Distributed Representations in Memory: Insights from Functional Brain Imaging

Jesse Rissman[1,3] and Anthony D. Wagner[1,2]

[1]Department of Psychology and [2]Neurosciences Program, Stanford University, Stanford, California, 94305; [3]Department of Psychology, University of California, Los Angeles, California 90095; email: rissman@psych.ucla.edu, awagner@stanford.edu

Annu. Rev. Psychol. 2012. 63:101–28

First published online as a Review in Advance on September 13, 2011

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi: 10.1146/annurev-psych-120710-100344

0066-4308/12/0110-0101$20.00

Keywords

episodic memory, working memory, encoding, retrieval, MVPA, multivariate

Abstract

Forging new memories for facts and events, holding critical details in mind on a moment-to-moment basis, and retrieving knowledge in the service of current goals all depend on a complex interplay between neural ensembles throughout the brain. Over the past decade, researchers have increasingly utilized powerful analytical tools (e.g., multivoxel pattern analysis) to decode the information represented within distributed functional magnetic resonance imaging activity patterns. In this review, we discuss how these methods can sensitively index neural representations of perceptual and semantic content and how leverage on the engagement of distributed representations provides unique insights into distinct aspects of memory-guided behavior. We emphasize that, in addition to characterizing the contents of memories, analyses of distributed patterns shed light on the processes that influence how information is encoded, maintained, or retrieved, and thus inform memory theory. We conclude by highlighting open questions about memory that can be addressed through distributed pattern analyses.

Contents

INTRODUCTION

There is broad consensus that the represented contents of a person's memories, as well as the cognitive processes that facilitate the formation, storage, and retrieval of these memories, depend on the coordinated activity of neural ensembles that are distributed across numerous cortical and subcortical brain regions (e.g., Eichenbaum & Cohen 2001, Fuster 2009, Jonides et al. 2008, Martin & Chao 2001, McClelland et al. 1995, McClelland & Rogers 2003, Schacter et al. 2007, Simons & Spiers 2003). Functional neuroimaging techniques, with their privileged capability of simultaneously measuring correlates of neural activity throughout the brain, have been productively applied to the study of learning and memory, supplementing and often extending the insights derived from lesion studies and neurophysiological recordings. The vast majority of this work has focused on localizing and functionally characterizing brain areas that support distinct aspects of our multifaceted mnemonic abilities (e.g., Badre & Wagner 2007, Binder et al. 2009, Carr et al. 2010, Davachi 2006, Ranganath 2006, Rugg & Yonelinas 2003, Wagner et al. 2005). While this research approach has shed considerable light on the differential contributions of distinct neural structures and networks to specific mnemonic operations, the past decade has witnessed the emergence of a powerful new approach for interrogating human brain function with functional neuroimaging. In particular, researchers have increasingly gained appreciation for the insights that can be gleaned by characterizing distributed activation patterns, rather than concentrating exclusively on peak regional effects. By leveraging novel statistical analysis techniques to extract the representational content of information-rich brain patterns (Haynes & Rees 2006, Norman et al. 2006, Tong & Pratte 2012), this approach has already advanced understanding of the neural and psychological mechanisms supporting memory, and it sets the stage for future discoveries.

This review aims to highlight ways in which pattern-based analyses of functional magnetic resonance imaging (fMRI) data have been utilized to capture and characterize the distributed neural representations that support human memory, as well as how leverage on these distributed representations has supported progress in addressing mechanistic questions about the workings of memory. We begin

by reviewing key neuroimaging findings that suggest that while particular categories and concepts often preferentially engage specific cortical regions over others, their neural representations are likely distributed and overlapping. We discuss how the ability to characterize elements of these distributed neural codes with fMRI has paved the way for a richer understanding of the cortical organization of perceptual and conceptual knowledge. Critically, these representations form the foundation of semantic memory—our database of accumulated factual knowledge about the world—and provide the building blocks for episodic memories—the contextually detailed records we store of specific life events. The ability to track the moment-to-moment activation state of such representations has proven a vital new tool with which to test theories of memory.

Given that theoretical accounts of memory generally posit an essential role for attentional control in the regulation of memory encoding, maintenance, and retrieval (e.g., Awh & Jonides 2001, Badre & Wagner 2007, Chun & Turk-Browne 2007, Mecklinger 2010, Race et al. 2009), we next review empirical demonstrations that an individual's goal state can serve to modulate the activation of distributed cortical representations associated with task-relevant and -irrelevant perceptual features or object categories. Many of the same general mechanisms that facilitate the top-down modulatory control of perception are likely central to the flexible goal-directed engagement of mnemonic processes (e.g., Rissman et al. 2009). We discuss recent experimental findings demonstrating that distributed neural populations in early visual processing areas are recruited in a targeted fashion to support the transient maintenance of relevant visual features, consistent with the hypothesis that short-term maintenance of perceptual content relies on the persistent activation of the same neural ensembles that support the perception of that content (e.g., Cowan 1993, Postle 2006, Ruchkin et al. 2003).

An analogous theoretical framework holds that long-term storage of episodic memories ultimately involves plastic changes in many of the same neural ensembles that were engaged during the initial processing of a given episode, with subsequent retrieval of episodic memories involving the reinstatement of these distributed neural codes, aided by the pattern competition mechanisms of the medial temporal lobe (for review, see Danker & Anderson 2010, Rugg et al. 2008). We consider how distributed pattern analyses have been exploited to (*a*) measure the activation of specific representational elements of an event memory, (*b*) track the cortical reinstatement of these representations during retrieval, (*c*) examine the intricate interplay between reactivation and subjective mnemonic experience, remembering and forgetting, and memory-based decision-making, and (*d*) test how the similarity of cortical patterns during encoding relates to later memory performance. This emerging line of research has helped elucidate the cascade of neural events that allow past experiences to influence present and future behavior. We conclude by highlighting open questions about the nature of memory that may be profitably examined through distributed pattern analyses.

BOLD: blood oxygenation level–dependent signal

DISTRIBUTED CORTICAL REPRESENTATIONS OF CATEGORIES AND CONCEPTS

Because our memories for events are partially built upon pre-existing cortical representations of perceptual and semantic features, we begin by selectively reviewing what functional neuroimaging has revealed about how such information is represented in the brain, emphasizing advances stemming from the application of analytical techniques for capturing the rich information represented within distributed blood oxygenation level–dependent (BOLD) fMRI activity patterns.

Characterizing the Cortical Activation Topography of Visual Object Categories

The field's efforts to use fMRI to examine putative distributed cortical representations began with a series of innovative studies by James

VTC: ventral temporal cortex

MVPA: multivoxel pattern analysis

Haxby and colleagues that provided evidence suggesting that the neural representations of stimuli from discrete visual object categories are more distributed and overlapping than previously thought (Haxby et al. 2001; Ishai et al. 1999, 2000). Haxby and colleagues hypothesized that, while certain patches of ventral temporal cortex (VTC) respond preferentially to individual visual categories, such as faces, houses, and chairs (see also Aguirre et al. 1998, Epstein & Kanwisher 1998, Kanwisher et al. 1997, Malach et al. 1995, Puce et al. 1995), the magnitude of the BOLD response observed in any given VTC voxel likely carries information about the degree to which the features represented by neurons within that voxel are present in the stimulus (for related data, see Martin & Chao 2001, Tanaka 1993). Accordingly, so long as exemplars within a category share more features with each other than they do with exemplars from different categories, then each visual category should have its own "neural signature"—a distributed VTC activation pattern that reflects the mean feature weightings for stimuli from the category. This framework allows for the existence of a virtually infinite number of category-specific cortical representations without the need to posit modularized cortical representations of individual categories.

Consistent with the distributed coding hypothesis, Haxby and colleagues (2001) demonstrated that by comparing the spatial correlation between VTC activity patterns measured during the perception of eight individual visual object categories, the category being viewed by an observer could be decoded with considerable accuracy. Importantly, decoding could succeed even when voxels with strong category preferences were excluded from analysis, indicating that information diagnostic of visual category is present in VTC well beyond focal category-selective regions. In addition, analyses restricted to voxels that responded maximally to a single category or a select group of categories also supported robust classification of the nonpreferred categories, suggesting that even focal regions that appear to show functional specialization for a given stimulus class may in fact contribute to the representation of other classes of visual stimuli (although see Spiridon & Kanwisher 2002 for an alternative perspective). Together, these results not only provided fMRI evidence for the distributed nature of visual object representations, but also served to foster appreciation for the rich, complementary information that can be garnered by characterizing activation "landscapes," relative to assessing the peaks and valleys of an activity map.

It did not take long before researchers began to apply more sophisticated multivariate pattern classification algorithms to the analysis of fMRI data—an approach that has become known as multivoxel pattern analysis or MVPA (see sidebar Multivoxel Pattern Analysis; Carlson et al. 2003, Cox & Savoy 2003, Haynes & Rees 2006, Mitchell et al. 2004, Norman et al. 2006). For example, Cox & Savoy (2003) used a support vector machine classifier to

MULTIVOXEL PATTERN ANALYSIS

Multivoxel pattern analysis (MVPA) typically begins with the division of each participant's fMRI data into training and test patterns, where "patterns" refer to brain activity measures extracted from those segments of fMRI data that one wishes to classify (e.g., individual time points or trial/block-specific activity estimates). Each training pattern is labeled as an example of a particular class. From the training patterns, the classifier formulates a model that can then be used to predict whether a new pattern (i.e., a test pattern) is likely to be an example of one class or another. In the model, some voxels are weighted more strongly than others, owing to their differential value in informing the classifier's predictions. To achieve stable results, the process of training and testing the classifier is typically repeated with different subsets of the total data set in an iterative fashion, known as cross-validation. The accuracy of the classifier's predictions provides an index of how robustly examples from different classes can be distinguished. Classification accuracy is often improved by limiting the number of voxels fed into the classifier (i.e., feature selection) since the inclusion of noisy or uninformative features can disrupt the classifier's ability to capture diagnostic patterns in the data.

achieve robust classification of the visual category of individual object stimuli, and they further demonstrated that the neural signatures of distinct categories are stable across scans collected more than a week apart. As with Haxby et al. (2001), Cox & Savoy also observed that the distributed activity patterns associated with certain visual object categories are more similar and hence more readily confusable with those of certain other categories; O'Toole et al. (2005) demonstrated that shared image-based attributes are a factor driving such neural similarity. Beyond distinguishing visual object categories, more recent studies have shown that activity patterns in the lateral occipital complex (LOC), an object-selective visual area just posterior to VTC, can facilitate classification of within-category exemplars, with these exemplar-level LOC representations generalizing across changes in stimulus size, location, and viewpoint (Cichy et al. 2011a, Eger et al. 2008). Other work has related distributed activation patterns in LOC to participants' judgments about the identity (Hsieh et al. 2010), category membership (Walther et al. 2009, Williams et al. 2007), or perceptual similarity (Haushofer et al. 2008, Weber et al. 2009) of viewed stimuli. Collectively, these studies illustrate how distributed pattern analyses provide a means to uncover neural representational structure and to relate neural representations to perception (see also Kriegeskorte et al. 2008).

Predicting Neural Representations of Perceptual and Semantic Content

Although decoding-based MVPA classification approaches have offered insight into the types of stimulus attributes that might be represented in distributed activity patterns (e.g., Kriegeskorte et al. 2008, O'Toole et al. 2005), they are inherently limited in their ability to characterize the underlying feature space. Moreover, despite their ability to infer aspects of a person's current experience from observed activity, decoding models typically lack the capacity to predict the activity patterns that should be associated with perceptual or cognitive experiences on which the classifier was not trained. Fortunately, the canonical MVPA-based decoding framework can be flipped around to allow for construction of theoretically guided forward prediction models (for review, see Naselaris et al. 2011). Generative classification approaches capture the relevant representational variables that mediate the mapping between stimuli and evoked activity patterns. Rather than simply decoding a finite set of states from a finite set of observed activity patterns, neural encoding models seek to learn the specific features represented within each voxel, and, in so doing, allow for the generative prediction of future activity patterns that should be associated with a potentially infinite number of stimuli.

