The Dissertation Committee for Margaret Lynn Schlichting

certifies that this is the approved version of the following dissertation:

When Memories Relate: Medial Temporal and Prefrontal

Contributions to Memory Integration and Inference

Committee:

<div>

Alison R. Preston, Supervisor

Jessica A. Church-Lang

Laura L. Colgin

Jarrod A. Lewis-Peacock

Russell A. Poldrack

David M. Schnyer

</div>

When Memories Relate: Medial Temporal and Prefrontal

Contributions to Memory Integration and Inference

by

Margaret Lynn Schlichting, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2015

Dedication

To my husband Mike,

for teaching me to celebrate the little victories

along the way.

Acknowledgements

First and foremost, I would like to thank my advisor Dr. Alison Preston for her unwavering support, guidance, and encouragement throughout my graduate career. She is not only a wonderfully talented scientist, but also a deeply caring mentor. I am incredibly fortunate to have been trained by such a fantastic researcher so invested in my success.

Thank you to my committee—Drs. Jessica Church-Lang, Laura Colgin, Jarrod Lewis-Peacock, David Schnyer, and Russell Poldrack—for their time and helpful comments on these studies. This dissertation was improved as a direct result of their feedback and for that I am extremely grateful.

This work would not have been possible without the generous support of Preston Lab members past and present. In particular, I would like to thank April Dominick and Amelia Wattenberger for their invaluable assistance with participant recruitment and data collection. Thank you to Jackson Liang and Michael Mack, who provided me with hours' worth of helpful feedback and practical guidance in carrying out these projects. Very special thanks to Dasa Zeithamova, who worked especially closely with me during the early stages of my graduate career. I often say that Dasa taught me everything I know; but in truth, she taught me something much more important: how to teach myself. Her intellect and attention to detail are unmatched, and much of my knowledge and accomplishments are thanks to her.

My friends and family have been incredible sources of support throughout graduate school. Thank you to my best friend Marsha Jacobskind and sisters Beth

When Memories Relate: Medial Temporal and Prefrontal Contributions to Memory Integration and Inference

by

Margaret Lynn Schlichting, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Alison R. Preston

Memory is of immeasurable importance to the human experience. It has been known for decades that memory for individual events is supported by the medial temporal lobes (MTL), which include the hippocampus and adjacent cortex. Emerging research suggests that by retrieving related prior experiences during new learning, connections can also be formed among memories to create knowledge that spans events. Such integration mechanisms might be further influenced by memory models stored in medial prefrontal cortex (PFC), which serve to guide learning-phase retrieval and integrate across related memories. In contrast, lateral PFC might maintain separate representations for related events, consistent with its role in resolving interference.

This dissertation used functional magnetic resonance imaging (fMRI) in humans to investigate the MTL and PFC mechanisms that link content across episodes. The first experiment investigated the contributions of hippocampal subfields to this encoding mechanism. Consistent with its hypothesized role in detecting inconsistencies and

integrating across memories, integration signatures were isolated to the hippocampal CA_1 subfield. The second experiment interrogated two aspects of offline processes that promote integration. First, neural evidence for reinstatement of initially learned content and enhanced hippocampal communication with content-specific visual regions during rest was associated with a behavioral index of integration. Moreover, enhanced functional connectivity between hippocampus and medial PFC both during and immediately following learning of related information was associated with better integration. This relationship was mirrored in hippocampal-medial PFC white matter integrity. The third experiment interrogated the nature of the hippocampal and prefrontal representations that underlie memory integration. Results revealed dissociable integration and separation signatures in hippocampus and PFC, highlighting how neural representations of memory elements can simultaneously promote integration across related events and protect from interference. In line with computational theory, these effects were also modulated by the manner in which events were experienced.

Taken together, these studies provide insight into the neural mechanisms supporting the dynamic interactions among related memories. More broadly, this dissertation represents an important shift in the scientific study of memory, from exploring memory for individual events to investigating how memories may be derived across experiences to support appropriate action in novel situations.

Table of Contents

xiii

List of Tables

List of Figures

BACKGROUND AND AIMS

1. Introduction

1.1. OVERVIEW

Memory is of immeasurable importance to the human experience. Accordingly, how our brains capture, store, and recall information has long received attention from psychologists and neuroscientists alike. Of particular relevance to the present dissertation is episodic memory, or memory for specific events. Normal episodic memory function is supported by the medial temporal lobe (MTL) region of the brain (Scoville, 1954; Scoville and Milner, 1957; Eichenbaum et al., 1992), which includes the hippocampus and immediately adjacent brain tissue (i.e., MTL cortex). It has been known for decades that individuals who suffer damage to the MTL through injury or disease exhibit profound memory impairments (Scoville, 1954; Scoville and Milner, 1957). However, the inner workings of this complex neural circuit are only recently starting to be unraveled. The research described here investigates the neural mechanisms of the MTL and its interactions with prefrontal cortex (PFC) that enable memories for related events to be combined across time.

Despite the discipline's longstanding interest in the study of memory, the vast majority of existing research has focused on MTL binding and prefrontal top-down control processes that link information presented within individual events. While this body of work has contributed much to our understanding of MTL and PFC memory mechanisms, it largely overlooks an important aspect of real-world memories: they are not isolated from one another. Rather, memories for related events are interactive,

forming complex memory networks that in turn support a wide range of cognitive abilities. This dissertation builds upon the current state of knowledge in the field to characterize how memories go beyond representing direct experience to be maximally adaptive and inform new behaviors. More broadly, this research represents an important shift in the scientific study of memory, from exploring memory for individual events to investigating how memories may be derived across experiences to support appropriate action in novel situations.

Below, I briefly describe the MTL and prefrontal structures and computations thought to be involved in the encoding and retrieval of individual events. I then review the literature documenting the involvement of hippocampus and PFC in novel inferences requiring consideration of multiple discrete events (Zeithamova et al., 2012b; Schlichting and Preston, 2015). Finally, I discuss open questions and outline the specific aims of the research presented in this dissertation.

1.2. THE MTL AND EVENT MEMORY

The MTL comprises the hippocampal formation and the adjacent perirhinal, parahippocampal, and entorhinal cortices (abbreviated PRc, PHc, and ERc, respectively) (**Fig. 1**). Its organization is hierarchical (Lavenex and Amaral, 2000), such that sensory input from PRc and PHc converges on lateral and medial ERc, respectively (Eichenbaum et al., 2007), which then send projections to the hippocampus. Both PRc and PHc have extensive anatomical connections with sensory association areas. PRc receives its primary input from unimodal visual areas in inferior temporal cortex thought to be involved in processing object features (Suzuki and Amaral, 1994; Eichenbaum et al.,

2007). In contrast, PHc receives input from polymodal association areas, including visual association cortex and posterior parietal cortex, which have been implicated in processing of spatial information (Suzuki and Amaral, 1994; Eichenbaum et al., 2007). Thus, the hippocampus receives converging information from ERc about both the "what" (e.g., stimulus features, from PRc) and "where" (e.g., spatial layout, from PHc) of an episode. The hippocampus sits at the top of this hierarchy as a convergence area (Van Hoesen et al., 1975; Amaral et al., 1987; Witter et al., 1989; Suzuki and Amaral, 1990; Witter and Amaral, 1991), receiving simultaneous input about these different aspects of experience. For this reason, the hippocampus has been implicated in binding disparate elements within an episode into a single memory trace (Davachi, 2006; Diana et al., 2007).
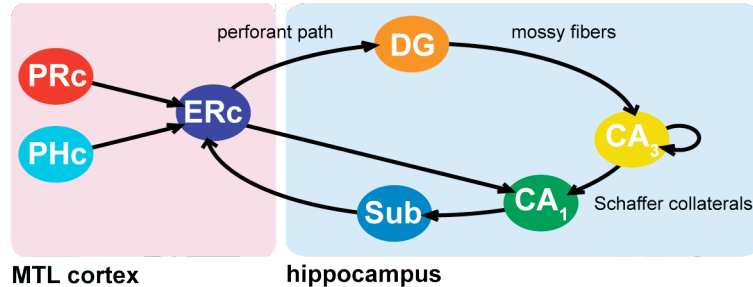
Figure 1. MTL Circuitry

Simplified depiction of MTL circuitry. Sensory information enters the hippocampus from ERc, which receives inputs from both PRc and PHc. MTL cortex (red box) includes PRc, PHc and ERc; hippocampus (blue box) comprises subfields DG, CA_1-CA_4, and subiculum (Sub).

The hippocampus may be further broken down into subfields: the dentate gyrus (DG), cornu ammonis (CA) fields 1-4 (for simplicity, the present discussion focuses on CA_1 and CA_3), and subiculum. Circuitry within the hippocampus follows a more or less

circular trajectory known as the trisynaptic pathway (**Fig. 1**). The majority of input to the hippocampus is received from ERc, which projects to DG via the perforant path (Amaral and Insausti, 1990; Lavenex and Amaral, 2000). DG is thought to be involved in *pattern separation*, a process by which memory representations for similar events are created in a sparse, orthogonalized manner (Treves and Rolls, 1994). This mechanism is thought to protect from interference, allowing retrieval of specific details among highly overlapping memories (McClelland et al., 1995; O'Reilly and Rudy, 2001; O'Reilly and Norman, 2002).

DG then projects to CA_3 via the mossy fibers. Theoretical and computational models (Marr, 1971; McNaughton and Morris, 1987) suggest that CA_3 is critical for the associative binding processes that link disparate elements experienced within an episode. Consistent with these models, human neuroimaging studies have shown that a combined dentate gyrus and $CA_{2,3}$ ($DG/CA_{2,3}$) region is active during encoding of associative information (Zeineh et al., 2003; Eldridge et al., 2005; Wolosin et al., 2012). This function of CA_3 is thought to be due to its extensive network of recurrent collaterals, in which CA_3 pyramidal neurons project to other cells in the region. Specifically, it is hypothesized that the recurrent collaterals enable the hippocampus to code for conjunctions among multiple stimuli and context that occur within a single event (Eichenbaum et al., 1996), represented in the brain as patterns of activation distributed throughout neocortex. These so-called conjunctive representations maintain information about individual elements in an experience while also coding the relationships among them (Eichenbaum et al., 1996). It is this representational scheme in CA_3 that is thought

to enable *pattern completion*, or retrieval of previously stored memory representations from partial or degraded input (Mizumori et al., 1989; O'Reilly and Rudy, 2001; Lisman and Grace, 2005; Bakker et al., 2008).

