

# UCLA

## UCLA Previously Published Works

### Title

Decoding fMRI Signatures of Real-world Autobiographical Memory Retrieval.

### Permalink

<https://escholarship.org/uc/item/5c35w25d>

### Journal

Journal of cognitive neuroscience, 28(4)

### ISSN

0898-929X

### Authors

Rissman, Jesse  
Chow, Tiffany E  
Reggente, Nicco  
[et al.](#)

### Publication Date

2016-04-01

### DOI

10.1162/jocn\_a\_00920

Peer reviewed

# Decoding fMRI Signatures of Real-world Autobiographical Memory Retrieval

Jesse Rissman<sup>1</sup>, Tiffany E. Chow<sup>1</sup>, Nicco Reggente<sup>1</sup>, and Anthony D. Wagner<sup>2</sup>

## Abstract

■ Extant neuroimaging data implicate frontoparietal and medial-temporal lobe regions in episodic retrieval, and the specific pattern of activity within and across these regions is diagnostic of an individual's subjective mnemonic experience. For example, in laboratory-based paradigms, memories for recently encoded faces can be accurately decoded from single-trial fMRI patterns [Uncapher, M. R., Boyd-Meredith, J. T., Chow, T. E., Rissman, J., & Wagner, A. D. Goal-directed modulation of neural memory patterns: Implications for fMRI-based memory detection. *Journal of Neuroscience*, 35, 8531–8545, 2015; Rissman, J., Greely, H. T., & Wagner, A. D. Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences, U.S.A.*, 107, 9849–9854, 2010]. Here, we investigated the neural patterns underlying memory for real-world autobiographical events, probed at 1- to 3-week retention intervals as well as whether distinct patterns are associated with different subjective memory states. For 3 weeks, participants ( $n = 16$ ) wore digital cameras that captured photographs of their

daily activities. One week later, they were scanned while making memory judgments about sequences of photos depicting events from their own lives or events captured by the cameras of others. Whole-brain multivoxel pattern analysis achieved near-perfect accuracy at distinguishing correctly recognized events from correctly rejected novel events, and decoding performance did not significantly vary with retention interval. Multivoxel pattern classifiers also differentiated recollection from familiarity and reliably decoded the subjective strength of recollection, of familiarity, or of novelty. Classification-based brain maps revealed dissociable neural signatures of these mnemonic states, with activity patterns in hippocampus, medial PFC, and ventral parietal cortex being particularly diagnostic of recollection. Finally, a classifier trained on previously acquired laboratory-based memory data achieved reliable decoding of autobiographical memory states. We discuss the implications for neuroscientific accounts of episodic retrieval and comment on the potential forensic use of fMRI for probing experiential knowledge. ■

## INTRODUCTION

Throughout day-to-day life, we constantly evaluate how elements of the present environment relate to our past experiences, as information retrieved from memory can guide selection of appropriate behaviors. An accumulating body of neuroimaging work has yielded insights into the functional contributions of frontoparietal and medial-temporal lobe structures to episodic retrieval (Kim, 2013; Hutchinson, Uncapher, & Wagner, 2009; Spaniol et al., 2009), and the particular profile of activation within and across these regions is closely linked to the subjective feeling of familiarity for a given retrieval cue and/or the recollection of associated contextual details (Rugg & Vilberg, 2013; Shimamura, 2011; Eichenbaum, Yonelinas, & Ranganath, 2007; Wagner, Shannon, Kahn, & Buckner, 2005; Squire, Stark, & Clark, 2004). Over the past decade, aided by the development and application of sophisticated multivoxel pattern analysis (MVPA) techniques (Tong & Pratte, 2012; Norman, Polyn, Detre, & Haxby, 2006), researchers have demonstrated that the distributed fMRI patterns associated with the act of memory retrieval

are sufficiently robust so as to be detectable on individual trials (Rissman & Wagner, 2012). For instance, a number of MVPA studies have showcased an ability to “read out” basic characteristics of retrieved mnemonic content, such as which of several contexts an item had been studied in or which of several candidate memories is currently being brought back to mind (e.g., Thakral, Wang, & Rugg, 2015; Kuhl & Chun, 2014; Leiker & Johnson, 2014; Chadwick, Hassabis, Weiskopf, & Maguire, 2010; Johnson, McDuff, Rugg, & Norman, 2009; Polyn, Natu, Cohen, & Norman, 2005). By indexing the reemergence of stimulus-specific activity patterns during associative retrieval, researchers have made progress in understanding the relationship between hippocampal signaling and neocortical reactivation (Leiker & Johnson, 2015; St-Laurent, Abdi, & Buchsbaum, 2015; Wing, Ritchey, & Cabeza, 2015; Bosch, Jehee, Fernandez, & Doeller, 2014; Gordon, Rissman, Kiani, & Wagner, 2014; Ritchey, Wing, LaBar, & Cabeza, 2013; Staresina, Henson, Kriegeskorte, & Alink, 2012) as well as characterizing the consequences of mnemonic competition and its resolution (Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015; Kuhl, Rissman, Chun, & Wagner, 2011). Other investigations have focused on decoding the cognitive processes engaged during retrieval, such as whether

<sup>1</sup>University of California, Los Angeles, <sup>2</sup>Stanford University

one's efforts are preferentially oriented toward recollecting source details or gauging the familiarity of probe items (Quamme, Weiss, & Norman, 2010), as well decoding the subjective outcome of retrieval, including one's confidence in a given memory judgment (Rissman, Greely, & Wagner, 2010) and the impact of one's retrieval goals on the neural signatures of recognition and novelty (Uncapher, Boyd-Meredith, Chow, Rissman, & Wagner, 2015).

Collectively, these studies have leveraged MVPA methods to generate valuable insights into the neural mechanisms of episodic retrieval. However, this body of work has almost entirely focused on memories for information studied in a laboratory setting—typically simple word or picture stimuli, but occasionally more complex stimuli such as brief video clips (e.g., St-Laurent et al., 2015; Buchsbaum, Lemire-Rodger, Fang, & Abdi, 2012; Chadwick, Hassabis, & Maguire, 2011; Chadwick et al., 2010). The emphasis on laboratory-encoded stimuli is sensible, given the high degree of experimental control that researchers can exert over the learning experience. That said, the constrained stimulus sets and encoding conditions employed by such studies may fail to adequately approximate the richness of the episodic memories formed in more naturalistic contexts as individuals freely navigate the world and engage in personally meaningful activities (Maguire, 2012; Cohen & Conway, 2008). Indeed, efforts to compare brain activation during the retrieval of laboratory-encoded and real-world event memories have noted some pronounced differences (McDermott, Szpunar, & Christ, 2009; Cabeza et al., 2004), including increased engagement of the hippocampus and ventromedial PFC (vmPFC) during autobiographical retrieval. This may stem from the greater degree to which spatiotemporal and self-referential contextual details are mentally reconstructed during the recall of real-world episodes (Rubin & Umanath, 2015; Conway, 2009; Cabeza & St Jacques, 2007; Hassabis & Maguire, 2007). To achieve a deeper understanding of human memory as it is actually used in day-to-day life, it may be necessary to relinquish some degree of experimental control in favor of task paradigms that offer enhanced ecological validity (e.g., St Jacques, Olm, & Schacter, 2013; Milton, Muhlert, Butler, Benattayallah, & Zeman, 2011; St Jacques, Conway, Lowder, & Cabeza, 2011).

To our knowledge, only three fMRI studies to date have applied MVPA techniques to characterize neural representations associated with the remembrance of naturalistically encoded autobiographical events. Two of these studies (which were based on a common fMRI data set) utilized verbal prompts to cue the reliving of a small set of recently experienced (~2 weeks old) and remote (~10 years old) event memories (Bonnici, Chadwick, & Maguire, 2013; Bonnici et al., 2012). Although the authors did not attempt to directly decode the age (i.e., temporal remoteness) of each memory, they did find that regions such as the vmPFC and posterior hippocampus showed heightened representational distinctiveness in

the neural patterns associated with more remote memories, which presumably rely more heavily on reconstructive processes during retrieval. A recent study by Nielson, Smith, Sreekumar, Dennis, and Sederberg (2015) examined the neural representation of spatial and temporal information in the hippocampus during the retrieval of real-world memories. In the scanner, participants viewed photographs captured by a GPS-enabled camera that they had worn over a 1-month period, and they attempted to vividly recall the depicted events. Activity patterns within the left anterior hippocampus were found to be sensitive to both spatial position (i.e., showing greater similarity across events encoded in nearby locations, relative to distant locations) and temporal distance (i.e., showing greater similarity across events encoded days apart, relative to weeks apart) revealing superimposed coding of these two critical mnemonic dimensions. Studies like these highlight the potential of MVPA methods to quantify the representational distinctiveness of individual memories in different regions of the brain. However, a number of open questions remain, including evaluation of the mnemonic contributions of other brain regions and the relationship between brain activity patterns and participants' subjective retrieval experiences.

In the present fMRI experiment, we sought to characterize the distributed brain activity patterns associated with the retrieval of real-world event memories. To gather a rich set of naturalistic stimuli, we deployed wearable digital camera devices to record the daily life events of research participants over the course of 3 weeks. Participants were then scanned 1 week later as we probed their memories by presenting them with brief sequences of photos depicting experiences from their own lives as well as some from the lives of other participants. Rather than focusing our fMRI analysis efforts on decoding the identity or representational content of individual memories, we instead chose to focus on decoding brain processes tied to participants' subjectively reported retrieval experiences. In doing so, we aimed to build on prior work demonstrating that the mnemonic outcome associated with a given retrieval attempt (i.e., whether a test probe is reported to be vividly remembered, perceived as familiar, or perceived as novel) can be reliably classified based on whole-brain fMRI activity patterns (Rissman et al., 2010). It is possible that the highly accurate decoding results of that study (with accuracies ranging from 70% to 90% depending on the mnemonic distinction in question) were inflated by the fact that all of the memories were of the same type (faces) and encoded very shortly (~1 hr) before scanning. If this were the case, then memories for a more heterogeneous set of real-world experiences probed at a wide range of retention intervals might have more variable neural signatures at the time of retrieval and be less amenable to classification. Indeed, prior fMRI studies have documented changes in the strength and anatomical distribution of activity levels over retention intervals of 1 month (Smith et al., 2010; Takashima et al.,

2006; Bosshardt et al., 2005). On the other hand, it is possible that memories for real-world experiences, by virtue of their heightened strength, enriched contextual associations, and personal relevance, might yield equal, if not better, classification performance than was obtained in the face memory study.