LOC: lateral occipital complex

Two recent studies of visual processing illustrate the power and potential of neural encoding models. In the first, Kay and colleagues (2008) developed a predictive model of early visual encoding based on extant evidence that early visual areas represent at least three low-level visual dimensions—spatial position, spatial frequency, and orientation. Given that any visual stimulus, however complex, can be compactly represented as a set of Gabor wavelets that together reflect the stimulus' attributes along these three dimensions (Daugman 1985), Kay et al. trained a classifier to learn the mapping between this Gabor wavelet feature space and fMRI activity levels, estimating the Gabor feature weightings for each voxel in early visual cortex as participants viewed over 1,000 randomly selected natural images. The resulting model's predictive power was evidenced by its remarkably accurate ability to forecast patterns of fMRI activity associated with viewing individual natural image stimuli that were not part of the training set. In the second study, Naselaris and colleagues (2009) went a step further, demonstrating that they could reconstruct the image being viewed from the brain activity pattern it elicits. Again, their generative model operated on an intermediate, or latent, feature space rather than on the manifest Cartesian space of the two-dimensional images themselves (see also Brouwer & Heeger 2009; cf. Miyawaki et al. 2008, Thirion et al. 2006).

Importantly, it characterized the responses of (*a*) early visual cortex according to a structural encoding model (i.e., Gabor wavelets) and (*b*) higher-level visual regions according to a semantic encoding model, thus incorporating explicit priors regarding the structure and semantic content of natural images. The semantic model, based on a category-level designation of each image, explained over half the variance in the voxel activity levels observed in anterior occipital cortex, and its inclusion dramatically improved the likeness of the reconstructed images to the observed images.

Generative classifier models have also been applied to characterize more abstract conceptual representations. For instance, the neuroscientific study of lexical semantics has centered on understanding the brain's scheme for interpreting what the words of a language denote. In a pioneering study, Mitchell and colleagues (2008) trained an encoding classifier to learn the mapping between whole-brain fMRI activity patterns associated with a set of concrete nouns and a latent feature space derived from the semantic properties of the nouns. In particular, guided by empirical and theoretical work suggesting that semantic representations of concrete entities are heavily linked to their sensorimotor attributes (Barsalou 2008, Farah & McClelland 1991, Martin & Chao 2001), Mitchell and colleagues constructed their model's semantic feature space around 25 verbs of perception and action. Each noun was then assigned a set of semantic feature weights based on the frequency of its textual co-occurrence with each verb. Impressively, by learning the neural correlates of these intermediate semantic features, the model could predict the future activation patterns elicited by test nouns. Moreover, the large-scale brain activity patterns that characterized each of the semantic features illustrated the highly distributed nature of conceptual representations while also revealing a number of focal regions that appeared to play a differential role in the representation of specific features (e.g., an area of right superior temporal sulcus often associated with the processing of biological motion was strongly linked to the semantic features for the verb "run," whereas a putative gustatory cortex area was associated with "eat"). In a subsequent study, Just and colleagues (2010) used a bottom-up factor analysis approach to reveal the semantic feature space (rather than relying on a preselected set of verb-based features), parsimoniously accounting for the multidimensional structure of noun-related activity patterns. Furthermore, they found that a classifier model trained on fMRI activity patterns from one group of participants could predict the activity patterns elicited by novel nouns read by another group of participants, suggesting that the neural organization of coarse-level semantic representations is partially shared across individuals (see also Chang et al. 2011; Pulvermüller et al. 2009; Shinkareva et al. 2008, 2011).

Taken together, the studies reviewed thus far illustrate how aspects of a person's perceptual experience and semantic cognition can be reliably decoded, or even reconstructed, from distributed fMRI activity patterns. We next consider how distributed pattern analyses have been used to test theories of attention and working memory.

ATTENTION, WORKING MEMORY, AND DISTRIBUTED CORTICAL REPRESENTATIONS

During everyday experiences, we frequently find ourselves bombarded with many more stimuli than we can simultaneously process. To be effective, we often must selectively attend to the subset of stimuli or stimulus features that are most relevant to our goals, using top-down control to regulate the processing of environmental stimuli based on current attentional priorities. At the neural level, considerable evidence indicates that the cortical representations of goal-relevant stimuli or stimulus features are up-regulated and/or sharpened, whereas representations of irrelevant stimuli/features are suppressed (e.g., Desimone & Duncan 1995, Gazzaley et al. 2005, Kastner & Pinsk 2004). In addition to regulating stimulus processing, top-down attentional processes also support the

generation and maintenance of mental images, with mental imagery serving to activate many of the same cortical regions that are involved in bottom-up stimulus processing (e.g., Kosslyn 2005, O'Craven & Kanwisher 2000). Likewise, the ability to maintain recently encountered stimuli in working memory (WM) is thought to depend on cortical regulation by top-down attentional control. In this section, we review some of the methodological strategies and key results that have emerged from research on the goal-directed attentional modulation of distributed fMRI activity patterns, beginning with the effects of attention during online stimulus processing and mental imagery, and then turning to studies of WM. Because the encoding and retrieval of representations in episodic memory are also modulated by top-down control (e.g., Race et al. 2009), many of the findings reviewed here are directly relevant to our later discussion of distributed representations in episodic memory.

Attentional Influences on Distributed Cortical Patterns

In the first fMRI study to apply MVPA techniques to examine the influence of goal-directed attention on distributed sensory representations, Kamitani & Tong (2005) examined whether the focus of attention—directed toward one of two superimposed oriented line gratings—could be decoded from distributed brain activity patterns measured from early visual areas (see also Haynes & Rees 2005). They reasoned that if different line orientations are associated with distinct neural signatures, then it should be possible to track the activation state of neural ensembles associated with a given line orientation and use this information to infer the degree to which an observer is allocating attention to that particular orientation. Indeed, Kamitani & Tong (2005) demonstrated that when subjects viewed a single oriented line grating, the elicited activity patterns in individual visual areas, including areas V1–V4, contained sufficient information to facilitate orientation decoding. Subsequently, to evaluate the influence of attention on these distributed neural representations, participants were scanned while viewing a "plaid" stimulus composed of two overlapping orthogonally oriented line gratings, one of which was cued to be task relevant. Critically, an MVPA classifier initially trained to differentiate the neural signatures of the two line orientations when each was presented alone was also able to decode which of the two line orientations was being attended when the stimuli were concurrently displayed. Distributed information about the attended orientation was present even at the earliest cortical level of visual processing (V1). Thus, despite equivalent bottom-up input, attentional signals served to bias neural patterns in favor of the task-relevant stimulus/feature (see sidebar Decoding Cortical Columns or Larger-Scale Maps?).

WM: working memory

DECODING CORTICAL COLUMNS OR LARGER-SCALE MAPS?

Kamitani & Tong (2005) hypothesized that fMRI-based orientation decoding capitalizes on slight biases in the distribution of orientation-tuned cortical columns within each voxel. The seemingly random spatial variance in fine-scale columnar architecture and its supporting microvasculature was posited to lead individual voxels in visual cortex to exhibit weak but consistent orientation tuning, which could be exploited by a pattern classifier supplied with activation states from many such voxels. This conjecture has been challenged by data indicating that V1 contains a topographic map of orientation at a much coarser spatial scale than previously realized (Freeman et al. 2011). This may explain why modest spatial smoothing of V1 activity patterns has little detrimental effect on orientation decoding (Freeman et al. 2011, Op de Beeck 2010; but see Kriegeskorte et al. 2010, Swisher et al. 2010). Moreover, the close correspondence between cortical maps of orientation and radial position could imply that previous demonstrations of orientation decoding were in fact capturing neural correlates of observers' preferential attention to positions along the long-axis of oriented gratings. Although the role of radial bias in orientation decoding remains a point of contention, the fact that reliable orientation information can be extracted from BOLD activity patterns in early visual cortex nevertheless provides a valuable means to investigate the neural substrates of visual attention and WM.

Subsequent studies have documented the attentional modulation of distributed cortical patterns across a variety of low-level and high-level stimulus materials, ranging from simultaneously presented motion fields (Kamitani & Tong 2006, Liu et al. 2011) to simultaneously presented visual objects (Macevoy & Epstein 2009, Reddy & Kanwisher 2006). Moreover, it is not only possible to decode which of multiple stimuli is currently being attended, but also what aspect of a given stimulus is being attended. For instance, distributed fMRI patterns across face-selective voxels in the fusiform and occipital cortices can be used to decode whether participants are preferentially attending to the race or the gender of a face (Chiu et al. 2011). Taken together, these studies provide powerful evidence that attentional priorities and expectations sculpt the distributed neural representations of visual stimuli, even at very early stages of cortical processing.

Researchers have also leveraged MVPA methods to decode the subjective contents of visual imagery, with initial results largely supporting prior univariate fMRI studies that demonstrate that self-generated mental images depend on the recruitment of the same neural populations that support stimulus perception (Kosslyn 2005). For instance, after training an MVPA classifier to differentiate the distributed cortical patterns associated with perception of the letters "X" and "O", Stokes and colleagues (2009, 2011) showed that the classifier could also succeed at decoding participants' imagery of these particular letters. Likewise, the category of imagined objects can be decoded from the same VTC voxel patterns that are engaged during the perception of stimuli from these categories (Cichy et al. 2011b, Reddy et al. 2010), and MVPA techniques can even reconstruct a coarse visual representation of what a participant is currently imagining based on fMRI activity patterns in retinotopic cortex (Thirion et al. 2006). Collectively, these studies of attention and mental imagery shed light on the specificity with which distributed cortical representations that support stimulus perception can be modulated by top-down attentional signals in a goal-directed fashion. This insight has provided powerful leverage on the nature of mnemonic representations and a means to exploit these representations to test mechanistic models of working memory.

Distributed Representations in Working Memory

One critical way in which memory serves as a bridge between our past and present is through the transient maintenance of just-experienced or just-retrieved stimuli. By allowing behaviorally relevant representations to remain active across brief intervals of time, WM facilitates a host of complex cognitive abilities (Baddeley 1992). From one theoretical perspective, recently dubbed the "sensory recruitment model" of WM (Serences et al. 2009), WM does not depend on neural systems specialized for transient memory maintenance; rather WM is an emergent product of sustained interactions between top-down control signals and neural representations of perceptual, conceptual, linguistic, affective, or other stimuli (e.g., Cowan 1993, D'Esposito 2007, Postle 2006, Ruchkin et al. 2003). In this model, the sustained allocation of attention to the neural ensembles (or a subset thereof) that are engaged during the neural encoding of encountered or retrieved stimuli serves to actively maintain these representations. The sensory recruitment model contrasts with the influential theoretical proposal that short-term maintenance involves the transfer of relevant stimulus representations to one or more dedicated storage buffers, putatively in prefrontal and/or parietal cortices (e.g., Baddeley 1992). From this WM systems perspective, the actively maintained neural representations of stimuli are distinct from those encoded during initial stimulus processing.

Some empirical support for the sensory recruitment model derives from demonstrations that during WM delay periods there is persistent firing of stimulus-selective VTC neurons (e.g., Fuster & Jervey 1981, Miyashita & Chang 1988) and sustained fMRI activation in sensory cortical areas thought to differentially represent

the maintained stimuli (e.g., Postle et al. 2003). Other data also suggest that the transient maintenance of neural representations in sensory regions involves top-down support from prefrontal and/or parietal cortices, presumably in the form of active neural communication between these regions (e.g., Fuster 2009, Gazzaley et al. 2004). That said, many fMRI studies that have reported sustained BOLD activity in sensory cortex during the delay period of WM tasks have documented relatively weak signal levels in these regions compared to the strong signals evoked during the stimulus encoding and decision stages of the tasks. Although low-amplitude BOLD activity during WM delays does not necessarily rule out a role for sensory areas in short-term maintenance of visual representations (see Rissman et al. 2004), traditional univariate fMRI analyses have been limited in their ability to relate delay period BOLD activity in sensory areas to the maintenance of specific stimuli or features.