CA_1 receives input from CA_3 representing information retrieved from memory (via the Schaffer collaterals), as well as direct projections from ERc representing current sensory experience (as part of the monosynaptic pathway; Amaral and Insausti, 1990; Duvernoy, 1998). The convergence of these distinct inputs on CA_1 is thought to enable one of its important hypothesized functions: signaling the degree of mismatch between current events and memory-based expectations (O'Reilly and Rudy, 2001; Lee et al., 2004). One possible function of this signal is to switch hippocampal processing between "encoding mode" and "retrieval mode" (Duncan et al., 2012). That is, when present experience is consistent with expectations (i.e., a "match"), the hippocampus is biased to retrieve prior experiences. In contrast, sensory input that violates expectations may trigger new encoding, in which the novel content is either incorporated into an existing memory trace (Shohamy and Wagner, 2008) or laid down in a new, pattern separated representation (Hasselmo and Schnell, 1994).

CA_1 then projects to subiculum, the major output structure of the hippocampus. Subiculum is thought to be primarily involved in episodic retrieval (Gabrieli et al., 1997; Zeineh et al., 2003; Eldridge et al., 2005). However, other studies have shown that it is modulated by stimulus novelty (Zeineh et al., 2000; Bakker et al., 2008), suggesting that it may also contribute to encoding (Preston et al., 2010). CA_1 and subiculum both send projections to ERc.

5

1.3. PREFRONTAL CONTRIBUTIONS TO MEMORY

The temporal lobes are densely anatomically connected with prefrontal regions through white matter pathways including the uncinate fasiculus and fornix (Simons and Spiers, 2003). Of particular relevance to this dissertation are the MTL connections with the ventromedial and ventrolateral aspects of PFC in particular.

While much work has documented the role of medial PFC (MPFC) in domains such as processing of value (McClure et al., 2004; De Araujo et al., 2005; Plassmann et al., 2008; Boorman et al., 2009; Chib et al., 2009; Hare et al., 2009; Kumaran et al., 2009) and emotion (Cacioppo et al., 1992; Cacioppo and Gardner, 1999; Barrett, 2006; Wager et al., 2008; Etkin et al., 2011), its role in episodic memory has just recently started to attract research attention (van Kesteren et al., 2010a, 2012; Roy et al., 2012; Preston and Eichenbaum, 2013; Schlichting and Preston, 2015). Medial PFC (MPFC) receives direct inputs from hippocampal area CA_1 (Rosene and Van Hoesen, 1977; Ferino et al., 1987; Barbas and Blatt, 1995; Roberts et al., 2007; Navawongse and Eichenbaum, 2013) and projects back to hippocampus indirectly via the nucleus reuniens of the thalamus (Vertes et al., 2007; Xu and Südhof, 2013). These anatomical connections are consistent with an important role of MPFC in MTL-based memory processes. Among other impairments, damage to ventral aspects of MPFC has been associated with memory distortions (Burgess and Shallice, 1996) and an inability to generalize across events (Ghosh et al., 2014; Warren et al., 2014). Moreover, MPFC has been shown to be engaged when strong memories are formed or used to guide encoding of new information (Tse et al., 2011; Zeithamova et al., 2012a). Such findings have led to the hypothesis that

MPFC is involved in promoting memory integration, which will be discussed in detail below.

In contrast to MPFC, ventrolateral PFC (i.e., inferior frontal gyrus [IFG]) receives only indirect projections from hippocampus through the fornix via ERc (Barbas and Blatt, 1995) and from posterior PHc (Roberts et al., 2007). Moreover, projections from IFG back to hippocampus are scarce. Reciprocal connections between MTL cortex and other areas of PFC (Rempel-Clower and Barbas, 2000) would, however, allow IFG to indirectly influence MTL encoding and retrieval given the dense interconnectedness of PFC regions (Roberts et al., 2007). Consistent with this notion, IFG engagement during encoding has been shown to be predictive of subsequent memory (Wagner et al., 1998; Uncapher and Rugg, 2005; Straube et al., 2009; Kuhl et al., 2012b). Moreover, patients with lateral PFC damage show impairments in recalling episodic details such as source information (Schacter et al., 1984; Shimamura et al., 1990; Simons et al., 2002). Thus, one hypothesized role of this region is a control process that selects relevant features for encoding and retrieval (Fletcher and Henson, 2001; Badre et al., 2005; Badre and Wagner, 2007). Such a process is thought to result in distinct memory traces for specific events, thereby resolving competition and protecting from interference (Paller and Wagner, 2002; Raposo et al., 2009; Ranganath, 2010; Kuhl et al., 2012a; Preston and Eichenbaum, 2013).

By engaging these distinct computations, the MTL-PFC circuit as a whole is known to support the encoding and retrieval of individual event memories. Yet, the episodic memory system in the real world is not tasked with simply encoding a series of

individual, isolated events. Rather, we make novel judgments that require simultaneous consideration of a number of highly interrelated experiences, demonstrating the importance of a memory system that takes advantage of commonalities across experiences. I next review empirical work suggesting that both the hippocampus and PFC are important for combining related memories across discrete episodes (Zeithamova et al., 2012b).

2. Linking Memories Across Time[1]

2.1. PROSPECTIVE USE OF MEMORY

The function of the episodic memory system is not simply to form and store memories for individual events, but rather to additionally code their relationships to one another. These representations can go on to support judgments that require inference based on knowledge acquired across multiple distinct experiences—as is often the case for decisions in our everyday lives. For example, imagine you encounter an unfamiliar woman leaving the house next door to walk her dog in a nearby park. Because the house was recently sold, you might conclude that the woman and her dog are your new neighbors. Several days later, you are in town and see the same dog again out for a walk, this time with a man. From the knowledge acquired on these two separate occasions, you may infer a relationship between the woman and man; for instance, you may deduce that they are a couple and recently moved into the house next door with their dog.

Successful inference may thus depend on our ability to recall detailed information from past events to determine the relationship among items experienced at different times. A growing body of literature indicates that such flexibility to combine experiences in novel ways to infer unobserved relationships between items or events crucially depends on the hippocampus and PFC. The ability to infer a relationship between two previously experienced events is complex, involving several distinct operations. While

---

[1] Portions of this chapter were adapted from Zeithamova D[+], Schlichting ML[+], Preston AR (2012b) The hippocampus and inferential reasoning: Building memories to navigate future decisions. Front Hum Neurosci 6:70 and Schlichting ML, Preston AR (2015) Memory integration: neural mechanisms and implications for behavior. Curr Opin Behav Sci 1:1–8. [+] Authors made equal contributions. D Zeithamova contributed to the writing of the first review paper. AR Preston contributed to the writing of both review papers.

some of these operations rely predominantly on brain structures outside the MTL such as PFC, others necessitate hippocampal processing. For example, for successful inference to ultimately take place, the arbitrary relations among previously unrelated elements within an event must be encoded (e.g., individual dog—woman and dog—man associations from the example above).

To extract new information about the relationship between these events, encoded associations must be retrieved and then manipulated, recombined, and recoded based on their content to support the inference itself (woman—man). The precise contribution of hippocampus and PFC to each of these processes remains an area of active investigation. Notably, while these operations are all requisite for inference, the relative role of the hippocampus and PFC in these different processes—and the relative timing of each— may depend on the particular demands of the task at hand.

An important factor contributing to success in a variety of inference tasks is the nature of the underlying hippocampal memory representations. Decades of research have characterized how the hippocampus builds rich, detailed records of individual events, or episodic memories (Eichenbaum and Cohen, 2001; Squire et al., 2004; Eichenbaum et al., 2007; Preston and Wagner, 2007). Hippocampal memory representations are well-suited for the particular demands of inferential reasoning tasks. As described previously, the hippocampus is thought to form conjunctive representations, which discretely code multiple event elements in terms of their relationships to one another (Cohen and Eichenbaum, 1993; O'Reilly and Rudy, 2001). Such conjunctive coding allows hippocampal representations to be flexible, as relevant details of past events can be

individually accessed as needed (i.e., through pattern completion) to support the types of novel decisions required in inferential reasoning.

Here, I discuss the developing literature linking hippocampal memory processes and representations to successful inference in a variety of tasks. By reviewing findings from both human and animal research, I highlight the hippocampal and prefrontal mechanisms that underlie mnemonic flexibility during both encoding and retrieval. I also argue that inference provides a useful means of exploring the adaptive nature of memory, whereby memory representations are used to successfully negotiate current behavior and anticipate future decisions and actions.

2.2. INFERENCE TASKS

The novel expression of learned information has many forms, ranging from generalization of conditioned responses to novel stimuli in animals (Pavlov, 1927) to transfer of a learned task structure to new perceptual settings in humans (Kumaran et al., 2009). Here, I focus specifically on the mechanisms supporting the novel application of memory during inferential reasoning tasks that require judgments about the relationships between items experienced across discrete episodes.

Several paradigms have been used to study the role of the hippocampus in inferential reasoning (**Fig. 2**), of which the most widely used in both animal (e.g., Dusek & Eichenbaum, 1997) and human research (e.g., Heckers, Zalesak, Weiss, Ditman, & Titone, 2004) is the transitive inference task (**Fig. 2A**). In this task, participants learn a set of overlapping premise relationships (e.g., A > B, B > C, C > D, D > E, E > F, where A > B indicates A is reinforced over B) via trial and error. During this initial training

11

phase, participants learn to select the correct (i.e., reinforced) item. Training typically continues until a criterion level of performance on premise associations is reached.