Although comparing classification accuracies across studies can provide some information about the relative ability to decode laboratory-based and real-world memories, a stronger test would be to evaluate whether a classifier trained to differentiate memory retrieval states based on fMRI data from our prior face memory study (Rissman et al., 2010) would be able to generalize its predictive power to brain patterns measured from an independent group of participants in the present study of real-world memory retrieval. There is reason to believe that classifier generalization across these types of memory tasks might be challenging to achieve. In addition to the aforementioned issue of heterogeneity, a recent meta-analysis comparing results of fMRI studies involving the retrieval of autobiographical memories with those involving the retrieval of laboratory-encoded memories reported surprisingly little neuroanatomical overlap (McDermott et al., 2009). Adding to the intrigue, recent reports of individuals with “highly superior autobiographical memory” (Patihis et al., 2013; LePort et al., 2012), as well as those with “highly deficient autobiographical memory” (Palombo, Alain, Soderlund, Khuu, & Levine, 2015), have provided striking demonstrations that autobiographical retrieval abilities are largely uncorrelated with one’s ability to perform standard laboratory-based memory tasks. Dissociations like these have led some to propose that retrieving autobiographical event knowledge is fundamentally different from other forms of episodic retrieval (Roediger & McDermott, 2013). Accordingly, should our classifier model show reasonable generalization performance across these two seemingly different memory tasks, it would highlight that important commonalities nevertheless exist.

Beyond our assessment of classification accuracy levels, which provide a useful assay of how well distinct mnemonic retrieval experiences can be predicted based on the underlying activity patterns, we also aim to evaluate which brain regions provide maximally diagnostic signals to each classifier model (i.e., importance maps). Of particular interest are the brain patterns tied to more subtle gradations in memory retrieval outcomes, such as the degree of recollection or the degree of familiarity reported by participants. To the extent that our binary classifier models are prone to settle on a unidimensional representation of memory strength, a classifier trained to differentiate strongly recollected events versus moderately recollected events might anchor on the very same neural signatures as a classifier trained to differentiate strongly familiar versus moderately familiar events. However, should the importance maps for these two classifications diverge, this would indicate that different brain regions

are driving the classifier’s predictions in each case and support a qualitative neurocognitive distinction between these memory states.

A final motivation for our study was to contribute to an emerging dialogue between neuroscientists, legal scholars, and the public regarding the potential use of fMRI as a memory detection technology (Schacter & Loftus, 2013; Shen & Jones, 2011; Bles & Haynes, 2008; Meegan, 2008). Previous fMRI studies have documented the high accuracy with which single-trial brain activity patterns can reveal whether a probe stimulus evokes a sense of recognition or novelty (Uncapher et al., 2015; Rissman et al., 2010). At the same time, these studies have noted serious limitations, including difficulty differentiating true versus false memories and the susceptibility to countermeasures (i.e., strategic efforts to conceal one’s memories). Despite these important boundary conditions that diminish the forensic value of fMRI as an objective tool for memory detection, fMRI measures could still hold potential as a means to quantify the strength of a memory, to supplement verbal reports of recognition, or perhaps even to assess the memories of individuals who are unable to communicate. By expanding the scope of earlier fMRI memory detection efforts, which used laboratory-encoded face stimuli (Uncapher et al., 2015; Rissman et al., 2010), our study has the potential to yield valuable data regarding the brain-based classification of real-world event memories.

## **METHODS**

### **Participants**

Sixteen participants (eight women; aged 18–22 years) took part in this experiment. Written informed consent was obtained in accordance with procedures approved by the institutional review board at Stanford University. All participants were right-handed native speakers of English, had normal or corrected-to-normal vision, and were prescreened for the presence of medical, neurological, or psychiatric illnesses and use of psychoactive medications. To provide some control over the nature of the daily life events experienced by our participants, enrollment was restricted to Stanford University undergraduate students who were residing on campus. Participants were remunerated with \$300 for their efforts over the course of their month-long enrollment period. One additional individual was enrolled in the study, but because of a camera malfunction, his photographs were too blurry for use in the experiment; his participation was discontinued before MRI scanning.

### **Procedure**

#### *Use of Wearable Cameras*

Each participant was provided a Vicon Revue digital camera (Vicon Motion Systems Ltd., Oxford, UK) for a 3-week

period. These small 0.3-megapixel necklace-mounted cameras contain sensors that detect changes in environmental factors, such as ambient light intensity, color, temperature, and movement. Wide-angle color photographs ( $640 \times 480$  pixels) are automatically taken whenever the sensors are triggered, with approximately 2–10 photos captured per minute. Importantly, the cameras lack LCD display screens, so participants had no means to review the photos being captured by their camera. Participants were encouraged to wear the camera in the “on” mode as much as possible each day, with the option of turning it off whenever they, or the people around them, desired privacy. Each week, participants returned to the laboratory to allow the experimenter to download the photos (approximately 5000–15,000 per week depending on participants’ wearing habits). Cameras were returned after 21 days of wearing, and an fMRI scanning session took place 6–9 days later (mean lag = 7.4 days). Before the fMRI session, participants had no knowledge of the specific goals of the experiment, although they were informed from the outset that the fMRI study would utilize images from their cameras as stimuli.

### *Selection of Photographic Stimuli*

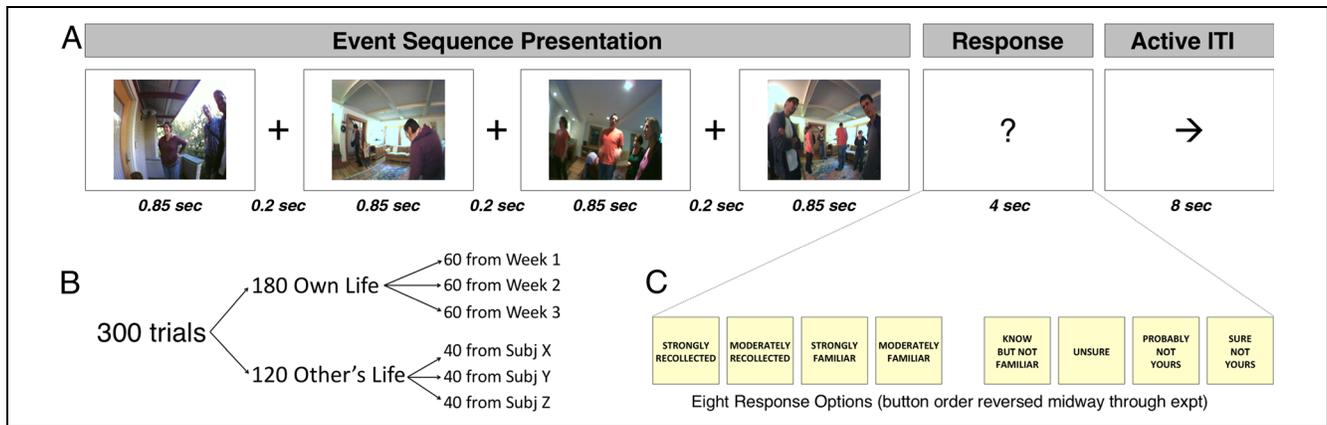
From the thousands of photos captured by each participant’s camera, we selected a set of 180 “event sequences” (60 from each week of camera wearing) to use as stimuli in the fMRI experiment. Each event sequence was composed of four photos captured within a 5-min interval that depicted the temporal unfolding of a potentially memorable episode from the participant’s day. The image content of the selected event sequences varied widely, with some events depicting the wearer in a stationary position (e.g., sitting in class, attending a concert or sporting event, eating at a restaurant), some events depicting the wearer on the move (e.g., entering or exiting a building, moving through a room, walking across campus, hiking on a trail, shopping at a store), and some events depicting a combination of these attributes. Many of the event sequences contained visible faces, whether of friends, acquaintances, or strangers, whereas other events contained primarily environmental features. Given the experimental requirement to create 60 event sequences per week for each participant, we occasionally had to break longer duration events (e.g., a picnic or party) into two or more qualitatively distinct subevents. Although it also was impossible to avoid the inclusion of multiple similar events (e.g., dining in the same cafeteria, studying in the same library, hanging out in the same place with the same group of friends), a concerted effort was made to select photos that had enough unique details to allow the episodes to be differentiated from one another. Given the variability of the life events captured by the cameras, we did not attempt to equate the selected event sequences for salience or other content-related attributes. Rather, we embrace this variance as an in-

herent feature of the stimulus set, serving to elicit a wide range of memory retrieval experiences from the participants and, as such, to bolster the ecological validity of the experiment. Although most of the selected photos were unedited, some were cropped to remove any depiction of the wearer’s own body, because such details might have provided participants with an easy cue to identify the images as being from their own camera. In addition, some photos were mildly edited to correct issues with coloration and exposure. By design, none of the participants were friends with each other, and we never encountered an instance where two concurrently enrolled participants came into direct contact with one another while wearing their cameras.

### *fMRI Task Design*

The fMRI experiment included 300 trials distributed across 10 scanning runs (30 trials/run). On each trial, participants were presented with a four-photo event sequence and asked to make a response indicating their memory for that event. Within each run, 18 of the trials featured event sequences that had been captured by the participant’s own camera (“Own Life” condition), with six of these event sequences drawn from each of the 3 weeks of camera wearing. The remaining 12 trials of each run featured event sequences that had been captured by other participants’ cameras (“Other’s Life” condition); for any given participant, these were drawn evenly from event sequences that had been created for three other randomly selected participants.<sup>1</sup> The presentation order of Own Life and Other’s Life trials was randomized. Across the 10 runs, participants encountered 180 Own Life trials and 120 Other’s Life trials.