The ability of MVPA techniques to sensitively index the activation state of distributed neural representations of specific stimuli in sensory cortex has caught the attention of researchers interested in delineating the structure and neural substrates of WM. For example, two contemporaneous fMRI studies exploited MVPA methods to directly test the sensory recruitment model of WM (Harrison & Tong 2009, Serences et al. 2009; **Figure 1**). In Harrison & Tong's (2009) study, each WM trial began with the sequential presentation of two distinct orientation gratings, followed by a cue indicating whether the task was to maintain the first or the second grating across a subsequent 11-second delay. Following the delay, participants were presented a third unique grating and judged which direction it was rotated relative to the maintained grating. Strikingly, although BOLD signal levels in visual cortex fell dramatically after stimulus encoding, analyses of the activity patterns in V1–V4 during the delay period revealed that there was sufficient information, within each visual region and temporally extended across the entire delay period, to accurately decode the content of WM. Importantly, since participants were cued as to which grating they should maintain only after both gratings had offset, the diagnostic brain activity patterns measured during the delay period were not attributable to residual hemodynamic responses evoked during stimulus encoding. That is, top-down influences of attention must have acted upon a stimulus-specific neural representation, maintaining the representation over the delay. Moreover, a classifier trained on fMRI data that captured purely stimulus-driven neural responses to each grating was subsequently able to successfully generalize its orientation predictions when applied to the delay period data from the WM task, providing further support for the sensory recruitment model.

Serences and colleagues (2009) also demonstrated that the orientation of a maintained line grating could be reliably decoded from delay period activity patterns in early visual cortex. In their experiment, the orientation gratings were presented on colored backgrounds, with task cues indicating whether the grating or the color hue should be maintained. MVPA revealed that delay period activity patterns only contained diagnostic information about the relevant stimulus dimension—when orientation was relevant, the classifier achieved above-chance decoding of orientation but not of color, with the converse being true when color was relevant. Moreover, delay period decoding was more robust when based on voxel patterns from V1, relative to those from later visual areas, suggesting that maintenance-related delay period activity can manifest itself at the earliest cortical stage of visual processing (though the experimental design left open the possibility that classification was partially based on the residual hemodynamic effects of attentional modulation that took place during stimulus encoding).

Harrison & Tong's (2009) and Serences and colleagues' (2009) data provide powerful demonstrations that, despite low signal amplitudes, sustained BOLD activity patterns associated with WM maintenance resemble the activity patterns associated with the bottom-up perception of the same stimuli, suggesting

PFC: prefrontal cortex

that the neural representations that support online sensory processing are also actively maintained in WM over delays (rather than being transferred to a separate WM buffer). Further extending this conclusion, Ester and colleagues (2009) showed that the orientation of lateralized gratings can be decoded not only from delay period activity in contralateral visual areas involved in the initial perception of the gratings, but also from ipsilateral visual areas. This suggests that sensory recruitment during visual WM maintenance may extend well beyond the retinotopic representation of the stimulus, with involvement of ipsilateral cortices potentially serving to bolster the fidelity of the maintained representation by incorporating additional feature-selective neural ensembles into the total pool of neurons operating in support of stimulus retention.

MVPA has also been used to investigate the active maintenance of distributed representations of content retrieved from long-term memory (Lewis-Peacock & Postle 2008). In this study, participants initially learned arbitrary cue-associate pairings of stimuli from three visual categories (faces, locations, and common objects) and were then scanned while recalling the learned associate of a given cue and maintaining this representation over an 11-s delay period. Critical associative pairs consisted of stimuli from two distinct classes (e.g., a face-location association), and MVPA examined BOLD activity patterns relating to neural representations of the presented cue and the retrieved associate. Importantly, when a classifier trained (on independent data) to differentiate between the neural patterns associated with the three stimulus categories was subsequently tested on the delay period data, it revealed relatively sustained activation of cortical patterns tied to the visual classes of both the cue and the retrieved associate, as compared with patterns tied to the third (irrelevant) stimulus class. The presence of sustained associate-related neural patterns documents the maintenance of internally generated (i.e., retrieved) representations that prospectively anticipate future events (Bar 2009, Schacter et al. 2007). Moreover, this study revealed that stimulus category-selective cortical patterns were widely distributed, extending from sensory cortical areas to prefrontal cortex (PFC). However, despite the presence of diagnostic voxels in PFC, the classifier achieved similar success when PFC voxels were excluded, but failed to yield above-chance decoding when exclusively trained on PFC voxels. These data further suggest that WM representations are not exclusively maintained within a PFC-mediated storage buffer. Rather, WM appears to depend on the targeted and sustained activation of cortical representations tied to the distinguishing features of the relevant memoranda.

It is important to note that no study to date has established a direct link between the sustained engagement of stimulus-selective cortical activity patterns and WM behavioral performance. To the extent that distributed cortical representations of stimuli are actively maintained to support goal-directed behavior that bridges short delays between perception and action (e.g., Fuster 2009), then one would expect the fidelity of these cortical representations to be closely related to participants' accuracy and/or response times on the WM tasks. Future studies, perhaps using challenging WM tasks that are structured to provide sensitive behavioral assays of performance (e.g., Curtis et al. 2004), may ultimately provide compelling evidence that delay period activity patterns support behavior, ruling out the possibility that they are an epiphenomenal consequence of back-propagating neural feedback from higher-level areas.

Decoding Putative Top-Down Control Signals in Frontoparietal Cortex

The studies of attention and WM discussed thus far have primarily been concerned with documenting the consequences of top-down control processes on the activation state of neural ensembles within posterior perceptual cortices. MVPA techniques can also be leveraged to gain insights into the putative

frontoparietal sources of these regulatory control signals. For instance, information about an individual's current attentional priorities can be extracted from fMRI activity patterns within dorsal regions of the frontal and parietal lobes. Whereas these regions have been commonly associated with the control of spatial attention and action intention (e.g., Bisley & Goldberg 2010, Corbetta & Shulman 2002), recent MVPA results have suggested that these areas may also support nonspatial feature-based attention, such as specifying which color or which motion direction happens to be relevant on a given trial (Liu et al. 2011) or specifying whether the processing of a face's gender should be prioritized over its race (Chiu et al. 2011). The role of these frontoparietal structures may also extend to the specification and maintenance of more abstract task sets, such as representing which stimulus-response mapping scheme (Bode & Haynes 2009, Woolgar et al. 2011), perceptual categorization rule (Li et al. 2007), or mathematical operation (Haynes et al. 2007) should be applied at a given moment in time.

Beyond examining the degree to which frontoparietal activity patterns reflect the neural coding of specific attentional priorities and/or task set configurations, researchers have used MVPA to identify activity patterns associated with the act of shifting one's attention between aspects of environmental stimuli or between representations held in WM (Esterman et al. 2009, Greenberg et al. 2010, Tamber-Rosenau et al. 2011). Although several frontal and parietal lobe structures exhibited activity patterns that could decode select types of attentional shifts, these studies converged in implicating the medial superior parietal cortex as playing a domain-general role in the transient reconfiguration of one's attentional set. We anticipate that further applications of MVPA to the study of attentional control and WM will serve to strengthen mechanistic understanding of how frontal and parietal cortical regions interact to specify current attentional priorities, to update these priorities as needed, and ultimately to modulate the activation state of neuronal ensembles that represent goal-relevant (or irrelevant) features.

INFORMATION CODING WITHIN THE HUMAN MEDIAL TEMPORAL LOBE

MTL: medial temporal lobe

PHC: parahippocampal cortex

PRC: perirhinal cortex

ERC: entorhinal cortex

As the preceding sections illustrate, MVPA techniques have provided unique leverage on the cortical representations of sensory features, perceptual categories, semantic content, and other higher-level cognitive states. With respect to memory theory, application of distributed pattern analyses has yielded compelling evidence in favor of the sensory recruitment model of WM. Given the demonstrated power of these techniques for revealing characteristics of neural representations, recent work has extended MVPA to test hypotheses regarding information coding in the human medial temporal lobe (MTL). By acquiring fMRI data with a higher spatial resolution than that afforded by standard whole-brain imaging parameters (see Carr et al. 2010), extant studies have attempted to characterize fine-grained voxel activity patterns within the specific anatomical subregions that comprise the MTL, including the hippocampus (dentate gyrus, CA1, CA3, and subiculum) and surrounding MTL cortical areas [parahippocampal cortex (PHC), perirhinal cortex (PRC), and entorhinal cortex (ERC)]. Much as researchers have investigated the representational structure of specific visual areas by determining the types of features that can be decoded from each area's distributed fMRI activity patterns, the application of MVPA methods to high-resolution MTL data has begun to yield insights into how event content is coded in distinct MTL subregions.

Illustrative of the approach, Diana and colleagues (2008) used MVPA to evaluate the hypothesis that certain MTL regions—specifically, the hippocampus and PHC—are selectively tuned to the representation of spatial information (Burgess et al. 2002, Epstein & Kanwisher 1998, O'Keefe & Nadel 1978), whereas other MTL regions—specifically,

PRC—are selectively tuned to the representation of complex visual objects (Bussey & Saksida 2007). In their high-resolution (hr-fMRI) study, pattern classification analyses were applied to hippocampal, PHC, and PRC data acquired while participants viewed stimuli from five categories (scenes, faces, toys, other common objects, and abstract shapes). Given that visual scenes inherently contain more spatial features than stimuli from the other four categories, neural structures that are highly specialized for topographical representation of space should differentiate scene from nonscene stimuli, while showing minimal sensitivity to the distinctions between the nonscene categories. In contrast to this prediction, however, Diana et al. (2008) observed that PHC activity patterns reliably distinguished between all five stimulus categories and that, even when scenes were excluded from analysis, above-chance decoding of the four nonscene visual categories was achieved. Beyond PHC, above-chance decoding was not observed when analyzing activity patterns in the hippocampus or PRC. Although these latter null results should be cautiously interpreted (see Preston et al. 2010 for hr-fMRI data demonstrating face and scene novelty and subsequent memory effects in PRC), the successful decoding of visual categories based on PHC activity patterns suggests that distributed neural representations within PHC carry information that distinguishes between multiple visual categories. At the same time, it should be emphasized that the features coded by PHC neural ensembles remain a subject of debate (Bar et al. 2008, Epstein 2008).

Although Diana et al. were unable to decode the viewing of complex scenes relative to other visual categories from hr-fMRI activity patterns in human hippocampus, recent hr-fMRI data indicate that it is possible to decode which of two complex scenes is being viewed based on distributed BOLD signals in the hippocampus (as well as in ERC and PHC) (Bonnici et al. 2011). Moreover, extensive neurophysiological data in rodents (Moser et al. 2008) and recent intracranial electrocorticography data in humans (Ekstrom et al. 2003) have revealed and characterized hippocampal "place cells" that are selectively tuned to specific environmental locations. Whereas the prevailing view from nonhuman animal work is that place cells are uniformly distributed throughout the hippocampus, without local anatomical asymmetries in location-selective tuning (e.g., Redish et al. 2001), Hassabis and colleagues (2009) examined whether it is possible to predict an individual's location within a virtual-reality environment based on distributed hr-fMRI activity patterns from human MTL. In this study, participants navigated two unique rooms, each consisting of four target positions. Decoding analyses were conducted using a "searchlight analysis" approach (Kriegeskorte et al. 2006), whereby a series of MVPA classifiers were serially trained and tested on the activation patterns within small spherical clusters of voxels, allowing evaluation of the representational content of relatively focal brain regions. The analyses revealed activation clusters within the posterior hippocampus that supported above-chance classification of an individual's location within a room and activation clusters within PHC that supported differentiation between the two rooms. Univariate analyses, on the other hand, failed to reveal activity differences associated with specific locations or rooms. From these results, the authors suggested that the ability of the hippocampus to discriminate individual locations within a room may reflect its role in the representation of an allocentric cognitive map of the room's layout, whereas PHC may extract contextual information from each room. Moreover, it was argued that the ability to decode spatial location from hr-fMRI data challenge the proposal that place cells are uniformly distributed, raising the possibility that location-selective hippocampal neurons in the human have sufficient consistency in their anatomical distribution to permit reliable location preferences to emerge at the voxel level.

