Title

Applications of Functional MRI in Memory Research

ISBN

9780429439957

Authors

Essoe, Joey K-Y
Rissman, Jesse

Publication Date

2018

Peer reviewed

22

APPLICATIONS OF FUNCTIONAL MRI IN MEMORY RESEARCH

Joey Ka-Yee Essoe and Jesse Rissman

Since its introduction 25 years ago, functional magnetic resonance imaging (fMRI) has provided researchers with a powerful tool to characterize the brain mechanisms underlying many facets of human cognition. The goal of this chapter is to highlight the ways in which fMRI methods can be, and have been, harnessed to deepen the understanding of human memory. We acknowledge that fMRI—which measures local changes in blood oxygenation levels as induced by fluctuations in neural activity—is but one of many functional neuroimaging techniques available to cognitive neuroscientists. Complementary tools such as positron emission tomography (PET), electroencephalography (EEG), and magnetoencephalography (MEG) have all been proven valuable in the quest to elucidate the neural correlates of memory formation, maintenance, and retrieval processes. However, in an effort to provide sufficient depth of coverage, we have chosen to focus exclusively on fMRI, which is currently the most widely used functional neuroimaging method. Our intention is to help readers with limited neuroimaging experience appreciate the important experimental design elements that one must consider when developing an fMRI study of memory, as well as the range of data analysis approaches that one can employ to gain insights into the contributions of individual brain regions and the functional interactions between regions. Please note that this area of research is replete with acronyms. We therefore list these acronyms in Table 22.1.

Experimental Design

Adapting a Cognitive Task Paradigm for the Scanner

The first major consideration when designing an fMRI study is how to structure the timing of task events to facilitate the measurement of brain activity associated with different cognitive processes of interest. We therefore begin by reviewing three commonly used experimental strategies for stimulus presentation: blocked, event-related, and mixed designs (Figure 22.1). Blocked designs examine the *sustained* blood-oxygen-level-dependent (BOLD) response across many successive trials of a given task, enabling between-task comparisons (e.g., encoding *versus* retrieval), whereas event-related designs examine the *transient* BOLD responses evoked during each trial, enabling comparisons between trial types (e.g., trials associated with remembered *versus* forgotten stimuli). Mixed designs incorporate a combination of both design characteristics, blocking trials of a given condition together for the examination of temporally sustained effects, while also allowing for analysis of trial-specific effects within individual task blocks.

Table 22.1 List of Acronyms

BOLD	Blood-oxygen-level-dependent
DCM	Dynamic causal modeling
EEG	Electroencephalography
ERS	Encoding-retrieval similarity
fMRI	Functional magnetic resonance imaging
ITI	Inter-trial interval
MEG	Magnetoencephalography
MVPA	Multi-voxel (or multivariate) pattern analysis
PET	Positron emission tomography
PLS	Partial least squares
PPI	Psychophysiological interactions
RDM	Representational dissimilarity matrix
ROI	Region-of-interest
RSA	Representational similarity analysis
SEM	Structural equation modeling
WM	Working memory

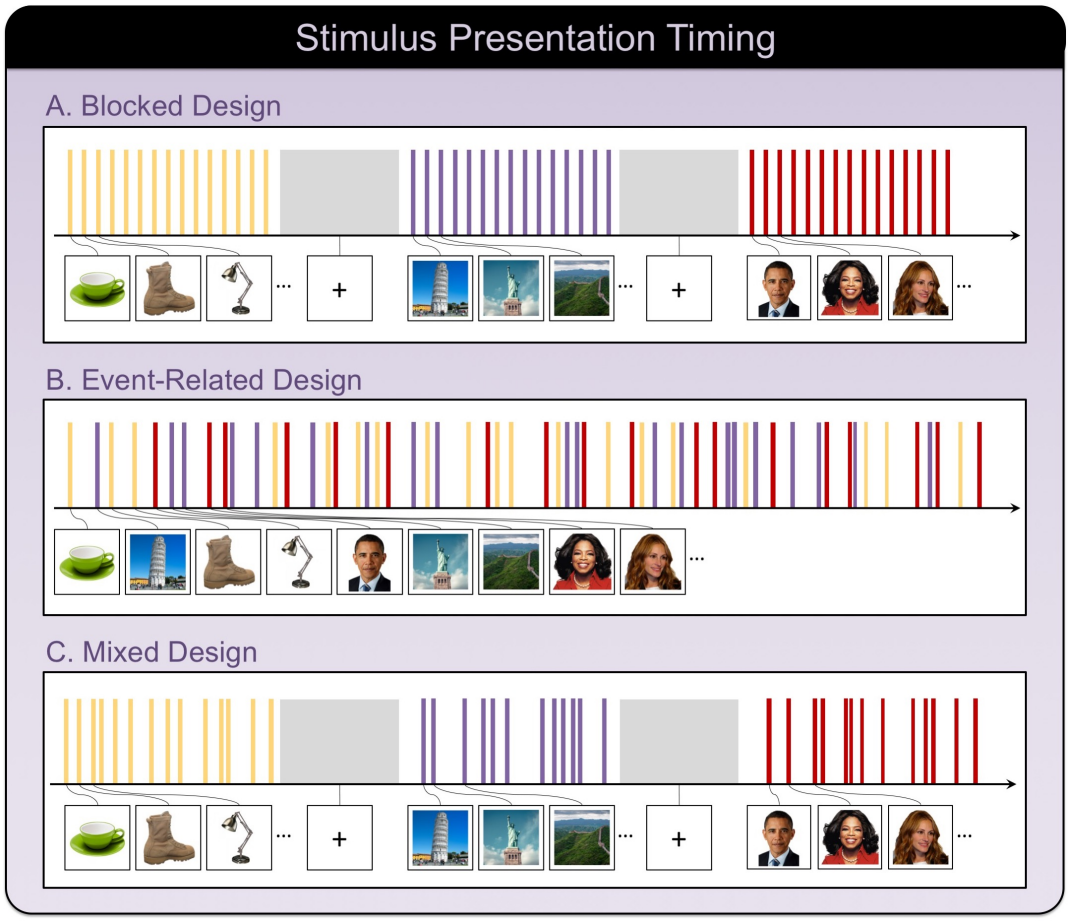

Figure 22.1 Stimulus presentation timing: In this example, the goal is to examine how the brain processes objects, scenes, and faces differently during encoding. This goal can be accomplished by any one of the three-stimulus presentation timing schema. The gray blocks represent baseline periods (which could involve resting fixation or an active baseline task), and the vertical bars represent the onsets of stimuli presentation (yellow for objects, purple for scenes, and red for faces). Regardless of design, the order of stimulus categories and/or items would be randomized or counterbalanced across participants. The blocked design (a) version of this experiment consists of multiple blocks, each comprised of 15 stimuli from the same category, with baseline blocks in between. The ITIs are fixed. In the event-related design (b) version, stimuli from all categories are intermixed and presented with jittered ITIs. The mixed design (c) version is the same as the blocked design, except with jittered, rather than fixed, ITIs to facilitate estimation of event-specific activity (e.g., to allow for analysis of subsequent memory effects or stimulus sub-categories, such as male/female or natural/manmade).

The blocked design (a) version of this experiment consists of multiple blocks, each comprised of 15 stimuli from the same category, with baseline blocks in between. The ITIs are fixed. In the event-related design (b) version, stimuli from all categories are intermixed and presented with jittered ITIs. The mixed design (c) version is the same as the blocked design, except with jittered, rather than fixed, ITIs to facilitate estimation of event-specific activity (e.g., to allow for analysis of subsequent memory effects or stimulus sub-categories, such as male/female or natural/manmade).

Blocked Designs

In *blocked designs* (Figure 22.1a), trials from a given task condition are grouped together and presented in a block (or epoch) typically lasting 12 to 60 s in duration. Over the course of the scanning session, participants will perform many such blocks of each task condition. Blocks of resting fixation are often interspersed between task blocks to allow the task-evoked BOLD signal to return to baseline level. The initial development of blocked designs for fMRI studies was heavily influenced by the design constraints associated with PET imaging, which lacks the temporal resolution to distinguish between brain signals evoked by closely spaced task events. Thus, the earliest fMRI studies of memory adopted blocked designs in an effort to compare the neural correlates associated with mnemonic processes that could be readily segregated into discrete task blocks, such as encoding versus retrieval (Gabrieli, Brewer, Desmond, & Glover, 1997), viewing of repeated versus novel images (Stern et al., 1996), and single-task versus dual-task working memory conditions (D'Esposito et al., 1995). As researchers came to appreciate that fMRI scans were capable of resolving cognitive events at a faster timescale than PET, the use of blocked designs became increasingly supplanted with event-related designs. That said, blocked designs continue to offer some advantages. Most notably, it increases statistical power. This owes largely to the fact that blocked designs integrate BOLD signal across many successively acquired brain volumes, enhancing the signal-to-noise ratio. Blocked design experiments are also easy to implement, and the data can be analyzed using a relatively simple model. For these reasons, some memory researchers whose questions do not depend on the ability to differentiate brain responses associated with individual trials or component stages of a cognitive task continue to utilize blocked designs in their work. This is especially the case for the popular N-back working memory task paradigm, in which blocks of high-load trials (e.g., 3-back or 2-back judgments) are compared to blocks of low-load trials (e.g., 1-back judgments).

Event-Related Designs

The most prevalent experimental design in modern fMRI studies is the *event-related design* (Figure 22.1b), in which trials from different conditions are presented in an intermixed sequence, and activity estimates are statistically derived for each trial type. Event-related designs first emerged in the mid-1990s (e.g., Buckner et al., 1996; Zarahn, Aguirre, & D'Esposito, 1997), inspired largely by the event-related averaging approaches that had long been employed in EEG event-related potential studies. A key innovation in event-related designs was the application of variable (or "jittered") inter-trial-intervals (ITIs). Because the BOLD response evoked by a momentary task event will typically take 4–8 s to reach its peak amplitude and around 16–20 s to return to baseline, a portion of this response may overlap in time with that of the ensuing task event. By systematically varying the degree of temporal spacing between events (or trials) through the use of jitter, one can computationally isolate (or "deconvolve") the event-related response associated with each unique trial type. The ability to do so is premised on the assumption that BOLD signals evoked by successive trials should be additive in an approximately linear fashion (Buckner, 1998; Glover, 1999).

For memory researchers, event-related designs offered a host of additional advantages over blocked designs. For instance, researchers may retrospectively sort individual encoding trials according to

whether the stimuli are later remembered or forgotten (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Wagner et al., 1998). Likewise, individual retrieval trials may be categorized based on participants' subjective reports (Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Konishi, Wheeler, Donaldson, & Buckner, 2000). Event-related designs also facilitate the estimation of region- and stimulus-specific hemodynamic response functions (i.e., the mapping between a brief burst of neural activity associated with an individual mental act and the slow rise and fall of BOLD signal that it evokes). That is, researchers may examine how the time course of BOLD activation within a given region varies across trial types (Miller & D'Esposito, 2012; Staresina, Cooper, & Henson, 2013) or brain regions (Druzgal & D'Esposito, 2003).

Mixed-Design

In the early 2000s, the *mixed design* (Figure 22.1c), or hybrid design, was introduced (Donaldson, Petersen, Ollinger, & Buckner, 2001) to allow researchers to simultaneously examine sustained and transient responses (for a review, see Petersen & Dubis, 2012). As the name denotes, the mixed design combines features of both event-related and blocked designs. Typically, trials of a given task condition are grouped together into blocks (e.g., semantic encoding condition and phonological encoding condition), whereas events within each block are presented with jittered spacing to allow for deconvolution of signals related to particular trial subtypes (e.g., subsequently remembered and forgotten stimuli). When modeling the effects in a mixed design, regressors can be included to estimate the magnitude of both transient responses (item-components) and sustained response (state-components). In this manner, mixed designs may reveal important brain activation characteristics that other designs may miss. For example, whereas many regions of prefrontal cortex show transient activations associated with retrieval success, some regions of frontopolar cortex fail to show transient engagement during individual trials, but rather show sustained activation throughout retrieval blocks that likely contributes to the maintenance of a retrieval-oriented attentional set (Velanova et al., 2003).

The Importance of Baseline

With fMRI, the raw signal intensity value for a particular region at a given moment in time is not a meaningful indicator of that region's neural activity level. To draw conclusions about a region's task-related activation, signals measured during the performance of one task must always be contrasted against those from another task, or against a baseline state. If a researcher is only interested in comparing the relative activity levels across two (or more) task conditions, then no baseline state is needed. But if one wishes to generate maps depicting brain activity for individual task conditions, then a baseline state is crucial. Traditionally, most fMRI studies have included periods of resting fixation as the baseline, either briefly interspersed between trials in the case of an event-related design, or as prolonged blocks of fixation in the case of a blocked design. However, despite the intuitive appeal of comparing task performance to wakeful rest, it has become increasingly clear that the brain is never truly at rest, and that the contents of participants' naturally wandering thoughts could influence the so-called baseline activity measure.