The structure of each trial (Figure 1) was as follows: The four constituent photos of an event sequence were sequentially presented for 850 msec each, with a 200-msec central fixation cross appearing between successive photos. After the offset of the fourth photo, a question mark appeared on the screen for 4 sec, turning from white to red during the final second to inform participants of the impending deadline for them to make a response. The response period required participants to depress one of eight buttons indicating their level of memory for the event sequence. To mitigate continued reminiscence or mind-wandering during the intertrial interval, participants performed an active baseline task (Stark & Squire, 2001). Specifically, after a 1-sec fixation cross, a series of five arrows appeared on the screen for 1 sec each, with 400 msec elapsing between successive arrows. Participants indicated the left/right direction of each arrow using their left and right index fingers, respectively. A red central fixation cross (1 sec) then signaled the impending onset of the next trial. The total trial onset asynchrony was held constant at 16 sec. The timing of stimulus presentation and response collection was controlled using the Psychophysics Toolbox (Brainard,



**Figure 1.** Experimental design. (A) On each trial, participants viewed a sequence of four photographs depicting the temporal unfolding of an event. Immediately thereafter, participants made a button-press response indicating their memory for that event. The eight response options were mapped to eight distinct buttons (C), the order of which was reversed midway through the scanning session. During the 8-sec intertrial interval (ITI), participants were tasked with judging the right/left direction of a series of five arrows. Participants performed 300 trials, the breakdown of which is illustrated in B.

1997) in MATLAB (The MathWorks, Natick, MA). Visual stimuli were projected onto a screen against an isoluminant gray background and viewed through a mirror.

Immediately before the scanning session, participants were provided with written instructions regarding the upcoming memory test, with emphasis on the critical distinctions between the eight different memory response options:

- Strongly recollected: You are able to recollect many details of this specific experience.
- Moderately recollected: You are able to recollect a few of the details surrounding this specific experience.
- Strongly familiar: This specific experience seems strongly familiar to you.
- Moderately familiar: This specific experience seems moderately familiar to you.
- Know but not familiar: You know that this was your experience, and yet the specific experience depicted in the photos does not seem particularly familiar to you.
- Unsure: You are unsure whether this was your experience.
- Probably not yours: You probably did not have this experience.
- Sure not yours: You are sure that you did not have this experience.

Additional instructions helped clarify the conceptual differences between recollection, familiarity, and knowing, using specific examples to illustrate these qualitatively different memory states. The experimenter then verbally confirmed the participant's understanding of the instructions and administered a practice version of the experiment. The practice version, conducted on a laptop computer, consisted of 12 trials (six Own Life event sequences drawn from surplus materials and six Other's Life event sequences), and participants made their re-

sponses on a keyboard with eight labeled keys. In the scanner, responses were made on two MR-compatible button boxes, one held in each hand; thumbs were not used. Before beginning the memory test, participants performed a button mapping training task to ensure mastery of the mappings. Halfway through the experiment (after completion of the first five scanning runs), the order of the button mappings was reversed, and participants received training on the new mappings before continuing. The goal of this manipulation was to enable training of MVPA classifiers on the neural signatures associated with unique memory states, independent of the specific motor response mappings. The order in which the two button mappings were administered was counterbalanced across participants.

#### *fMRI Data Acquisition*

Whole-brain imaging was conducted on a 3.0-T Signa MRI system (GE Healthcare Systems, Milwaukee, WI). Functional images were collected using a T2\*-weighted 2-D gradient-echo spiral-in/out pulse sequence (repetition time [TR] = 2.0 sec, echo time = 30 msec, flip angle = 75°, field of view = 21 cm, in-plane resolution = 3.44 mm<sup>2</sup>). Each functional volume consisted of 30 contiguous 3.8-mm thick slices acquired parallel to the AC-PC plane. Functional data were collected across 10 runs of 248 volumes each. The six initial volumes from each run were discarded to allow for T1 equilibration. To aid with spatial registration, anatomical images coplanar with the functional data were collected at the start of the experiment using a T2-weighted flow-compensated spin-echo sequence, and T1-weighted whole-brain spoiled gradient recalled (SPGR) 3-D anatomical image (voxel size = 0.86 × 0.86 × 1.0 mm) was acquired after the fifth functional run (during the second button mapping training session).

### *fMRI Data Preprocessing*

Physiological noise correction was applied during reconstruction of the functional images using respiratory data measured from a pneumatic belt strapped around the upper abdomen during scanning. This consisted of removal of time-locked respiratory artifacts using RETROICOR (Glover, Li, & Ress, 2000) and removal of low-frequency respiratory effects using RVHRCOR (Chang & Glover, 2009). The reconstructed images were then preprocessed using SPM5 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). Functional images were corrected for differences in slice acquisition timing, followed by motion correction using a two-pass six-parameter rigid-body realignment procedure. The T2-weighted coplanar anatomical image was then coregistered to the mean functional image, and the T1-weighted whole-brain SPGR image was in turn coregistered to the T2-weighted image. The SPGR image was then segmented by tissue type, and the gray matter image was warped to a gray matter template image in Montreal Neurological Institute space. The resulting nonlinear transformation parameters were applied to all functional images, which then were resampled into 3-mm isotropic voxels and smoothed with an 8-mm FWHM kernel.

### *MVPA Analysis*

Pattern classification analyses were implemented in MATLAB using the Princeton MVPA Toolbox ([code.google.com/p/princeton-mvpa-toolbox](http://code.google.com/p/princeton-mvpa-toolbox)) and custom code. Within each run, each voxel's time series was detrended to remove linear and quadratic trends, high-pass filtered to remove frequencies below 0.01 Hz, and  $z$  scored. To reduce the 2480-volume fMRI time series to a single brain activity measure for each of the 300 trials, the four TRs acquired 6–14 sec after the onset of each trial (i.e., TRs = 4–7), corresponding to the peak window of task-related activation, were extracted and averaged. Trials for which the global activity level deviated by more than  $\pm 3$  SD from the mean were deemed to be outliers and discarded before analysis. A common 55,761-voxel inclusive mask was applied to the spatially normalized data of all participants to exclude the cerebellum and motor, premotor, and somatosensory cortices. This masking, coupled with the reversal of the button mappings halfway through the scanning session, prevented the classifier from exploiting brain activity differences related to the motor responses associated with distinct mnemonic states.

In each classification analysis, we assessed how accurately the classifier could discriminate between trials from two or more distinct mnemonic conditions. Owing to the relatively low number of incorrectly performed trials (i.e., Misses and False Alarms), only correct trials (i.e., Hits and Correct Rejections [CRs]) were included in the analyses. Except where otherwise indicated, separate classifier models were trained and tested on each participant's data using a 10-fold cross-validation procedure. Trials were

randomly divided into 10 balanced subsets, with each subset containing an equal number of trials from each class. Trials from nine of these subsets were used for classifier training, and the held-out trials were used as a test set for assessing generalization performance. This process was iteratively repeated with each of the 10 subsets of held-out trials. Balancing the number of trials from each class prevented the classifier from developing a bias to identify trials as belonging to the more plentiful class and ensured a theoretical null hypothesis classification accuracy rate of 50% and area under the curve (AUC) of 0.5. An additional set of analyses with shuffled class labels confirmed that chance classification performance converged around these values. For any given classification, participants with fewer than 15 trials per class were excluded, because having an insufficient number of training examples can result in unstable classifier performance (Pereira, Mitchell, & Botvinick, 2009). To further ensure the stability of our results, all classification analyses were repeated 20 times, each using a different randomly sampled subset of trials, and the results were then averaged.

In addition to the standard within-participant classification analyses, several across-participant classification analyses were conducted. In one such analysis, we trained the classifier on the pooled data from all but one participant and tested its ability to predict the condition labels of brain patterns measured in the held-out participant. This leave-one-participant-out cross-validation scheme was iterated until each participant's data served as the test set. In another analysis, we trained a classifier on the combined data from all 16 participants in this study and tested it on data from 16 unique participants who performed a face recognition memory experiment in a different 3-T scanner (methods and results from that experiment were previously reported in Rissman et al., 2010). We also ran an analysis with the reverse training/testing designation (i.e., training the classifier on data from the face memory experiment and testing it on data from the present experiment). Note that, in all of these across-participant analyses, the classifier was always trained and tested on brain patterns from individual trials.

All classifications utilized a regularized logistic regression (RLR) algorithm, which we have found to perform well in similar experimental paradigms (Uncapher et al., 2015; Rissman et al., 2010). This algorithm implemented a multiclass logistic regression function using a softmax transformation of linear combinations of features (i.e., voxels) with an additional ridge penalty term as a Gaussian prior on the feature weights. This penalty term provided  $L_2$  regularization, enforcing small weights. During classifier training, the RLR algorithm learned the set of weights ( $\beta$  values) that maximized the log likelihood of the data; weights were initialized to zero, and optimization was implemented with conjugate gradient minimization using the gradient of the log likelihood combined with the  $L_2$  penalty. The  $L_2$  penalty was set to be half of the additive inverse of a user-specified parameter, multiplied by the square of the  $L_2$  norm of the weight vector for each class,

added over classes. We elected to set this parameter to a fixed value of 100 for all within-participant classification analyses and 10,000 for all across-participant classification analyses. Other than the use of a large anatomical mask (described above), no additional feature selection was performed. As with our prior work (Rissman et al., 2010), here too, we found that restricting the classifier to a subset of voxels based on their within-training set univariate effects did not typically improve classification accuracy. This outcome likely reflects the ability of the classifier to effectively reduce the weighting of features (i.e., voxels) that provide little relevant information to the classifier (cf. Chu et al., 2012).

For all binary (i.e., two-class) analyses, classification performance was summarized by an AUC metric. After fitting the RLR model parameters using the training set data, each brain activity pattern from the test set was fed into the model and yielded an estimate of the probability of that trial being from Class A or Class B (by construction, these two values always sum to 1). These probability values were concatenated across all cross-validation testing folds and then ranked. The classifier's true positive [ $P(\text{Class A} \mid \text{Class A})$ ] rate and false positive [ $P(\text{Class A} \mid \text{Class B})$ ] rate were calculated across all possible decision boundaries yielding a receiver operating characteristic curve. The area under this curve can be formally interpreted as the probability that a randomly chosen member of one class has a smaller estimated probability of belonging to the other class than has a randomly chosen member of the other class. That is, the AUC indexes the mean accuracy with which a randomly chosen pair of Class A and Class B trials could be assigned to their correct class.

For multiclass analyses (i.e., those conducted on more than two classes of trials), the classifier computed an exhaustive set of binary Class N versus Class  $\sim N$  analyses and returned the probability of each trial being a member of each class. The class with the maximal probability estimate was designated as the classifier's guess, and the accuracy of these guesses was aggregated across trials to form a single accuracy value for each participant. Multiclass decoding performance was further summarized by confusion matrices, which illustrate the complete probabilistic relationship between the classifier's guesses and the true class labels.