It should be noted, however, that Hassabis et al. (2009) did not directly evaluate whether hr-fMRI was critically necessary to successfully decode location from MTL activity

patterns, and thus it is unclear whether their analyses exploited fine-scale irregularities in the distribution of location-selective neurons or whether information about a person's location was coded at a coarser scale. Bearing on this issue, a recent fMRI study by Rodriguez (2010), which also used a virtual navigation task, revealed that standard-resolution fMRI activity patterns in the hippocampus could be used to predict in which of four locations a participant was currently located. Given the increased coarseness with which hippocampal BOLD activity was sampled in the Rodriguez study, it is possible that location decoding in both studies relied not on hippocampal maps of allocentric space, but rather on more abstract hippocampal representations of the visuosemantic qualities and/or internally generated verbal labels associated with each goal location in the virtual environments. Future work will be needed to critically examine whether hr-fMRI affords advantages for location-based decoding, and if so, what this indicates about the nature of the underlying MTL representations.

Taken together, the preceding studies highlight ways in which MVPA techniques provide leverage on the nature of information coding in specific MTL subregions. At the same time, these studies do not address whether distributed MTL patterns can be used to differentiate between complex individual events. Promising new data from Chadwick and colleagues (2010) suggest that distributed analyses of hr-fMRI data from the MTL may ultimately enable decoding of rich episodic memories. In this experiment, participants recalled one of three brief movie clips on each retrieval trial (each clip had been viewed prior to scanning); the resulting fMRI data were submitted to searchlight classification analysis. Impressively, activity patterns within the hippocampus, ERC, and parahippocampal gyrus each independently supported above-chance decoding of the retrieved episode, demonstrating that, with sufficient variance in the perceptual and/or semantic content of events, MTL voxel patterns contain information that differentiates between complex episodes (**Figure 2**). Although it remains to be seen whether the decoding of rich, multiattribute event memories is further facilitated by simultaneously considering MTL activation patterns along with distributed patterns in cortical and subcortical structures beyond the MTL, Chadwick et al.'s (2010) findings suggest content-based biases in the distributed coding of event memories in the MTL.

Given these initial successes in information decoding from human MTL, it is important to emphasize that future work is needed to determine whether and how distributed MTL activity patterns are linked to behavioral performance on tasks requiring category discrimination and spatial navigation, as well as those assaying memory encoding, consolidation, and retrieval. We also anticipate that MVPA techniques will ultimately provide leverage on the role of hippocampal subregions in pattern separation (i.e., creating distinctive neural codes for highly similar events) and pattern completion (i.e., retrieving multiple event details associated with a partial cue). For instance, using MVPA to quantify the representational similarity of neural patterns elicited by pairs of events (e.g., Kriegeskorte et al. 2008) could provide a measure of how the two events are represented within distinct components of the hippocampal circuit, such as the CA3 and CA1 subfields (for initial MVPA findings bearing on pattern separation within MTL, see Bonnici et al. 2011).

DISTRIBUTED REPRESENTATIONS IN EPISODIC MEMORY

Beyond providing leverage on information coding within the MTL, distributed pattern analyses have yielded new insights into the psychological and neural processes supporting the encoding and retrieval of episodic memories. In this section we review how distributed pattern analyses have been used to (*a*) measure the activation of specific representational elements of an event memory, (*b*) track the cortical reinstatement of these representations during retrieval, and (*c*) examine the intricate interplay between reactivation and subjective mnemonic

experience, remembering and forgetting, and memory-based decision-making. MVPA methods have also shed light on how the similarity of across-event encoding patterns relate to later memory performance. As we emphasize, the ability of distributed analyses to quantify the strength of cortical representations, as well as the similarity between representations, has provided novel purchase on central theoretical questions.

Cortical Reinstatement and Event Recollection

In the first MVPA study of episodic memory, Polyn and colleagues (2005) tested two critical predictions of the contextual reinstatement hypothesis of memory retrieval (Tulving & Thompson 1973)—namely that the act of recalling an event from memory involves the targeted reactivation of stored representations of the properties (attributes) of the event, which, in turn, serve as additional cues that guide and constrain subsequent mnemonic searches. Some support for the first prediction has come from fMRI studies, implementing univariate analyses, demonstrating that the cortical regions active during episodic retrieval tend to mimic those active during event encoding, suggesting that retrieval is associated with the reinstatement of cortical representations that were present during event encoding (e.g., Danker & Anderson 2010, Kahn et al. 2004). For instance, regions of auditory and visual sensory cortex are respectively reactivated during the cued retrieval of auditory and visual memories (Nyberg et al. 2000, Wheeler et al. 2000). Polyn and colleagues (2005) expanded upon this earlier work, using MVPA to index the engagement of content-sensitive cortical activation patterns during encoding and then examining the reemergence of these cortical patterns during recall. Importantly, to the extent that the reactivation of encoding-related activity patterns constitutes neural evidence for the psychological construct of contextual reinstatement, Polyn and colleagues further predicted that cortical reactivation would temporally precede the recall of items from memory. In their experiment, participants were scanned while encoding famous faces, famous locations, and common objects, and subsequently freely recalling the names of as many items as possible. Based on the fMRI encoding data, a classifier was trained to characterize the activity patterns that distinguished the three stimulus categories; subsequently, the classifier quantified the re-engagement of these category-sensitive activity patterns during recall. Strikingly, the results indicated that the cortical activity pattern associated with the encoding of a particular stimulus category was reactivated prior to participants' behavioral expressions that they had successfully retrieved an exemplar from the category (for related single-unit neurophysiology data, see Gelbard-Sagiv et al. 2008). Although these results do not specify what attributes were reinstated, nor whether reinstatement depended on strategic processes, the temporal dynamics of the observed cortical reactivation is consistent with the hypothesis that subsequent retrieval depends upon the internal generation of effective retrieval cues.

Given the demonstration that cortical reinstatement accompanies event recall, Johnson and colleagues (2009) sought to determine whether reinstatement is a specific marker of event recollection or whether reinstatement also occurs during familiarity-based recognition decisions. To do so, participants were scanned as they encoded visual words under one of three orienting task contexts. During a subsequent recognition test, participants indicated whether they "remembered" details surrounding each word's encoding presentation or, absent the experience of "remembering," participants indicated their confidence that the item was old or new; the latter responses were argued to reflect recognition decisions based on gradations in item familiarity (e.g., Yonelinas et al. 2005). During analysis, a classifier was trained to distinguish the activity patterns associated with each of the three encoding contexts and then applied to the retrieval data. Consistent with Polyn et al., Johnson and colleagues observed robust cortical reinstatement

during "remembered" items, as revealed by the classifier's ability to decode the item's encoding context from the retrieval data. Importantly, the classifier also demonstrated above-chance context decoding for test items recognized as old but for which participants were unwilling to respond "remembered."

Based on this latter finding, Johnson et al. (2009) argued that the cortical reinstatement of contextual details is not sufficient to produce the subjective experience of recollection, which could have important implications for psychological and neural theories of recognition memory (Eichenbaum et al. 2007, Mayes et al. 2007, Wixted & Mickes 2010, Wixted & Squire 2011, Yonelinas et al. 2010). For example, it has been argued that recollection-based recognition depends on pattern completion processes, whereas familiarity-based recognition depends on pattern matching between retrieval cues and stored representations (e.g., Gonsalves et al. 2005, Norman & O'Reilly 2003). However, to the extent that cortical reinstatement subserves familiarity-based recognition, this would suggest that familiarity also depends, at least in part, on pattern completion. It should be noted, though, that subjective reports of the bases for recognition decisions likely depend on a signal-detection decision process, whereby the amount of recollected event details is weighed relative to an internally calibrated decision threshold (Dunn 2008, Rotello et al. 2004, Wixted & Mickes 2010). Although Johnson et al. encouraged participants to adopt a lenient threshold for warranting a "remembered" response, it is possible that participants subjectively experienced some amount of recollection even on those trials for which they ultimately reported recognition in the absence of "remembering." Although future work is needed to determine whether cortical reinstatement contributes to pure familiarity-based recognition decisions, Johnson et al. illustrate how MVPA methods, by providing an index of cortical reinstatement, can provide unique leverage on pressing, and long-debated, theoretical issues. Their approach also sets the stage for investigating the ways in which frontoparietal circuits "read out" retrieved mnemonic evidence, integrating this evidence to guide memory-based decisions (e.g., Dobbins et al. 2002, Donaldson et al. 2010, Wagner et al. 2005).

Decoding Mnemonic States

Although the preceding studies focused on measuring cortical reinstatement during retrieval, other studies have used MVPA techniques to characterize the neural signatures of distinct cognitive states associated with memory retrieval. For instance, Quamme and colleagues (2010) examined the neural processes that support the psychological construct of "listening for recollection"—an internally directed attentional state posited to promote recollection of event details and bias mnemonic decision-making toward the reliance on recollected details over perceived familiarity. To this end, a classifier was trained to distinguish between neural signatures of familiarity-oriented versus recollection-oriented retrieval and was then used to index the relative engagement of the two retrieval orientations during individual trials from an independent retrieval task. The latter task required participants to differentiate old items from highly similar lures (e.g., a plural version of a word that had initially been studied in the singular form; Hintzman et al. 1992) and thus emphasized the recollection of event details, since targets and similar lures would both elicit familiarity. Strikingly, a searchlight MVPA approach revealed a region of right inferior parietal cortex that exhibited a prestimulus activity profile consistent with a putative role in "listening for recollection." Moreover, increased engagement of this recollection-related activity pattern was associated with reduced false recognition of the similar lures, suggesting that this region plays a role in promoting detailed episodic retrieval or in biasing attention toward recollected content during decision-making. Although right-lateralized parietal activation is not commonly reported in univariate fMRI studies of episodic retrieval (e.g., Wagner et al. 2005), Quamme et al.'s innovative methodological

approach highlights a promising avenue for investigating how goal-specific attentional states gate retrieval and influence the weighing of evidence during memory-based decisions.

Quamme et al.'s (2010) study is grounded by a rich behavioral literature documenting the active nature of episodic retrieval. Indeed, extensive evidence indicates that retrieval goals can render certain features of a past experience more relevant than others, with attentional processes serving to enhance the processing of prioritized content (e.g., Jacoby et al. 2005). For example, during source memory retrieval, people can adopt single-agenda or multiagenda source monitoring strategies (Johnson et al. 1993)—the former emphasize monitoring of a single source detail (e.g., deciding whether a stimulus was studied in a particular source context), whereas the latter emphasize monitoring of multiple potential sources. Recently, McDuff et al. (2009) had participants perform a source retrieval task that emphasized either single- or multiagenda monitoring; in both conditions, a particular source was defined as the "target" source to which participants were to respond "yes." When a classifier that had been trained to differentiate between three distinct encoding contexts was applied to the retrieval data, cortical reinstatement of the target source context was revealed to be more robust during single- relative to multiagenda source monitoring. This outcome suggests that retrieval processes were focused on recovering information related to the target source. In contrast, cortical reinstatement of the actual source context (i.e., when it diverged from the target source), although robust in both conditions, was selectively associated with participants' behavioral performance during multiagenda monitoring. Along with Quamme et al.'s data, these findings indicate that a person's retrieval goals influence the probability that cortical representations of encoded details will be reinstated, as well as the manner in which reinstated details are weighed during mnemonic decision-making.