The use of a resting baseline poses particular problems for memory studies. Even early PET studies noted the striking correspondence in brain activity between memory retrieval tasks and wakeful rest and suggested that rest is actually comprised of "a mixture of freely wandering past recollection, future plans, and other personal thoughts and experiences" (Andreasen et al., 1995). Stark and Squire (2001) drew further attention to this point by observing that fMRI studies of memory that used resting fixation as the baseline state were less likely to find task-related activity in the hippocampus than similarly structured experimental paradigms that used an *active baseline* task. By

keeping participants' minds occupied between trials using a simple task (such as indicating whether periodically presented arrows are pointing left or right), an active baseline task can mitigate, if not entirely prevent, mind-wandering.

Practically speaking, an active baseline task should only engage cognitive mechanisms that are not of interest to the experiment. For example, in a verbal memory experiment, one might use an active baseline task that involves mathematical computations or perceptual changes such as moving dots. It is advisable that the active baseline tasks require participants to produce behavioral responses such as button presses to ensure engagement.

fMRI Analysis Approaches: Univariate Versus Multivariate

There are many approaches to fMRI data analysis, but they can be generally sorted into two distinct classes. *Univariate analyses* involve independent statistical tests that assess the level of brain activity in each brain "voxel" (the term for the 3-dimensional pixels of which MRI images are comprised), yielding statistical parametric maps of task-related activation or activity estimates within individual regions-of-interest (ROIs). *Multivariate analyses* involve running statistical tests that explicitly take advantage of the fact that activity levels are being measured throughout the entire brain virtually simultaneously—by exploiting the non-independence of these signals to characterize functional interactions between brain regions, or to extract the informational content of distributed brain activity patterns. Both classes of analysis can provide valuable insights into the neural underpinnings of cognition, but they answer fundamentally different research questions, so it is important to understand the virtues and limitations of each approach.

Univariate Approach

Description

The univariate approach, which has long been the most prevalent in fMRI research, is designed to identify the functional specialization of individual brain regions by measuring how their mean BOLD activation level changes under different conditions, such as stimuli types or cognitive processing demands. Univariate analyses can be used to generate whole-brain maps depicting the statistical evidence for task-related effects at each brain voxel. These maps can be contrasted across task conditions, and then a statistical threshold can be applied (i.e., to exclude voxels that failed to reach significance) to reveal focal clusters of activation exhibiting reliable effects. For example, univariate analyses can be used to localize brain areas specialized in face processing, such as the fusiform face area (FFA), or scene processing, such as the parahippocampal place area (PPA). Activity within these functionally defined ROIs can then be assessed during a memory task involving face and scene stimuli, which may provide insights into the effects of task goals, attentional control, and memory load on stimulus encoding and maintenance (e.g., Ranganath, DeGutis, & D'Esposito, 2004; Rissman, Gazzaley, & D'Esposito, 2009).

Implementation

Univariate fMRI analysis operates on the BOLD time-series data from each voxel using a general linear model framework. The experimenter specifies a design matrix, comprised of a set of predictor variables structured to explain the observed variance within the time-series, in order to identify voxels that are sensitive to various components of the experimental task(s). The most important predictors include a model of the expected activity for each task condition, given the timing and duration of individual task events (or blocks), and the assumed hemodynamic response functions

that translate neural activity to BOLD signal. Other predictors may be entered to explain nuisance factors, such as activity fluctuations that might be correlated with subject head motion. The output of the analysis is a set of activity parameter estimates (betas) at each voxel, reflecting the amplitude of that voxel's activation for each task condition. When applied to the whole brain (i.e., "mass-univariate" analysis), researchers can conduct voxel-wise brain mapping. This identifies voxels (or clusters of voxels) throughout the brain that show a significant activity increase for one condition (e.g., viewing faces) relative to another (e.g., viewing scenes). These analyses are initially performed on the data from individual subjects, and then random-effects statistical contrasts are conducted on the data from multiple subjects (typically 20 to 30 per study) to identify effects that are significant at the group level. Because thousands of statistical tests are conducted (one at each voxel) in a whole brain analysis, the maps must be corrected for multiple comparisons in order to control the familywise error rate. A variety of techniques have been proposed, but cluster-based thresholding procedures are the most common (Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994; Woo, Krishnan, & Wager, 2014). For a critical review of the general linear model approach to fMRI analysis and consideration of the viability of its underlying assumptions, the reader is referred to Monti (2011) and Poline and Brett (2012).

Researchers often supplement whole brain voxel-wise analyses with ROI-based analyses, which can more sensitively interrogate the signal properties in a specific region, or set of regions, wherein task-related effects are anticipated (Poldrack, 2007). Such ROI analyses can be particularly useful, for example, in high-resolution fMRI studies focused on characterizing memory-related effects within the small subfields of the hippocampus (Carr, Rissman, & Wagner, 2010; De Shetler & Rissman, 2017). ROI analyses typically involve several steps: (1) defining ROIs hypothesized to show an effect, (2) extracting the activity from all voxels within each region, (3) averaging across voxels within each ROI, and (4) conducting statistical analysis on the averaged data to identify which ROIs show reliable activity differences between conditions, or show activation levels that correlate with relevant behavioral variables of interest.

An important consideration for any ROI-based analysis is that the ROIs must be defined in a manner that is statistically independent from the tests of interest (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). In other words, if one is to test a hypothesis relating to the differential activation level of the ROI across task conditions, the ROI cannot be defined using a contrast that included one or more of these task conditions. Rather, it must be defined using data from an orthogonal contrast (e.g., from a separately acquired scanner run) or specified anatomically (e.g., from defining the boundaries of a region in each individual subject or at the group level using a spatially normalized anatomical atlas).

Applications

EPISODIC MEMORY ENCODING

The *subsequent memory paradigm* (Figure 22.2) is a popular approach to address an important question in learning and memory: "What happens in the brain during learning that leads some memories to be successfully formed and retained, while others are forgotten?" This calls for fMRI scans during encoding, and the data are then analyzed based on participants' subsequent memory performance. This analytic approach is grounded in the *difference due to memory* effect, first documented in EEG event-related potential studies in the 1980s (Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980). Specifically, individual encoding trials are retrospectively labeled according to whether the participant would later remember or forget the given stimulus, and brain activation associated with each trial type is then compared. The first two fMRI experiments of subsequent memory identified regions of prefrontal cortex and the medial temporal lobes that showed significantly elevated

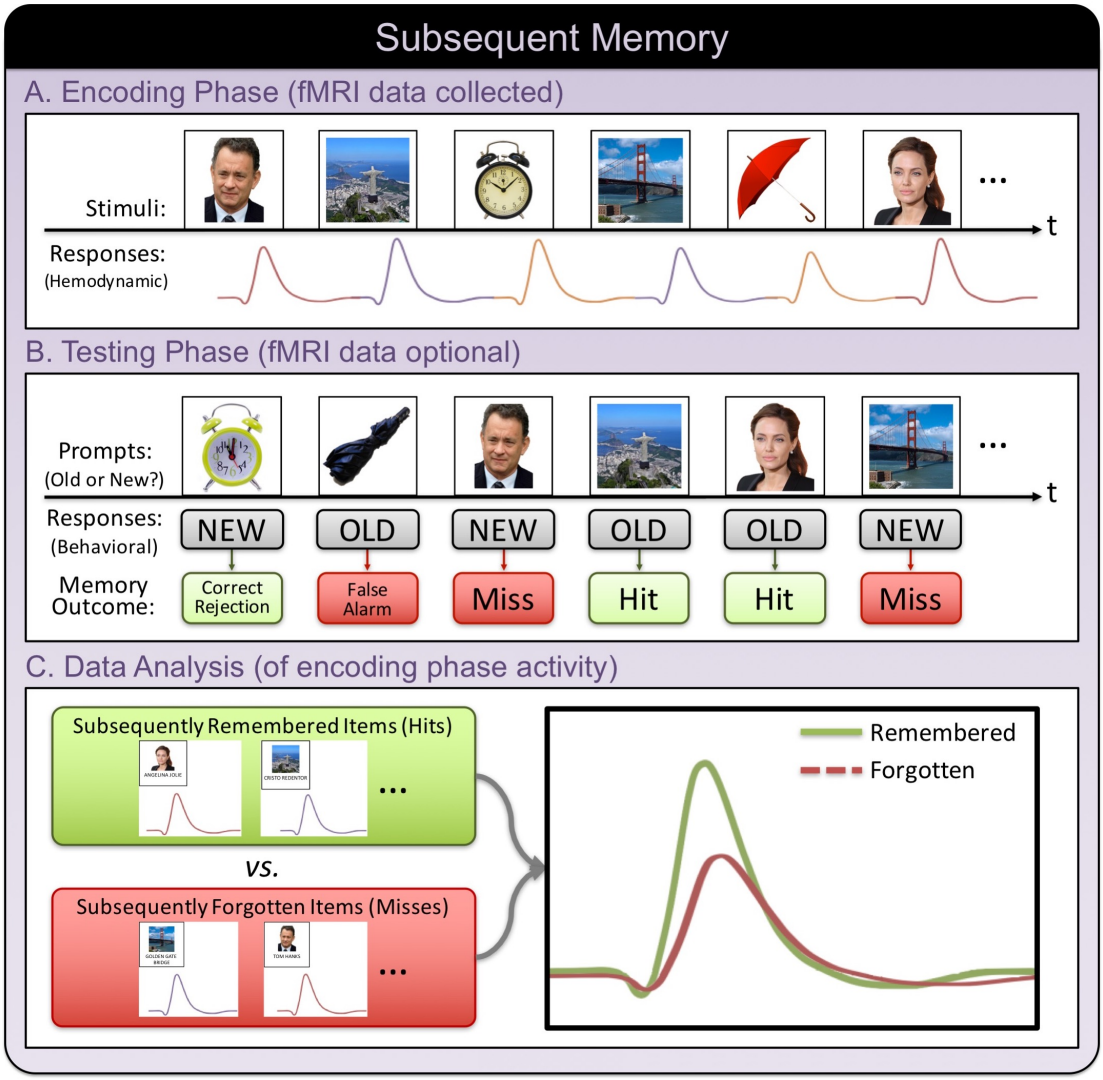

Figure 22.2 The subsequent memory paradigm: The goal of this example is to examine the differences in encoding activity that subsequently lead to different memory outcomes. (a) Encoding phase: During fMRI scans, participants would be shown images using blocked, event-related, or mixed design. (b) Testing phase: Participants would be shown previously seen images along with unstudied foil images (foil prompts have been omitted from the illustration for visual simplicity). (c) Analysis phase: Participants' behavioral responses during the testing phase can be used, retroactively, to categorize each encoding trial according to its subsequent memory outcome. Event-related activity associated with subsequently remembered and forgotten items can be separately estimated in each brain voxel, or within regions-of-interest. Time course plots can then be extracted to illustrate the mean hemodynamic response associated with each trial type.

BOLD signal during successfully encoded verbal stimuli (Wagner et al., 1998) and visual stimuli (Brewer et al., 1998). Many ensuing studies confirmed and expanded upon these initial findings, showing, for instance, that activity within different medial temporal lobe regions can predict if the contextual details associated with an item will be subsequently recollected, or whether the item will merely be recognized as familiar (Davachi, Mitchell, & Wagner, 2003). Other work has shown that reduced activation (or "deactivations") in certain brain regions during encoding, such as those within the brain's default mode network, can also be highly predictive of subsequent memory

(Daselaar, Prince, & Cabeza, 2004). For a meta-analysis of fMRI studies of subsequent remembering and forgetting, see Kim (2011).

EPISODIC MEMORY RETRIEVAL

Neuroimaging data collected during retrieval allow us to examine brain activity related to retrieval success and fidelity (Rugg & Vilberg, 2013). In these experiments, researchers generally have subjects perform a learning session and then collect fMRI data during a memory test. Depending on the goals of the experiment, the retention interval—the amount of time elapsed between the learning and testing sessions—varied from minutes (e.g., an encoding scan followed by a retrieval scan) to hours (e.g., encoding session conducted in a behavioral testing room before the retrieval scan) or months/years (e.g., experiment involving multiple visits or probing memory for real-world autobiographical events). The neural correlates of retrieval success can be examined by presenting previously learned and novel material (usually words and/or images) during a scan, and asking participants to make memory judgments. These judgments require participants to distinguish old versus new items (e.g., Konishi et al., 2000), make remember versus know judgments (e.g., Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000), rate stimulus familiarity on a graded scale (e.g., Montaldi, Spencer, Roberts, & Mayes, 2006), report the recollection of contextual source details (e.g., Kahn, Davachi, & Wagner, 2004), determine whether an item is correctly paired with its learned associate (e.g., De Shetler & Rissman, 2017; Giovanello, Schnyer, & Verfaellie, 2004), or any combinations thereof. Thereafter, fMRI data can be analyzed based on the participant's judgment or response, such as hits (studied items correctly identified as old), misses (studied items incorrectly identified as new), correct rejections (non-studied items correctly identified as new), and false alarms (non-studied items incorrectly identified as old). Furthermore, participant's confidence ratings, familiarity strength ratings, or reports of source details can also be factored into the analysis. Retrieval success effects are typically defined by contrasting activity for hits against that of correct rejections or misses. Some researchers are interested in examining the putative activity differences between true and false memories, for instance, by comparing hits against false alarms (Cabeza, Rao, Wagner, Mayer, & Schacter, 2001; Okado & Stark, 2003; Slotnick & Schacter, 2004).