Notably, multiple types of representations may support learning in this task. For instance, knowledge of reinforcement histories alone may guide memory for the end items of the hierarchy (A is always rewarded, while F is never rewarded) and individual conditional associations may support memory for the inner pairs in the hierarchy (e.g., C is rewarded in the context of D, but not in the context of B). Alternatively, all items may be represented simultaneously as an ordered hierarchy of relationships (A > B > C > D > E > F) that concisely represents trained associations as well as information about the relationships between items that were not directly trained (e.g., B > D). To assess which types of representations support performance as well as how they may depend on the hippocampus, knowledge of the premise associations is tested together with novel, untrained combinations of items during the critical test phase. Novel test trials include inferential pairs with one degree of separation between items (e.g., B ? D, C ? E), as well as pairs with two degrees of separation (e.g., B ? E), depending on the total number of items in the hierarchy. Novel non-inferential pairings consisting of the end items of the hierarchy (A ? F) are also tested.
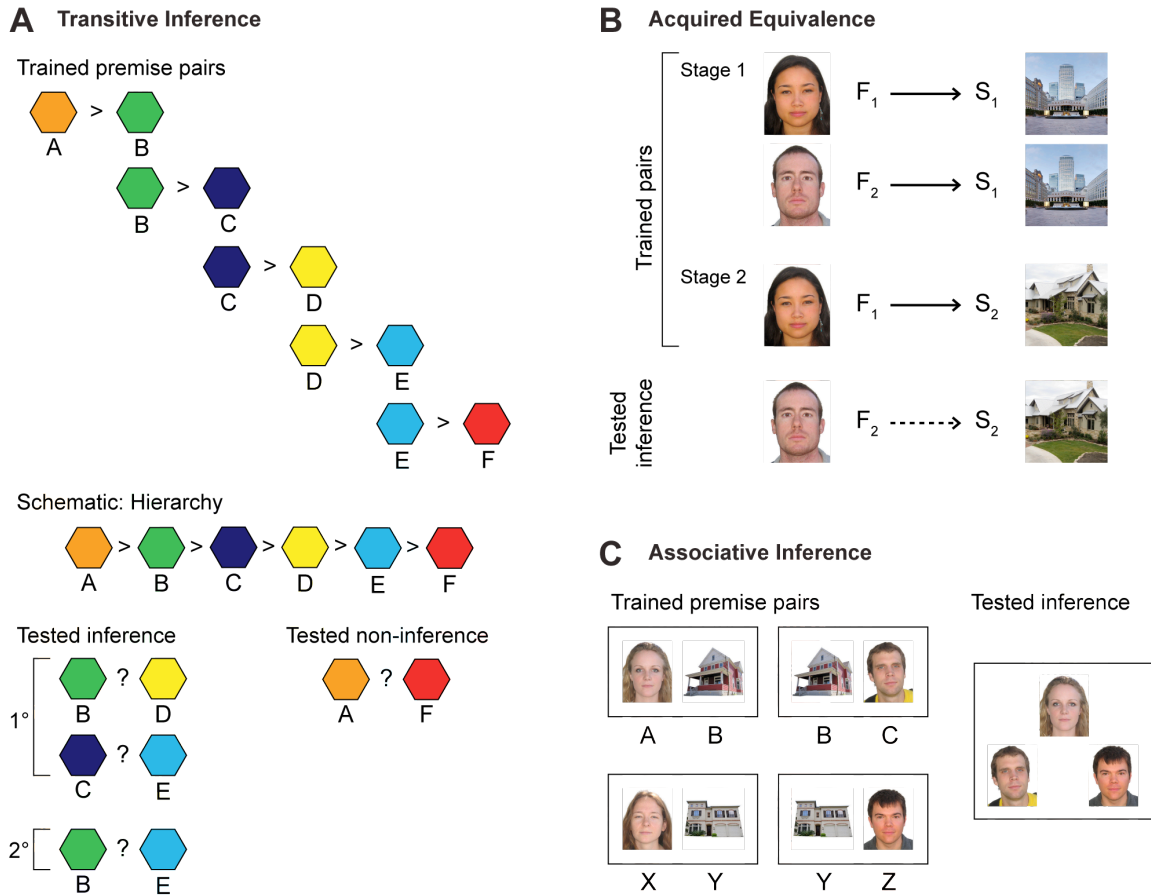
Figure 2. Inference Tasks

(**A**) Transitive inference task with six elements. A set of overlapping training pairs forms an ordered hierarchy of relationships. Participants learn each individual training pair via feedback-based learning (e.g., A > B) and are then tested on novel inference and novel non-inference judgments. Items in inferential probe trials may be separated by one element in the hierarchy (e.g., B ? D, indicated as 1°) or two elements (e.g., B ? E, indicated as 2°). Novel non-inferential probes test knowledge of the relationship between the end items of the hierarchy (A ? F). (**B**) Acquired equivalence task. In stage one of training, participants are trained via feedback to associate two faces (F_1 and F_2) with a particular scene (S_1). In stage two, participants learn to select a second scene (S_2) when cued with one of the faces (F_1). Inference is then measured as the proportion of trials on which participants choose S_2 when cued with F_2. The schematic depicts trained stimulus–response relationships (solid black arrows) and inferential relationships (dashed black arrows). (**C**) Associative inference task. Participants learn an overlapping set of associations (here, face–house associations), in which two stimuli (a man and a woman) are associated with a common third item (a house). Novel inference trials evaluate knowledge for the indirect relationship between items (who lives together in the same house).

13

Critically, correct performance on the inferential test trials can only be achieved by considering the overarching hierarchy of relationships because both items (e.g., B and D) are reinforced at the same rate during training. In contrast, success on non-inferential test trials involving the end items can be based solely on the reinforcement histories and does not require reference to the hierarchical relationships. This difference in task demands may be reflected by differing degrees of hippocampal recruitment for inferential and non-inferential test probes. Specifically, hippocampal engagement may be unique to inferential judgments that require reference to an ordered hierarchy of stimuli.

A second task used to examine the role of the hippocampus in inference is the acquired equivalence task (**Fig. 2B**) (Myers et al., 2003; Shohamy and Wagner, 2008). In this task, participants learn a set of stimulus–response associations via feedback training, such as learning to select $Scene_1$ when cued with $Face_1$. In the first stage, the stimulus–response relationships are organized such that two distinct cue stimuli are associated with the same response (e.g., $Face_1$—$Scene_1$, $Face_2$—$Scene_1$). In the second training stage, additional information is learned about one of those cue items (e.g., $Face_1$—$Scene_2$) that can also be inferred to be true for the second, equivalent item (e.g., $Face_2$). Thus, unlike the transitive inference task, the overlapping associations in the acquired equivalence task do not form a logical ordered hierarchy. Inference in the acquired equivalence task is assessed by testing participants' knowledge of novel, untrained associations ($Face_2$–$Scene_2$) that can be inferred through transfer of learned associations ($Face_1$—$Scene_2$) to the equivalent item. While premise pairs may be encoded as inflexible stimulus–response associations, only representations that encode discrete event elements together with the

relationships among them—such as conjunctive representations formed by the hippocampus—are thought to support the acquired equivalence judgment.

Another commonly used inferential reasoning paradigm is the associative inference task (**Fig. 2C**) (e.g., Bunsey and Eichenbaum, 1996; Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a; Schlichting and Preston, 2014; Schlichting et al., 2014). In associative inference, stimuli are organized into groups of three and presented to participants as overlapping pairs (e.g., AB, BC and XY, YZ, where AB denotes a studied association between items A and B) using either feedback-based or observational training. Inferential performance is then assessed by asking participants to make judgments about the relationship between elements of overlapping pairs that were not explicitly studied together (e.g., AC, XZ). As in other inferential tasks, the premise associations could be encoded as unitized representations, such as an "AB" unit during observational learning or an A—B stimulus–response association during feedback training. However, like acquired equivalence judgments, associative inference would be supported only by conjunctive hippocampal representations that enable flexible access to individual event details. Because the associative inference task can employ observational learning procedures, it provides an additional flexibility in research design. While training in the transitive inference and acquired equivalence tasks is typically limited to a small set of overlapping associations learned across multiple exposures, the associative inference task can be performed using a larger number of associations and single-trial learning procedures, wherein each trained association is only seen once during the learning phase (Zeithamova and Preston, 2010; Schlichting and Preston, 2014;

15

Schlichting et al., 2014). Such rapid acquisition of arbitrary information is characteristic of the type of learning that occurs during daily episodic experiences and provides a means for studying how inference is performed with limited direct experience.