In related work, Rissman and colleagues (2010) used MVPA to examine whether the mnemonic states of recollection, graded item familiarity, and perceived novelty are associated with distinguishable activity patterns, and whether the emergence of these patterns depends on retrieval orientation (explicit versus implicit). Using fMRI data from a face memory task, separate MVPA classifiers were trained to identify activity patterns associated with subjective recognition states (irrespective of memory accuracy), as well as activity patterns that might reveal an item's true old/new status (irrespective of subjective recognition). Analyses revealed a remarkably accurate ability to classify whether a given face was subjectively experienced as old or new, as well as whether recognition was associated with vivid recollection, or a strong versus weak sense of familiarity (**Figure 3**). Perhaps most strikingly, a participant's subjective memory state could be decoded from her/his brain patterns even when using a classifier that had been trained on brain patterns from other participants, suggesting a high degree of neuroanatomical consistency across individuals and a relatively coarse coding of the cortical patterns associated with perceived oldness and novelty. In contrast to this robust classification of subjective memory states, the ability to decode whether or not a particular face had actually been previously experienced was rather limited (when controlling for subjective memory state or when participants adopted an implicit retrieval orientation); for example, discrimination between true and false recognition was only modestly above chance. Moreover, whereas distributed activity patterns in frontal, parietal, and MTL areas provided highly diagnostic information about subjective memory states, the ability to distinguish true from false recognition was limited to perceptual cortical regions (e.g., fusiform cortex), consistent with univariate data suggesting that true and false memories often differ most in their perceptual qualities (Schacter & Slotnick 2004).

Rissman et al.'s (2010) results have implications for memory theory and for possible forensic extensions of fMRI-based MVPA memory decoding. First, from a neuroscientific perspective, classifier-derived "importance maps" revealed that widely distributed and

coarsely coded neural patterns in frontal, parietal, occipitotemporal, and MTL regions putatively underlie the subjective experiences of novelty, familiarity, and recollection. Second, the finding that mnemonic classification performance was substantially diminished when test probes were processed under an implicit retrieval orientation (i.e., when participants were not instructed to reflect on their memories for the faces) further emphasizes the profound influence that goal states exert on mnemonic retrieval processes. Third, from a forensic perspective, these data highlight the potential power of distributed fMRI analyses for decoding a person's recognition of specific stimuli while raising concerns about whether these methods are adequate to uncover a person's true experiential history.

Beyond Single-Event Memory

Memory for the past is often reinstated during the encoding and retrieval of subsequent events that share attributes (i.e., overlap) with the past event (O'Reilly & McClelland 1994). Recently, such reinstatement has been argued to foster the building of integrative multievent representations that support across-event generalization (Shohamy & Wagner 2008) and protect memories from interference-driven forgetting (Kuhl et al. 2010). Exploiting the ability of MVPA techniques to measure content-specific cortical reactivation, Kuhl and colleagues (2011) examined the behavioral consequences of reinstating neural representations of competing (past) memories during the attempted retrieval of subsequently acquired target memories. In this experiment, participants initially encoded and recalled a set of arbitrary associations between words and pictures of either famous faces or scenes (A-B associations). Subsequently, for a subset of the words, participants encoded a new (C) associate, drawn from the opposite category as the old (B) associate. Thus, when participants were later challenged to recall the most recent (C) associate of each word, the relative degree of face- or scene-related reactivation in VTC served as a quantitative index of the selectivity with which participants were able to bring the target associate back to mind. This classifier-derived measure of reactivation fidelity was found to predict overall retrieval success as well as the phenomenological experience of remembering specific event details, with the fidelity of reactivation substantially diminished during competitive retrieval trials (**Figure 4**). Moreover, lower-fidelity reactivation of target (C) memories was associated with a greater likelihood of later remembering the competing (B) events, raising the possibility that the failure to selectively reactivate VTC representations associated with the target memory reflected the retrieval of both the target and its competitor (activation in frontoparietal cortical regions independently supported this conclusion). As such, Kuhl et al.'s data indicate that one consequence of reinstating older memories when attempting to remember newer memories is reduced forgetting of the past (i.e., reduced retroactive interference). Future research, exploiting MVPA approaches, promises to further reveal whether failures of selective retrieval result in the encoding of integrated multievent representations that confer additional benefits (e.g., fostering mnemonic consolidation) as well as costs (e.g., fostering across-event memory blending that gives rise to memory errors and distortion).

Distributed Activity During Event Encoding

In addition to providing leverage on the psychological and neural mechanisms subserving episodic retrieval, MVPA methods also have utility for testing hypotheses about event encoding. For example, following in the tradition of numerous fMRI studies that have used univariate analyses to examine the relationship between encoding activity in frontoparietal and MTL regions and later memory behavior (for recent meta-analyses of such studies, see Kim 2011, Uncapher & Wagner 2009), Watanabe and colleagues (2011) demonstrated that multivoxel patterns within MTL are predictive of whether visually presented pseudowords will be subsequently recognized or forgotten.

From these data alone, it is unclear whether the predictive value of the classifier was driven by diagnostic information contained within distributed activity patterns per se or whether it capitalized on the fact that many MTL voxels tended to show greater BOLD signal on trials associated with later recognition. Indeed, a voxel-wise univariate analysis of the data (liberally thresholded) revealed greater parahippocampal activity during the encoding of subsequently remembered items. Future studies are needed to assess whether MVPA analyses offer increased sensitivity for documenting neural signatures of memory formation within the MTL and beyond. If so, then distributed pattern analyses may provide novel leverage on pressing issues, such as how to characterize the differential computations subserved by the hippocampus and MTL cortical regions during encoding (e.g., Brown & Aggleton 2001, Davachi et al. 2003, Eichenbaum et al. 2007, Mayes et al. 2007, Wixted & Squire 2011).

Approaching episodic encoding from a theory-driven perspective, Jenkins & Ranganath (2010) examined whether MVPA methods can predict subsequent recall of the temporal context of a given event memory. The ability to remember when an event occurred is thought to depend, at least in part, on the fact that contextual cues inevitably drift from one moment to the next, a phenomenon that allows successively encoded events to each be associated with a partially unique temporal context (see Polyn et al. 2009). In Jenkins & Ranganath's (2010) experiment, participants were scanned while encoding visual objects and later were asked to estimate the approximate time at which each stimulus had been presented. Univariate analyses revealed that regions of the PFC and hippocampus exhibited greater activity immediately following the encoding of stimuli for which participants subsequently provided the most accurate estimates of the time of encounter. MVPA was then used to test the hypothesis that memory for temporal context would be most accurate when neural activity patterns associated with temporally adjacent stimuli were maximally distinctive. This prediction was motivated by theory suggesting that increased trial-to-trial drift in temporal context would result in individual events being associated with a greater number of trial-unique temporal context cues, thus allowing participants to more accurately estimate the time of encounter. In support of this hypothesis, activity patterns within rostrolateral PFC (RLPFC) showed a greater degree of trial-to-trial distinctiveness for items that later received an accurate estimate of temporal occurrence than those that received an inaccurate estimate. In other words, when the RLPFC activity pattern observed during a given trial was compared with the activity patterns observed during the trials that preceded or followed it, the multivariate distance between these patterns was found to progressively increase with temporal lag, with the overall magnitude of pattern dissimilarity predicting subsequent temporal memory. Importantly, this finding held even when activity patterns from RLPFC were mean centered, indicating that this putative neural correlate of temporal context indexed the pattern itself and not temporal drift in the overall BOLD signal. Jenkins and Ranganath speculated that the RLPFC's representation of temporal context might be tied to its proposed role in continuously updating high-level rule representations that specify which items and relationships are relevant in a given behavioral context; this in turn might serve to segment ongoing experience into discrete episodes. Regardless of whether this interpretation turns out to be correct, this study highlights how MVPA approaches can assess mechanistic theories of episodic memory. Future studies should examine whether temporally drifting multivoxel patterns can also account for serial position effects in free recall behavior, given that the development of the temporal context model has been heavily influenced by free recall output patterns (Howard & Kahana 2002).

The effects of representational similarity during encoding on later memory were also examined by Xue and colleagues (2010). Rather than focus on context, this study evaluated

whether the degree of neural pattern similarity across multiple encounters with a given item is predictive of later memory for the item. These authors sought to test the encoding variability hypothesis (e.g., Bower 1972), which posits that repeated exposures to a given stimulus will benefit subsequent memory to the extent that the features encoded during each exposure are nonredundant with those encoded during other exposures. At the neural level, this hypothesis might predict that the less similar the distributed cortical pattern is across an item's encoding exposures, the higher the likelihood the item will be later remembered (e.g., Wagner et al. 2000). In apparent contradiction to this prediction, Xue and colleagues observed that later-remembered stimuli were associated with more similar distributed activity patterns across study encounters than were later-forgotten stimuli. This positive relationship between neural pattern similarity and subsequent memory was observed in many brain regions, including areas of the prefrontal, parietal, occipitotemporal, and MTL cortices. These findings raise the possibility that when common attributes are attended across successive encounters with an item, the mnemonic representation of the item is strengthened. This conclusion would appear to stand in conflict with leading computational theories of memory, which demonstrate that, at least with respect to context, greater encoding variability gives rise to superior subsequent remembering (e.g., Howard & Kahana 2002, Raaijmakers & Shiffrin 1992).

Given the centrality of encoding variability in models of memory, including its role in explaining empirically robust behavioral phenomena, such as the spacing effect, it seems likely that Xue et al.'s (2010) findings will motivate follow-up studies that more fully examine the circumstances in which increased neural variability may help or hinder memory performance (Kuhl et al. 2012). As such, this study is the latest to illustrate how analyses of distributed activity patterns are affording leverage on increasingly more precise mechanistic hypotheses, leading to novel theoretical advances. We expect the coming years will bring considerable progress in delineating how encoding computations relate to later memory performance, with much of this progress stemming from the use of distributed pattern analyses to quantify the similarity between cortical representations of encoded events as well as between cortical representations of retrieval cues and of encoded stimuli.

CONCLUSIONS AND FUTURE DIRECTIONS

As we have sought to highlight, functional neuroimaging research over the past decade has been revolutionized by use of machine learning techniques to extract the representational content of distributed brain activity patterns. While traditional univariate statistical analysis approaches have informed, and will continue to inform, our understanding of the functional contributions of specific brain regions, MVPA approaches have opened new avenues for experimentation and have begun to offer fresh insights into the psychological and neural underpinnings of human cognition. Our goal in this review has been to discuss and critically evaluate some of the ways that researchers have applied MVPA methods to investigate the mechanisms of human memory. Although use of these methods to gain leverage on the workings of memory is at a relatively early stage, we believe their promise is clear, as evidenced by the many insights derived from their application over the past decade. Below we summarize some of these key insights, and we conclude by highlighting open questions that can be profitably addressed through future applications of distributed pattern analyses.

SUMMARY POINTS

1. MVPA techniques offer a powerful means for characterizing and quantifying the strength of information representation in the brain.

2. Cortical representations of stimuli are often highly distributed. By identifying the "neural signatures" of particular stimuli or classes of stimuli, MVPA can evaluate how the brain carves up the sensory world, how frontoparietal mechanisms subserve the representation and implementation of attentional priorities, as well as how moment-to-moment fluctuations of attention affect stimulus processing and memory.
3. The transient maintenance of information in working memory involves recruitment of the same cortical ensembles that mediate the perceptual representation of the information.
4. Activity patterns within subregions of the medial temporal lobe carry information about attributes of an observer's environment, allowing neural decoding of the observer's spatial location and the category of viewed stimuli.
5. Recalling a past event often involves the reinstatement of cortical activity patterns that were elicited during the initial encoding of the event. Through measuring cortical reinstatement, MVPA methods are beginning to shed new light on the psychological and neural processes underlying free recall, source monitoring, the subjective experience of recollection and familiarity, and competition-laden retrieval.
6. Retrieval goals substantially influence the pattern of cortical activity elicited by a retrieval cue. Retrieval oriented toward recollection of particular attributes of past experience fosters cortical reinstatement of those event details; cortical patterns that differentiate old and new stimuli vary depending upon whether memory is probed explicitly or implicitly.
7. Encoding-related distributed activity patterns in neocortex and the medial temporal lobe are informative predictors of subsequent memory performance. The similarity of distributed neural patterns across events influences later retrieval success.
8. Extant studies highlight how MVPA techniques can test mechanistic models of memory as well as how these methods can elucidate factors promoting remembering, or alternatively, increasing the likelihood of forgetting.