Researchers can isolate state-components (task-related sustained activities, or "brain modes") and item-components (trial-evoked transient activities) by using mixed designs. To do this, experimenters present trials in blocks, interleaved with resting or baseline blocks. Based on the task timing, item-components can be deconvolved from the data. Thereafter, the state-components—such as the retrieval state (Donaldson et al., 2001) and the encoding state (Otten, Henson, & Rugg, 2002)—can be computed by contrasting the task blocks with the resting/baseline blocks.

WORKING MEMORY

Although some working memory (WM) task paradigms, such as the N-back, can be well accommodated by blocked designs, many WM tasks necessitate event-related designs. These paradigms typically feature temporally extended trials in which participants are first presented with one or more stimuli to encode, then tasked with holding this information in mind over a brief delay interval (usually 6 to 12 s), and finally probed to evaluate the accuracy of their memory. It is of great interest to researchers to examine the evolution of brain activity across the component phases of each trial: encoding, maintenance, and retrieval. One commonly used approach for modeling activity during these successive task phases involves constructing a general linear model with separate regressors whose onsets and offsets are temporally positioned to capture BOLD signal variance attributable to each task phase. Oftentimes researchers will position the onset of the maintenance phase regressor near the middle of the maintenance interval to avoid collinearity (i.e., shared variance) with the

regressors modeling the preceding encoding phase and ensuing retrieval phase (Postle, Zarahn, & D'Esposito, 2000; Zarahn et al., 1997). Occasionally multiple regressors will be used to model the early, middle, and late phases of the maintenance period (Linden et al., 2003). Estimation of the model parameters will yield separate beta map for each task condition combination (e.g., low load or high-load trials) and trial phase. These maps can then be contrasted across subjects to identify regions with significant effects of interest (e.g., elevated activities during delay periods for high- relative to low-load trials).

Interrogation of effects within ROIs can also provide valuable insights. Researchers can do so by plotting either the beta estimates or mean BOLD activation time course within a given ROI for each condition (Linden et al., 2003; Rissman et al., 2009; Xu & Chun, 2006). As with episodic memory studies, fMRI activity from WM trials can also be sorted and analyzed as a function of participants' behavioral performance, such as how activation levels during encoding and/or maintenance phases would predict retrieval success (Curtis, Rao, & D'Esposito, 2004; Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002).

REPETITION SUPPRESSION AND PRIMING EFFECTS

Another approach that has been utilized in many neuroimaging studies of memory is to look for evidence of *repetition suppression* in the BOLD signal. Repetition suppression is the reduction of neural responses after repeat exposure to the same stimuli. This phenomenon was initially observed in the firing rate of individual neurons (e.g., Lueschow, Miller, & Desimone, 1994), and researchers quickly found that activity reductions in response to repeated stimuli could also be observed with fMRI (e.g., Grill-Spector et al., 1999), even though each voxel represents the integrated activity level of hundreds of thousands of neurons. Repetition suppression has been interpreted to constitute a neural marker of *priming*—the increased processing efficiency that stimuli enjoy after repeated exposure (Henson, Shallice, & Dolan, 2000; Schacter & Buckner, 1998) or overlapping semantic representations (Rissman, Eliassen, & Blumstein, 2003). In the context of memory studies, behavioral priming effects have long been interpreted as expressions of implicit memory because the magnitude of behavioral facilitation for a repeated stimulus does not necessitate conscious awareness of the fact that it was previously encountered (Tulving & Schacter, 1990). However, there is some evidence that fMRI repetition suppression effects can be predictive of both implicit and explicit memory. For instance, Turk-Browne, Yi, and Chun (2006) found that repetition suppression in visual brain areas was associated with both behavioral facilitation to repeated scene stimuli and participants' subsequent recognition memory for these scenes. That said, others have reported dissociable neural signatures of priming and explicit memory (Schott et al., 2006), and it is likely that repetition suppression is not a monolithic construct but rather may reflect different underlying neural mechanisms depending on the specific brain region being queried and the demands of the task (Barron, Garvert, & Behrens, 2016; Grill-Spector, Henson, & Martin, 2006; Schacter, Wig, & Stevens, 2007).

Multivariate Approaches

Aided by advances in computing technology, computationally intensive multivariate analysis approaches have advanced in leaps and bounds over the past decades. Amongst multivariate analyses, (1) *functional connectivity analysis* identifies networks of regions that show correlated BOLD signal fluctuations indicative of inter-regional communication; (2) *effective connectivity analysis* takes this a step further by modeling the directionality of information flow between regions; (3) *multi-voxel pattern analysis* (MVPA) uses machine learning algorithms to decode spatial patterns of brain activity associated with different classes of stimuli or mental states; and (4) *representational similarity analysis*

(RSA) quantifies the degree of similarity (or dissimilarity) of brain activity patterns across trials or task conditions.

Connectivity Analysis

Whereas univariate analyses aim to establish functional segregation or localization (i.e., identifying brain regions "responsible" for certain cognitive processes), connectivity analyses seek to examine how multiple regions work together—or functional integration (Bassett & Sporns, 2017). Our focus will be on analyses aimed at measuring the degree of functional communication between regions based on examination of fMRI BOLD effects. This is distinct from structural connectivity analyses, which aim to characterize the anatomical connections (i.e., white matter pathways) that link discrete brain regions.

fMRI connectivity approaches are often categorized as analyses of either *functional connectivity* or *effective connectivity*. The former class of analyses seek to measure the statistical dependency of activations across different brain regions using techniques like correlation and regression, without regard to the directionality of this dependency. The latter class of analyses seeks to more explicitly model how brain regions interact, with an emphasis on characterizing how one region influences another. This is usually done by applying a series of theoretically informed models to the data and then identifying the model that best explains the data. In other words, functional connectivity tests *whether* the activity fluctuations observed in different brain regions are dependent in some way, whereas effective connectivity tries to explain the *how* they are dependent. For a review, commentary, and technical explanation of these approaches, see Friston (2011).

A note on terminology: An alternative categorization of connectivity analyses is "directed" versus "undirected." This distinction, most often used in graph theory-based approaches, refers to whether inference was made about the directionality of the effect. Whereas most functional connectivity analyses do not infer directionality, notable exceptions exist and are sometimes called "directed functional connectivity" (Friston, Moran, & Seth, 2013), for example, those grounded in Granger causality (Roebroeck, Formisano, & Goebel, 2005).

Although the following section focuses on task-based connectivity analyses, connectivity measured during wakeful rest can also be relevant to investigations of memory. By indexing the intrinsic network dynamics of a person's brain (as a trait characteristic), resting state connectivity levels can be highly predictive of individual differences in behavioral performance on WM tasks (e.g., Alavash, Doebler, Holling, Thiel, & Giessing, 2015; Stevens, Tappon, Garg, & Fair, 2012; Zou et al., 2013) or on long-term episodic memory tasks (e.g., Ferreira et al., 2013; Salami, Pudas, & Nyberg, 2014; Sheldon, Farb, Palombo, & Levine, 2016).

FUNCTIONAL CONNECTIVITY

In the memory literature, functional connectivity analyses have been used to test predictions about the role of inter-regional communication in a wide variety of mnemonic processes. Functional connectivity analyses often begin by the researchers identifying a "seed" ROI whose connectivity they are interested in examining and comparing across distinct task conditions. Various analysis procedures, several of which we describe below, can then be used to estimate the seed region's connectivity with every other voxel in the brain, yielding whole-brain maps of connectivity effects. If researchers have hypotheses about the functional interactions of a relatively small number of regions, they can compute the functional connectivity between each pair of regions and examine how these values change across task conditions. Finally, graph theoretical analysis techniques make it possible to analyze the functional connectivity properties of much larger-scale networks containing dozens, or even hundreds, of individual regions.

PSYCHOPHYSIOLOGICAL INTERACTIONS

Description First introduced by Friston et al. (1997), psychophysiological interactions (PPI) analysis measures task context-dependent functional connectivity. This analysis identifies voxels whose connectivity with the seed region changes when the task context changes (e.g., different task conditions). In other words, psychophysiological interactions analysis examines how task conditions (the psychological factor) and seed region's activity (the physiological factor) interact with one another to result in changes in other regions' activity.

Implementation In short, this is done by creating a physiological vector, reflecting the mean activity of the seed ROI at each point in time, and a psychological vector indicating which time points belongs to the condition of interest (or indicating the contrast between two task conditions). Then an interaction vector can be generated by the element-by-element multiplication of the physiological and psychological vectors. All three vectors, along with any additional nuisance vectors (e.g., head movement parameters) can then be entered into a general linear model as regressors, and the resulting beta values for the interaction regressor can be interpreted as maps of regions exhibiting task-dependent connectivity with the seed. For a helpful tutorial on psychophysiological interactions analysis, see O'Reilly, Woolrich, Behrens, Smith, and Johansen-Berg (2012).

Applications As psychophysiological interactions analysis is a versatile and widely used functional connectivity technique, it has been applied to many areas of memory research. To give just a few examples, it has been used to evaluate (1) goal-dependent changes in parietal lobe connectivity based on whether participants are focused on word identity versus word order in a verbal WM task (Majerus et al., 2006), (2) changes in hippocampal connectivity during encoding as function of subsequent memory outcome and whether participants engage in shallow versus deep stimulus processing (Schott et al., 2013), (3) large-scale connectivity changes for cortical and hippocampal regions within the brain's "core recollection network" during episodic retrieval predictive of successful recollection (King, de Chastelaine, Elward, Wang, & Rugg, 2015), and (4) task-dependent reconfiguration of prefrontal connectivity with posterior regions as a function of whether participants were cued to engage in episodic retrieval, analogical reasoning, or visuo-spatial processing of word arrays (Westphal, Reggente, Ito, & Rissman, 2016).

BETA SERIES CORRELATION ANALYSIS

Description Rissman, Gazzaley, and D'Esposito (2004) introduced the beta series correlation approach as a method for measuring correlated fluctuations in trial-to-trial activity across regions. The method is especially well suited for obtaining separate estimates of the functional connectivity for each stage of a multi-stage cognitive task (such as the encoding, maintenance, and retrieval phases of a WM task).

Implementation In this approach, parameter estimates (beta values) from event-related fMRI data are first derived from a general linear model that models each stage of every trial with a separate regressor. The resulting betas are then binned based on task stage and/or experimental conditions to form condition-specific beta series. Finally, correlations are computed between the beta series of the seed ROI and that of every other voxel in the brain, to yield a map of the seed's functional connectivity for each condition. Regions whose beta series are correlated in a given condition are inferred to be functionally dependent under that condition, and conditions can be contrasted against one another (much like with psychophysiological interactions) to reveal task-dependent effects. The method can alternatively be implemented by computing the pairwise correlations between the beta series of a set of ROIs.

Applications The beta series correlation method has made it possible for memory researchers to characterize how functional interactions between fronto-parietal regions and stimulus-selective sensory regions evolve over the course of a typical delayed recognition WM trial, such as one requiring the brief maintenance of a face stimulus (Gazzaley, Rissman, & D'Esposito, 2004). The method can also be useful for examining how connectivity levels increase or decrease between regions as function of (1) the number of stimuli that need to be maintained (Fiebach, Rissman, & D'Esposito, 2006; Rissman, Gazzaley, & D'Esposito, 2008), (2) the task-relevance of the stimuli (Gazzaley et al., 2007), (3) the ability to overcome irrelevant distractors (Clapp, Rubens, Sabharwal, & Gazzaley, 2011), (4) dynamic changes in participants' focus of attention during maintenance (Nee & Jonides, 2014), and (5) the likelihood that participants will be able to subsequently remember the stimuli on a later test (Ranganath, Heller, Cohen, Brozinsky, & Rissman, 2005; Ritchey, Dolcos, & Cabeza, 2008).

PARTIAL LEAST SQUARES

Description Partial least squares (PLS) aims to characterize the covariance structure between two or more matrices of experimental variables, with the goal of deriving a set of orthogonal latent variables that optimally relate the original matrices using the fewest dimensions. First adopted as a functional connectivity approach to examine across-subject covariance patterns in PET data (McIntosh, Bookstein, Haxby, & Grady, 1996), it has since been productively applied to event-related fMRI data, where it can take advantage of the higher-resolution temporal fluctuations that drive covariance between brain networks (McIntosh, Chau, & Protzner, 2004).