2.3. CRITICAL ROLE FOR THE HIPPOCAMPUS IN INFERENTIAL REASONING

Converging evidence from animal and human research indicates that the hippocampus is necessary for successful performance in inferential reasoning tasks. In a series of animal lesion studies, Eichenbaum and colleagues trained rats on overlapping odor–odor associations using both the associative inference (Bunsey and Eichenbaum, 1996) and transitive inference paradigms (Dusek and Eichenbaum, 1997). The hippocampal system was damaged prior to training, either by lesion to the hippocampus proper (Bunsey and Eichenbaum, 1996) or through disconnection of the hippocampus from its cortical and subcortical pathways (Dusek and Eichenbaum, 1997). In both tasks, hippocampal lesions impaired performance on inferential probe trials that tested knowledge of the untrained relationship between stimuli, while acquisition of the trained associations was unimpaired (**Figs. 3A** and **3B**).
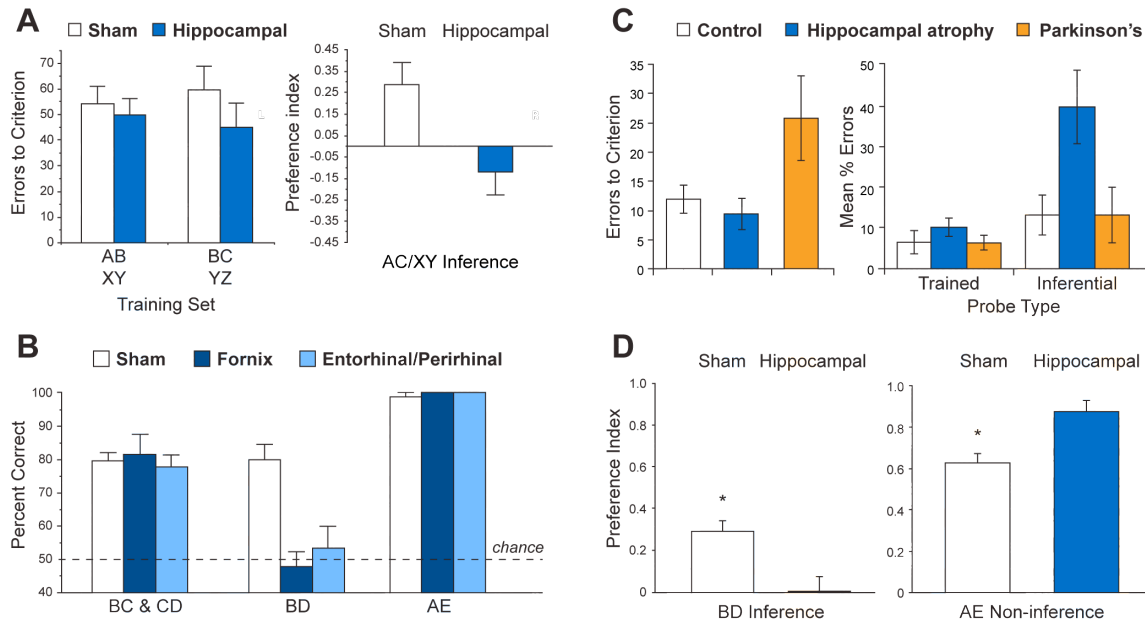
Figure 3. Critical Role of Hippocampus in Inference

(**A**) Associative inference task. Left panel: mean number of errors to criterion on training of two sets of overlapping associations (e.g., AB, BC). Right panel: Inference performance as measured by a preference index for indirectly related item (e.g., selecting C when cued with A). White bars denote sham operated control rats; blue bars denote rats with hippocampal lesion. Hippocampally lesioned rats learn individual relationships between item pairs at a rate similar to control rats, but fail on the inference test. Adapted from Bunsey and Eichenbaum (1996). (**B**) Transitive inference task. Rats with lesions disconnecting the hippocampus from its subcortical (fornix, dark blue bars) or cortical (entorhinal/perirhinal, light blue bars) target structures performed similarly to sham operated control rats (white bars) on trained associations (BC, CD) and novel non-inferential probe trials (AE). However, lesioned rats were severely impaired on inferential probe trials (BD). Adapted from Dusek and Eichenbaum (1997). (**C**) Mean performance of control participants (white bars), patients with hippocampal atrophy (blue bars) or Parkinson's disease patients (orange bars) in an acquired equivalence task. Patients with hippocampal atrophy reached a criterion level of performance at a rate similar to control participants. In contrast, Parkinson's disease patients required more extensive training. However, at test, hippocampal patients were severely impaired on inferential probe trials relative to both control participants and patients with Parkinson's disease. Adapted from Myers et al. (2003). (**D**) Post-training hippocampal lesions (blue bars) impaired transitive inference judgments (BD) in mice, but enhanced performance on novel non-inferential probe trials (AE) involving the end items of the hierarchy relative to sham operated animals (white bars). Adapted from DeVito et al. (2010a).

Similar impairments in transitive inference have been observed in non-human

primates with lesions to the hippocampal system (Buckmaster et al., 2004). Notably, both

lesioned rats (Dusek and Eichenbaum, 1997) and lesioned monkeys (Buckmaster et al., 2004) performed perfectly on novel non-inferential test trials that probed knowledge about the relationship between end items of the hierarchy (e.g., A ? F), suggesting that making judgments about novel combinations of familiar items does not itself require hippocampal processing. Rather, these findings indicate that the hippocampus plays an essential role in judgments that require the flexible manipulation of learned relationships among items when simple comparisons of reinforcement history do not suffice.

Evidence for an essential role of the hippocampus in the acquired equivalence task is somewhat inconsistent across species. Rodent research has shown impaired performance on a spatial variant of the acquired equivalence task in animals with lesions to ERc but not in animals with hippocampal lesions (Coutureau et al., 2002). In contrast, neuropsychological research in humans indicates that hippocampal lesions critically impact performance in the acquired equivalence task (Myers et al., 2003). Patients with hippocampal atrophy acquire stimulus–response associations at a rate similar to control patients and demonstrate intact memory for these trained associations during the test phase (**Fig. 3C**). However, these patients are impaired on inferential judgments testing knowledge of the untrained, equivalent relationships between items. In contrast, as a result of damage to the basal ganglia, Parkinson's disease patients demonstrate slower acquisition of the premise associations during the feedback-based learning phase, but intact performance on both trained and inferential probes at test. This double dissociation demonstrates the critical, specialized role of the hippocampus in decisions that require the flexible application of learned knowledge to novel situations in humans.

These initial neuropsychological and animal lesion studies provide substantial evidence for the critical role of the hippocampus in inference across a variety of experimental paradigms. One finding common across these experiments is intact learning of the explicitly trained associations despite hippocampal damage. This may suggest that while the hippocampus is not required for the acquisition of individual premise associations during encoding (when trained across multiple repetitions using feedback-based learning procedures), it is essential for retrieving and recombining event elements during inference tests. An alternate possibility is that hippocampal lesions induce changes in encoding strategy, resulting in a different representational form for the trained associations that does not allow for the flexible recombination of information during inference itself (Bunsey and Eichenbaum, 1996; Frank et al., 2003). From these initial lesion and neuropsychological studies alone, it is not possible to determine the precise stage—encoding or retrieval—or the precise mechanism of hippocampal involvement in inference, as hippocampal damage was present prior to the initial training phase.

More recent findings indicate that multiple hippocampal mechanisms contribute to successful inference, including processes engaged during initial encoding, flexible retrieval, and offline rest periods following encoding. I discuss these ideas by reviewing evidence from functional magnetic resonance imaging (fMRI) studies in humans demonstrating changes in hippocampal activation during different phases of inferential reasoning tasks along with convergent findings from computational modeling and more recent animal lesion studies.

2.4. RETRIEVAL PROCESSES THAT SUPPORT INFERENTIAL REASONING

Inferential reasoning is traditionally thought of as a logical process in which novel relationships are deduced from knowledge about premise relationships. In the example with the man, woman, and dog, you have no direct knowledge about either the relationship between the man and the woman or where the woman lives. When faced with a misplaced piece of mail addressed to the house next door, you might retrieve information acquired during two previous events—that "the woman in the house next door owns the dog" and "the dog belongs to the man"—and recombine that knowledge to conclude that "the man is my new neighbor." In doing so, you determine that you can deliver the mail to him the next time you see him in the neighborhood. In this way, inferential reasoning is accomplished at the time of retrieval when faced with a novel judgment.

2.4.1. Hippocampal Processes

While inferential reasoning itself has not been traditionally conceptualized as a function of the hippocampus, recent research highlights a hippocampal role in several of the processes contributing to this ability. In the example above, successful inference requires the initial encoding of associations, retrieval of these associations through individual elements when faced with a misplaced envelope, and subsequent recombination of information to yield a solution. Such retrieval-based processes that allow for the flexible use of previous experience are hypothesized to rely on conjunctive representations in the hippocampus that code event details individually in terms of their relationships to one another (Cohen and Eichenbaum, 1993; O'Reilly and Rudy, 2001).

This discrete, relational coding allows event elements to be individually addressable and retrievable from partial input through pattern completion (Marr, 1971; McClelland et al., 1995). In inference tasks used in the laboratory, novel test probes (e.g., AC test trials in associative inference) would trigger hippocampal pattern completion, leading to the retrieval of previously encountered, overlapping memories (i.e., A was paired with B, B was paired with C) that could then be recombined to support successful inferential judgments (**Fig. 4A**).

Several human neuroimaging studies have provided evidence that the hippocampus plays an important role in successful inference at the time of retrieval (Heckers et al., 2004; Preston et al., 2004; Zalesak and Heckers, 2009). In one such study, participants were trained on overlapping face—house associations (AB, BC) and non-overlapping face—face associations (DE) in a modified version of the associative inference paradigm (Preston et al., 2004). Participants were then tested on the learned associations (AB, BC, DE) as well as novel face—face associations (AC) that required mediation through explicitly learned common houses (**Fig. 5A**). While a region in posterior hippocampus was engaged for all associative retrieval trials, bilateral anterior hippocampus was uniquely engaged during inferential (AC) memory probes. This pattern of response was specific to the anterior hippocampus, providing support for its role in the flexible use of memory at retrieval.
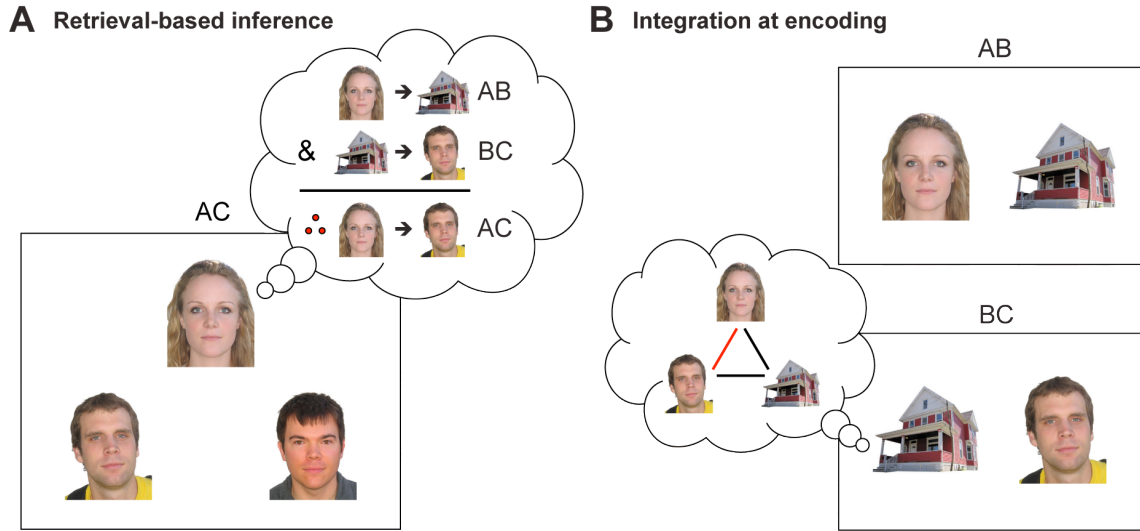
Figure 4. Retrieval and Encoding Processes Supporting Inference