FUTURE ISSUES

1. Generative classification approaches offer an exciting avenue for future research into the neural mechanisms of working memory and episodic retrieval. The utilization of a latent feature space circumvents the need to train a classifier on a predefined set of stimuli or contexts, allowing for the interrogation of a far greater range of internally represented mnemonic content (and potentially even facilitating a rudimentary reconstruction of individual memories).
2. Because MVPA techniques can provide a neural metric of interstimulus similarity, they offer a means of drawing on neural data to test computational models of memory (e.g., how experience drives representational differentiation in semantic memory, how item-item similarity gives rise to memory errors, and how subregions of the hippocampus support pattern separation).

3. The utility of memory often rests in its ability to guide subsequent thought and behavior. By quantifying the strength of mnemonic evidence elicited by a retrieval cue, distributed pattern analyses pave the way for researchers to examine how such evidence is monitored and accumulated in the service of decision-making and action. Such an approach may also reveal the presence of nonconscious mnemonic evidence that unknowingly shapes our interpretations of the world.
4. Progress in understanding future thinking (simulating possible future events and actions; Schacter et al. 2007) and prospective memory (remembering to initiate a planned behavior at some point in time; McDaniels & Einstein 2007) may come from efforts to decode the contents of simulated events and intended actions (e.g., Haynes 2011).
5. A complex interplay between midbrain, striatal, and medial temporal lobe structures is thought to support the enhanced encoding of motivationally salient events (Lisman & Grace 2005, Shohamy & Adcock 2010). Future research can exploit the ability of MVPA to decode neural representations of reward value (Kahnt et al. 2010) and of experienced or perceived affective states (Peelen et al. 2010, Rolls et al. 2009) to relate trial-to-trial variance in these dimensions of stimulus salience to neuromodulation and memory outcomes.
6. MVPA could potentially provide an assay of memory replay during sleep (O'Neill et al. 2010). By relating the replay of specific representational content within the hippocampus and neocortex to subsequent memory outcomes, distributed pattern analyses may provide a unique window onto consolidation processes.
7. It is possible to decode aspects of an individual's mental state using a classifier trained exclusively on the data from other individuals (Clithero et al. 2011; Davatzikos et al. 2005; Just et al. 2010; Poldrack et al. 2009; Rissman et al. 2010; Shinkareva et al. 2008, 2011). Such observations raise the possibility that classifiers can be used to identify individuals who deviate from the group in some fundamental way—e.g., processing strategy, stage of neural development, or neurological health.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Supported by the National Institute of Mental Health (5R01-MH080309, 5R01-MH076932) and the MacArthur Foundation's Law and Neuroscience Project.

LITERATURE CITED

Aguirre GK, Zarahn E, D'Esposito M. 1998. An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron* 21:373–83

Awh E, Jonides J. 2001. Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci.* 5:119–26

Baddeley A. 1992. Working memory. *Science* 255:556–59

Badre D, Wagner AD. 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45:2883–901

Bar M. 2009. The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364:1235–43

Bar M, Aminoff E, Schacter DL. 2008. Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *J. Neurosci.* 28:8539–44

Barsalou LW. 2008. Grounded cognition. *Annu. Rev. Psychol.* 59:617–45

Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19:2767–96

Bisley JW, Goldberg ME. 2010. Attention, intention, and priority in the parietal lobe. *Annu. Rev. Neurosci.* 33:1–21

Bode S, Haynes J-D. 2009. Decoding sequential stages of task preparation in the human brain. *NeuroImage* 45:606–13

Bonnici HM, Kumaran D, Chadwick MJ, Weiskopf N, Hassabis D, Maguire EA. 2011. Decoding representations of scenes in the medial temporal lobes. *Hippocampus*. In press

Bower G. 1972. Stimulus-sampling theory of encoding variability. In *Coding Processes in Human Memory*, ed. AW Melton, E Martin, pp. 85–124. Washington, DC: Winston

Brouwer GJ, Heeger DJ. 2009. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 29:13992–4003

Brown MW, Aggleton JP. 2001. Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.* 2:51–61

Burgess N, Maguire EA, O'Keefe J. 2002. The human hippocampus and spatial and episodic memory. *Neuron* 35:625–41

Bussey TJ, Saksida LM. 2007. Memory, perception, and the ventral visual-perirhinal-hippocampal stream: thinking outside of the boxes. *Hippocampus* 17:898–908

Carlson TA, Schrater P, He S. 2003. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15:704–17

Carr VA, Rissman J, Wagner AD. 2010. Imaging the human medial temporal lobe with high-resolution fMRI. *Neuron* 65:298–308

Chadwick MJ, Hassabis D, Weiskopf N, Maguire EA. 2010. Decoding individual episodic memory traces in the human hippocampus. *Curr. Biol.* 20:544–47

Chang K-MK, Mitchell T, Just MA. 2011. Quantitative modeling of the neural representation of objects: how semantic feature norms can account for fMRI activation. *NeuroImage* 56:716–27

Chiu Y-C, Esterman M, Han Y, Rosen H, Yantis S. 2011. Decoding task-based attentional modulation during face categorization. *J. Cogn. Neurosci.* 23:1198–204

Chun MM, Turk-Browne NB. 2007. Interactions between attention and memory. *Curr. Opin. Neurobiol.* 17:177–84

Cichy RM, Chen Y, Haynes J-D. 2011a. Encoding the identity and location of objects in human LOC. *NeuroImage* 54:2297–307

Cichy RM, Heinzle J, Haynes J-D. 2011b. Imagery and perception share cortical representations of content and location. *Cerebral. Cortex*. In press

Clithero JA, Smith DV, Carter RM, Huettel SA. 2011. Within- and cross-participant classifiers reveal different neural coding of information. *NeuroImage* 56:699–708

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3:201–15

Cowan N. 1993. Activation, attention, and short-term memory. *Mem. Cognit.* 21:162–7

Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19:261–70

Curtis CE, Rao VY, D'Esposito M. 2004. Maintenance of spatial and motor codes during oculomotor delayed response tasks. *J. Neurosci.* 24:3944–52

Danker JF, Anderson JR. 2010. The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychol. Bull.* 136:87–102

Daugman JG. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* 2:1160–69

Davachi L. 2006. Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* 16:693–700
Davachi L, Mitchell JP, Wagner AD. 2003. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc. Natl. Acad. Sci. USA* 100:2157–62
Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, et al. 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *NeuroImage* 28:663–68
Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18:193–222
D'Esposito M. 2007. From cognitive to neural models of working memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362:761–72
Diana RA, Yonelinas AP, Ranganath C. 2008. High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. *Hippocampus* 18:536–41
Dobbins IG, Foley H, Schacter DL, Wagner AD. 2002. Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. *Neuron* 35:989–96
Donaldson DI, Wheeler ME, Petersen SE. 2010. Remember the source: dissociating frontal and parietal contributions to episodic memory. *J. Cogn. Neurosci.* 22:377–91
Dunn JC. 2008. The dimensionality of the remember-know task: a state-trace analysis. *Psychol. Rev.* 115:426–46
Eger E, Ashburner J, Haynes J-D, Dolan RJ, Rees G. 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. *J. Cogn. Neurosci.* 20:356–70
Eichenbaum H, Cohen NJ. 2001. *From Conditioning to Conscious Recollection: Memory Systems of the Brain.* New York: Oxford Univ. Press
Eichenbaum H, Yonelinas AP, Ranganath C. 2007. The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.* 30:123–52
Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, et al. 2003. Cellular networks underlying human spatial navigation. *Nature* 425:184–88
Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature* 392:598–601
Epstein RA. 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* 12:388–96
Ester EF, Serences JT, Awh E. 2009. Spatially global representations in human primary visual cortex during working memory maintenance. *J. Neurosci.* 29:15258–65
Esterman M, Chiu Y-C, Tamber-Rosenau BJ, Yantis S. 2009. Decoding cognitive control in human parietal cortex. *Proc. Natl. Acad. Sci. USA* 106:17974–79
Farah MJ, McClelland JL. 1991. A computational model of semantic memory impairment: modality specificity and emergent category specificity. *J. Exp. Psychol.: Gen.* 120:339–57
Freeman J, Brouwer GJ, Heeger DJ, Merriam EP. 2011. Orientation decoding depends on maps, not columns. *J. Neurosci.* 31:4792–804
Fuster JM. 2009. Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosci.* 21:2047–72
Fuster JM, Jervey JP. 1981. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* 212:952–55
Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M. 2005. Top-down enhancement and suppression of the magnitude and speed of neural activity. *J. Cogn. Neurosci.* 17:507–17
Gazzaley A, Rissman J, D'Esposito M. 2004. Functional connectivity during working memory maintenance. *Cogn. Affect. Behav. Neurosci.* 4:580–99
Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I. 2008. Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322:96–101
Gonsalves BD, Kahn I, Curran T, Norman KA, Wagner AD. 2005. Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. *Neuron* 47:751–61
Greenberg AS, Esterman M, Wilson D, Serences JT, Yantis S. 2010. Control of spatial and feature-based attention in frontoparietal cortex. *J. Neurosci.* 30:14330–39
Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458:632–35
Hassabis D, Chu C, Rees G, Weiskopf N, Molyneux PD, Maguire EA. 2009. Decoding neuronal ensembles in the human hippocampus. *Curr. Biol.* 19:546–54
Haushofer J, Livingstone MS, Kanwisher N. 2008. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biol.* 6:e187

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–30

Haynes J-D. 2011. Decoding and predicting intentions. *Ann. N. Y. Acad. Sci.* 1224:9–21

Haynes J-D, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8:686–91

Haynes J-D, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7:523–34

Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. 2007. Reading hidden intentions in the human brain. *Curr. Biol.* 17:323–28

Hintzman DL, Curran T, Oppy B. 1992. Effects of similarity and repetition on memory: registration without learning? *J. Exp. Psychol.: Learn. Mem. Cogn.* 18:667–80

Howard MW, Kahana MJ. 2002. A distributed representation of temporal context. *J. Math. Psychol.* 46:269–99

Hsieh P-J, Vul E, Kanwisher N. 2010. Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. *J. Neurophysiol.* 103:1501–7

Ishai A, Ungerleider LG, Martin A, Haxby JV. 2000. The representation of objects in the human occipital and temporal cortex. *J. Cogn. Neurosci.* 12(Suppl. 2):35–51

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. 1999. Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. USA* 96:9379–84

Jacoby LL, Shimizu Y, Daniels KA, Rhodes MG. 2005. Modes of cognitive control in recognition and source memory: depth of retrieval. *Psychon. Bull. Rev.* 12:852–57

Jenkins LJ, Ranganath C. 2010. Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *J. Neurosci.* 30:15558–65

Johnson JD, McDuff SGR, Rugg MD, Norman KA. 2009. Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. *Neuron* 63:697–708

Johnson MK, Hashtroudi S, Lindsay DS. 1993. Source monitoring. *Psychol. Bull.* 114:3–28

Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, Moore KS. 2008. The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59:193–224

Just MA, Cherkassky VL, Aryal S, Mitchell TM. 2010. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE* 5:e8622

Kahn I, Davachi L, Wagner AD. 2004. Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. *J. Neurosci.* 24:4172–80

Kahnt T, Heinzle J, Park SQ, Haynes J-D. 2010. The neural code of reward anticipation in human orbitofrontal cortex. *Proc. Natl. Acad. Sci. USA* 107:6010–15

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8:679–85

Kamitani Y, Tong F. 2006. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* 16:1096–102

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17:4302–11

Kastner S, Pinsk MA. 2004. Visual attention as a multilevel selection process. *Cogn. Affect. Behav. Neurosci.* 4:483–500

Kay KN, Naselaris T, Prenger RJ, Gallant JL. 2008. Identifying natural images from human brain activity. *Nature* 452:352–55

Kim H. 2011. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *NeuroImage* 54:2446–61