Implementation The most commonly employed variant of partial least squares for fMRI studies is known as spatiotemporal partial least squares, which aims to relate the covariance in BOLD signal between brain voxels to aspects of the experimental design matrix, using an analytic procedure called singular value decomposition. Relative to seed-based psychophysiological interactions and beta series correlation (which separately assess the statistical dependency between the seed's time-series and that of each individual voxel in the brain), partial least squares operates in a more classically multivariate manner and accounts for the observed covariance structure of the entire brain in a single step. In this sense, partial least squares is conceptually similar to other data-driven analysis methods like principal components analysis or independent components analysis, but it adds the important constraint that it only concerns itself with brain networks patterns that covary in some way with the experimental design matrix. Rather than requiring the specification of a priori contrasts between task conditions, the contrasts that explain the most variance in brain connectivity will emerge from internal model comparison. The researcher may then interpret the resulting latent variables, which typically indicate the relative weightings of each task condition (design scores) and the degree to which each voxel is a member of a network whose connectivity profile adheres to those weights (brain scores). An alternative variant of partial least squares, called seed partial least squares, can incorporate a seed ROI's BOLD time-series as one of the input matrices to yield estimates of seed-specific connectivity effects. For a detailed review and tutorial of partial least squares applications for neuroimaging, see Krishnan, Williams, McIntosh, and Abdi (2011).

Applications Its ability to parse brain activities with multiple input matrices makes partial least squares a useful tool in discovering the functional properties of distinct brain networks, or for examining task-related functional connectivity changes within known brain networks. For instance, Spreng, Stevens, Chamberlain, Gilmore, and Schacter (2010) used partial least squares to show that depending on the participant's current task goals, the brain's "fronto-parietal control network" can flexibly adjust which other brain networks it communicates, such that it couples more strongly with

the default mode network (thought to be involved in internally focused mentation) during autobiographical planning, but couples more strongly with the dorsal attention network (thought to be involved in externally oriented attention) during visuo-spatial planning. Another study used partial least squares to showcase the striking overlap between the brain network associated with remembering past events, and one associated with imagining future events (Addis, Pan, Vu, Laiser, & Schacter, 2009). Interestingly, the authors further fractionated this core network into dissociable sub-networks, and they showed that the posterior subnetwork (which included hippocampus, parahippocampal gyrus, and regions of visual cortex) was disproportionately engaged during retrospective event recall. A later partial least squares study found that the timing and spatial distribution of hippocampal connectivity during autobiographical event recall changed as a function of subjective vividness and temporal remoteness of the memories (Sheldon & Levine, 2013).

GRAPH ANALYSIS

Description Graph analysis of brain networks uses the mathematical principles of graph theory to treat the brain as a complex system composed of a large set of individual regions (referred to as *vertices* or *nodes*) linked together in some way by connections (referred to as *edges*). The structure, or topology, of the network may be mathematically evaluated based on the observed connectivity matrix. This allows for the assessment of a vast array of global network properties such as *modularity* (reflecting the tendency of a network's nodes to cluster together into a set of close-nit communities called modules) and *efficiency* (reflecting the number of nodes that typically need to be traversed for any one node to communicate with another node), as well as local network properties that pertain to each individual node, such as *degree* (reflecting the number of connections that link that node to the rest of the network; nodes with a high degree are often considered to be *hubs*). Since its initial introduction as a tool for fMRI connectivity modeling (Salvador et al., 2005), graph analysis has become a widely used method for characterizing the human brain "connectome" during the resting state, and it has also begun to provide valuable insights into the ways that large-scale brain networks reconfigure their connectivity properties during cognitive tasks. For recent reviews on graph analysis and its applications, the reader is referred to Sporns and Betzel (2016) and Bassett and Mattar (2017).

Implementation To run a graph analysis, the researcher must first decide on a set of nodes that adequately include all of the brain regions that one is interested in modeling (this could vary from dozens to hundreds of nodes). The central coordinates of these nodes are often defined based on an anatomical atlas or using a publically available functional parcellation (e.g., Power et al., 2011), and the BOLD time-series of each node is extracted. The connectivity between all pairs of nodes is then estimated for each task condition, and any connections of non-interest can be discarded (i.e., set to zero), as well as any connections whose functional connectivity strength falls below a specified threshold (this is needed to ensure sufficient sparsity). Suprathreshold connections can then either be binarized (i.e., set to one) or left as scalar values, and the graph properties of the network and its constituent nodes can be estimated by a set of algorithms (e.g., using the Brain Connectivity Toolbox; Rubinov & Sporns, 2010) and statistically compared against a set of randomly weighted networks (Fornito, Zalesky, & Breakspear, 2013).

Applications Although the application of graph analysis to fMRI studies of memory is still in its early days, a number of interesting finding have already begun to emerge. For example, two large-scale networks—the fronto-parietal control network and default mode network—that do not typically interact with one another in a cooperative fashion have been found to strengthen their coupling with one another during episodic retrieval relative a pseudo-resting condition (Fornito, Harrison, Zalesky, & Simons, 2012) and relative to two other closely matched and comparably demanding

non-episodic memory tasks (Westphal, Wang, & Rissman, 2017). As a consequence of these networks working together, they exhibit a significantly less modularized organization during episodic retrieval, which is conducive to improved memory performance (Westphal et al., 2017). Related work using graph analysis has shown that successful retrieval (relative to forgetting) is associated with pronounced changes in the connectivity profile of the hippocampus, including an enhancement of its hub-like characteristics (Geib, Stanley, Dennis, Woldorff, & Cabeza, 2017; Schedlbauer, Copara, Watrous, & Ekstrom, 2014). Hippocampal communication efficiency with other brain networks during recollection has also been found to increase on trials where participants report that their recall is vivid relative to when they report it as dim (Geib, Stanley, Wing, Laurienti, & Cabeza, 2017).

EFFECTIVE CONNECTIVITY

Effective connectivity takes a somewhat more mechanistic approach to the characterization of interregional interactions. The researcher first must specify a circuit model indicating the putative connections between a set of ROIs (or nodes). Then this model is applied to the observed data to test how a given region's activities affect those of other regions and how this relationship changes across task conditions or performance (Stephan, Li, Iglesias, & Friston, 2015). The major challenge in modeling cause and effect relationships in neuroimaging data is that fMRI does not directly measure neuronal activity, but rather only measures an indirect and inherently noisy proxy of neuronal population activity (BOLD signal). Therefore, noise modeling must be carefully handled. The most common approaches are structural equation modeling and dynamic causal modeling, and they address the noise issues in different ways. Structural equation modeling models the noise and state/signal separately and operates on the covariance rather than directly on the data. Dynamic causal modeling includes a hemodynamic forward model to deduce the neuronal level response from the observed hemodynamic response. For a detailed comparison of these two techniques, see Penny, Stephan, Mechelli, and Friston (2004).

STRUCTURAL EQUATION MODELING

Description McIntosh and Gonzalez-Lima (1994) first applied structural equation modeling (SEM) to neuroimaging for PET data, and shortly afterward, Büchel and Friston (1997) adapted the method for use with fMRI data. In this approach experimenters first specify a set of ROIs (or nodes) and the connections between them (often determined from the neuroanatomical literature). Then the causal relations between these nodes are estimated within the constraints of the specified model. In fMRI research, the terms "path analysis" or "path model" are sometimes used synonymously with structural equation modeling; however, path analysis is a special case of structural equation modeling in which only observed variables are modeled and thus does not involve the estimation of latent variables (Schlosser, Wagner, & Sauer, 2006).

Implementation A structural equation model consists of two parts: the observed and model-implied covariance matrices. The *observed covariance matrix* is produced by (1) identifying the relevant nodes (typically a set of ROIs derived either from univariate analyses or from an anatomical atlas), (2) extracting the time-series data from these nodes (their BOLD signals across relevant time windows), and (3) computing the covariance matrices from the data across time or participants. The *model-implied covariance matrix* is produced by (1) creating a hypothesized functional path model, which consists of a system of linear equations, each representing a path (the relationship between two given nodes), and (2) determining the model-implied covariance matrix, that is, what the covariance would be if the model is correct. When implementing structural equation models, *path coefficients* are estimated by minimizing the difference between the *observed matrices* and the *model-implied*

matrices, and model fit is evaluated. When multiple models were being tested, model comparison is conducted. Thereafter, higher-level statistical analyses can be conducted, such as those comparing between conditions or groups.

Applications The use of structural equation modeling in fMRI data has declined in recent years, owing at least in part to its inefficiency in estimating connections that are bi-directional or reciprocal and its inability to incorporate precise temporal information (Friston, 2011; Schlosser et al., 2006). Indeed, the senior author of the first paper to adopt structural equation modeling for fMRI data recently concluded that it is probably inappropriate for electrophysiology and fMRI data but remains useful for non-time-series data (Friston, 2011). However, structural equation modeling has yielded some intriguing findings regarding changing connectivity between prefrontal and parietal regions as a function of increasing verbal WM load (Honey et al., 2002). Other work has shown that the functioning of these fronto-parietal circuits during verbal WM is notably altered in patients with schizophrenia (Schlosser et al., 2003). And, a recent WM training study found that a strengthening of the path from left dorsolateral prefrontal cortex to the left inferior parietal lobule during training was correlated with improved verbal WM performance (Shen, Zhang, Yao, & Zhao, 2015).

DYNAMIC CAUSAL MODELING

Description Dynamic causal modeling (DCM) was first introduced by Friston, Harrison, and Penny (2003). Like structural equation modeling, this approach involves estimating experimentally induced changes in the directional flow of information processing between a set of nodes, but it also incorporates a sophisticated biophysical model of the relationship between neural activity and the BOLD response and uses Bayesian inversion to dynamically identify effective connectivity that would cause the observed data.

Implementation A dynamic causal model consists of three parts: input (deterministic sensory input/stimuli), states (observed brain activity of various regions, times, and/or conditions), and output (behavioral response). Dynamic interactions are approximated between these states, yielding the following parameters: (1) one for how the input affected the states, or evoked responses; (2) one for how the states couple with one another, interpreted as effective connectivity; and (3) one for how the input affects the coupling, which is interpreted similarly as psychophysiological interactions.

Applications Because dynamic causal modeling requires a deterministic sensory input that is predefined, it is most applicable to experiments using direct sensory stimuli to evoke the cognitive processes of interest (Schlosser et al., 2006). For example, Staresina, Cooper, and Henson (2013) presented participants with item or scene images to cue the retrieval of previously encoded item or scene paired associates (Figure 22.3). Using a simple 3-node dynamic causal model, they showed that the hippocampus bi-directionally waylays information between the perirhinal (PrC) and parahippocampal (PhC) cortices during memory retrieval in a directionally specific manner. Namely, when the cue was a scene and the retrieval target was an object, the parahippocampal cortex drives the activation of the perirhinal cortex by way of the hippocampus; however, when the cue was an object and the retrieval target was a scene, the direction of information flow within this circuit was reversed. Dynamic causal modeling has also been used to help researchers elucidate (1) the interactions between multiple neural networks during autobiographical memory retrieval (St Jacques, Kragel, & Rubin, 2011), (2) the importance of connectivity between the hippocampus and the amygdala—modulated by the orbitofrontal cortex—during the retrieval of contextual information with emotional valence (A. P. Smith, Stephan, Rugg, & Dolan, 2006), and (3) the effects of WM

