hit > correct rejection) > (known hit > correct rejection)]. To illustrate the direction of the effects contributing to this significant interaction ($p < .05$ corrected), we color coded the regions based on whether the interaction was due to known cr > hit (green), unknown hit > cr (blue), or both (yellow) at $p < .03$ uncorrected. Smoothed for visualization purposes.

lateral temporal regions, bilateral temporal pole, and PRC (see Figure 3D).

Results from an interaction contrast [(unknown hit > correct rejection) > (known hit > correct rejection)] complemented the exclusive masking analysis (Figure 4). That is, suprathreshold clusters in the interaction contrast were the same regions exhibiting RE for unknown, but not known, statements and RS for known, but not unknown, statements.

In sum, activation patterns showed a clear dissociation between unknown and known trials. Unknown statements exhibited greater RE effects in regions commonly associated with recollection, such as the ventral parietal cortex, posterior midline, and anterior PFC. In contrast, known statements exhibited greater RS effects in regions associated with familiarity, such as PRC, lateral temporal cortex, and IFG.

3.2.2 Brain-behavior correlations.

Next we examined whether the magnitude of the neural repetition effects related to behavior, defined as the difference in oldness ratings for old and new statements (i.e., recognition memory performance). This analysis revealed a striking dissociation. Treating regions that exhibited RE effects for both known and unknown statements as a single ROI, greater RE for unknown statements was correlated across participants with memory performance for unknown statements ($r = 0.54$, $p < .01$), but not between RE and memory for known statements ($r = 0.04$, $p > .05$). This difference was statistically significant ($Z = 2.45$, $p < .01$; Figure 5A). Consistent with this, regions that exhibited RE for unknown statements only were also correlated with memory performance for unknown statements ($r = 0.71$, $p < .001$; Figure 5B).

On the other hand, the opposite pattern was found in regions that exhibited RS effects for both known and unknown statements. RS and memory was correlated for known statements (r =

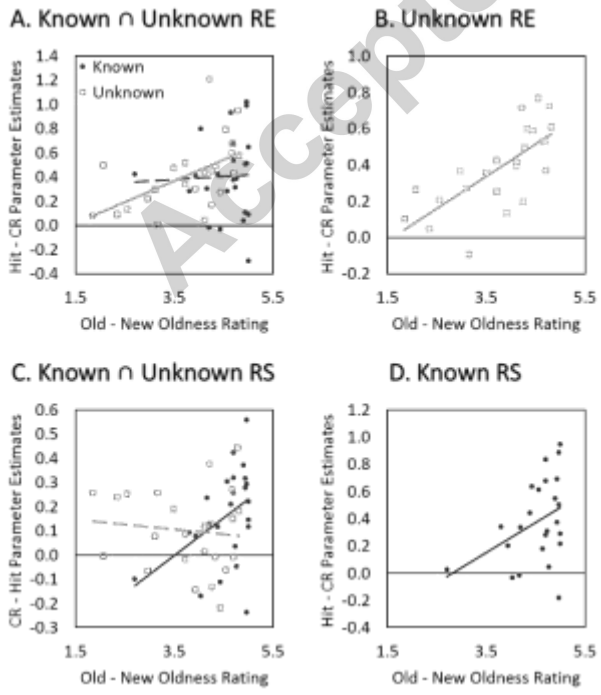


Figure 5. Regions exhibiting repetition enhancement (RE) for both known and unknown statements correlated with memory performance for unknown, but not known, statements (A). Regions exhibiting RE for unknown statements only correlated with memory performance for unknown statements (B). Regions exhibiting repetition suppression (RS) for both known and unknown statements correlated with memory performance for known, but not unknown, statements (C). Regions exhibiting RS for known statements only correlated with memory performance for known statements (D).