(**A**) Retrieval-based inference through recall and recombination of individual memories. When encountering a novel inferential probe (e.g., AC), the individual elements may trigger hippocampal pattern completion mechanisms, leading to the retrieval of the previously encountered overlapping associations (AB, BC) that can be then recombined to answer novel questions. In this example, when having to select which of the two men lives with the woman, one can recall that the woman lives in the red house, and that the man on the left also lives in the red house. Therefore, the woman lives with the man on the left. (**B**) Integration of overlapping events during encoding. When encountering an event that overlaps with prior experience (e.g., experiencing BC after encountering AB), the overlapping element (B) may trigger hippocampal pattern completion, reactivating the prior memory. The current experience may then be encoded in the context of the reactivated memory to form an integrated (A-B-C) representation that combines elements from both events. In this example, the prior memory for the woman living in the red house may be reactivated when learning about the man living in the same house. The current and reactivated experiences can then be combined to form a novel association that the man and the woman live together.
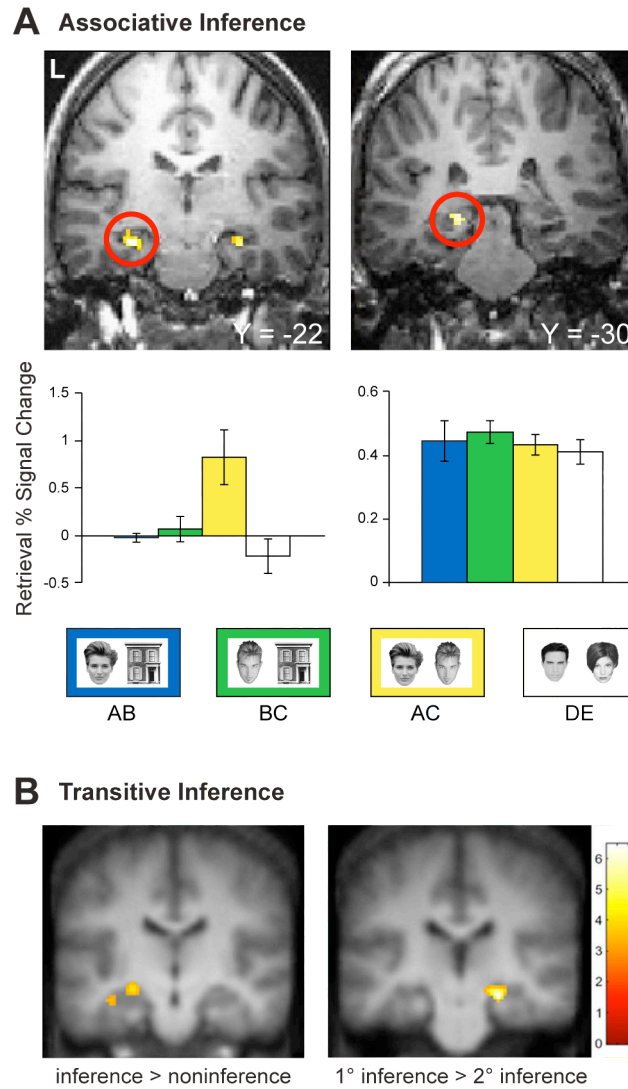
Figure 5. Retrieval Activation During Inference

(**A**) Bilateral anterior hippocampus demonstrated selective activation during novel inferential probe trials (AC) at retrieval in an associative inference task. In contrast, a posterior region of the hippocampus demonstrated equivalent activation during associative retrieval of overlapping trained associations (AB, BC), non-overlapping trained associations (DE), and inferential probe trials. Adapted from Preston et al. (2004). (**B**) During transitive inference, left hippocampal activation demonstrated greater retrieval activation for novel inference trials relative to non-inference trials, while right hippocampus showed increased activation during inference trials in which items were separated by one element in the hierarchy (1°) compared with items separated by two hierarchical elements (2°). Adapted from Zalesak and Heckers (2009).

Hippocampal engagement during retrieval-based inference has also been observed in transitive inference tasks (Heckers et al., 2004; Zalesak and Heckers, 2009). In one such study, participants were presented with pairs of stimuli and trained to select the rewarded item from each pair (e.g., A > B). Two sets of pairs were used: overlapping pairs (A > B, B > C, C > D, D > E) forming an ordered hierarchy and non-overlapping pairs (a > b, c > d, e > f, g > h) with no hierarchical structure (Heckers et al., 2004). Novel test probes consisting of the elements from overlapping pairs (e.g., A ? C, B ? D) produced greater hippocampal engagement than did novel test probes comprised of items from non-overlapping associations (e.g., c ? f). However, this study did not directly address the role of hippocampus in inferential test trials, as there was no comparison made between inferential probe trials requiring reference to the stimulus hierarchy (e.g., B ? D) and non-inferential probe trials containing the hierarchy's end items (i.e., A ? E).

In a follow-up experiment (Zalesak and Heckers, 2009), participants were trained on a six-element hierarchy and tested on inferential relationships between inner members of the hierarchy (i.e., B ? D, C ? E, B ? E) and non-inferential pairs including the end items (e.g., A ? D, C ? F), which could be solved based solely on knowledge of the reinforcement histories. Left hippocampus showed greater activation during retrieval of inferential relative to non-inferential probe trials (**Fig. 5B**). Additionally, the degree of hippocampal activation during inferential probe trials was related to the relative distance between items in the transitive hierarchy. Greater activation was observed in right hippocampus for inferential judgments with one degree of separation between probe items (i.e., B ? D, C ? E) relative to inferential judgments with two degrees of separation

between probe items (i.e., B ? E), suggesting that hippocampal activation tracks the degree of relational processing required for successful inference.

While these human fMRI studies demonstrate that the hippocampus is engaged during inferential judgments, they cannot demonstrate whether such engagement is necessary for successful performance. However, in a recent animal lesion study, selective hippocampal damage induced after acquisition of overlapping memories severely impaired performance on inference judgments in a transitive inference task, providing direct evidence for the essential role of the hippocampus beyond the initial training phase (**Fig. 3D**; DeVito, Kanter, & Eichenbaum, 2010). Interestingly, animals with hippocampal lesions performed more accurately than did control animals on the non-inferential probe trials involving end items. These data show that animals with hippocampal lesions relied primarily upon the reinforcement histories of individual stimulus elements (most salient for end items) rather than the relative contingencies of reinforcement between items. Presumably, hippocampal lesions induced after encoding specifically eliminated memory for the relationships between items that would be needed for successful inferential performance, while leaving intact information about the reinforcement patterns of individual stimuli.

2.4.2. Prefrontal Processes

Collectively, the above reviewed findings indicate that hippocampal memory representations are accessed during inferential judgments. However, additional evidence suggests that other areas—namely, regions of PFC—are recruited in concert with the hippocampus in service of successful inference. For example, activation in the

25

dorsolateral PFC and MPFC has been observed during transitive inference in humans (Acuna et al., 2002b), and lesions to MPFC produce selective impairments on the critical inference judgments in the transitive inference (DeVito et al., 2010b; Koscik and Tranel, 2012) and acquired equivalence tasks (Iordanova et al., 2007).

A number of human studies have implicated lateral aspects of the PFC in particular in inferential reasoning (e.g., Bunge et al., 2009, 2005; Cho et al., 2010; Christoff et al., 2001; Goel, 2007; Goel and Dolan, 2001). This region is generally thought to be involved in resolving interference and selecting appropriate responses among highly similar, competing alternatives (Badre and Wagner, 2005; Thompson-Schill et al., 2005). To aid in this function, lateral PFC is thought to represent similar events as distinct, pattern separated memory traces (Paller and Wagner, 2002; Jenkins and Ranganath, 2010; Ranganath, 2010; Kuhl et al., 2012a), thereby facilitating later retrieval of specific episodic details. Consistent with this idea, a number of studies have shown that lateral PFC is recruited during recollection of details of prior events (Mitchell and Johnson, 2009; Raposo et al., 2009). For example, one study showed engagement of lateral PFC associated with retrieval of distinctive features of individual experiences that were necessary for successful task performance. Interestingly, the researchers also reported that greater activation in this region was associated with less evidence for behavioral interference (Raposo et al., 2009), suggesting the important role of lateral PFC in retrieving appropriate information among a set of highly interrelated memories.