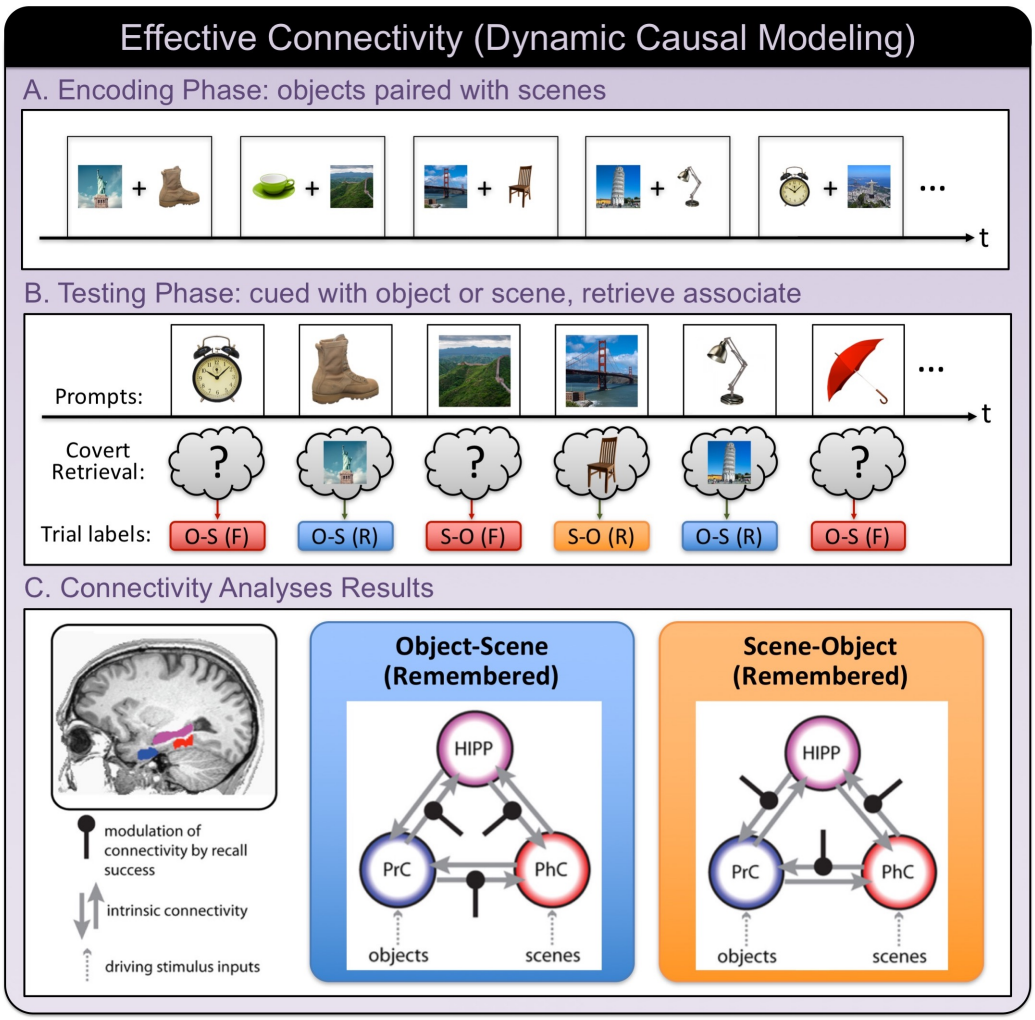

Figure 22.3 Effective Connectivity Analysis (Dynamic Causal Modeling): This example experiment is based on Staresina, Cooper, and Henson (2013), and panel C partially reproduces Figure 22.4 in that publication. The goal of this experiment is to characterize the flow of information within the medial temporal lobe during associative memory retrieval. (a) Participants first encode a set of associations between arbitrarily paired objects and scenes. (b) During scanning, participants are prompted with studied objects and scenes and instructed to covertly retrieve a visual image of the paired associate. Trials can be categorized based on whether an object cues retrieval of a scene (O-S) or a scene cues retrieval of an object (S-O), as well as whether participants reported remembering (R) or forgetting (F) the associate. (c) A 3-node dynamic causal model allows the evaluation of information flow between the object-selective perirhinal cortex (PrC), the scene-selective parahippocampal cortex (PhC), and the hippocampus. Models with various parameter settings can then be compared. In this example, the best-fitting model showed that recall success was associated with stronger connectivity from the PrC to the PhC (both directly, and via the hippocampus) during object-cued retrieval of scenes, whereas the reverse was true for scene-cued retrieval of objects.

load on the effective connectivity between fronto-parietal regions for numeric information (Ma et al., 2012) and verbal information (Dima, Jogia, & Frangou, 2014).

Distributed Pattern Analyses

As reviewed above, functional connectivity analyses constitute one important way that researchers have exploited the inherently multivariate nature of fMRI data to go beyond brain mapping

and draw inferences about the functional communication between brain regions. Another significant way in which researchers have taken advantage of the multivariate nature of fMRI data is through the application of analytic techniques that emphasize the richness of the information represented within spatially distributed patterns of activity, rather than concentrating exclusively on peak regional effects. This distributed pattern analysis approach, which includes multi-voxel pattern analysis (MVPA) and representational similarity analysis (RSA), has become increasingly influential in the neuroimaging field and is especially useful for memory research (Rissman & Wagner, 2012). The conceptual distinction between MVPA and RSA is this: Although both operate on distributed patterns, MVPA is used to differentiate (decode) brain states based on the predictions of a classifier model, whereas RSA merely measures the similarities (or dissimilarities) between them.

MULTI-VOXEL PATTERN ANALYSIS

DESCRIPTION

Haxby and colleagues (2001) first introduced the MVPA approach (Figure 22.4). They demonstrated that different categories of visual objects (e.g., faces, houses, shoes, chairs, cats, etc.) each evoked a distinctive pattern of fMRI BOLD activity in visual association regions of the ventral temporal cortex. Haxby showed that it was possible to infer which category of object a participant was viewing simply by evaluating the similarity of the brain activity pattern to the characteristic "neural signature" of each visual category (in this sense, Haxby's seminal paper can also be considered the first RSA study because the classification of brain patterns was based solely on the assessment of pattern similarity). Shortly thereafter, this general analysis approach was formalized using a pattern classification framework derived from machine learning (Cox & Savoy, 2003).

Since its inception, memory researchers have harnessed the power of MVPA in a number of creative ways to provide novel insights into the mechanisms of learning and remembering. This method is sometimes sensationalized as "mind reading" (K. Smith, 2013; Wardlaw et al., 2011). Although this is undeniably true in a limited and specific sense, the scope of this claim remains bounded by the fact that the classifiers must be provided with known, defined, and finite categories—at least for now. Should it be achievable, it would have implications beyond the field of cognitive sciences, to that of forensics and ethics. For a review and commentary on decoding and mind/brain reading, and potential ethical issues that might arise, see Haynes and Rees (2006). For further discussion of methods, applications, and results interpretation, see Norman, Polyn, Detre, and Haxby (2006), Tong and Pratte (2012), and Chadwick, Bonnici, and Maguire (2012).

IMPLEMENTATION

Typically, MVPA begins with dividing the fMRI data into training versus testing patterns. Each "pattern" is a vectorized representation of the BOLD activation levels across voxels within a particular region of the brain (or sometimes even across the entire brain) for a given time point or trial in the experiment. The experimenter must label each training pattern as an example of a particular class (i.e., trial type). These training patterns are then fed as inputs to a *multivariate classifier algorithm*, such as a support vector machine or regularized logistic regression, which formulates a model that can then be used to predict whether new patterns (i.e., test patterns that were not used in the classifier's training) are more 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. For more stable results, this process is often repeated, each time with different subsets of the data used as the training and testing patterns, through a procedure known as cross-validation. Classification accuracy is often improved by reducing the number of voxels fed into the classifier (i.e., feature selection)

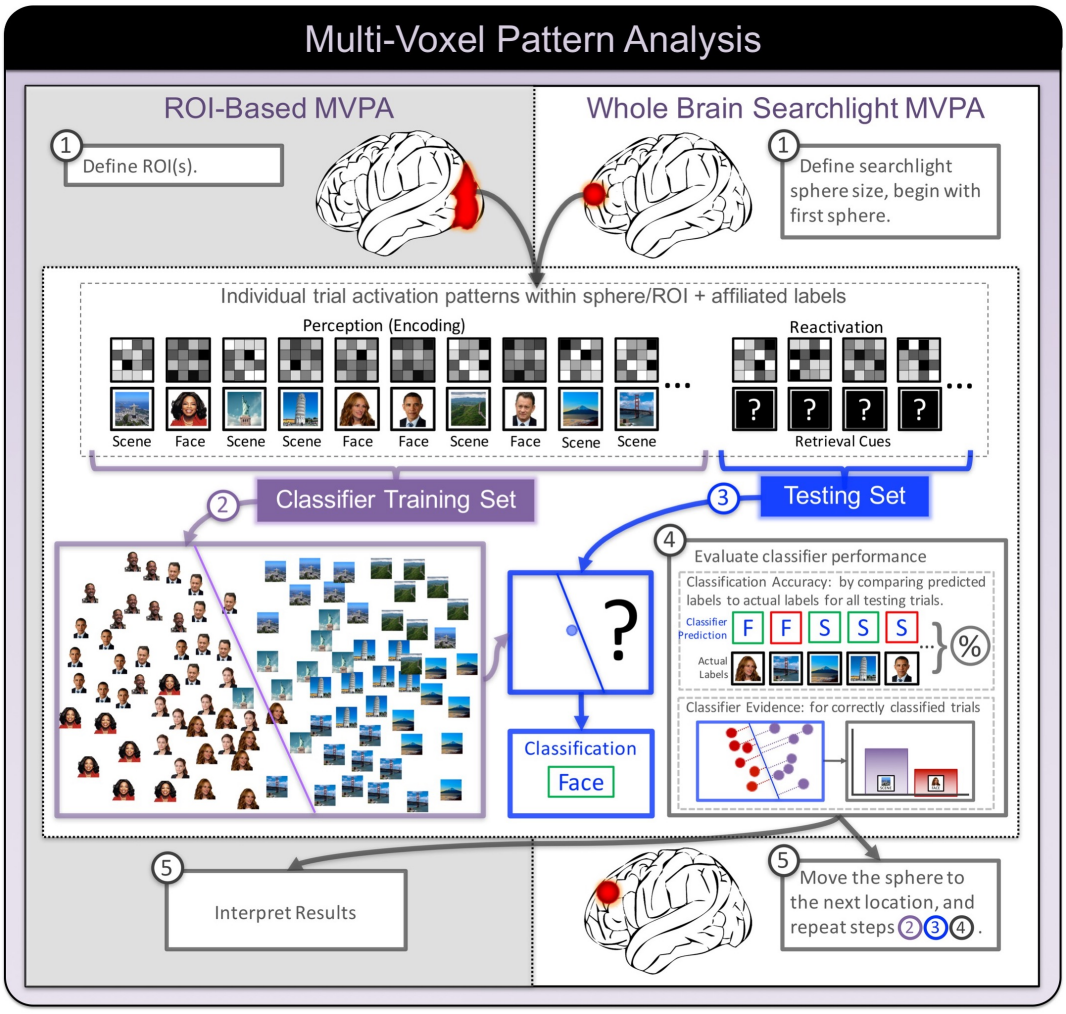

Figure 22.4 Multi-voxel pattern analysis (MVPA): This example MVPA application illustrates a scenario where one wishes to train a classifier to distinguish the brain patterns associated with two visual categories (faces and scenes) based on fMRI data acquired during perception (encoding) of face and scene stimuli, and then test the classifier's ability to predict which stimulus category participants are bringing to mind during each retrieval trial based on the brain patterns evoked in response to an associative retrieval cue (e.g., a word or object that had previously been associated with a face or scene). (1) The classifier can either be trained and tested using the brain patterns within a specified region-of-interest (ROI), or whole brain searchlight MVPA can be conducted to map areas containing local voxel activity patterns that are reliably able to distinguish between the classes of stimuli. (2) The data are divided into training and testing sets (in this case based on encoding and retrieval, but in many applications it might be useful to divide the data based on runs using a leave-one-run-out cross-validation approach). Data within the training set are labeled trial by trial according to their class membership (e.g., face or scene), and the classifier then derives a high-dimensional decision boundaries for these classes. (3) After this, the withheld testing set trials would be submitted to the classifier without the labels. The classifier identifies each trial's "place" in the decision space and outputs a classification (which category the classifier thinks that the trial belongs to). (4) Thereafter, the overall classification accuracy can be computed for this specific region or sphere. One can also evaluate the "classifier evidence" for individual predictions based on how far a given test pattern falls from the decision boundary. For instance, if face retrieval tends to be more vivid than scene retrieval, the classifier might show stronger evidence scores for face trials in the testing set. (5) This concludes an ROI-based analysis, whereas a searchlight analysis would store the classification result at the central voxel of the searchlight sphere and then move the sphere one voxel over and repeat the procedure until each voxel in the brain has served as the center of the searchlight sphere.

because including noisy or uninformative features can impair the classifier's ability to capture diagnostic patterns in the data. Although the multivariate nature of MVPA can make it difficult to interpret the contributions of individual voxels to classification performance, inspection of classification "importance maps" may provide clues into which voxels most strongly influence the classifier's predictions. Researchers may also conduct classifications within different ROIs and compare classification accuracy to evaluate the informational content of each region. This procedure can be extended to map informational content throughout the entire brain through an approach known as *searchlight analysis*, which involves running thousands of separate classifiers, each trained and tested on only a small cluster of voxels (Etzel, Zacks, & Braver, 2013; Kriegeskorte, Goebel, & Bandettini, 2006).

APPLICATIONS

Cortical Reinstatement Many theories of memory posit that the act of retrieving a memory involves the partial reactivation—or *reinstatement*—of the cortical representations that were activated during the initial formation of that memory. Initial fMRI evidence suggestive of neural reinstatement came from univariate analyses demonstrating that many of the same stimulus-selective cortical regions that were active during the initial encoding of a memory appear to be reactivated during its retrieval (for review, see Danker & Anderson, 2010). The advent of MVPA allowed cortical reinstatement to be quantified with far more precision because researchers can train a classifier to learn the brain activity patterns associated with the stimulus encoding and then test the classifier on a set of retrieval trials to evaluate the degree to which the retrieval patterns matched the encoding patterns (Levy & Wagner, 2013).

In the first MVPA study of episodic memory, Polyn, Natu, Cohen, and Norman (2005) trained a classifier to differentiate the activity patterns associated with faces, objects, and scenes during encoding and found these encoding pattern were indeed reinstated during a free recall test. Furthermore, this reinstatement typically preceded participants' behavioral responses by several seconds, suggesting that recall may be facilitated by the internal generation of effective retrieval cues. Later, Johnson, McDuff, Rugg, and Norman (2009) found that cortical reinstatement is not only apparent on trials in which participants report the subjective experience of contextual recollection, but that it also can be observed (albeit to a lesser degree) on trials in which participants only reported a sense of item familiarity. The authors argued that reinstatement in and of itself may not be sufficient for high-fidelity memory recall.