To visualize the anatomical distribution of informative voxels, classification importance maps were derived based on the logistic regression  $\beta$  weights yielded during each classifier training cycle; these  $\beta$  weights were averaged across each of the 10 cross-validation iterations and then across each of the 20 rounds of trial-count-balanced classifications. By convention, a positive weight value indicates that a voxel's activity magnitude on each trial was positively correlated with the probability of that trial being from Class A, whereas a negative weight value indicates the opposite relationship (i.e., increased activity leading to a prediction of Class B). These  $\beta$  weights were then multiplied by each voxel's mean activity level for Class A trials (which, owing to our trial balancing and  $z$

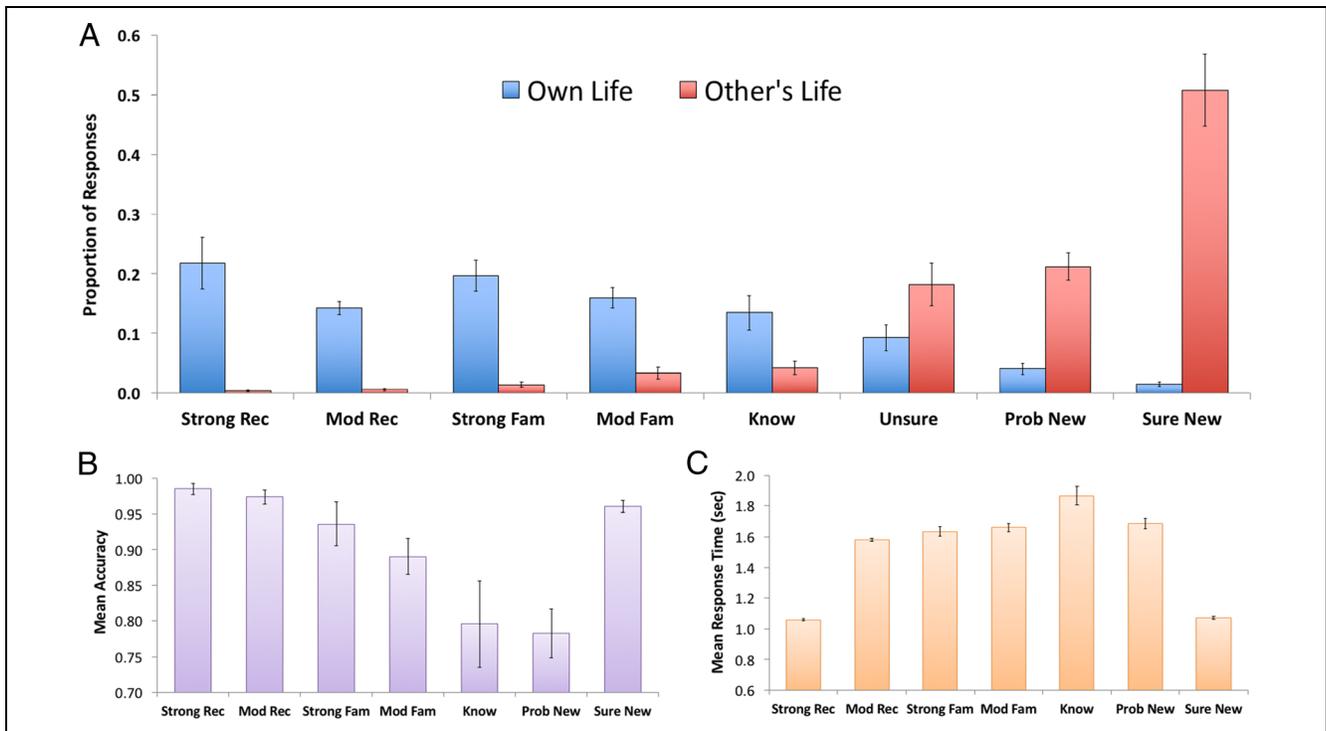
scoring procedure, is always the additive inverse of its mean activity level for Class B trials) and rescaled by a constant factor of 10,000 (to aid in later visualization). Voxels with positive values for both activity and weight were given positively signed importance values, voxels with negative activity and weight were given negatively signed importance values, and voxels for which the activity and weight had opposite signs were assigned importance values of zero (Johnson et al., 2009; McDuff, Frankel, & Norman, 2009). Random effects  $t$  tests were used to reveal regions whose mean importance values reliably differed from zero. Except where otherwise indicated, importance maps were thresholded at  $p < .05$  (corrected) based on the combination of a voxel height threshold of  $p < .005$  (two tailed) and a minimum cluster extent threshold of 45 voxels. These thresholds were derived based on Monte Carlo simulations implemented in the AFNI program *3dClustSim*; spatial smoothness for the simulations was estimated using the AFNI program *3dFWHMx*, based on a null hypothesis importance map derived from a classification analysis that used shuffled class labels (averaged across 50 different shuffled iterations). For visualization, thresholded importance maps were projected onto the left and right hemisphere inflated PALS cortical surface templates using Caret software ([www.nitrc.org/projects/caret](http://www.nitrc.org/projects/caret)).

## RESULTS

### Behavioral Results

At test, participants were asked to differentiate Own Life from Other's Life events. Overall, they were correct on 0.80 of trials, were incorrect on 0.07 of trials, and were unsure on 0.13 of trials. When excluding unsure responses from analysis, mean hit rate was 0.93, and the mean false alarm rate was 0.13, indicating that participants rarely indicated false recognition of Other's Life events (mean  $d' = 2.87$ ; above chance,  $t(15) = 16.88$ ,  $p < 10^{-10}$ ). Figure 2 depicts the full distribution of behavioral responses across the eight response options, along with the mean accuracy and RT associated with each. One participant (s01) only indicated Familiarity responses on 0.04 of Own Life events, whereas he indicated Recollection responses on 0.88 of these events; because it is unclear if this participant properly appreciated the subjective distinction between Recollection and Familiarity, his data are excluded from Figure 2 as well as all behavioral and fMRI analyses involving contrasts between or within these response types.

Collapsing across Strong and Moderate response subtypes, participants were significantly more accurate (0.98 vs. 0.94;  $t(14) = 2.58$ ,  $p = .021$ ) and faster (1.32 vs. 1.60 sec;  $t(14) = 4.84$ ,  $p = .0002$ ) when indicating Recollection than when indicating Familiarity. Familiarity responses were more accurate than Know (0.84) responses ( $t(14) = 2.28$ ,  $p = .038$ ), although mean RTs did not reliably differ ( $p = .22$ ). When comparing Strong



**Figure 2.** (A) Distribution of behavioral responses to Own Life events (blue) and Other's Life events (red). (B) Mean accuracy associated with each response option. (C) Mean response times for correct trials. Note that "Unsure" trials are not included in B and C because we cannot assess the correctness of such responses.

Recollection (Strong Rec) and Moderate Recollection (Mod Rec) responses, accuracies did not differ ( $p = .44$ ), but RTs were significantly faster for Strong Rec ( $t(14) = 11.08, p < 10^{-7}$ ). When comparing Strong Familiarity (Strong Fam) and Moderate Familiarity (Mod Fam) responses, Strong Fam responses were more accurate ( $t(14) = 2.70, p = .017$ ), but RTs did not differ ( $p = .31$ ). Accuracies for Sure New and Probably New (Prob New) responses significantly differed ( $t(15) = 5.05, p = .0001$ ), as did RTs ( $t(15) = 8.85, p < 10^{-6}$ ). Notably, neither the accuracy of participants' responses nor their use of the memory rating scale differed as a function of whether the event sequences being tested had been captured during the first, second, or third week of camera wearing (all  $ps > .1$ ; Table 1).

### fMRI Results

A series of MVPA decoding analyses were performed to evaluate how accurately multivariate classifiers could dis-

criminate fMRI activity patterns associated with distinct memory retrieval experiences, and importance maps were generated to determine which regions were most diagnostic for specific classifications. First, we trained and tested whole-brain classifier models (excluding motor, premotor, and cerebellar regions) on data from individual participants. Given the relatively low number of incorrect trials (i.e., Misses and False Alarms), all classification analyses were restricted to data from correctly performed trials (i.e., Hits and CRs). We also excluded Hit trials for which participants indicated a Know response, as such responses may reflect autobiographical semantic knowledge (i.e., recognizing a personally relevant object in the photos, such as one's bicycle), rather than autobiographical episodic memory (Tulving, 1989). Moreover, although the average Know response rate was 0.14 for Own Life trials, the degree of Know responding varied widely across participants; importantly, seven participants had an insufficient number of Know hits (i.e.,  $<15$  trials) to warrant its inclusion as a stand-alone condition of interest.

**Table 1.** Behavioral Performance for Own Life Events Captured during Each Week of Camera Wearing

	Overall Hit Rate	Proportion Rec Hits	Proportion Fam Hits	Proportion Know Hits
Week 1	0.93	0.37	0.46	0.16
Week 2	0.93	0.39	0.44	0.17
Week 3	0.94	0.40	0.43	0.17