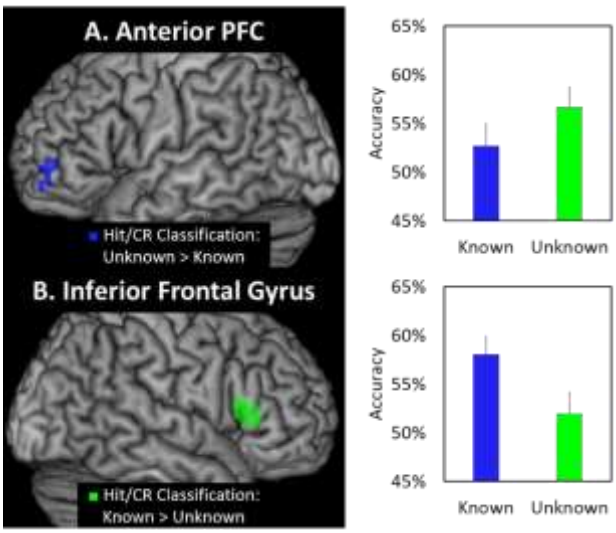


Figure 6. Left anterior PFC exhibited greater classification accuracy for unknown than known statements (A). Right inferior frontal gyrus exhibited greater classification accuracy for known than unknown statements (B). Smoothed for visualization purposes.

-0.43, $p < .05$), but not unknown statements ($r = 0.11$, $p > .05$), and this difference was also statistically significant ($Z = -1.99$, $p < .05$; Figure 5C). Moreover, regions that exhibited RS for known statements were marginally correlated with memory performance for known statements ($r = -0.39$, $p = .06$; Figure 5D).[2]

The dissociation in memory correlations for unknown and known statements cannot be explained by activation differences between these conditions: The dissociation occurred even in regions that showed RE or RS effects for both known and unknown statements. Together, these results suggest that the magnitude of neural repetition differentially correlated with recognition memory performance during episodic retrieval. Specifically, performance correlated with RE for unknown statements, whereas performance correlated with RS for known statements.

[2] The correlations were reduced when removing an outlier with low memory performance in regions exhibiting RS for both known and unknown trials ($r = -.32$, $p = .13$) and in regions exhibiting RS for known trials only ($r = -.33$, $p = .11$).

3.2.3 Pattern classification of episodic memory.

Lastly, using MVPA, we examined whether activity patterns beyond neural repetition effects differed between known and unknown statements. Specifically, using a searchlight analysis, we directly contrasted whether any regions better classified hits and correct rejections for known compared to unknown statements, and vice versa. This analysis revealed that a cluster in anterior PFC (xyz = -27 46 -1; cs = 30; Z = 3.70; Figure 6A) successfully classified hits and correct rejections to a greater extent for unknown than known statements, whereas a cluster in IFG (xyz = 59 27 7; cs = 31; Z = 3.70; Figure 6B) exhibited greater classification accuracy for known than unknown statements. Additionally, the anterior PFC cluster showing greater classification accuracy for unknown statements overlapped with univariate regions showing RE for unknown statements, whereas the IFG cluster showing greater classification accuracy for known statements overlapped with univariate regions showing RS for known statements. These results suggest that activity in regions associated with familiarity and recollection carry information for known and unknown statements, respectively. Crucially, the univariate and MVPA effects were uncorrelated across participants within the overlapping voxels for both the anterior PFC (r = .28, p = .19) and IFG (r = .33, p = .12) clusters, suggesting that the MVPA results are not merely a byproduct of the univariate activity differences (Jimura & Poldrack, 2012).

4. Discussion

The present research investigated the neural correlates of how knowledge impacts episodic retrieval (measured during a recognition memory test), with the goal of identifying the contributions of recollection and familiarity, two processes through which episodic retrieval can occur. Our results revealed striking effects of knowledge, as measured by dissociations between

unknown and known statements across three methods (univariate activity, correlations with recognition memory performance, and memory representations). Our results suggest that known statements were retrieved through familiarity, whereas unknown statements were retrieved through recollection.