Accordingly, one possible mechanism by which lateral PFC might contribute to inference is by enabling formation and retrieval of memories for individual premise

associations that underlie novel decisions. That is, lateral PFC may support retrieval, manipulation, and recombination of individual event memories, thus enabling novel inference via logical recombination (**Fig. 5A**). This hypothesis has received support from empirical work (Zeithamova and Preston, 2010) employing the associative inference task. Specifically, the researchers demonstrated that while encoding activation in lateral PFC—in particular, IFG—predicted subsequent memory for individual episodes (i.e., AB, BC), it did not significantly relate to later inference performance (i.e., AC judgments). This suggests that lateral PFC is important for forming memory traces for individual episodes, though it does not necessarily code the relationships among events. Interestingly, however, the degree of IFG activation during the test phase itself tracked performance on those inference judgments. Moreover, the researchers observed enhanced functional coupling between IFG and MTL structures during inferential relative to direct test trials, consistent with this region's hypothesized role in recombining individual event memories in service of novel decisions.

Despite this evidence that both MTL and lateral PFC are engaged during inference decisions, the specific role of each region is yet to be determined. One possibility is that the hippocampus supports memory for directly experienced associations, while PFC is important for retrieving, manipulating and recombining individual event memories to allow for relational reasoning (Robin and Holyoak, 1995; Christoff et al., 2001). Some evidence for this view comes from a non-mnemonic inferential reasoning task in which all premise relationships were concurrently displayed along with the probe judgment (Wendelken and Bunge, 2010). In this study, the

hippocampus was similarly engaged during both probe trials requiring consideration of a single premise relationship and probe trials requiring inference across multiple premise relationships. However, lateral PFC was preferentially engaged during inferential judgments.

In this non-mnemonic inference task, the hippocampus may be processing individual relationships among premise elements, an operation required for both probe trial types. Only those probe trials requiring inference across multiple premises necessitate manipulation and recombination through prefrontal processing to support inference. An alternative possibility is that hippocampus may also contribute to the actual manipulation of relational information (Hannula et al., 2006; Hannula and Ranganath, 2008), creating novel combinations of elements extending beyond prior experience. Findings demonstrating hippocampal engagement unique to inferential judgments (**Fig. 5A**; Preston et al., 2004) may reflect such processes. While the respective contributions of the hippocampus and PFC during inferential reasoning remains an active area of study, there is substantial evidence suggesting that these regions form a network for relational processing (DeVito et al., 2010b; Ranganath, 2010) that is engaged during successful inference.

2.5. ENCODING PROCESSES THAT SUPPORT INFERENTIAL REASONING

2.5.1. Hippocampal Processes

Initial studies of inference primarily focused on the role of hippocampus in flexible retrieval processes in which novel probes (e.g., B ? D in transitive inference) trigger recall of directly experienced memories (B > C, C > D), with successful inference

28

(B > D) being supported by flexible recombination of recalled memories. This focus on retrieval-based processes describes how memories are recombined or modified after they are initially encoded.

However, more recent neuroimaging studies support the notion that hippocampal encoding plays an equally important role in successful inference. In one of the first human neuroimaging studies on inference, Nagode and Pardo (2002) used positron emission tomography (PET) to examine the relationship between hippocampal encoding activation and successful transitive inference. Greater hippocampal activation was observed during the intermediate relative to initial stages of training, suggesting experience-dependent changes in hippocampal recruitment during encoding. In a second experiment, participants were first trained on non-overlapping portions of the transitive hierarchy (A > B, C > D, E > F, G > H) followed by training on the "bridging," or overlapping, pairs (B > C, D > E, F > G) that connected the initially learned associations. Hippocampal activation was greater during training of the bridging pairs relative to training on the non-overlapping pairs in the hierarchy.

A similar pattern of findings was shown by Greene and colleagues, who scanned participants during both training and test phases of the transitive inference task using fMRI (Greene et al., 2006). During the latter part of the learning phase, left hippocampal activation was greater for inner training pairs (B > C, C > D) relative to outer training pairs (A > B, D > E), with this difference in activation correlating with individual differences in inferential performance (**Fig. 6A**). Greater test-phase activation was also observed in hippocampus for inferential test probes (B ? D) among participants who were

successful at inference ("performers") relative to unsuccessful participants ("non-performers"), suggesting that the way in which associations are initially encoded relates to the ability to later use them flexibly. Together, the results of these neuroimaging studies demonstrate a contribution of hippocampal encoding processes to novel inferential judgments.
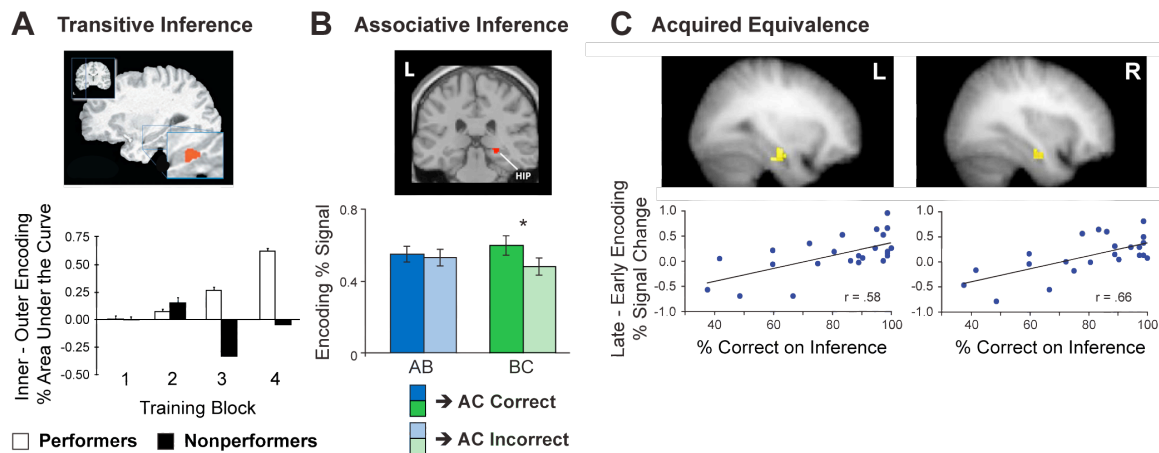
Figure 6. Encoding Activation in Reasoning Tasks

(**A**) Left hippocampal activation increased across training block for inner pairs in the transitive hierarchy (B > C) relative to outer pairs (A > B), but only for those participants who were successful on the inferential test. Adapted from Greene et al. (2006). (**B**) In an associative inference task, right hippocampal activation during encoding of overlapping associations (BC) was greater for trials in which the corresponding inference judgment (AC) was later correct relative to trials on which the inference judgment was later incorrect. Hippocampal activation during initially acquired associations (AB) was not related to subsequent inferential performance. Adapted from Zeithamova and Preston (2010). (**C**) Activation in left and right hippocampus during the training phase of an acquired equivalence task was correlated with individual differences in inference performance. Specifically, increases in bilateral hippocampal activation from the early to late portion of the training phase were associated with superior performance on inferential probe trials. Adapted from Shohamy and Wagner (2008).

Animal research is also consistent with the idea that hippocampal encoding processes are important for successful inference. As previously discussed, hippocampal

damage that occurs after initial learning eliminates memory representations that code the relationships among elements in a transitive hierarchy, while leaving intact information about the reinforcement histories of individual items (DeVito et al., 2010a). Presumably, animals with post-training lesions demonstrate significant impairments on inference tests due to loss of the representations of critical relationships among event elements formed by the hippocampus during encoding. Recent animal research indicates that hippocampal lesions performed prior to learning produce even greater deficits in inferential performance than do post-training lesions (Van der Jeugd et al., 2009), highlighting the critical role of hippocampus during the encoding process.

While the above described human neuroimaging and animal lesion studies established the importance of the hippocampal encoding processes in successful inference, the precise mechanisms by which the hippocampus contributes to performance cannot be determined from these data alone. Two hippocampal encoding mechanisms have been proposed to underlie subsequent inference: (1) elemental encoding of individual premise associations (McClelland et al., 1995) and (2) integrative encoding. As previously expressed, elemental encoding is critical for successful inference and refers to the initial encoding of experience such that memories can be later accessed through individual event details. Such conjunctive representations formed by hippocampus during encoding are essential for retrieval-based inference processes, as they enable access to necessary details when faced with a novel judgment about items not directly experienced together.

In addition to the flexible, elemental encoding of individual associations, additional evidence suggests that the hippocampus may also support inferential judgments by dynamically integrating newly encountered information into existing memory networks at the time of learning—a process termed *integrative encoding* (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010). According to this hypothesis, the hippocampus not only forms relationships among elements within an individual experience, but also links elements across discrete experiences. During integrative encoding, new experiences are not encoded in isolation, but in the context of internally generated memory representations of prior overlapping events (**Fig. 4B**). Through reactivation of previous experience (Eichenbaum, 2000; O'Reilly and Rudy, 2001; Shohamy and Wagner, 2008; Nadel et al., 2012), hippocampal processing may allow for integration of information across distinct experiences (McKenzie and Eichenbaum, 2011; Nadel and Hardt, 2011; Nadel et al., 2012) in anticipation of future use. This constructive, or prospective, nature of memory (Klein et al., 2002; Buckner, 2010; Addis and Schacter, 2012) dates back to Tolman's concept of a "cognitive map" (Tolman, 1948) and has been proposed as a key mechanism underlying successful inferential reasoning (O'Reilly and Rudy, 2001; Shohamy and Wagner, 2008; Zeithamova and Preston, 2010; Zeithamova et al., 2012a). As related memories have already been integrated during encoding, this mechanism—in contrast to flexible recombination at retrieval—predicts that inference judgments may not require any additional processing during test.