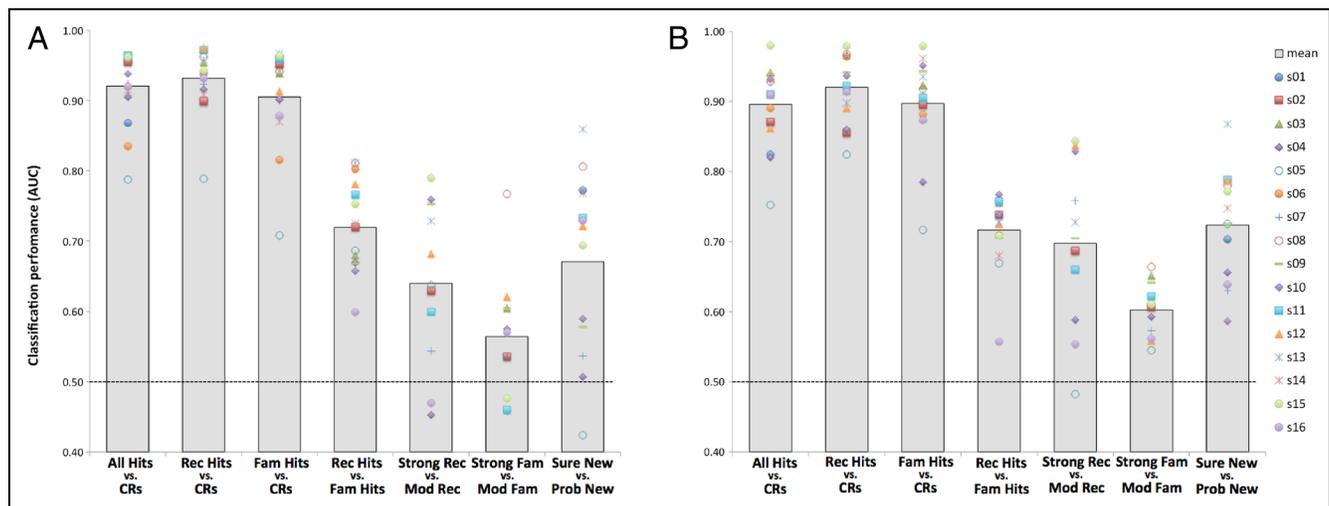
The results of the within-participant classification analyses are reported in Figure 3A. We first collapsed across more subtle mnemonic distinctions, assessing the neural discriminability of all recognized Own Life events (i.e., Hits, collapsed across Recollection and Familiarity) versus correctly rejected Other's Life events (i.e., CRs). Classification of Hits versus CRs was extremely accurate (mean AUC = 0.920;  $t(15) = 32.62, p < 10^{-14}$ ), with robust decoding observed for every participant's data (AUC range = 0.79–0.97). Notably, when only the top 10% of the classifier's most "confidently" made guesses were considered for each participant (cf. Uncapher et al., 2015; Rissman et al., 2010), mean classification performance rose to AUC = 0.987, with perfect performance (AUC = 1.0) obtained in 9 of 16 participants. This indicates that the subset of test trials that the classifier deemed to be the most paradigmatic examples of Hits and CRs were nearly always true examples of those classes. When the Hits-versus-CRs classification analysis was rerun 50 times for each participant with randomly shuffled class labels, decoding performance (mean AUC = 0.5003, range = 0.49–0.52) converged on the theoretical null hypothesis level (0.5), indicating that no insidious biases were present in our analysis workflow.

We next examined whether the classifier's ability to distinguish Hits and CRs diminished as the probed memories became more temporally remote. This was not the case; separate classifier models using only the Hit trials from the first, second, or third week of camera wearing showed roughly equivalent decoding performance (AUCs = 0.870, 0.880, and 0.903, respectively;  $F(2, 28) = 2.46, p = .10$ ). In an attempt to decode the temporal remoteness of a memory, we trained a classifier to discriminate Week 1 Hits versus Week 3 Hits, but performance did not reach significance (AUC = 0.545,  $p = .13$ ). This was also the case

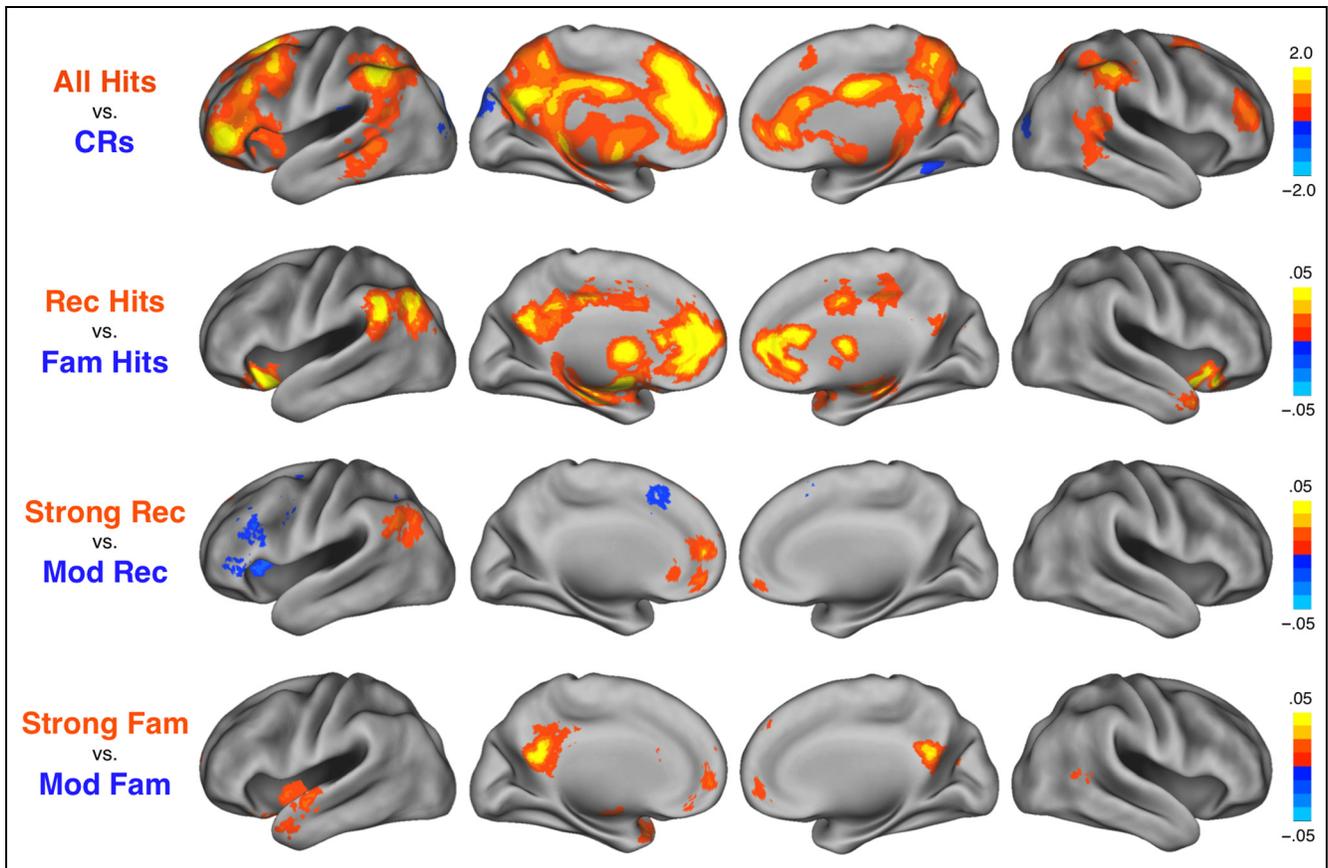
when the analysis was restricted to the Recollection Hits (AUC = 0.549,  $p = .16$ ). Thus the remainder of our reported analyses combine events captured across all 3 weeks of camera-wearing.

When Hits were broken down by whether participants indicated Recollection or Familiarity, the classifier showed a modest advantage for decoding Rec Hits versus CRs (AUC = 0.928) as compared with Fam Hits versus CRs (AUC = 0.905; difference:  $t(14) = 2.45, p = .028$ ). Furthermore, Rec Hits could be reliably discriminated from Fam Hits (AUC = 0.719;  $t(14) = 12.90, p < 10^{-8}$ ). Importantly, this was also the case when Mod Rec Hits were contrasted with Strong Fam Hits (AUC = 0.583;  $t(11) = 3.57, p = .0044$ ), despite these two trial types being closely matched on accuracy and RT. Within mnemonic categories, gradations in retrieval strength could also be decoded, although this distinction was more robust for Strong versus Mod Rec (AUC = 0.640;  $t(10) = 4.01, p = .0025$ ) than for Strong versus Mod Fam (AUC = 0.563;  $t(10) = 2.39, p = .038$ , uncorrected). Finally, high confidence CRs (Sure New) were readily discriminable from low confidence CRs (Prob New; AUC = 0.671;  $t(12) = 4.71, p < 10^{-3}$ ). For each of these binary classification schemes, comparable decoding performance was obtained when the classifier model was trained on the pooled data from all but one participant and tested on the held-out data from that participant (Figure 3B). Indeed, within-participant and across-participant decoding performance did not reliably differ for any of the classification schemes reported in Figure 3, when Bonferroni correcting for seven paired comparisons.

To evaluate which brain regions most strongly and consistently contributed to the success of our within-participant classification analyses, we generated importance maps reflecting the mean weighting of individual voxels (Figure 4). Importance maps for the Hits-versus-CRs classification



**Figure 3.** Classification results. (A) Classifier performance from within-participant analysis, using a 10-fold cross-validation scheme. (B) Classifier performance from across-participant analysis, using a leave-one-participant-out scheme. In both plots, vertical bars represent the mean AUC value for each two-class discrimination. Colored markers indicate AUC values from individual participants; only participants with 15 or more trials per class are included in each analysis. The dashed line represents chance-level decoding (AUC = 0.5).



**Figure 4.** Classification importance maps. Group-averaged maps of classifier importance values are shown for four binary classifications of interest. All maps are masked to only include voxels whose importance values significantly differ from zero at  $p < .005$  (two tailed). The upper two maps also impose a minimum cluster extent threshold of 45 voxels to yield a whole-brain corrected threshold of  $p < .05$ . Warm colors indicate voxels for which increased activity biased the classifier to predict Condition A, and cool colors indicate voxels for which increased activity biased the classifier to predict Condition B. Importance values are shown in raw arbitrary units.

revealed an extensive set of frontoparietal regions, including both lateral and medial areas, that were positively predictive of Hits and a sparser set of visual cortical regions, including bilateral occipital lobe and right inferior temporal lobe, that were positively predictive of CRs.

Relative to the Hits-versus-CRs classification, maps for the Rec Hits versus Fam Hits classification implicated some of the same regions (especially along the medial wall), but there were a number of notable differences in the overall pattern. For instance, the lateral frontal regions that were positively predictive of Hits in the previous analysis were not diagnostic of the Rec versus Fam distinction, nor were the bilateral regions of the intraparietal sulcus. Rather, within the frontal lobe, only medial frontal areas, together with the bilateral anterior insula, were predictive of Rec Hits. Within the parietal lobe, diagnostic voxels associated with Rec Hits were found in the left angular and supramarginal gyri as well as in the retrosplenial cortex (RSC)/posterior cingulate cortex (PCC). Bilateral regions of the hippocampus and parahippocampal cortex were also strongly implicated in the discriminability of Rec Hits versus Fam Hits. The absence of significant negative effects in the importance maps

indicates that engagement of these recollection-related regions was more consistently informative to the classifier than engagement of familiarity-preferring regions.