Gordon, Rissman, Kiani, and Wagner (2014) examined the relationship between encoding strength and cortical reinstatement. While in the scanner, participants encoded a set of descriptive adjectives, each arbitrarily paired with a cue to imagine a person or a scene associated with that adjective. Then, during the second half of the scanning session, they were again presented with each adjective and asked to recall whether they had previously imagined it with a person or scene (i.e., the source context). The MVPA classifier was trained to discriminate person versus scene imagery during a subset of the encoding trials, and it was then tested on the remaining encoding trials to yield trial-by-trial estimates of encoding strength. The trained classifier was also applied to the retrieval trials to provide estimates of cortical reinstatement. Encoding strength was found to predict both the probability that a trial's source context would later be recalled and the magnitude of cortical reinstatement during retrieval.

Other work has shown that the lack of robust cortical reinstatement is also informative in certain experimental contexts. For instance, Kuhl, Rissman, Chun, and Wagner (2011) reasoned that low fidelity reinstatement of a target memory may be a marker of mnemonic competition during retrieval. They had participants learn a set of associations between individual words and images of either faces or scenes. For some of these words, participants were then tasked with learning a new paired associate (from the opposite category). When participants were later asked to recall the most recently learned image associate, the degree to which they reactivated the appropriate

category-selective cortical patterns was substantially diminished for targets that had a competitor. Moreover, the weaker the cortical reinstatement was for a target, the more likely its competitor would be subsequently remembered. Interestingly, as decoding become more ambiguous (interpreted as increased competition between the two retrieved memories), fronto-parietal regions become more engaged, putatively to help resolve the competition.

Decoding Mnemonic States In addition to classifying between trial conditions, MVPA can also be used to decode mnemonic states. For example, Quamme, Weiss, and Norman (2010) used MVPA to identify the right supramarginal gyrus to be involved in supporting the maintenance of an internally directed attentional state that prepares the mind to make a recollection, or the "listening for recollection" state. More recently, Richter, Chanales, and Kuhl (2016) used cross-subject MVPA to successfully decode between an encoding state, a retrieval state, and integration state (conducive to building a link between a new item and an already-learned paired-associate). They found that these three states could be robustly discriminated from the underlying brain activity patterns, and that the degree to which participants' brains were in an integration state during learning could predict behavioral expressions of successful memory integration. Furthermore, the trained classifier could reliably decode specific instances of spontaneous memory integration in an independent sample of subjects.

In a related line of work, Rissman, Greely, and Wagner (2010) reported that MVPA classifiers could achieve remarkably accurate decoding of participants' subjective retrieval states, such as whether a given face was experienced as old or new, and whether recognition was associated with vivid recollection, or a strong versus weak sense of familiarity. In contrast to this robust classification of subjective states, the ability to decode if a particular face had actually been previously experienced was rather limited, as was decoding of faces' old/new status when recognition was assessed implicitly rather than explicitly.

Going beyond a standard laboratory-based paradigm, Rissman, Chow, Reggente, and Wagner (2016) examined decoding of memories for real-world events from the participants' own lives. Participants wore digital cameras for three weeks, then image sequences captured by their cameras were shown during a scan, along with those from other participants, and participants judged their level of memory for each event. The results showed near-perfect classification between correctly recognized versus correctly rejected events, regardless of retention interval (the temporal remoteness of the event). In addition to successfully differentiating recollection from familiarity as well as different levels of subjective memory strength, they found dissociable brain maps for these mnemonic states. Interestingly, when they applied the classifier that they had trained on the data from their earlier laboratory-based face memory experiment to the data from the autobiographical event memory experiment, decoding performance remained robust, suggesting that these mnemonic retrieval states are relatively stable across participants, experimental paradigms, retention intervals, and stimulus types.

Reading Out the Contents of Working Memory Much as theories of episodic memory emphasize the importance of reactivation of encoding-related activity patterns, theories of WM emphasize the persistent activation of cortical patterns representing the to-be-maintained content (Lee & Baker, 2016; Postle, 2016). Given that MVPA methods are well suited for quantifying the representation of stimulus-specific activity patterns over time, researchers have used MVPA to "read out" the WM contents. For example, Harrison and Tong (2009) trained a classifier model to discriminate the BOLD activity patterns in visual cortex associated with distinct orientation gratings. They then applied this classifier to fMRI data from a WM task in which participants were shown two distinct orientation gratings on each trial and then cued to hold one of these gratings in memory across an 11-s delay period, after which they judged whether a probe grating matched the one held in

memory. The fMRI analyses showed that even though BOLD signal levels in the visual cortex dropped dramatically after stimulus encoding, the classifier accurately decoded the orientation of the grating held in memory based on delay period activity patterns in visual areas V1 to V4. Serences et al. (2009) conducted a similar experiment, except that they used orientation gratings on colored backgrounds, and participants were cued to maintain either the orientation or the color. Their MVPA analysis showed that the maintenance period only contained diagnostic information about the relevant dimension, and most robustly in area V1. That is, when orientation information is maintained, the classifier can decode between the orientations, but not the colors, and vice versa. These two experiments supported the sensory recruitment hypothesis of WM by demonstrating that BOLD activity patterns during WM maintenance resemble those evoked during bottom-up perception. This suggests that neural patterns that support online sensory processing are also active during WM maintenance of the same stimuli, rather than transferring the processing to a separate WM buffer.

Later studies showed that visual WM content could be decoded beyond V1 to V4. Christophel, Hebart, and Haynes (2012) found that fMRI activity patterns within posterior parietal cortex contained sufficient information to allow a classifier to decode between colorful abstract stimuli, whereas patterns within frontal cortex did not. Results like these support a fronto-parietal network model of visual WM, in which parietal regions contribute to the maintenance of visual WM feature information, whereas frontal regions exert top-down control for accessing the stored contents. However, the evidence for this model is mixed, with some studies failing to show reliable WM decoding in parietal cortex (Riggall & Postle, 2012) and others reporting reliable decoding throughout visual, frontal, and parietal cortices (Ester, Sprague, & Serences, 2015).

REPRESENTATIONAL SIMILARITY ANALYSIS

DESCRIPTION

Although RSA has its roots in the early fMRI pattern analysis work of Haxby and colleagues (2001), the RSA approach was formally introduced by Kriegeskorte, Mur, and Bandettini (2008) and has since become a popular alternative to, or complement of, classifier-based MVPA. Rather than attempting to decode mental states, the goal of RSA is merely to characterize the similarity structure of a set of brain activity patterns. Researchers will typically use RSA to evaluate how pattern similarity within specific regions changes as a function of stimulus characteristics, task conditions, or behavioral performance. RSA can also provide a valuable tool to test how well brain activity patterns adhere to the predictions of various computational models.

IMPLEMENTATION

Much like MVPA, RSA begins by extracting the BOLD activity patterns within a given ROI for each trial of the experiment, but instead of training a classifier model, these patterns are simply correlated with one another to yield a matrix of pattern similarity values. These similarity values may then be summarized for trial pairs within and across individual trial types or task conditions. Oftentimes, the values in this matrix are all subtracted from 1 to create a *representational dissimilarity matrix* (RDM) reflecting the distinctiveness of BOLD patterns. Individual cells of the RDM may then be statistically contrasted against each other (e.g., to evaluate whether a region shows significantly greater dissimilarity between trials of different conditions than for trials of the same condition), or the entire RDM may be compared to an RDM derived based on participants' behavior (e.g., subjective stimulus similarity ratings) or a theoretical/computational model. Although RSA procedures are typically applied within individual ROIs, the technique may also be applied to characterize

local pattern similarity throughout the brain using a searchlight mapping procedure. For a step-by-step conceptual tutorial to RSA, see Kriegeskorte et al. (2008), and for further details pertaining to implementation, see Nili et al. (2014).

APPLICATIONS

Memory Encoding During learning, individual stimuli are often re-studied several times, and there has been debate about whether encoding should be facilitated or hindered by representing each stimulus in a similar fashion upon repeated encounters. Xue et al. (2010) applied an RSA approach to quantify the neural similarity across multiple encounters of a given stimulus, and they found that several prefrontal, parietal, and visual association areas showed heightened similarity for stimuli that later went on to be remembered versus those that were later forgotten. In a related study, LaRocque et al. (2013) examine the roles of medial temporal lobes regions and reported that subsequent memory could be predicted by the degree to which perirhinal cortex and parahippocampal cortex activity patterns were more similar and the degree to which hippocampal patterns were more dissimilar. These findings supported the notion that the hippocampus is responsible for pattern separation (differentiating the neural representations of similar stimuli to ensure their distinctiveness in memory), whereas perirhinal cortex and parahippocampal cortex encode highly overlapping representations of similar stimuli. Favila, Chanales, and Kuhl (2016) expanded upon these findings by demonstrating that lower representational overlap in the hippocampus is conducive to subsequent learning by virtue of preventing interference between similar memories. The degree of hippocampal pattern similarity/dissimilarity between events has also been found to predict participants' later judgments of the events' temporal proximity to one another (Ezzyat & Davachi, 2014).

MEMORY RETRIEVAL

Whereas most MVPA-based analyses of episodic retrieval have assessed the accuracy with which category-specific or context-specific activation patterns could be decoded (i.e., by training the classifier to differentiate broad classes of trials), RSA methods have shown promise in their ability to capture event-specific pattern similarity effects. For instance, Kuhl and Chun (2014) measured the pattern similarity between activity patterns evoked during a cued recall task (in which participants recalled a target image in response to a word associate) and those that evoked during a visual recognition task. Although these two trial types contained no perceptual information in common, fMRI patterns within a number of regions—most notably the angular gyrus—showed greater similarity for trials that required retrieval of the same exemplar than those that involved retrieval of different exemplars. RSA approaches have also been useful for querying the degree to which a region's activity patterns are influenced by the spatiotemporal relationships between retrieved memories. Deuker, Bellmund, Schröder, and Doeller (2016) reported that the pattern similarity in the hippocampus across retrieval trials scales with temporal and spatial distance of objects encoded in a virtual city. Along the same line of inquiry, but using real-world, personal episodic memories (cued by photographs of the participant's own life-logged images), Nielson et al. (2015) reported that the anterior hippocampus pattern similarity across retrieval trials scaled with both temporal and spatial distance of the event being retrieved.

ENCODING-RETRIEVAL SIMILARITY

In similar fashion to MVPA studies of cortical reactivation, RSA techniques can provide a powerful means to index the degree to which activity patterns observed during retrieval mimic those observed during encoding (Figure 22.5). But as an advantage over a classifier-based MVPA approach, which

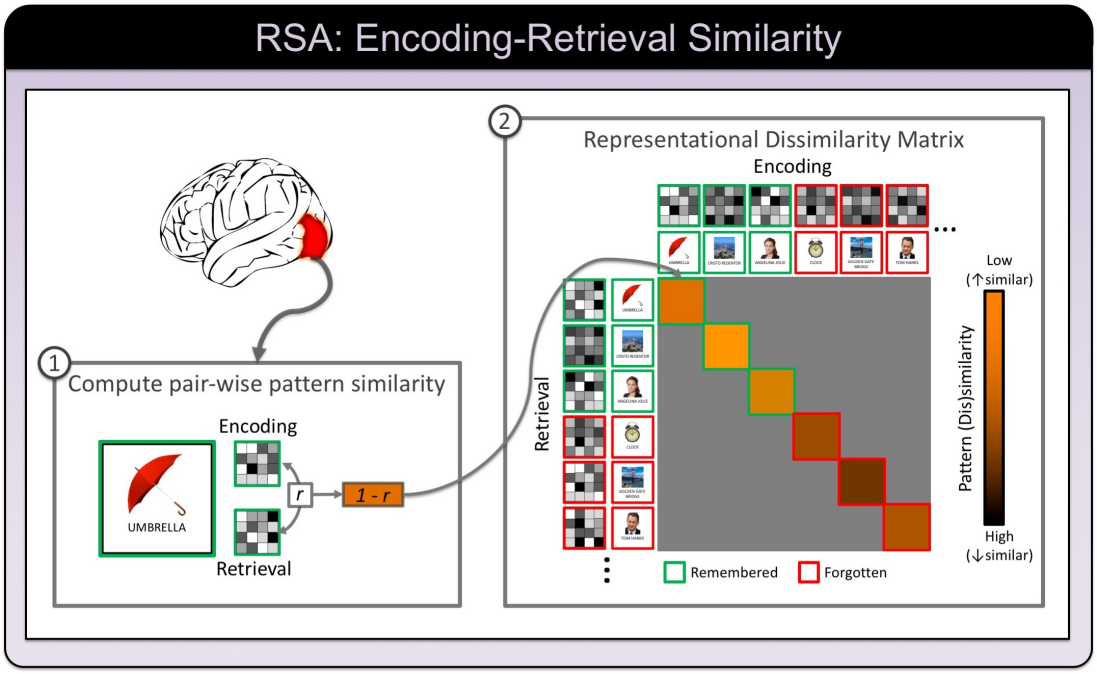

Figure 22.5 Representational Similarity Analysis: Encoding-Retrieval Similarity (ERS): This example builds upon the subsequent memory example experiment (Figure 22.2), with testing phase fMRI data collection. The goal of this example analysis is to examine, within a given ROI, whether the degree of similarity between encoding-related and retrieval-related activity is greater for items that were successfully remembered. In this analysis, the images are categorized based on whether they were subsequently remembered or forgotten. (1) Then a correlation (*r*) is computed between the encoding and retrieval activation pattern for each stimulus. (2) After the pairwise dissimilarity (1—*r*) is computed for each stimulus, a representational dissimilarity matrix can be used to plot the results, and relevant cells of this matrix can be contrasted to evaluate whether encoding-retrieval similarity (ERS) differs significantly as a function of memory outcome.