Kosslyn SM. 2005. Mental images and the brain. *Cogn. Neuropsychol.* 22:333–47

Kriegeskorte N, Cusack R, Bandettini P. 2010. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *NeuroImage* 49:1965–76

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* 103:3863–68

Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4

Kuhl BA, Rissman J, Chun MM, Wagner A. 2011. Fidelity of neural reactivation reveals competition between memories. *Proc. Natl. Acad. Sci. USA* 108:5903–8

Kuhl BA, Rissman J, Wagner AD. 2012. Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia*. In press

Kuhl BA, Shah AT, Dubrow S, Wagner AD. 2010. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* 13:501–6

Lewis-Peacock JA, Postle BR. 2008. Temporary activation of long-term memory supports working memory. *J. Neurosci.* 28:8765–71

Li S, Ostwald D, Giese M, Kourtzi Z. 2007. Flexible coding for categorical decisions in the human brain. *J. Neurosci.* 27:12321–30

Lisman JE, Grace AA. 2005. The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* 46:703–13

Liu T, Hospadaruk L, Zhu DC, Gardner JL. 2011. Feature-specific attentional priority signals in human cortex. *J. Neurosci.* 31:4484–95

Macevoy SP, Epstein RA. 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.* 19:943–47

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, et al. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. USA* 92:8135–39

Martin A, Chao LL. 2001. Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11:194–201

Mayes A, Montaldi D, Migo E. 2007. Associative memory and the medial temporal lobes. *Trends Cogn. Sci.* 11:126–35

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102:419–57

McClelland JL, Rogers TT. 2003. The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.* 4:310–22

McDaniels MA, Einstein GO. 2007. *Prospective Memory: An Overview and Synthesis of an Emerging Field*. Thousand Oaks, CA: Sage

McDuff S, Frankel HC, Norman KA. 2009. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *J. Neurosci.* 29:508–16

Mecklinger A. 2010. The control of long-term memory: brain systems and cognitive processes. *Neurosci. Biobehav. Rev.* 34:1055–65

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X. 2004. Learning to decode cognitive states from brain images. *Machine Learn.* 57:145–75

Mitchell TM, Shinkareva SV, Carlson A, Chang K-M, Malave VL, et al. 2008. Predicting human brain activity associated with the meanings of nouns. *Science* 320:1191–95

Miyashita Y, Chang HS. 1988. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331:68–70

Miyawaki Y, Uchida H, Yamashita O, Sato M-a, Morito Y, et al. 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60:915–29

Moser EI, Kropff E, Moser MB. 2008. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* 31:69–89

Naselaris T, Kay KN, Nishimoto S, Gallant JL. 2011. Encoding and decoding in fMRI. *NeuroImage* 56:400–10

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63:902–15

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev.* 110:611–46

Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10:424–30

Nyberg L, Habib R, McIntosh AR, Tulving E. 2000. Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. USA* 97:11120–24

O'Craven KM, Kanwisher N. 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* 12:1013–23

O'Keefe J, Nadel L. 1978. *The Hippocampus as a Cognitive Map*. Oxford, UK: Oxford Univ. Press

O'Neill J, Pleydell-Bouverie B, Dupret D, Csicsvari J. 2010. Play it again: reactivation of waking experience and memory. *Trends Neurosci.* 33:220–29

Op de Beeck HP. 2010. Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage* 49:1943–48

O'Reilly RC, McClelland JL. 1994. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus* 4:661–82

O'Toole AJ, Jiang F, Abdi H, Haxby JV. 2005. Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* 17:580–90

Peelen MV, Atkinson AP, Vuilleumier P. 2010. Supramodal representations of perceived emotions in the human brain. *J. Neurosci.* 30:10127–34

Poldrack RA, Halchenko YO, Hanson SJ. 2009. Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychol. Sci.* 20:1364–72

Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. *Science* 310:1963–66

Polyn SM, Norman KA, Kahana MJ. 2009. A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* 116:129–56

Postle BR. 2006. Working memory as an emergent property of the mind and brain. *Neuroscience* 139:23–38

Postle BR, Druzgal TJ, D'Esposito M. 2003. Seeking the neural substrates of visual working memory storage. *Cortex* 39:927–46

Preston AR, Bornstein AM, Hutchinson JB, Gaare ME, Glover GH, Wagner AD. 2010. High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. *J. Cogn. Neurosci.* 22:156–73

Puce A, Allison T, Gore JC, McCarthy G. 1995. Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.* 74:1192–99

Pulvermüller F, Kherif F, Hauk O, Mohr B, Nimmo-Smith I. 2009. Distributed cell assemblies for general lexical and category-specific semantic processing as revealed by fMRI cluster analysis. *Hum. Brain Mapp.* 30:3837–50

Quamme JR, Weiss DJ, Norman KA. 2010. Listening for recollection: a multi-voxel pattern analysis of recognition memory retrieval strategies. *Front Hum. Neurosci.* 4:61

Raaijmakers JG, Shiffrin RM. 1992. Models for recall and recognition. *Annu. Rev. Psychol.* 43:205–34

Race EA, Kuhl BA, Badre D, Wagner AD. 2009. The dynamic interplay between cognitive control and memory. In *The Cognitive Neurosciences*, ed. MS Gazzaniga, pp. 705–24. Cambridge, MA: MIT Press

Ranganath C. 2006. Working memory for visual objects: complementary roles of inferior temporal, medial temporal, and prefrontal cortex. *Neuroscience* 139:277–89

Reddy L, Kanwisher N. 2006. Coding of visual objects in the ventral stream. *Curr. Opin. Neurobiol.* 16:408–14

Reddy L, Tsuchiya N, Serre T. 2010. Reading the mind's eye: decoding category information during mental imagery. *NeuroImage* 50:818–25

Redish AD, Battaglia FP, Chawla MK, Ekstrom AD, Gerrard JL, et al. 2001. Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *J. Neurosci.* 21:RC134

Rissman J, Gazzaley A, D'Esposito M. 2004. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage* 23:752–63

Rissman J, Gazzaley A, D'Esposito M. 2009. The effect of non-visual working memory load on top-down modulation of visual processing. *Neuropsychologia* 47:1637–46

Rissman J, Greely HT, Wagner AD. 2010. Detecting individual memories through the neural decoding of memory states and past experience. *Proc. Natl. Acad. Sci. USA* 107:9849–54

Rodriguez PF. 2010. Neural decoding of goal locations in spatial navigation in humans with fMRI. *Hum. Brain Mapp.* 31:391–97

Rolls ET, Grabenhorst F, Franco L. 2009. Prediction of subjective affective state from brain activations. *J. Neurophysiol.* 101:1294–308

Rotello CM, Macmillan NA, Reeder JA. 2004. Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model. *Psychol. Rev.* 111:588–616

Ruchkin DS, Grafman J, Cameron K, Berndt RS. 2003. Working memory retention systems: a state of activated long-term memory. *Behav. Brain Sci.* 26:709–28; discussion 728–77
Rugg MD, Johnson JD, Park H, Uncapher MR. 2008. Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspective. *Prog. Brain Res.* 169:339–52
Rugg MD, Yonelinas A. 2003. Human recognition memory: a cognitive neuroscience perspective. *Trends Cogn. Sci.* 7:313–19
Schacter DL, Addis DR, Buckner RL. 2007. Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* 8:657–61
Schacter DL, Slotnick SD. 2004. The cognitive neuroscience of memory distortion. *Neuron* 44:149–60
Serences JT, Ester EF, Vogel EK, Awh E. 2009. Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 20:207–14
Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA. 2011. Commonality of neural representations of words and pictures. *NeuroImage* 54:2418–25
Shinkareva SV, Mason RA, Malave VL, Wang W, Mitchell TM, Just MA. 2008. Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE* 3:e1394
Shohamy D, Adcock RA. 2010. Dopamine and adaptive memory. *Trends Cogn. Sci.* 14:464–72
Shohamy D, Wagner AD. 2008. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* 60:378–89
Simons JS, Spiers HJ. 2003. Prefrontal and medial temporal lobe interactions in long-term memory. *Nat. Rev. Neurosci.* 4:637–48
Spiridon M, Kanwisher N. 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35:1157–65
Stokes M, Saraiva A, Rohenkohl G, Nobre AC. 2011. Imagery for shapes activates position-invariant representations in human visual cortex. *NeuroImage* 56:1540–45
Stokes M, Thompson R, Cusack R, Duncan J. 2009. Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J. Neurosci.* 29:1565–72
Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon C-H, et al. 2010. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J. Neurosci.* 30:325–30
Tamber-Rosenau BJ, Esterman M, Chiu Y-C, Yantis S. 2011. Cortical mechanisms of cognitive control for shifting attention in vision and working memory. *J. Cogn. Neurosci.* In press
Tanaka K. 1993. Neuronal mechanisms of object recognition. *Science* 262:685–88
Tong F, Pratte MS. 2012. Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63:In press
Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline J-B, et al. 2006. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *NeuroImage* 33:1104–16
Tulving E, Thompson DM. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80:352–73
Uncapher MR, Wagner AD. 2009. Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiol. Learn. Mem.* 91:139–54
Wagner AD, Maril A, Schacter DL. 2000. Interactions between forms of memory: when priming hinders new episodic learning. *J. Cogn. Neurosci.* 12(Suppl. 2):52–60
Wagner AD, Shannon BJ, Kahn I, Buckner RL. 2005. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* 9:445–53
Walther DB, Caddigan E, Fei-Fei L, Beck DM. 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* 29:10573–81
Watanabe T, Hirose S, Wada H, Katsura M, Chikazoe J, et al. 2011. Prediction of subsequent recognition performance using brain activity in the medial temporal lobe. *NeuroImage* 54:3085–92
Weber M, Thompson-Schill SL, Osherson D, Haxby J, Parsons L. 2009. Predicting judged similarity of natural categories from their neural representations. *Neuropsychologia* 47:859–68
Wheeler ME, Petersen SE, Buckner RL. 2000. Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc. Natl. Acad. Sci. USA* 97:11125–29
Williams MA, Dang S, Kanwisher NG. 2007. Only some spatial patterns of fMRI response are read out in task performance. *Nat. Neurosci.* 10:685–86

Wixted JT, Mickes L. 2010. A continuous dual-process model of remember/know judgments. *Psychol. Rev.* 117:1025–54

Wixted JT, Squire LR. 2011. The medial temporal lobe and the attributes of memory. *Trends Cogn. Sci.* 15:210–17

Woolgar A, Thompson R, Bor D, Duncan J. 2011. Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage* 56:744–52

Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA. 2010. Greater neural pattern similarity across repetitions is associated with better memory. *Science* 330:97–101

Yonelinas AP, Aly M, Wang W-C, Koen JD. 2010. Recollection and familiarity: examining controversial assumptions and new directions. *Hippocampus* 20:1178–94

Yonelinas AP, Otten LJ, Shaw KN, Rugg MD. 2005. Separating the brain regions involved in recollection and familiarity in recognition memory. *J. Neurosci.* 25:3002–8