We next examined the importance maps associated with gradations of recollection (Strong Rec vs. Mod Rec) and familiarity (Strong Fam vs. Mod Fam). Because these classifications were based on a select subset of each participant's Hit trials, several participants lacked the requisite trial counts (a minimum of 15 trials per class) to be included in one or both analyses, leaving us with only 11 participants for each analysis. This substantially reduced the statistical power of the group  $t$  tests, as did the fact that lower classification accuracy levels are typically associated with noisier and more variable importance maps. Given the lower power, when we applied our stringent criteria for whole-brain corrected significance ( $p < .005$ , two tailed; cluster extent  $\geq 45$  voxels), no clusters achieved significance in the Strong Rec versus Mod Rec map, and only the RSC/PCC clusters achieved significance in the Strong Fam versus Mod Fam map. For exploratory purposes, we then rendered each map at the same voxel-level threshold ( $p < .005$ , two tailed), but without the cluster extent requirement (Figure 4,

bottom). For the classification of recollection strength, voxels that were positively predictive of Strong Rec were found in the left angular gyrus and bilateral vmPFC, whereas voxels that were positively predictive of Mod Rec were found in the left lateral and dorsomedial PFC. For the classification of familiarity strength, voxels that were positively predictive of Strong Fam were most prominent in bilateral RSC/PCC but also seen in the left anterior temporal pole, left insula, right posterior middle temporal gyrus, and bilateral vmPFC. Although, in the surface rendering, this vmPFC region appears to overlap with that seen in the Strong Rec versus Mod Rec map, in actuality (i.e., in volumetric space), these maps only share two overlapping voxels. No regions showed signal changes that were positively predictive of Mod Fam.

The results described thus far were all derived from binary classification analyses where the model was trained to discriminate between trials from two distinct classes. To determine whether a classifier could reliably predict a trial's mnemonic status out of a larger set of possible options, we trained a new classifier model to differentiate six classes of trials, including two levels of recollection, two levels of familiarity, and two levels of novelty. Again, incorrectly performed trials (i.e., Misses and False Alarms) and Know trials were excluded, and classifications were only run on data from the eight participants who had at least 15 trials of each of the six classes. The results of this analysis are reported in the form of a confusion matrix (Figure 5), reflecting the distribution of the classifier's guesses for each of the six trial types. The six-

way classification achieved an accuracy level of 34.9%, which was significantly better than chance-level guessing (empirically estimated to be 16.4% based on rerunning the analysis with shuffled class labels;  $t(7) = 7.22, p < 10^{-3}$ ). For each column of the confusion matrix (reflecting the participant's actual response), the classifier's modal guess was always the correct guess. Perhaps more importantly, however, the distribution of the classifier's incorrect guesses followed an orderly profile reflecting a hierarchy of memory states. For instance, when Strong Rec trials were misclassified, they were most often misclassified as Mod Rec, and vice versa for Mod Rec trials. Strong Fam trials were almost equally likely to be misclassified as Mod Rec or Mod Fam but were rarely misclassified as New. In contrast, for Mod Fam trials, the classifier incorrectly guessed Prob New with a similar frequency as its guesses of Mod Rec and Strong Fam. In addition, for trials where participants indicated that photos from someone else's life were Sure New, the classifier rarely misclassified these events as recognized but rather erred by predicting the wrong level of novelty decision confidence.

In a final set of analyses, we evaluated the degree to which the brain patterns associated with autobiographical retrieval in the present paradigm resemble those previously found to be diagnostic of memory states in a laboratory-based face recognition memory task (Rissman et al., 2010). To this end, we trained a classifier model on the face memory data collected from the 16 independent participants in Rissman et al. (2010) and tested its ability to decode memory states (Hits vs. CRs and Rec Hits vs. Fam Hits) within the 16 participants in this autobiographical memory study. We also conducted the reverse analysis, training on these data and testing on the data from our earlier laboratory-based memory study. Note that, in the laboratory-based study, Rec Hits versus Fam Hits was operationalized as Rec Hits versus High-Confidence Fam Hits (that study's response options only included one level of recollection but two levels of familiarity); in this study, Rec Hits included both Strong and Mod Rec, and Fam Hits included both Strong and Mod Fam. The results of the across-experiment classifications are reported in Table 2. A classifier trained on data from the laboratory-based memory study and tested on the autobiographical memory study was able to reliably decode both Hits versus CRs ( $t(15) = 7.62, p < 10^{-7}$ ) and Rec Hits versus Fam Hits ( $t(14) = 7.91, p < 10^{-7}$ ), with a significant performance advantage for the former ( $t(14) = 3.16, p = .004$ ). A classifier trained on data from the autobiographical memory study and tested on the laboratory-based memory study was also able to reliably decode Hits versus CRs ( $t(15) = 7.22, p < 10^{-7}$ ) and Rec Hits versus Fam Hits ( $t(14) = 5.95, p < 10^{-5}$ ), with a marginally significant performance advantage for the latter ( $t(14) = 2.02, p = .053$ ). Interestingly though, the levels of across-study classification performance, when applied to the present autobiographical data, were substantially lower than that observed using within-study classification (e.g., Hits vs.

		Participant's Actual Response					
		Strong Rec	Mod Rec	Strong Fam	Mod Fam	Prob New	Sure New
Classifier's Guess	Strong Rec	43%	25%	16%	13%	6%	7%
	Mod Rec	20%	28%	22%	17%	6%	6%
	Strong Fam	12%	19%	24%	18%	9%	5%
	Mod Fam	10%	15%	21%	25%	8%	6%
	Prob New	7%	6%	10%	17%	42%	29%
	Sure New	8%	7%	6%	10%	29%	47%
Overall accuracy = 35% (chance = 16.4%)							

**Figure 5.** Confusion matrix for six-way classification. Percentages indicate the distribution of the classifier's guesses for each trial type. Values along the diagonal indicate the probability of the classifier indicating the correct class label. Results are averaged across classifications run on data from eight participants who had at least 15 trials in each of these six classes.

**Table 2.** Across-experiment Classification Performance (AUC)

	<i>Train Laboratory-based</i>			<i>Train Real-world</i>		
	↓			↓		
	<i>Test Real-world</i>			<i>Test Laboratory-based</i>		
	<i>Mean</i>	<i>Std Dev</i>	<i>Range</i>	<i>Mean</i>	<i>Std Dev</i>	<i>Range</i>
Hits vs. CRs	0.75	0.09	0.54–0.90	0.67	0.07	0.53–0.79
Rec Hits vs. Fam Hits	0.69	0.06	0.55–0.79	0.75	0.12	0.53–0.98

The left-hand column reports results from two classification analyses where the classifier was trained on data from a laboratory-based recognition memory fMRI study (Rissman et al., 2010) and tested on data from this study. The right-hand column reports results from the reverse analysis scheme.

CRs: mean within-study AUC = 0.92, mean across-study AUC = 0.75).

## DISCUSSION

Our experimental protocol, featuring the use of wearable cameras, allowed us to “noninvasively” catalog a diverse array of potentially memorable events from our participants’ day-to-day lives without drawing particular attention to the encoding of these memories. Although we had little control over what activities participants engaged in during their 3 weeks of camera wearing, we gained ecological validity in that we could later probe their memories for real-world autobiographical events and measure the associated brain activity patterns. While in the scanner, participants evaluated brief sequences of never-before-viewed photographs of their own events, interspersed with foil sequences that were captured by other participants’ cameras. MVPA methods were used to characterize the degree to which various neural signatures of memory retrieval could be reliably decoded on individual trials and provided information about which brain areas carried diagnostic signals.

Our analyses revealed several notable results. First, when we trained a classifier model to discriminate brain patterns associated with correct recognition of one’s own events versus CRs of someone else’s events, it was able to achieve near-perfect decoding accuracy (mean AUC = 0.92). This performance level exceeded the level observed in our previous face memory study (mean AUC = 0.83; Rissman et al., 2010), suggesting that the heterogeneity of the real-world photographic probes did not diminish our model’s ability to hone in on a highly consistent neural signature of recognition. By having participants make their responses using an 8-point memory rating scale that included two levels of recollection, familiarity, and novelty, we were able to show that our models could reliably determine whether participants were subjectively experiencing each memory state to a stronger or weaker degree. The neural signatures of these distinct memory states were sufficiently consistent across participants to yield comparable accuracy levels even when classifier models were trained and tested on data from different partici-

pants. Moreover, a six-way classification analysis showed that a classifier could infer which of these six states the participant was currently experiencing with accuracy levels well above chance. Importantly, classification errors followed a predictable pattern, illustrating that the classifier had acquired sensitivity to the full range of memory retrieval outcomes and their associated neural representations. Finally, we found that a classifier model trained on data from our prior laboratory-based face recognition task (Rissman et al., 2010), which required a 5-point response from an independent group of participants, could largely succeed at decoding the mnemonic status of events encountered in our experiment, which required an 8-point response. The converse was also true. Thus, the neural signatures associated with episodic retrieval appear to be relatively well preserved when probing for rich autobiographical events at 1- to 3-week retention intervals and for constrained laboratory-based memories of faces probed at a brief retention interval.

Given that our classifier models were trained and tested on whole-brain data (excepting motor-related regions), the neural signatures that the classifier learns as diagnostic for a given mnemonic distinction end up being broadly distributed throughout the brain. Accordingly, even weakly diagnostic voxels may make some small contribution to the classifier’s success, and the diagnosticity of individual voxels can be highly variable across participants. Thus, assessment of thresholded group-averaged importance maps will always be an imperfect approximation of the neural signatures of each state. Still, these maps can be informative by showcasing which regions are most consistently diagnostic. For the classification of Hits versus CRs, the importance maps revealed a prominent contribution of lateral frontal areas and the intraparietal sulcus, both with a left hemisphere bias as is typical in fMRI studies of retrieval success (i.e., old > new) effects (Hutchinson et al., 2014; Kim, 2013; McDermott et al., 2009). Although these frontoparietal regions are highly diagnostic of recognition, one should not assume that they collectively constitute the site of the memory engram. Rather, these regions likely contribute to the cognitive and attentional control processes needed for memory search, monitoring, and mnemonic evidence accumulation. Midline regions, including large swaths of the medial PFC and parietal

cortex, together with the hippocampus, were also diagnostic of Hits. In contrast, voxels diagnostic of CRs were most prominent in visual areas, including the occipital lobe and right inferior temporal cortex, presumably responding to the novelty of the people, places, and activities depicted in others' photos. When a classifier was trained to differentiate recollection and familiarity (Rec Hits vs. Fam Hits), the resulting importance map predominantly highlighted a set of recollection-preferring regions—including medial PFC, ventral posterior parietal cortex inclusive of angular gyrus, the RSC/PCC, and the hippocampus—sometimes referred to as the “core recollection network” (Rugg & Vilberg, 2013). This network shows considerable overlap with the default mode network, presumably owing to the fact that both undirected mentation and episodic retrieval entail self-referential processing and representation of information retrieved from (or constructed based on) memory (Andrews-Hanna, Saxe, & Yarkoni, 2014; Kim, 2012; Spreng & Grady, 2010).