While reverse inference has its limitations (Hutzler, 2014; Poldrack, 2006), our interpretations build on decades of research that associates brain networks with particular memory processes. That is, prior studies strongly implicate the ventral parietal cortex, posterior midline, and anterior PFC in context-rich recollection (Kim, 2013; Rugg & Vilberg, 2013; Skinner & Fernandes, 2007; Spaniol et al., 2009). Critically, in our study, episodic retrieval of unknown, but not known, statements primarily relied on this recollection network and consistent with its role in memory retrieval, the magnitude of RE correlated with memory performance. The lack of pre-existing semantic representations for unknown statements may necessitate recollection-based retrieval (Yonelinas, 2002). Indeed, the MVPA results suggest that known and unknown items were represented differently: A left anterior PFC cluster that overlapped with univariate regions showing RE for unknown trials exhibited greater classification accuracy for unknown than known statements. Furthermore, the novelty of the unknown statements may also attract attention at encoding (Knight, 1996; Lisman & Grace, 2005), which could later be recollected (e.g., *I remember reading this sentence because I thought the word "caman" sounded funny*). Consistent with this possibility, unknown statements (M = 3.80, SEM = 0.19) received numerically higher interest ratings than known statements (M = 3.47, SEM = 0.14) during encoding ($t(23) = 1.56$, $p = .13$, $d = .32$). The lack of hippocampal activity, which is typically associated with novelty processing (Knight, 1996; Lisman & Grace, 2005), does not preclude the possibility that novelty mediates the link between a lack of knowledge and

recollection, as new statements (correct rejections) are also novel (i.e., hit > correct rejection contrasts rarely show RE in the hippocampus; Kim, 2013).

In contrast, episodic retrieval of known statements relied on regions associated with familiarity, including perirhinal cortex (Daselaar, Fleck, & Cabeza, 2006; Kafkas & Montaldi, 2014; Montaldi, Spencer, Roberts, & Mayes, 2006; Wang, Ranganath, & Yonelinas, 2014), lateral and anterior temporal regions (Daselaar et al., 2006; Kafkas & Montaldi, 2014; Montaldi et al., 2006; Wang et al., 2014; Yonelinas, Otten, Shaw, & Rugg, 2005), and lateral—including IFG—and medial PFC (Daselaar et al., 2006; Kafkas & Montaldi, 2014; Montaldi et al., 2006; Wang et al., 2014; Yonelinas et al., 2005). Interestingly, common RS effects in left PRC also extended into left anterior hippocampus. While this result might seems to contradict the role of the hippocampus in recollection, it is consistent with recent frameworks that emphasize functional variability along the hippocampus' long axis (Poppenk et al., 2013; Ranganath & Ritchey, 2012). It is also consistent with evidence that both PRC and anterior hippocampus are sensitive to stimulus familiarity (Daselaar et al., 2006; Rugg, Henson, & Robb, 2003; Vilberg & Rugg, 2009; Wang & Giovanello, 2016).

These regions typically exhibit RS during familiarity-based retrieval, consistent with our finding of a correlation between the magnitude of RS and memory performance. Greater RS effects for known than unknown statements may reflect the retrieval of a pre-existing semantic representation that is more accessible following repetition (Henson, 2003; Henson et al., 2002; Reggev, Bein, & Maril, 2016). In other words, the pre-existing representations may make the known items easy to process, and more quickly retrieved, engendering a feeling of fluency which often is interpreted as familiarity (Bruett & Leynes, 2015; Ozubko & Yonelinas, 2014; Rajaram & Geraci, 2000; Woollams, Taylor, Karayanidis, & Henson, 2008). Paralleling findings

implicating different memory processes, a right IFG cluster that overlapped with univariate regions showing RS for known trials exhibited greater classification accuracy for known than unknown statements. In addition, known statements may be relatively more unitized (i.e., holistic concepts), which are more easily retrieved based on familiarity than unknown statements, which consist of arbitrary associations (e.g., shinty and caman; Greve, Van Rossum, & Donaldson, 2007; Parks & Yonelinas, 2015). Unknown statements, on the other hand, lack pre-existing representations and thus familiarity-based retrieval may be insufficient, thus requiring more effortful, recollection-based retrieval of the novel statements. The reaction time data are consistent with this interpretation; retrieval times were faster for known than unknown statements, suggesting a relatively automatic process (Yonelinas, 2002). The pre-existing representations of known information likely allowed participants to respond based on the relative familiarity of known statements. However, such a process would fail for unknown statements, which would require recruitment of a more controlled, recollection-based memory search for the novel semantic representations.[3]