For example, let us return to the scenario with your new neighbors, the woman, man, and dog. When you initially see the woman walking her dog in the park near your house, you form a memory for the event that represents the relationship among the woman, the dog, and the park. Upon seeing the dog a second time with the man, the familiar element (the dog) may serve as a cue for hippocampal pattern completion, leading to the reactivation of your prior experience with the woman and dog. The new event (the man walking the dog in town) is then encoded in the presence of the reactivated information about your first experience with the dog. In this way, a link between the woman, the man, the dog, and the house next door can be formed during encoding, despite the fact that you have never seen the woman with the man (**Fig. 7**). Therefore, when you receive the misplaced piece of mail addressed to the house next door, no new recombination of information is required; rather, you can directly retrieve your memory that the woman and man are your new neighbors and determine that you can deliver the mail to either of them when you see them around the neighborhood. Importantly, elemental encoding of individual associations and integrative encoding are not mutually exclusive. For example, integration of new information into an existing memory depends on elemental encoding of the initial memory such that it can be reactivated when the overlapping element (the dog) is encountered again in the second episode.
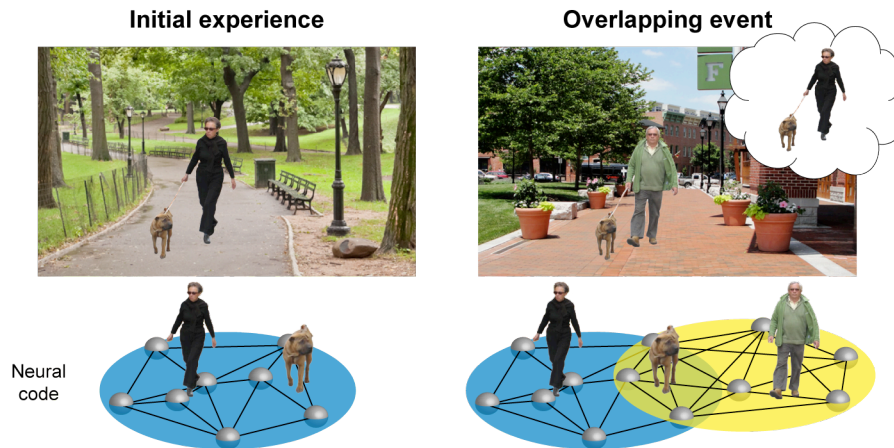
Figure 7. Schematic Depiction of Memory Integration

Example overlapping events that might lead to integration and their associated neural codes. One day while walking in the park, you encounter a woman and her dog (initial experience, top panel). Connections are formed among a group of simultaneously activated neurons, coding the woman–dog association (blue network). A few days later, you encounter the same dog in town, this time with a man (overlapping event, bottom panel). The dog (overlapping element) triggers reactivation of your initial experience in the park (woman–dog association). Such reactivation enables connections to be formed among neural representations of the woman, dog, and man, linking the related events across time (overlapping blue and yellow networks). The resulting integrated memories are hypothesized to support novel judgments that require consideration of both events; here, for instance, you may infer a relationship between the woman and the man despite never having seen them together.

Electrophysiological studies in rodents have shown hippocampal-mediated replay of prior event sequences in new spatial contexts (Karlsson and Frank, 2009) and sequential activation of hippocampal place cells for never-experienced spatial trajectories that represent a shortcut through a well-learned environment (Gupta et al., 2010). Furthermore, in environments with overlapping elements, individual hippocampal neurons demonstrate experience-dependent generalized firing patterns that respond in multiple similar locations (Singer et al., 2010; McKenzie et al., 2013) or to the overlapping features themselves (Wood et al., 1999). Such generalized firing patterns

suggest that hippocampal neurons develop representations that code the similarities between events. By representing features common to multiple events similarly, hippocampal codes can capture regularities shared across different experiences and, in doing so, may act as "nodes" that link distinct behavioral episodes (**Fig. 7**) (Eichenbaum et al., 1999).

Behavioral work in humans suggests that reactivating related memories immediately prior to a new learning experience increases the likelihood that new content will be integrated into existing memories (Hupbach et al., 2007). Using neuroimaging, researchers have also related the degree of reactivation of prior experience during encoding of new overlapping events to evidence for integration (Kuhl et al., 2010; Zeithamova et al., 2012a). In one study, the evidence for hippocampus-mediated reactivation of prior memories was associated with greater retention of the reactivated information (Kuhl et al., 2010), demonstrating that reactivating memories during new learning helps reduce forgetting of past events. In an associative inference paradigm, another study demonstrated that reactivation of existing knowledge during new learning of overlapping associations was associated with superior inference performance, suggesting that a similar mechanism might underlie successful inferential reasoning (Zeithamova et al., 2012a).

The notion that related events are integrated during encoding is consistent with the symbolic distance effect, which refers to increased accuracy and decreased reaction time for judgments comparing items farther apart in a stimulus hierarchy. The symbolic distance effect is often (Frank et al., 2005; Zalesak and Heckers, 2009), although not

universally (Moses et al., 2006), observed in transitive inference tasks. Integration during encoding would promote the formation of a single hierarchical representation of overlapping stimuli (A > B > C > D > E > F) rather than the formation of representations of each individual pair (A > B, B > C, C > D, D > E, E > F). If the transitive inference task was solved by retrieval and recombination of individual learned relationships at test, it should be easier to infer the relationship between items with one degree of separation (e.g., B ? D) than between items with two degrees of separation (e.g., B ? E), as the former requires retrieval of just two representations (B > C and C > D) while the latter requires retrieval of three (B > C, C > D, and D > E). A single hierarchical representation has the advantage of directly encoding the relationships between distant stimuli along with the directly learned relationships, making indirectly learned relationships readily available at the time of test. In contrast to the retrieval-based mechanism, this representational structure would predict that inferences about items far apart in the hierarchy (e.g., B > E) would be easier than inferences for items that are close together (e.g., B > D) because their relative positions in the hierarchy are more distinct.

Recent neuroimaging studies provide more direct evidence for an integrative encoding process in the hippocampus in the acquired equivalence (Shohamy and Wagner, 2008) and associative inference paradigms (Zeithamova and Preston, 2010; Zeithamova et al., 2012a). In the acquired equivalence task, increases in hippocampal activation across the learning phase were associated with individual differences in inferential performance, even when accounting for performance differences on trained associations (**Fig. 6C**). Moreover, no difference in test-phase activation between inferential and

36

trained probes was observed, and reaction times for inferential probe trials and trained associations at test did not differ for successful participants (Shohamy and Wagner, 2008). These findings suggest that performance in the acquired equivalence task is dependent upon the formation of the inferential relationships during encoding that are immediately available when probed with novel combinations of items at test.

Hippocampal encoding activation also predicted success in two studies that employed the associative inference task (Zeithamova and Preston, 2010; Zeithamova et al., 2012a). In one study, changes in hippocampal and MPFC activation over learning were related to performance on inference judgments across participants, even when accounting for premise pair memory (Zeithamova et al., 2012a). In a second study, interrogation of trial-by-trial neural engagement revealed that hippocampal activation during encoding of overlapping associations (BC), but not initially acquired associations (AB), differentiated between subsequently correct and incorrect inferential judgments (AC; **Fig. 6B**) (Zeithamova et al., 2012a). Importantly, while both hippocampal and lateral PFC responses were implicated in successful inference performance at the time of test, this pattern of encoding activation was observed in MTL and MPFC (for a more in-depth discussion of MPFC and inference, see below). Thus, these findings highlight the importance of a hippocampal encoding mechanism whereby overlapping experiences are integrated into a network of related memories as they are learned.

It is noteworthy that a contradiction exists among neuroimaging studies of inferential reasoning: while several studies have demonstrated enhanced hippocampal engagement during inferential reasoning probes (Heckers et al., 2004; Preston et al.,

2004; Zalesak and Heckers, 2009; Wendelken and Bunge, 2010; Zeithamova and Preston, 2010), others have failed to find evidence for retrieval-based hippocampal processing during inference itself, instead linking inferential ability primarily to encoding processes (Nagode and Pardo, 2002; Shohamy and Wagner, 2008). While it is important to not over-interpret the null test-phase findings in these studies, several possible explanations may be worth exploring in future research to help make sense of these apparent empirical contradictions. For example, the number of training trials during the learning phase of inference tasks may have a major influence on the type of hippocampal representation recruited in service of successful inference. In many cases, a limited set of overlapping associations (e.g., five in transitive inference) are trained across many repetitions (e.g., Nagode & Pardo, 2002). In other cases, a large number of overlapping associations are learned in a single exposure (Zeithamova and Preston, 2010; Schlichting et al., 2014) or relatively few repetitions (Preston et al., 2004; Zeithamova et al., 2012a). Integrative encoding may fully support inference during learning of a limited set of experiences across many repetitions, as there are multiple opportunities to build and strengthen the links between different trained associations as they are learned. In contrast, such hippocampus-mediated integration may not be sufficient in tasks that utilize large stimulus sets and single trial learning procedures, leading to the additional recruitment of retrieval-based hippocampal mechanisms. For example, both an integrative encoding signature and increased retrieval activation for inferential probes were observed in a single trial associative inference task (Zeithamova and Preston, 2010), suggesting that both processes may be recruited in a single task. Other task demands may similarly

influence when one hippocampal mechanism—integration during encoding or recombination at retrieval—is favored over another, including feedback-based vs. observation learning and interleaved vs. blocked presentation of training associations (see **Chapter 7**). Understanding how the dynamics of the task influence the recruitment of the hippocampus at different stages of learning may provide key insights into computational properties of the hippocampus and its functional role in inferential reasoning.