If, instead of considering the distinction between recollection and familiarity, we consider the distinctions within recollection and familiarity, an intriguing profile of results emerges. Despite the fact that each classification compared a stronger memory with a weaker memory, the voxels that showed significant importance in the Strong Rec versus Mod Rec classification had essentially no overlap with those that showed importance in the Strong Fam versus Mod Fam classification (only two voxels in the entire brain showed common effects). Although some interpretive caution is warranted, given that these maps are based on data from a restricted set of subjects ( $n = 11$ ) and displayed at an uncorrected threshold, the neuroanatomical divergence is striking. When attempting to decode the strength of recollection, the classifier heavily anchored on activity levels in the left angular gyrus and vmPFC as predictive of Strong Rec. Previous fMRI studies have found the angular gyrus to be particularly sensitive to the amount of information that participants recollect (Vilberg & Rugg, 2007, 2009), and activity patterns within this region show greater content-specific reinstatement effects when more information is recalled (Leiker & Johnson, 2014) or when recall is more vivid (Kuhl & Chun, 2014). Such findings suggest a plausible role for this region in transiently buffering (Vilberg & Rugg, 2008) or binding (Shimamura, 2011) the reinstated mnemonic content represented within other brain regions. The vmPFC is thought to play a critical role in self-referential processing (Denny, Kober, Wager, & Ochsner, 2012; Sajonz et al., 2010), which likely explains its prominent involvement when participants attempt to project themselves back into and mentally “re-live” events depicted in their photographs (St Jacques et al., 2011). Thus, it is sensible that both regions are highly diagnostic of recollection strength in our data. Intriguingly, regions of the left lateral PFC and dorsomedial PFC showed the opposite profile, with their engagement being predictive of Mod Rec. It is possible that, when our participants experienced a mod-

erate amount of recollection, they attempted to use the few details they could recollect to cue their memory in hopes of recovering additional contextual information; these regions may potentially contribute to that strategic search and monitoring process (Dobbins, Foley, Schacter, & Wagner, 2002). For the decoding of familiarity strength, the most diagnostic brain region was bilateral RSC/PCC. The fact that this region was diagnostic of familiarity strength and not recollection strength was surprising, considering that this region was positively diagnostic of recollection in the Rec Hits versus Fam Hits analysis and is often associated with recollection in the literature (Rugg & Vilberg, 2013). However, prior fMRI studies and meta-analyses have also implicated this region in familiarity (Horn et al., 2015; Qin et al., 2012; Montaldi, Spencer, Roberts, & Mayes, 2006). Moreover, Binder, Desai, Graves, and Conant (2009) have suggested that PCC may be a critical hub for linking episodic and semantic information. In our study, reports of strong familiarity may have been influenced by personal semantics (Renoult, Davidson, Palombo, Moscovitch, & Levine, 2012), such as the recognition that the depicted event is consistent with the kind of thing one tends to do. The presence of diagnostic signals in the anterior temporal lobe, also a component of the semantic network (Binder et al., 2009), may likewise be attributable to participants' reliance on personal semantic knowledge in their assessment of familiarity. Taken together, the divergence of these maps suggests that the respective classifiers are preferentially relying on different brain regions to provide information about memory strength in the domain of recollection, than in the domain of familiarity. This affirms a qualitative, rather than purely quantitative, distinction between these expressions of retrieval.

A particularly compelling aspect of the present results was our ability to train a six-way classifier model to discriminate between strong and moderate expressions of recollection, familiarity, and novelty based on the underlying brain activity patterns. Although the six-way classification accuracy of 35% was far from perfect, the result is impressive when one considers that the empirically validated chance accuracy was 16.4%. Furthermore, when the classifier guessed incorrectly, its guesses reflected a gradation in activity patterns that appeared to parallel the gradation within each mnemonic state. For example, the classifier tended to confuse strong recollection with moderate recollection and vice versa, rather than confusing these trial types with familiarity or novelty. Similarly, strong and moderate levels of perceived novelty were rarely confused with recollection or familiarity, indicating that participants' decision confidence was clearly dissociated from their memory state. To our knowledge, our data constitute the first demonstration that brain patterns measured during retrieval can be used to decode the specific level of memory associated with a given probe stimulus out of a wide array of potential options.

Another noteworthy result was our demonstration that fMRI activity patterns measured in response to real-world

photographic probes in this study were sufficiently similar to those measured in response to face stimuli in our previous study (Rissman et al., 2010) so as to allow a classifier model trained on one data set to reliably predict the mnemonic status of individual trials in the other data set. This remarkable across-experiment generalization was found despite the fact that the two data sets were acquired on different scanners with different participants making judgments on very different types of stimuli in tasks requiring distinct memory-to-response mappings. Although across-experiment decoding performance did not achieve the accuracy levels observed in the respective within-experiment decoding analyses, the reliability of the effects suggests that at least a subset of the neural processes engaged during recognition (facilitating Hits vs. CRs decoding) and recollection (facilitating Rec vs. Fam decoding) are relatively well preserved across these two distinct tasks.

Our successful across-experiment classification results appear to challenge the view that memory judgments made in response to laboratory-encoded stimuli engage fundamentally different mechanisms than judgments made in response to probes that evoke the retrieval of one's own personal past (Roediger & McDermott, 2013). One reason that a recent meta-analysis (McDermott et al., 2009) of laboratory-based and autobiographical memory studies showed such limited overlap between the regions engaged during these two types of tasks may be that most of the included laboratory-based studies used old/new item recognition judgments whereas most of the included autobiographical studies required judgments that necessarily evoked recollection. Thus, the apparent dissociation may in fact be driven by the disparate nature of the mnemonic retrieval processes engaged during laboratory-based and autobiographical memory tasks. Indeed, a meta-analysis of Remember/Know judgments based exclusively on data from standard laboratory-based memory studies (Kim, 2010) reported a map of Remember > Know effects that looks strikingly similar to McDermott et al.'s (2009) meta-analytic map of autobiographical retrieval studies. Conversely, a recent meta-analysis of standard old/new region effects (i.e., Hits > CRs), again based exclusively on laboratory-based memory studies (Kim, 2013), yielded a map that looks virtually identical to McDermott et al.'s (2009) meta-analytic map of laboratory-based memory studies. Because our across-experiment classification analyses inherently matched the two studies on the mnemonic distinction of interest (either Hits vs. CRs or Rec vs. Fam), this may have enabled our classifier model to hone in on the relevant neural patterns that underlie these respective mnemonic distinctions, distinctions that appear common to real-world and laboratory-based memories. Future work will be needed to isolate conditions under which retrieval-related brain activity patterns evoked during laboratory-based and autobiographical memory tasks may diverge, because neuropsychological dissociations clearly exist between these two classes of memory (Palombo et al., 2015; Patihis et al., 2013; LePort

et al., 2012) and our across-experiment classification levels were lower than our within-experiment levels.

The present investigation was motivated in part by a desire to determine how accurately the presence or absence of individual event memories could be detected based on analysis of the evoked fMRI activity pattern. One goal was thus to weigh in on potential forensic uses of brain scans as a means for detecting experiential knowledge (Meegan, 2008). Our classification results showcase extraordinarily accurate predictions of whether participants were viewing photos from their own past experiences as well as a reasonably accurate ability to infer the specific subjective nature of one's retrieval state. The fact that our classifier models could predict mnemonic retrieval outcomes even in participants whose data the classifier was never trained on highlights the robustness of these neural signatures and their potential applicability. However, there are several reasons to be cautious when interpreting these results. For one, our use of an explicit memory task prevented us from evaluating whether we could detect neural signatures of autobiographical retrieval triggered automatically in response to salient probe stimuli. Furthermore, because participants performed so accurately on our task, this left us with very few false memories and forgotten memories, precluding what might have been an interesting investigation into illusory memory and implicit expressions of memory, respectively. By restricting our analyses to correctly performed trials, the memory states that the classifier was trained to decode were inextricably linked to participants' subjective retrieval experiences. Other recent results suggest that these subjective states could be willfully manipulated when participants were given incentives to do so (Uncapher et al., 2015). Thus, the present results should not be taken as evidence that brain-based memory detection procedures are ready for applied uses, especially not in judicial or forensic contexts. Nonetheless, these memory decoding data add to the growing number of demonstrations indicating that there is a wealth of information contained within single-trial fMRI activity patterns that, when analyzed with the right techniques, can reveal key features of one's mnemonic state.

## Acknowledgments

This study was supported by a grant from the John D. and Catherine T. MacArthur Foundation to Vanderbilt University, with a subcontract to Stanford University. Its contents do not necessarily represent official views of either the John D. and Catherine T. MacArthur Foundation or the MacArthur Foundation Research Network on Law and Neuroscience ([www.lawneuro.org](http://www.lawneuro.org)). We thank Hank Greely, Owen Jones, Nita Farahany, and Melina Uncapher for helpful discussions. We are grateful for Kevin Hardekopf's assistance with reviewing and processing participants' photographs and for Daniel Lin and Ziwei Zhang's assistance with fMRI data visualization.

Reprint requests should be sent to Jesse Rissman, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563, or via e-mail: [rissman@psych.ucla.edu](mailto:rissman@psych.ucla.edu).