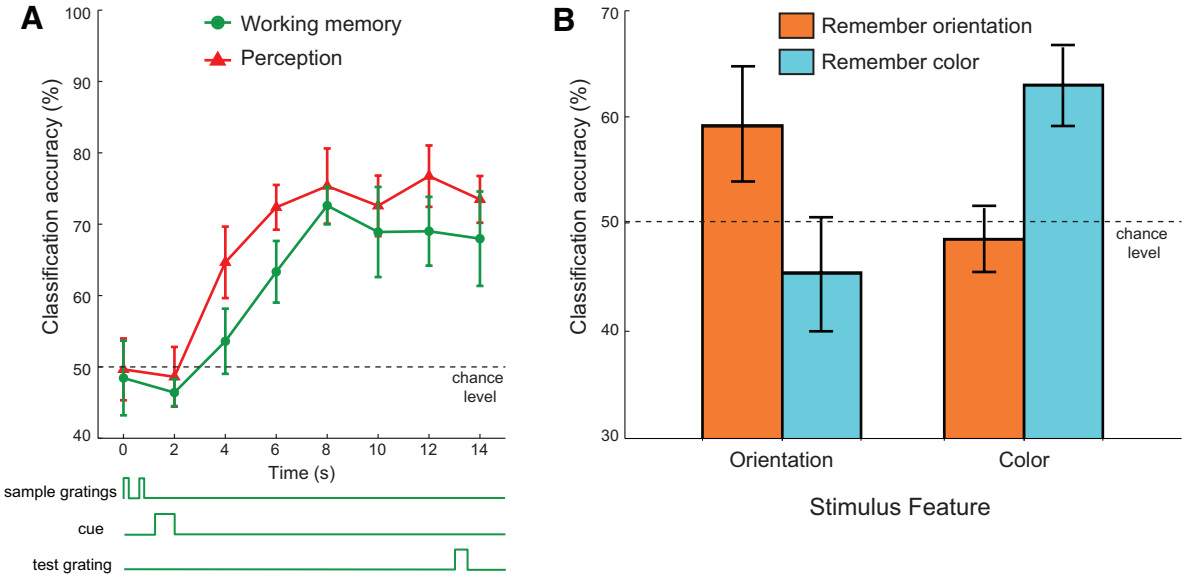

Figure 1

Recruitment of feature-specific visual representations during working memory maintenance. (*A*) Data illustrating that the orientation of a to-be-remembered line grating can be decoded from multivoxel activity patterns measured from visual areas V1–V4 throughout the entire duration of the working memory delay period. The diagram below the graph depicts the presentation times of the two sample gratings, the ensuing cue (indicating which of the two gratings should be maintained), and the final test grating (upon which participants make their memory-based judgment). The classifier's ability to decode which orientation was maintained in working memory (*green circles*) was statistically indistinguishable from its ability to decode the orientation of a perceived grating that was presented throughout the entire trial (*red triangles*). Adapted with permission from Harrison & Tong (2009). (*B*) Delay period activity patterns from V1 contain sufficient information to decode which of two orientations a participant was maintaining in working memory (on trials in which participants were cued to remember orientation) as well as to decode which of two colors participants were maintaining (on trials in which participants were cued to remember color). In both cases, decoding performance was at chance for the irrelevant stimulus dimension. Adapted with permission from Serences et al. (2009).

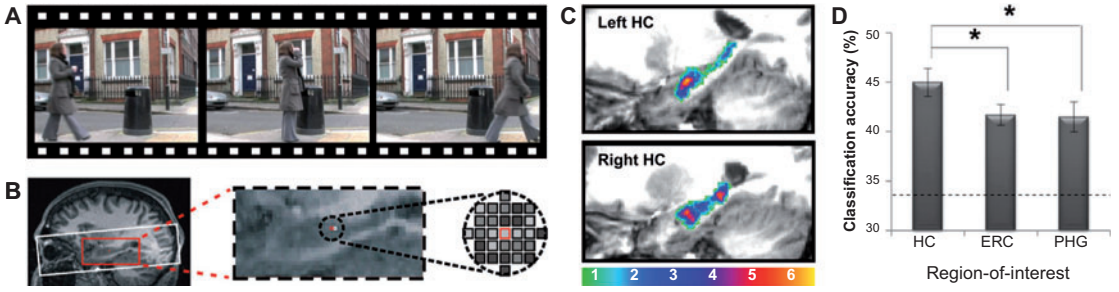

Figure 2

Decoding the content of episodic retrieval from medial temporal lobe activity patterns. (*A*) Selected frames from one of three movie clips viewed by participants prior to scanning. During each trial of the scanning session, participants closed their eyes and attempted to recall one of the three clips as vividly as possible. (*B*) Illustration of the spherical searchlight analysis approach. High-resolution fMRI data were collected from the medial temporal lobe, and classification analyses were run on the data from small spherical cliques of voxels to evaluate the accuracy with which local activity patterns could be used to decode which episode was recalled on each trial. (*C*) Frequency heat maps for the left and right hippocampi illustrating the number of participants (out of 10) for whom searchlights centered at each voxel showed above-chance mnemonic decoding performance. High across-participant consistency was observed in bilateral anterior and right posterior hippocampus (HC). (*D*) Comparison of classification performance within the HC, entorhinal cortex (ERC), and parahippocampal gyrus (PHG) revealed above-chance (dashed line: 33%) classification in all three ROIs, with decoding accuracy being significantly higher within the HC. Adapted with permission from Chadwick et al. (2010) and Hassabis et al. (2009).

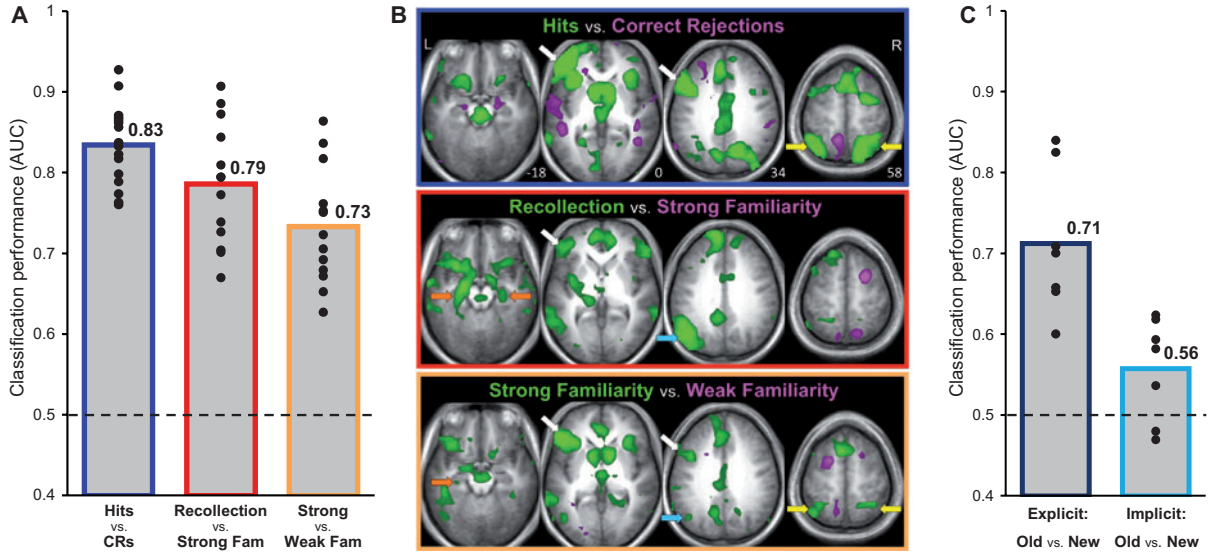

Figure 3

Decoding the mnemonic status of individual stimuli. (*A*) Classifier performance, as indexed by the mean area under the receiver operating characteristic curve [area under the curve (AUC)], is plotted for whole-brain classifier models trained to differentiate recognized studied faces (hits) from unrecognized novel faces [correct rejections (CRs)], hits associated with subjective reports of contextual recollection from those for which participants only indicated a strong feeling of familiarity, and hits associated with strong versus weak familiarity. Neural discriminability was well above chance (*dashed line*) for each classification; individual participant classification results are indicated by the black dots. (*B*) Group mean importance maps highlight lateral frontoparietal and medial temporal lobe voxels wherein greater activity drove the classifier toward a class A prediction (*green*) or class B prediction (*violet*). Comparisons across the importance maps suggest that bilateral hippocampus (*orange arrows*) and left angular gyrus (*blue arrow)* were associated with the classifier's prediction of Recollection, whereas these regions were less important for the classification of Strong versus Weak Familiarity. Rather, classification of item recognition strength appeared to depend on dorsal posterior parietal cortex (*yellow arrows*) and left lateral prefrontal cortex regions (*white arrows*) that were also observed for the Hits versus Correct Rejections classification. (*C*) Results from a second functional magnetic resonance imaging experiment reveal that a classifier's ability to discriminate old faces from new faces was dramatically diminished when recognition was probed implicitly (participants made male/female judgments rather than memory judgments) relative to old/new decoding performance during explicit retrieval conditions. Adapted with permission from Rissman et al. (2010).

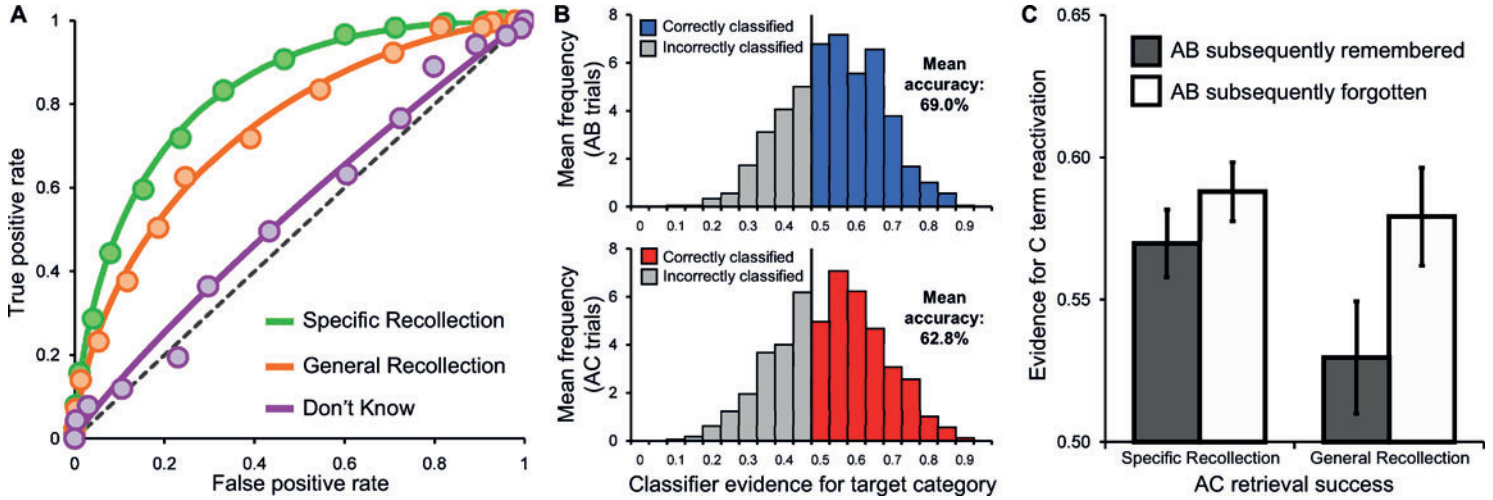

Figure 4

Selective cortical reinstatement of content-specific activity patterns scales with the specificity of episodic retrieval, diminishes with interference from competing memories, and predicts subsequent memory outcomes. (*A*) Receiver operating characteristic curves depict the ability of a multivoxel pattern analysis classifier, trained to discriminate face- versus scene-related activity patterns measured from ventral occipitotemporal cortex during event encoding, to index the reinstatement of these patterns during cued associative retrieval. The degree of neural reinstatement tracked participants' phenomenological retrieval experience such that decoding performance was most robust when participants reported recalling the specific face or scene associated with a given cue word (Specific Recollection; AUC = 0.83), significantly lower when they only were able to recall the generic category (General Recollection; AUC = 0.75), and no better than chance when participants reported that they could not recall whether the associate was a face or scene (Don't Know; AUC = 0.54). (*B*) Neural evidence for selective reactivation of the target category was diminished during competitive retrieval (AC trials) relative to noncompetitive retrieval (AB trials). Histograms depict the mean distribution of trial-specific estimates of target category reinstatement and illustrate that the classifier's predictions were less heavily skewed toward the target category when interference was present from an overlapping association. (*C*) Weaker reactivation of the C term (i.e., the target associate) during AC retrieval was linked to an increased likelihood that the competing AB associate would later be remembered in a postscan memory test. This subsequent memory effect, which was observed regardless of whether AC retrieval yielded Specific or General Recollection, suggests that lower fidelity (i.e., less selective) reactivation during AC retrieval may in fact reflect the coactivation of both the target (C term) and competing (B term) associations. Adapted with permission from Kuhl et al. (2011).