At first glance, our findings seem to contradict prior behavioral work, which suggests that knowledge benefits memory by encouraging recollection. One major differences is that those studies operationalized knowledge as expertise, which may be encoded and retrieved differently

[3] Single-process models of episodic retrieval (e.g., Squire, Wixted, & Clark, 2007; Wixted, 2007) posit that regions associated with recollection and familiarity reflect "strong" (i.e., high confidence) and "weak" (i.e., low confidence) memories, respectively. This possibility, however, is inconsistent with our results, as our analyses were restricted to high-confidence responses, and recognition performance was better for known than unknown statements.

than general knowledge (Kalakoski & Saariluoma, 2001; Kawamura et al., 2007). In addition, those studies relied almost exclusively rely on remember-know judgments. The issue is that people do not have perfect insight into the processes underlying their behavior (e.g., Benjamin, Bjork, & Schwartz, 1998); for example, when later asked to retrieve the presentation colors of studied words, people are above chance for items labeled as "known" (Wais, Mickes, & Wixted, 2008). Furthermore, relatively small changes in remember-know instructions change how participants apply the labels (McCabe & Geraci, 2009; Migo, Mayes, & Montaldi, 2012). While remember-know judgments are useful in some cases, our results suggest that they do not fully capture how knowledge supports episodic memory.

The present study highlights how neuroimaging can lend insight into understanding cognitive processes. While behavioral evidence suggests that knowledge supports retrieval of specific events through recollection, our fMRI data intriguingly implicate a different mechanism. Specifically, three converging analyses indicate that recollection-related regions supported the episodic retrieval of unknown stimuli, whereas familiarity-related regions supported the episodic retrieval of known stimuli. Our interpretation is consistent with our reaction time data, as known statements were retrieved quicker than unknown statements, consistent with an automatic and familiarity-based process for the former, and a controlled and recollection-based process for the latter. Our results emphasize the benefit of fMRI in understanding how we remember. Together, our results suggest that familiarity is sufficient to support the retrieval of known stimuli, whereas the retrieval of unknown stimuli may require recollection.

References

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55.

Bird, C. M., & Burgess, N. (2008). The hippocampus supports recognition memory for familiar words but not unfamiliar faces. *Current Biology, 18*(24), 1932-1936.

Brandt, K. R., Cooper, L. M., & Dewhurst, S. A. (2005). Expertise and recollective experience: Recognition memory for familiar and unfamiliar academic subjects. *Applied Cognitive Psychology, 19*(9), 1113-1125.

Bruett, H., & Leynes, P. A. (2015). Event-related potentials indicate that fluency can be interpreted as familiarity. *Neuropsychologia, 78*, 41-50.

Daselaar, S. M., Fleck, M. S., & Cabeza, R. E. (2006). Triple Dissociation in the Medial Temporal Lobes: Recollection, Familiarity, and Novelty. *J Neurophysiol, 96*, 1902-1911.

Dew, I. T. Z., & Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: behavioral and neural evidence. *Annals of the New York Academy of Sciences, 1224*(1), 174-190.

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS One, 10*(4), e0121945.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & cognition, 24*(4), 523-533.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review, 102*(2), 211.

Fisher, R. (1990). *Statistical Methods, Experimental Design and Scientific Inference*. Oxford: Oxford University Press.

Furman, O., Mendelsohn, A., & Dudai, Y. (2012). The episodic engram transformed: Time reduces retrieval-related brain activity but correlates it with memory accuracy. *Learning & Memory, 19*(12), 575-587.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & cognition, 16*(4), 309-313.