## Note

1. One participant (s08) did not have his camera turned on enough during Weeks 2 and 3 of the camera-wearing period, and thus we could only generate 36 event sequences per week for those weeks. To compensate for this reduction in stimulus materials, we were able to generate 90 viable event sequences from his Week 1 photos. Nonetheless, with only 162 event sequences in total, we reduced this participant's memory test to nine runs, each with 10 events from Week 1, four events from Week 2, and four events from Week 3. Because of this imbalance, this participant's data were excluded from any classification analyses that subdivided events based on their week of occurrence.

## REFERENCES

- Andrews-Hanna, J. R., Saxe, R., & Yarkoni, T. (2014). Contributions of episodic retrieval and mentalizing to autobiographical thought: Evidence from functional neuroimaging, resting-state connectivity, and fMRI meta-analyses. *Neuroimage*, *91*, 324–335.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767–2796.
- Bles, M., & Haynes, J. D. (2008). Detecting concealed information using brain-imaging technology. *Neurocase*, *14*, 82–92.
- Bonnici, H. M., Chadwick, M. J., Lutti, A., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2012). Detecting representations of recent and remote autobiographical memories in vmPFC and hippocampus. *Journal of Neuroscience*, *32*, 16982–16991.
- Bonnici, H. M., Chadwick, M. J., & Maguire, E. A. (2013). Representations of recent and remote autobiographical memories in hippocampal subfields. *Hippocampus*, *23*, 849–854.
- Bosch, S. E., Jehee, J. F., Fernandez, G., & Doeller, C. F. (2014). Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *Journal of Neuroscience*, *34*, 7493–7500.
- Bosshardt, S., Degonda, N., Schmidt, C. F., Boesiger, P., Nitsch, R. M., Hock, C., et al. (2005). One month of human memory consolidation enhances retrieval-related hippocampal activity. *Hippocampus*, *15*, 1026–1040.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Buchsbaum, B. R., Lemire-Rodger, S., Fang, C., & Abdi, H. (2012). The neural basis of vivid memory is patterned on perception. *Journal of Cognitive Neuroscience*, *24*, 1867–1883.
- Cabeza, R., Prince, S. E., Daselaar, S. M., Greenberg, D. L., Budde, M., Dolcos, F., et al. (2004). Brain activity during episodic retrieval of autobiographical and laboratory events: An fMRI study using a novel photo paradigm. *Journal of Cognitive Neuroscience*, *16*, 1583–1594.
- Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*, 219–227.
- Chadwick, M. J., Hassabis, D., & Maguire, E. A. (2011). Decoding overlapping memories in the medial temporal lobes using high-resolution fMRI. *Learning and Memory*, *18*, 742–746.
- Chadwick, M. J., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2010). Decoding individual episodic memory traces in the human hippocampus. *Current Biology*, *20*, 544–547.
- Chang, C., & Glover, G. H. (2009). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *Neuroimage*, *47*, 1448–1459.
- Chu, C., Hsu, A. L., Chou, K. H., Bandettini, P., Lin, C., & Alzheimer's Disease Neuroimaging Initiative. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage*, *60*, 59–70.
- Cohen, G., & Conway, M. A. (2008). *Memory in the real world* (3rd ed.). New York: Psychology Press.
- Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, *47*, 2305–2313.
- Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*, 1742–1752.
- Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron*, *35*, 989–996.
- Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123–152.
- Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*, *44*, 162–167.
- Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cerebral Cortex*, *24*, 3350–3364.
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*, 299–306.
- Horn, M., Jardri, R., D'Hondt, F., Vaiva, G., Thomas, P., & Pins, D. (2015). The multiple neural networks of familiarity: A meta-analysis of functional imaging studies. *Cognitive, Affective & Behavioral Neuroscience*. doi:10.3758/s13415-05-0392-1.
- Hutchinson, J. B., Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic retrieval: Convergent and divergent effects of attention and memory. *Learning and Memory*, *16*, 343–356.
- Hutchinson, J. B., Uncapher, M. R., Weiner, K. S., Bressler, D. W., Silver, M. A., Preston, A. R., et al. (2014). Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. *Cerebral Cortex*, *24*, 49–66.
- Johnson, J. D., McDuff, S. G., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, *63*, 697–708.
- Kim, H. (2010). Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. *Neuroimage*, *50*, 1648–1657.
- Kim, H. (2012). A dual-subsystem model of the brain's default network: Self-referential processing, memory retrieval processes, and autobiographical memory retrieval. *Neuroimage*, *61*, 966–977.
- Kim, H. (2013). Differential neural activity in the recognition of old versus new events: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, *34*, 814–836.
- Kuhl, B. A., & Chun, M. M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *Journal of Neuroscience*, *34*, 8051–8060.
- Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 5903–5908.

- Leiker, E. K., & Johnson, J. D. (2014). Neural reinstatement and the amount of information recollected. *Brain Research*, *1582*, 125–138.
- Leiker, E. K., & Johnson, J. D. (2015). Pattern reactivation co-varies with activity in the core recollection network during source memory. *Neuropsychologia*, *75*, 88–98.
- LePort, A. K., Mattfeld, A. T., Dickinson-Anson, H., Fallon, J. H., Stark, C. E., Kruggel, F., et al. (2012). Behavioral and neuroanatomical investigation of highly superior autobiographical memory (HSAM). *Neurobiology of Learning and Memory*, *98*, 78–92.
- Maguire, E. A. (2012). Studying the freely-behaving brain with fMRI. *Neuroimage*, *62*, 1170–1176.
- McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, *47*, 2290–2298.
- McDuff, S. G. R., Frankel, H. C., & Norman, K. A. (2009). Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *Journal of Neuroscience*, *29*, 508–516.
- Meegan, D. V. (2008). Neuroimaging techniques for memory detection: Scientific, ethical, and legal issues. *American Journal of Bioethics*, *8*, 9–20.
- Milton, F., Muhlert, N., Butler, C. R., Benattayallah, A., & Zeman, A. Z. (2011). The neural correlates of everyday recognition memory. *Brain and Cognition*, *76*, 369–381.
- Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, *16*, 504–520.
- Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, 11078–11083.
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430.
- Palombo, D. J., Alain, C., Soderlund, H., Khuu, W., & Levine, B. (2015). Severely deficient autobiographical memory (SDAM) in healthy adults: A new mnemonic syndrome. *Neuropsychologia*, *72*, 105–118.
- Patihis, L., Frenda, S. J., LePort, A. K., Petersen, N., Nichols, R. M., Stark, C. E., et al. (2013). False memories in highly superior autobiographical memory individuals. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 20947–20952.
- Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, *45*(Suppl. 1), S199–S209.
- Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*, 1963–1966.
- Qin, P., Liu, Y., Shi, J., Wang, Y., Duncan, N., Gong, Q., et al. (2012). Dissociation between anterior and posterior cortical regions during self-specificity and familiarity: A combined fMRI-meta-analytic study. *Human Brain Mapping*, *33*, 154–164.
- Quamme, J. R., Weiss, D. J., & Norman, K. A. (2010). Listening for recollection: A multi-voxel pattern analysis of recognition memory retrieval strategies. *Frontiers in Human Neuroscience*, *4*, 61.
- Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences*, *16*, 550–558.
- 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, U.S.A.*, *107*, 9849–9854.
- Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: Insights from functional brain imaging. *Annual Review of Psychology*, *63*, 101–128.
- Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral Cortex*, *23*, 2818–2828.
- Roediger, H. L., 3rd, & McDermott, K. B. (2013). Two types of event memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 20856–20857.
- Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, *122*, 1–23.
- Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*, 255–260.
- Sajonz, B., Kahnt, T., Margulies, D. S., Park, S. Q., Wittmann, A., Stoy, M., et al. (2010). Delineating self-referential processing from episodic memory retrieval: Common and dissociable networks. *Neuroimage*, *50*, 1606–1617.
- Schacter, D. L., & Loftus, E. F. (2013). Memory and law: What can cognitive neuroscience contribute? *Nature Neuroscience*, *16*, 119–123.
- Shen, F. X., & Jones, O. D. (2011). Brain scans as evidence: Truths, proofs, lies, and lessons. *Mercer Law Review*, *62*, 861–883.
- Shimamura, A. P. (2011). Episodic retrieval and the cortical binding of relational activity. *Cognitive, Affective & Behavioral Neuroscience*, *11*, 277–291.
- Smith, J. F., Alexander, G. E., Chen, K., Husain, F. T., Kim, J., Pajor, N., et al. (2010). Imaging systems level consolidation of novel associate memories: A longitudinal neuroimaging study. *Neuroimage*, *50*, 826–836.
- 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*, 1765–1779.
- Spreng, R. N., & Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience*, *22*, 1112–1123.
- Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279–306.
- St Jacques, P. L., Conway, M. A., Lowder, M. W., & Cabeza, R. (2011). Watching my mind unfold versus yours: An fMRI study using a novel camera technology to examine neural differences in self-projection of self versus other perspectives. *Journal of Cognitive Neuroscience*, *23*, 1275–1284.
- St Jacques, P. L., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 19671–19678.
- St-Laurent, M., Abdi, H., & Buchsbaum, B. R. (2015). Distributed patterns of reactivation predict vividness of recollection. *Journal of Cognitive Neuroscience*, *27*, 2000–2018.
- Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic reinstatement in the medial temporal lobe. *Journal of Neuroscience*, *32*, 18150–18156.
- 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.
- Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., et al. (2006). Declarative memory

- consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 756–761.
- Thakral, P. P., Wang, T. H., & Rugg, M. D. (2015). Cortical reinstatement and the confidence and accuracy of source memory. *Neuroimage*, *109*, 118–129.
- Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, *63*, 483–509.
- Tulving, E. (1989). Remembering and knowing the past. *American Scientist*, *77*, 361–367.
- Uncapher, M. R., Boyd-Meredith, J. T., Chow, T. E., Rissman, J., & Wagner, A. D. (2015). Goal-directed modulation of neural memory patterns: Implications for fMRI-based memory detection. *Journal of Neuroscience*, *35*, 8531–8545.
- Vilberg, K. L., & Rugg, M. D. (2007). Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia*, *45*, 2216–2225.
- Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia*, *46*, 1787–1799.
- Vilberg, K. L., & Rugg, M. D. (2009). Left parietal cortex is modulated by amount of recollected verbal information. *NeuroReport*, *20*, 1295–1299.
- Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*, 445–453.
- Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*, 582–589.
- Wing, E. A., Ritchey, M., & Cabeza, R. (2015). Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *Journal of Cognitive Neuroscience*, *27*, 679–691.