operates on categories of stimuli, RSA can measure the encoding-retrieval similarity (ERS) of individual items. For example, Wing, Ritchey, and Cabeza (2015) conducted an fMRI experiment in which participants first encoded a large set of scene stimuli with verbal labels. During the retrieval period, participants were to covertly retrieve the scene cued by the labels and report the quality of the recall. Later they underwent a recognition test with the learned scenes against three exemplars. This design allowed the researchers to measure the relationship between ERS and recognition outcome for individual scenes (item level) and for all scenes (set level). They found that successful recognition scaled with occipito-temporal ERS at the item level, but not set level, whereas ventrolateral prefrontal ERS showed recognition-predictive effects at both levels. Using a similar design, Danker, Tompary, and Davachi (2016) found that cortical ERS correlates with univariate hippocampal activation during encoding for a given item. In another experiment that used high-resolution fMRI to measure ERS within medial temporal lobes and hippocampal subfields, Tompary, Duncan, and Davachi (2016) found evidence that individual episodic memories are reinstated within the CA1 subfield of the hippocampus as well as in the perirhinal cortex. Participants with better overall memory performance also showed more pronounced modulation of ERS during successful remembering at the level of individual trials. The important link between ERS and retrieval success was also observed by Mack and Preston (2016), who found that hippocampal and perirhinal cortex ERS predicted the speed of participants' memory-based decisions.

Similar to ERS, Bird, Keidel, Ing, Horner, and Burgess (2015) examined the pattern similarity between encoding and covert retrieval practice, using video clips as memoranda. One week later, they tested participants' memory for the details of the videos. They found that pattern similarity between encoding and retrieval practice in the posterior cingulate cortex predicted long-term retention of complex information. Retrieval practice can also adversely impact competing memory. Using an event-specific RSA approach, Wimber, Alink, Charest, Kriegeskorte, and Anderson (2015) found that the repeated retrieval of a given target memory suppresses the specific cortical patterns of its competitors. Strikingly, not only did this pattern suppression predict subsequent forgetting of the competitor, but it also correlated with univariate activation of prefrontal regions known for resolving retrieval competition.

Conclusion

Over the past couple of decades, fMRI has proven to be a powerful and versatile tool for learning and memory research. Exciting new developments lead us to predict that fMRI will continue to increase in usefulness in the foreseeable future (Poldrack & Farah, 2015). First, computing technology continues to grow cheaper and more powerful, enabling increasingly sophisticated analyses. Second, ultra-high field MRI scanners (7T or more) are becoming increasingly available, improving spatial resolution and signal-to-noise ratio. For memory researchers, high field scanning has allowed for the unprecedented isolation of BOLD effects within the thin laminar structures of the hippocampus and entorhinal cortex (Maass et al., 2014). Third, advances in multi-channel head coil technology and MRI pulse sequences have facilitated parallel imaging approaches that allow multiple slices to be acquired simultaneously (Feinberg & Setsompop, 2013). This vastly improves temporal resolution with which the BOLD signal can be sampled, improving the estimation of event-related time courses and the robustness of functional and effective connectivity analyses. Fourth, with the Human Connectome Project (Van Essen et al., 2013) nearing completion, and the broader open data movement gathering momentum, many data sets (including many incorporating memory tasks), connectivity maps, and fully documented toolboxes and scripts (for visualization, preprocessing, and analyses) have become accessible to all researchers (Milham, 2012; Nichols et al., 2017). This affords novices access to fMRI data as a learning tool and experts the ability to widen their skillsets and make new discoveries. Most importantly, it increases the accountability and thus integrity in the scientific field in general (Poldrack et al., 2017).

In addition to these developments, fMRI researchers are becoming increasingly adept at applying advanced machine learning techniques to fMRI data. Two notable examples are the use of generative models and the real-time decoding of fMRI data for neurofeedback-based reinforcement. *Generative models* (a.k.a. forward encoding models), which characterize the tuning of individual voxels to specific perceptual or semantic features, have the potential to predict the multi-voxel activity patterns that should be associated with any potential stimulus (Naselaris, Kay, Nishimoto, & Gallant, 2011). These models have provided exquisite characterization of abstract semantic concepts (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016) and can even facilitate the reconstruction of images or movies of what a participant is currently viewing (Nishimoto et al., 2011) or imagining (Naselaris, Olman, Stansbury, Ugurbil, & Gallant, 2015) in the scanner. It is not hard to envision ways that such generative models could be productively applied to provide deeper insights into the nature of cortical memory representations. *Neurofeedback reinforcement* involves the real-time analysis of fMRI data as they are acquired—often with the use of MVPA-based classification. This can be a powerful tool for so-called closed-loop brain training (Sitaram et al., 2017), including enhancement of attentional control (deBettencourt, Cohen, Lee, Norman, & Turk-Browne, 2015), perceptual

learning (Shibata, Watanabe, Sasaki, & Kawato, 2011), and perhaps most remarkably, fear extinction learning without forcing participants to consciously confront fear-evoking stimuli (Koizumi et al., 2016). We expect that this recent marriage of real-time fMRI analysis and MVPA decoding techniques will continue to spur new advances in cognitive neuroscience and potentially also novel avenues of treatment for a range of neurological/psychiatric disorders. In sum, we feel that this is an exciting time for neuroimaging research, and we hope that our chapter has provided a helpful overview of the various experimental design and data analysis procedures available to researchers to study the neural mechanisms of learning and memory.

References

Addis, D. R., Pan, L., Vu, M. A., Laiser, N., & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, *47*, 2222–2238.

Alavash, M., Doebler, P., Holling, H., Thiel, C. M., & Giessing, C. (2015). Is functional integration of resting state brain networks an unspecific biomarker for working memory performance? *NeuroImage*, *108*, 182–193.

Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, G. L., . . . Hichwa, R. D. (1995). Remembering the past: Two facets of episodic memory explored with positron emission tomography. *American Journal of Psychiatry*, *152*, 1576–1585.

Barron, H. C., Garvert, M. M., & Behrens, T. E. (2016). Repetition suppression: A means to index neural representations using BOLD? *Philosophical Transaction of the Royal Society of London B Biological Science*, *371*(1705), 1–14.

Bassett, D. S., & Mattar, M. G. (2017). A network neuroscience of human learning: Potential to inform quantitative theories of brain and behavior. *Trends in Cognitive Sciences*, *21*, 250–264.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*, 353–364.

Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J., & Burgess, N. (2015). Consolidation of complex events via reinstatement in posterior cingulate cortex. *Journal of Neuroscience*, *35*, 14426–14434.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, *281*, 1185–1187.

Buchel, C., & Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7, 768–778.

Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping*, *6*, 373–377.

Buckner, R. L., Bandettini, P. A., OCraven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., . . . Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceeding of the National Academy of Sciences of the United States of America*, *93*, 14878–14883.

Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proceeding of the National Academy of Sciences of the United States of America*, *98*, 4805–4810.

Carr, V. A., Rissman, J., & Wagner, A. D. (2010). Imaging the human medial temporal lobe with high-resolution fMRI. *Neuron*, *65*, 298–308.

Chadwick, M. J., Bonnici, H. M., & Maguire, E. A. (2012). Decoding information in the human hippocampus: A user's guide. *Neuropsychologia*, *50*, 3107–3121.

Christophel, T. B., Hebart, M. N., & Haynes, J. D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience*, *32*, 12983–12989.

Clapp, W. C., Rubens, M. T., Sabharwal, J., & Gazzaley, A. (2011). Deficit in switching between functional brain networks underlies the impact of multitasking on working memory in older adults. *Proceeding of the National Academy of Sciences of the United States of America*, *108*, 7212–7217.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading" : Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*, 261–270.

Curtis, C. E., Rao, V. Y., & D'Esposito, M. (2004). Maintenance of spatial and motor codes during oculomotor delayed response tasks. *Journal of Neuroscience*, *24*, 3944–3952.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, *136*, 87–102.

Danker, J. F., Tompary, A., & Davachi, L. (2016). Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cerebral Cortex*, *27*, 3515–3524.

Daselaar, S. M., Prince, S. E., & Cabeza, R. (2004). When less means more: Deactivations during encoding that predict subsequent memory. *NeuroImage*, *23*, 921–927.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceeding of the National Academy of Sciences of the United States of America*, *100*, 2157–2162.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*, 279–281.

De Shetler, N. G., & Rissman, J. (2017). Dissociable profiles of generalization/discrimination in the human hippocampus during associative retrieval. *Hippocampus*, *27*, 115–121.

deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, *18*, 470–475.

Deuker, L., Bellmund, J. L. S., Schroder, T. N., & Doeller, C. F. (2016). An event map of memory space in the hippocampus. *Elife*, *5*, e16534.

Dima, D., Jogia, J., & Frangou, S. (2014). Dynamic causal modeling of load-dependent modulation of effective connectivity within the verbal working memory network. *Human Brain Mapping*, *35*, 3025–3035.

Donaldson, D. I., Petersen, S. E., Ollinger, J. M., & Buckner, R. L. (2001). Dissociating state and item components of recognition memory using fMRI. *NeuroImage*, *13*, 129–142.

Druzgal, T. J., & D'Esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience*, *15*, 771–784.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, *3*, 1149–1152.

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, *87*, 893–905.

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269.

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*, 1179–1189.

Favila, S. E., Chanales, A. J., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, 7, 11066.

Feinberg, D. A., & Setsompop, K. (2013). Ultra-fast MRI of the human brain with simultaneous multi-slice imaging. *Journal of Magnetic Resonance*, *229*, 90–100.

Ferreira, L., Regina, A., Kovacevic, N., Martin, M., Amaro, E., McIntosh, A., & Busatto, G. (2013). Global functional connectivity is related to age and memory performance in healthy adults: A resting-state fMRI study. *Journal of the Neurological Sciences*, *333*, e726.

Fiebach, C. J., Rissman, J., & D'Esposito, M. (2006). Modulation of inferotemporal cortex activation during verbal working memory maintenance. *Neuron*, *51*, 251–261.

Fornito, A., Harrison, B. J., Zalesky, A., & Simons, J. S. (2012). Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceeding of the National Academy of Sciences of the United States of America*, *109*, 12788–12793.

Fornito, A., Zalesky, A., & Breakspear, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage*, *80*, 426–444.

Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain Connectivity*, *1*, 13–36.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*, 218–229.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Friston, K. J., Moran, R., & Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology*, *23*, 172–178.

Friston, K. J., Worsley, K. J., Frackowiak, R. S., Mazziotta, J. C., & Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, *1*, 210–220.

Gabrieli, J. D., Brewer, J. B., Desmond, J. E., & Glover, G. H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science*, *276*, 264–266.

Gazzaley, A., Rissman, J., Cooney, J., Rutman, A., Seibert, T., Clapp, W., . . . D'Esposito, M. (2007). Functional interactions between prefrontal and visual association cortex contribute to top-down modulation of visual processing. *Cereb Cortex, 17 Supplement 1*, i125–135.

Gazzaley, A., Rissman, J., & D'Esposito, M. (2004). Functional connectivity during working memory maintenance. *Cognitive, Affective, and Behaviroal Neuroscience*, *4*, 580–599.

Geib, B. R., Stanley, M. L., Dennis, N. A., Woldorff, M. G., & Cabeza, R. (2017). From hippocampus to whole-brain: The role of integrative processing in episodic memory retrieval. *Human Brain Mapping*, *38*, 2242–2259.

Geib, B. R., Stanley, M. L., Wing, E. A., Laurienti, P. J., & Cabeza, R. (2017). Hippocampal contributions to the large-scale episodic memory network predict vivid visual memories. *Cereb Cortex*, *27*, 680–693.

Giovanello, K. S., Schnyer, D. M., & Verfaellie, M. (2004). A critical role for the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus*, *14*, 5–8.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–429.

Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cereb Cortex*, *24*, 3350–3364.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*, 187–203.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*, 523–534.

Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *19*, 3962–3972.

Henson, R. N., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, *287*, 1269–1272.