Greve, A., Van Rossum, M. C., & Donaldson, D. I. (2007). Investigating the functional interaction between semantic and episodic memory: Convergent behavioral and electrophysiological evidence for the role of familiarity. *Neuroimage, 34*(2), 801-814.

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2014). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in neuroinformatics, 8*.

Henson, R. N. A. (2003). Neuroimaging studies of priming. *Prog Neurobiol, 70*(1), 53-81.

Henson, R. N. A., Cansino, S., Herron, J. E., Robb, W. G. K., & Rugg, M. D. (2003). A familiarity signal in human anterior medial temporal cortex? *Hippocampus, 13*, 259-262.

Henson, R. N. A., Shallice, T., Gorno-Tempini, M., & Dolan, R. (2002). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cereb Cortex, 12*(2), 178-186.

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *Neuroimage, 84*, 1061-1069.

Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia, 50*(4), 544-552.

Kafkas, A., & Montaldi, D. (2014). Two separate, but interacting, neural systems for familiarity and novelty detection: A dual-route mechanism. *Hippocampus, 24*(5), 516-527.

Kalakoski, V., & Saariluoma, P. (2001). Taxi drivers' exceptional memory of street names. *Memory & cognition, 29*(4), 634-638.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kawamura, S., Suzuki, S., & Morikawa, K. (2007). The effect of expertise in hiking on recognition memory for mountain scenes. *Memory, 15*(7), 768-775.

Kim, H. (2013). Differential neural activity in the recognition of old versus new events: An Activation Likelihood Estimation Meta-Analysis. *Human brain mapping, 34*(4), 814-836.

Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature, 383*(6597), 256-259.

Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron, 46*(5), 703-713. doi: 10.1016/j.neuron.2005.05.002

Long, D. L., & Prat, C. S. (2002). Memory for Star Trek: The role of prior knowledge in recognition revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(6), 1073.

McCabe, D. P., & Geraci, L. D. (2009). The influence of instructions and terminology on the accuracy of remember–know judgments. *Conscious Cogn, 18*(2), 401-413.

Meade, M. L., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory, 17*(1), 39-48.

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Conscious Cogn, 21*(3), 1435-1455.

Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus, 16*(5), 504-520.

Ozubko, J. D., & Yonelinas, A. P. (2014). The disruptive effects of processing fluency on familiarity-based recognition in amnesia. *Neuropsychologia, 54*, 59-67.

Paller, K. A., Voss, J. L., & Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends Cogn Sci, 11*(6), 243-250.

Parks, C. M., & Yonelinas, A. P. (2015). The importance of unitization for familiarity-based learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 881.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci, 10*(2), 59-63.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends Cogn Sci, 17*(5), 230-240.

Poppenk, J., Köhler, S., & Moscovitch, M. (2010). Revisiting the novelty effect: when familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1321.

Rajaram, S. (1996). Perceptual effects on remembering: recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 365.

Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology Learning Memory and Cognition, 26*(4), 1070-1074.

Rajimehr, R., Young, J. C., & Tootell, R. B. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. *Proceedings of the National Academy of Sciences, 106*(6), 1995-2000.

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience, 13*(10), 713-726.

Rawson, K. A., & Van Overschelde, J. P. (2008). How does knowledge promote memory? The distinctiveness theory of skilled memory. *J Mem Lang, 58*(3), 646-668.

Reder, L. M., Victoria, L. W., Manelis, A., Oates, J. M., Dutcher, J. M., Bates, J. T., . . . Gyulai, F. (2013). Why it's easier to remember seeing a face we already know than one we don't preexisting memory representations facilitate memory formation. *Psychol Sci, 24*(3), 363-372.

Reggev, N., Bein, O., & Maril, A. (2016). Distinct neural suppression and encoding effects for conceptual novelty and familiarity. *J Cogn Neurosci*.

Rissman, J., Greely, H. T., & Wagner, A. D. (2010). Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences, 107*(21), 9849-9854.

Rugg, M. D., Henson, R. N., & Robb, W. G. (2003). Neural correlates of retrieval processing in the prefrontal cortex during recognition and exclusion tasks. *Neuropsychologia, 41*(1), 40-52.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology, 23*(2), 255-260. doi: 10.1016/j.conb.2012.11.005

Särkkä, S., Solin, A., Nummenmaa, A., Vehtari, A., Auranen, T., Vanni, S., & Lin, F.-H. (2012). Dynamic retrospective filtering of physiological noise in BOLD fMRI: DRIFTER. *Neuroimage, 60*(2), 1517-1527.

Simons, J. S., Gilbert, S. J., Owen, A. M., Fletcher, P. C., & Burgess, P. W. (2005). Distinct roles for lateral and medial anterior prefrontal cortex in contextual recollection. *Journal of Neurophysiology, 94*(1), 813-820.

Simons, J. S., Owen, A. M., Fletcher, P. C., & Burgess, P. W. (2005). Anterior prefrontal cortex and the recollection of contextual information. *Neuropsychologia, 43*(12), 1774-1783.

Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia, 45*(10), 2163-2179.

Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia, 47*(8), 1765-1779.

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nature reviews. Neuroscience, 8*(11), 872.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381-404). London: Academic.

Tulving, E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences, 7*(02), 223-238.

Vallee-Tourangeau, F., Anthony, S., & Austin, N. (1998). Strategies for generating multiple instances of common and ad hoc categories. *Memory, 6*(5), 555-592.

Van Overschelde, J. P., & Healy, A. F. (2001). Learning of nondomain facts in high-and low-knowledge domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1160.

Vilberg, K. L., & Rugg, M. D. (2009). An investigation of the effects of relative probability of old and new test items on the neural correlates of successful and unsuccessful source memory. *Neuroimage, 45*(2), 562-571. doi: 10.1016/j.neuroimage.2008.12.020

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., . . . Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281*(5380), 1188-1191.

Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *J Cogn Neurosci, 20*(3), 400-405.

Wang, W. C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2016). On Known Unknowns: Fluency and the Neural Mechanisms of Illusory Truth. *Journal of cognitive neuroscience, 28*(5), 739-756.

Wang, W. C., & Giovanello, K. S. (2016). The role of medial temporal lobe regions in incidental and intentional retrieval of item and relational information in aging. *Hippocampus, 26*(6), 693-699.

Wang, W. C., Ranganath, C., & Yonelinas, A. P. (2014). Activity reductions in perirhinal cortex predict conceptual priming and familiarity-based recognition. *Neuropsychologia, 52*, 19-26. doi: 10.1016/j.neuropsychologia.2013.10.006

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review, 114*(1), 152.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychon Bull Rev, 11*(4), 616-641.

Woollams, A. M., Taylor, J. R., Karayanidis, F., & Henson, R. N. (2008). Event-related potentials associated with masked priming of test cues reveal multiple potential contributions to recognition memory. *J Cogn Neurosci, 20*(6), 1114-1129.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *J Mem Lang, 46*(3), 441-517.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *The Journal of Neuroscience, 25*(11), 3002-3008.

Acknowledgments

We thank Shaina Garrison, Elena Lagon, and Rachel James for their help with developing stimuli, recruiting participants, and collecting data.

Funding

This research is supported by grant R01 AG034580 (RC) and fellowship F32 AG049574 (WCW) from the National Institute on Aging, and a National Science Foundation Graduate Research Fellowship (NMB).

Highlights

- Participants retrieved old and new known and unknown factual statements.

- Retrieval was quicker for known than unknown statements.

- Recollection-related regions were recruited during retrieval of unknown statements.

- Familiarity-related regions were recruited during retrieval of known statements.

- Suggests that known but not unknown statements can be retrieved based on familiarity.