Honey, G. D., Fu, C. H., Kim, J., Brammer, M. J., Croudace, T. J., Suckling, J., . . . Bullmore, E. T. (2002). Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. *NeuroImage*, *17*, 573–582.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*, 453–458.

Johnson, J. D., McDuff, S. G., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, *63*, 697–708.

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: Implications for models of recognition memory. *Journal of Neuroscience*, *24*, 4172–4180.

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage*, *54*, 2446–2461.

King, D. R., de Chastelaine, M., Elward, R. L., Wang, T. H., & Rugg, M. D. (2015). Recollection-related increases in functional connectivity predict individual differences in memory accuracy. *Journal of Neuroscience*, *35*, 1763–1772.

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., . . . Lau, H. (2016). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature Human Behavior*, *1*, 0006.

Konishi, S., Wheeler, M. E., Donaldson, D. I., & Buckner, R. L. (2000). Neural correlates of episodic retrieval success. *NeuroImage*, *12*, 276–286.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceeding of the National Academy of Sciences of the United States of America*, *103*, 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*, 535–540.

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, *56*, 455–475.

Kuhl, B. A., & Chun, M. M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *Journal of Neuroscience*, *34*, 8051–8060.

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceeding of the National Academy of Sciences of the United States of America*, *108*, 5903–5908.

LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *Journal of Neuroscience*, *33*, 5466–5474.

Lee, S. H., & Baker, C. I. (2016). Multi-voxel decoding and the topography of maintained information during visual working memory. *Frontiers in Systems Neuroscience*, *10*, 2.

Levy, B. J., & Wagner, A. D. (2013). Measuring memory reactivation with functional MRI: Implications for psychological theory. *Perspectives on Psychological Science*, *8*, 72–78.

Linden, D. E., Bittner, R. A., Muckli, L., Waltz, J. A., Kriegeskorte, N., Goebel, R., . . . Munk, M. H. J. (2003). Cortical capacity constraints for visual working memory: Dissociation of fMRI load effects in a fronto-parietal network. *NeuroImage*, *20*, 1518–1530.

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex*, *4*, 523–531.

Ma, L., Steinberg, J. L., Hasan, K. M., Narayana, P. A., Kramer, L. A., & Moeller, F. G. (2012). Working memory load modulation of parieto-frontal connections: Evidence from dynamic causal modeling. *Human Brain Mapping*, *33*, 1850–1867.

Maass, A., Schütze, H., Speck, O., Yonelinas, A., Tempelmann, C., Heinze, H. J., . . . Düzel, E. (2014). Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding. *Nature communications*, *5*, 5547.

Mack, M. L., & Preston, A. R. (2016). Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. *NeuroImage*, *127*, 144–157.

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., . . . Maquet, P. (2006). The left intraparietal sulcus and verbal short-term memory: Focus of attention or serial order? *Neuroimage*, *32*, 880–891.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, *3*, 143–157.

McIntosh, A. R., Chau, W., & Protzner, A. B. (2004). Spatiotemporal analysis of event-related fMRI data using partial least squares. *NeuroImage*, *23*, 764–775.

Mclntosh, A., & Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, *2*, 2–22.

Milham, M. P. (2012). Open neuroscience solutions for the connectome-wide association era. *Neuron*, *73*, 214–218.

Miller, B. T., & D'Esposito, M. (2012). Spatial and temporal dynamics of cortical networks engaged in memory encoding and retrieval. *Frontiers in Human Neuroscience*, *6*, 109.

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

Monti, M. M. (2011). Statistical analysis of fMRI time-series: A critical review of the GLM approach. *Frontiers in Human Neuroscience*, *5*, 28.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*, 400–410.

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228.

Nee, D. E., & Jonides, J. (2014). Frontal-medial temporal interactions mediate transitions among representational states in short-term memory. *Journal of Neuroscience*, *34*, 7964–7975.

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., . . . Yeo, B. T. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*, 299–303.

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceeding of the National Academy of Sciences of the United States of America*, *112*, 11078–11083.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, *10*, e1003553.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, *21*, 1641–1646.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*, 604–609.

Okado, Y., & Stark, C. (2003). Neural processing associated with true and false memory retrieval. *Cognitive, Affective, and Behavioral Neuroscience*, *3*, 323–334.

Otten, L. J., Henson, R. N. A., & Rugg, M. D. (2002). State-related and item-related neural correlates of successful memory encoding. *Nature Neuroscience*, *5*, 1339–1344.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Modelling functional integration: A comparison of structural equation and dynamic causal models. *NeuroImage*, *23*, S264-S274.

Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron*, *35*, 975–987.

Petersen, S. E., & Dubis, J. W. (2012). The mixed block/event-related design. *NeuroImage*, *62*, 1177–1184.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*, 67–70.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafo, M. R., . . . Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*, 115–126.

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature*, *526*, 371–379.

Poline, J. B., & Brett, M. (2012). The general linear model and fMRI: Does love last forever? *NeuroImage*, *62*, 871–880.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*, 1963–1966.

Postle, B. R. (2016). How does the brain keep information "in mind"? *Current Directions in Psychological Science*, *25*, 151–156.

Postle, B. R., Zarahn, E., & D'Esposito, M. (2000). Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. *Brain Research Protocols*, *5*, 57–66.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., . . . Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*, 665–678.

Quamme, J. R., Weiss, D. J., & Norman, K. A. (2010). Listening for recollection: A multi-voxel pattern analysis of recognition memory retrieval strategies. *Frontiers in Human Neuroscience*, 4, 61.

Ranganath, C., DeGutis, J., & D'Esposito, M. (2004). Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Cognitive Brain Research*, *20*, 37–45.

Ranganath, C., Heller, A., Cohen, M. X., Brozinsky, C. J., & Rissman, J. (2005). Functional connectivity with the hippocampus during successful memory formation. *Hippocampus*, *15*, 997–1005.

Richter, F. R., Chanales, A. J. H., & Kuhl, B. A. (2016). Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *NeuroImage*, *124*, 323–335.

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *Journal of Neuroscience*, *32*, 12990–12998.

Rissman, J., Chow, T. E., Reggente, N., & Wagner, A. D. (2016). Decoding fMRI signatures of real-world autobiographical memory retrieval. *Journal of Cognitive Neuroscience*, *28*, 604–620.

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, *15*, 1160–1175.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*, 752–763.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2008). Dynamic adjustments in prefrontal, hippocampal, and inferior temporal interactions with increasing visual working memory load. *Cerebral Cortex*, *18*, 1618–1629.

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*, 1637–1646.

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

Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: Insights from functional brain imaging. *Annual Review of Psychology*, *63*, 101–128.

Ritchey, M., Dolcos, F., & Cabeza, R. (2008). Role of amygdala connectivity in the persistence of emotional memories over time: An event-related fMRI investigation. *Cerebral Cortex*, *18*, 2494–2504.

Roebroeck, A., Formisano, E., & Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, *25*, 230–242.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52, 1059–1069.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*, 255–260.

Salami, A., Pudas, S., & Nyberg, L. (2014). Elevated hippocampal resting-state connectivity underlies deficient neurocognitive function in aging. *Proceeding of the National Academy of Sciences of the United States of America*, *111*, 17654–17659.

Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, *15*, 1332–1342.

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. *Psychophysiology*, *17*, 568–576.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*, 185–195.

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology*, *17*, 171–176.

Schedlbauer, A. M., Copara, M. S., Watrous, A. J., & Ekstrom, A. D. (2014). Multiple interacting brain areas underlie successful spatiotemporal memory retrieval in humans. *Scientific Reports*, *4*, 6431.

Schlösser, R., Gesierich, T., Kaufmann, B., Vucurevic, G., Hunsche, S., Gawehn, J., & Stoeter, P. (2003). Altered effective connectivity during working memory performance in schizophrenia: A study with fMRI and structural equation modeling. *Neuroimage*, *19*, 751–763.

Schlösser, R. G. M., Wagner, G., & Sauer, H. (2006). Assessing the working memory network: Studies with functional magnetic resonance imaging and structural equation modeling. *Neuroscience*, *139*, 91–103.

Schott, B. H., Richardson-Klavehn, A., Henson, R. N., Becker, C., Heinze, H. J., & Duzel, E. (2006). Neuroanatomical dissociation of encoding processes related to priming and explicit memory. *Journal of Neuroscience*, *26*, 792–800.

Schott, B. H., Wüstenberg, T., Wimber, M., Fenker, D. B., Zierhut, K. C., Seidenbecher, C. I., . . . Richardson-Klavehn, A. (2013). The relationship between level of processing and hippocampal—cortical functional connectivity during episodic memory formation in humans. *Human Brain Mapping*, *34*, 407–424.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*, 207–214.

Sheldon, S., Farb, N., Palombo, D. J., & Levine, B. (2016). Intrinsic medial temporal lobe connectivity relates to individual differences in episodic autobiographical remembering. *Cortex*, *74*, 206–216.

Sheldon, S., & Levine, B. (2013). Same as it ever was: Vividness modulates the similarities and differences between the neural networks that support retrieving remote and recent autobiographical memories. *NeuroImage*, *83*, 880–891.

Shen, J., Zhang, G., Yao, L., & Zhao, X. (2015). Real-time fMRI training-induced changes in regional connectivity mediating verbal working memory behavioral performance. *Neuroscience*, *289*, 144–152.

Shibata, K., Watanabe, T., Sasaki, Y., & Kawato, M. (2011). Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science*, *334*, 1413–1415.

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., . . . Birbaumer, N. (2017). Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*, *18*(2), 86.

Slotnick, S. D., & Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience*, *7*, 664–672.

Smith, A. P., Stephan, K. E., Rugg, M. D., & Dolan, R. J. (2006). Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. *Neuron*, *49*, 631–638.

Smith, K. (2013). Brain decoding: Reading minds. *Nature*, *502*, 428–430.

Sporns, O., & Betzel, R. F. (2016). Modular brain networks. *Annual Review of Psychology*, *67*, 613–640.

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W., & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, *53*, 303–317.

St Jacques, P. L., Kragel, P. A., & Rubin, D. C. (2011). Dynamic neural networks supporting memory retrieval. *NeuroImage*, *57*, 608–616.

Staresina, B. P., Cooper, E., & Henson, R. N. (2013). Reversible information flow across the medial temporal lobe: The hippocampus links cortical modules during memory retrieval. *Journal of Neuroscience*, *33*, 14184–14192.

Stark, C. E., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceeding of the National Academy of Sciences of the United States of America*, *98*, 12760–12766.

Stephan, K. E., Li, B. J., Iglesias, S., & Friston, K. J. (2015). Inferring effective connectivity from fMRI Data. *Biological Magnetic Resonance*, *30*, 365–386.

Stern, C. E., Corkin, S., González, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., . . . Rosen, B. R. (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *93*, 8660–8665.

Stevens, A. A., Tappon, S. C., Garg, A., & Fair, D. A. (2012). Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. *PLoS One*, *7*, e30468.

Tompary, A., Duncan, K., & Davachi, L. (2016). High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. *Hippocampus*, *26*, 995–1007.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, *63*, 483–509.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

Turk-Browne, N. B., Yi, D. J., & Chun, M. M. (2006). Linking implicit and explicit memory: Common encoding factors and shared representations. *Neuron*, *49*, 917–927.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79.

Velanova, K., Jacoby, L. L., Wheeler, M. E., McAvoy, M. P., Petersen, S. E., & Buckner, R. L. (2003). Functional-anatomic correlates of sustained and transient processing components engaged during controlled retrieval. *Journal of Neuroscience*, *23*, 8460–8470.

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

Wardlaw, J. M., O'Connell, G., Shuler, K., DeWilde, J., Haley, J., Escobar, O., . . . Schafer, B. (2011). "Can it read my mind?" —What do the public and experts think of the current (Mis)uses of neuroimaging? *Plos One*, *6*, e25829.

Westphal, A. J., Reggente, N., Ito, K. L., & Rissman, J. (2016). Shared and distinct contributions of rostrolateral prefrontal cortex to analogical reasoning and episodic memory retrieval. *Human Brain Mapping*, *37*, 896–912.

Westphal, A. J., Wang, S., & Rissman, J. (2017). Episodic memory retrieval benefits from a less modular brain network organization. *Journal of Neuroscience*, *37*, 3523–3531.

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*, 582–589.

Wing, E. A., Ritchey, M., & Cabeza, R. (2015). Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *Journal of Cognitive Neuroscience*, *27*, 679–691.

Woo, C. W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage*, *91*, 412–419.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.

Xue, G., Dong, Q., Chen, C. S., Lu, Z. L., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *330*, 97–101.

Zarahn, E., Aguirre, G., & D'Esposito, M. (1997). A trial-based experimental design for fMRI. *NeuroImage*, *6*, 122–138.

Zou, Q. H., Ross, T. J., Gu, H., Geng, X. J., Zuo, X. N., Hong, L. E., . . . Yang, Y. (2013). Intrinsic resting-state activity predicts working memory brain activation and behavioral performance. *Human Brain Mapping*, *34*, 3204–3215.