2.5.2. Prefrontal Processes

Recent work has highlighted the importance of MPFC in encoding of events that relate to prior knowledge (**Fig. 8**) (van Kesteren et al., 2012). One possible explanation for this region's involvement in encoding-phase memory updating lies in its pattern of anatomical connectivity. As described previously, MPFC is directly connected to the hippocampus, receiving inputs primarily from the anterior portion of CA_1 (Barbas and Blatt, 1995; Cavada et al., 2000). MPFC also has extensive connections with a diverse set of sensory, limbic, and subcortical structures (Cavada et al., 2000), suggesting that it might be important for combining across episodic memories, represented in the brain across distributed cortical and subcortical networks. Consistent with this idea, recent studies have observed encoding-phase engagement of MPFC during inference tasks (Zeithamova et al., 2012a), perhaps reflecting the importance of this region for memory integration during encoding.
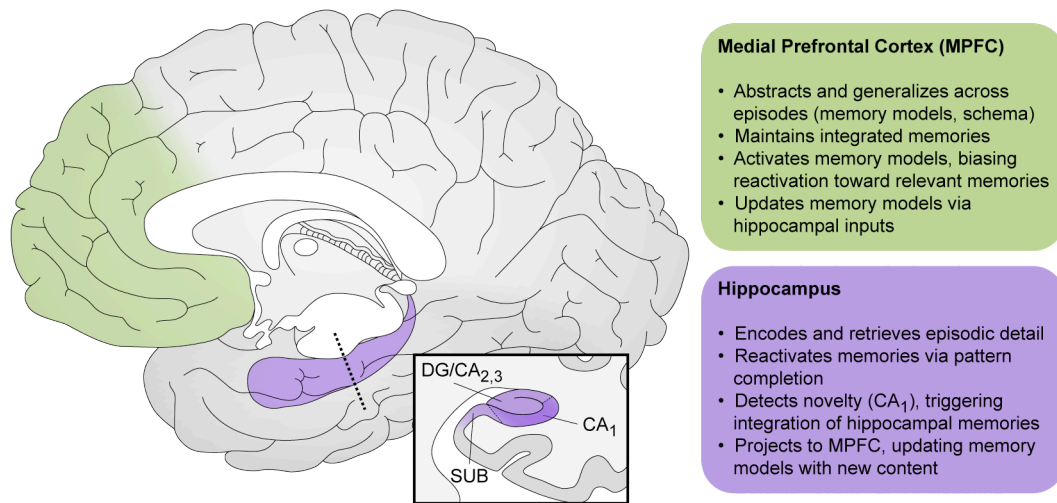
Figure 8. Proposed Roles of Human MPFC and Hippocampus in Memory Integration

Locations and hypothesized functions of regions critical for memory integration in the human brain. Green, medial prefrontal cortex; purple, hippocampus. Here, I intentionally provide a broad definition of MPFC due to high variability in the precise location of effects reported across studies. For instance, I include anterior cingulate cortex, which has been implicated in memory integration (Wang et al., 2012) and the formation of memory models (Roy et al., 2012). Inset, cross section through the hippocampus (purple) highlighting area CA_1 (dark purple portion). Approximate hippocampal subfield boundaries are indicated with thin dashed lines. Location of cross section along hippocampal axis is indicated with a thick dashed line. MPFC, medial prefrontal cortex; CA_1, Cornu ammonis field 1; $DG/CA_{2,3}$, dentate gyrus and Cornu ammonis fields 2 and 3; SUB, subiculum.

As mentioned previously, existing work in rodents (Iordanova et al., 2007; DeVito et al., 2010b) and humans (Koscik and Tranel, 2012) with brain damage also suggest the critical role of MPFC in inference. For instance, using a transitive inference paradigm, DeVito and colleagues (DeVito et al., 2010b) found that rodents with lesions to MPFC were selectively impaired on inference trials; interestingly, they performed just as well as did control animals on non-inferential judgments. While these studies demonstrate the critical role of MPFC in inference, the specific task phase and mechanism by which MPFC supports inference is unknown from these data alone.

MPFC and Memory Schemas

The notion that the hippocampus and MPFC contribute to inferential reasoning by integrating new experiences into existing memory networks to form links between distinct events is conceptually related to an emerging body of literature on the role of the hippocampus in the formation of schemas. Schemas are knowledge frameworks that capture regular patterns in the environment by abstracting information across experiences (Bartlett, 1932) and represent features common to multiple different events while discarding idiosyncratic details. For example, a "restaurant schema" may contain commonly experienced elements such as sitting at a table, ordering from a menu, and paying the bill, but not one-time elements such as the waiter spilling water on you.

Schemas guide behavior by providing a set of expectations for a given experience. Like integrated memory representations, schemas also contain information derived from multiple events that may support inferential decisions. Specifically, schemas represent relationships among elements commonly associated with certain types of situations, despite the fact that these elements have not necessarily been experienced together. Moreover, encoding new events in the context of a reactivated schema may provide an additional mechanism for inferential reasoning. For example, a person may come to your table at the end of your meal and inquire about the quality of the food and service. In the absence of an introduction, you may infer that this person is the owner or manager of the restaurant because your restaurant schema contains information about who is likely to ask for feedback about your dining experience.

Like the integrative encoding processes discussed in the previous section, schemas build knowledge representations from multiple individual events and may thus involve neural mechanisms similar to the hypothesized integrative encoding processes that support inference. Prior reports have implicated hippocampal-MPFC interactions in the generation (van Kesteren et al., 2010a) and use of schemas during inference (Kumaran et al., 2009). For instance, one neuroimaging study in humans (Kumaran et al., 2009) revealed that the degree of hippocampal-MPFC functional coupling tracked schema formation and inference. Schema-dependent speeded learning has also been observed in rodent research, which demonstrated that reactivation of an existing task schema (in this case, a familiar spatial layout) allowed for rapid acquisition of new flavor–place associations in a single trial (Tse et al., 2007, 2011). Without an existing schema, such associative learning required repeated training across multiple weeks. Importantly, rats with hippocampal lesions failed to show facilitated learning of new information in the presence of reactivated schemas, highlighting a critical role for this region in the rapid incorporation of new information into existing knowledge frameworks.

A recent model (van Kesteren et al., 2012) suggests that when new information can be integrated into existing schemas, MPFC may guide online predictions based on the established knowledge. The hippocampus is then though to mediate detection of associative mismatches between current experience and retrieved content (Shohamy and Wagner, 2008), perhaps through CA_1 processes (Duncan et al., 2012). This model suggests that hippocampal-MPFC interactions would be maximal during encoding of

42

events that relate to prior experience (i.e., as in learning BC after learning AB in the associative inference task), reflecting the reorganization and updating of existing schemas with new content. In contrast to the interference resolution in lateral PFC (e.g., IFG) thought to be accomplished by maintaining *separate* memory traces, MPFC may resolve interference among memories by aiding in the formation of *integrated* representations.

Work in the domains of schemas and inferential reasoning has been carried out in isolation from one another. While direct evidence linking schema formation and updating to integrative encoding processes observed in inference tasks is lacking, the striking similarities between these two literatures suggest that investigating the commonalities between these research areas might reveal further neural similarities. For example, one recent fMRI study (Zeithamova et al., 2012a) has shown that functional connectivity between hippocampus and MPFC increases across learning of events overlapping with prior experiences in the associative inference task, suggesting a role for these regions in updating existing memories with new information during encoding. In addition, changes in activation across the encoding phase in both MPFC and hippocampus were linked specifically to subsequent inference success across participants. These results suggest that inference may rely on encoding processes mediated by the hippocampal-MPFC network. That is, in addition to forming schemas, these regions may also aid in the formation of integrated networks of episodic experiences that span multiple events, thereby facilitating novel decisions.

While one important characteristic of schemas is the loss of idiosyncratic details that code the differences among events, it remains unknown whether the same is true of

integrated memory representations. Anecdotal evidence from the acquired equivalence paradigm suggests that some event details may also be lost during integration, as participants failed to recognize inferential probe trials as novel pairings of stimuli (Shohamy and Wagner, 2008). This finding suggests that details about directly experienced events may sometimes be lost in favor of an abstracted, generalized framework that codes consistencies among distinct stimulus-response relationships. However, whether a similar loss of detailed event information is typical in other inference paradigms, especially those that utilize rapid acquisition procedures (e.g., single-trial learning), is not known. More research is needed to understand how the processes that support inference are related to those implicated in the formation and use of schemas. Consideration of how task dynamics influence the type of representational structure formed may provide important insights into how the hippocampus codes overlapping event information and interacts with MPFC to support mnemonic flexibility.

2.5.3. Factors Promoting Integration at Encoding

A number of studies have investigated the various factors that influence integration. For instance, while there is evidence that integration can occur in the absence of conscious awareness (Shohamy and Wagner, 2008; Wimmer and Shohamy, 2012; Henke et al., 2013; Munnelly and Dymond, 2014), studies have shown that integration may be facilitated when subjects become aware of the task structure (either via instructional manipulations or spontaneously) (Kumaran and Melo, 2013). In fact, one experiment (Kumaran and Melo, 2013) demonstrated that such knowledge specifically benefitted judgments that spanned episodes with no effect on memory for the individual

episodes themselves, suggesting that integration does not necessarily emerge with learning of the underlying experiences. One possibility is that awareness constrains prefrontal control processes, which in turn biases hippocampal reactivation during learning toward task-relevant memories, allowing for integration across events.

As described previously, it has been hypothesized that being reminded of related memories prior to a new learning experience also increases the likelihood of integration, as the reactivated memories become labile and readily updated. Consistent with this idea, behavioral work in humans (Hupbach et al., 2007) found more intrusions from a second learned list (List 2) when recalling the initial list (List 1) if participants had been reminded of List 1 before encoding List 2. This finding was recently replicated in rodents using "lists" of ordered feeder locations (Jones et al., 2012), with animals who learned two lists in the same relative to different spatial contexts producing more intrusions. These findings are consistent with the proposal that integration occurs via reactivation of prior memories; here, this work further highlights that reminding the learner of the original encoding context may encourage integration.

Other factors hypothesized to impact integration include (1) the nature of the underlying memory representations—with more distributed as opposed to localized representations proposed to promote integration (Schiller and Phelps, 2011); and (2) the degree of competition between new content and prior memories (i.e., whether or not the two memories can coexist), with integration preferentially occurring in cases when competition is minimal (Hupbach, 2011).

2.6. OFFLINE MECHANISMS OF MEMORY INTEGRATION

Numerous empirical studies (Tambini et al., 2010; Jadhav et al., 2012; Deuker et al., 2013; Staresina et al., 2013) and theoretical accounts (Marr, 1970; McClelland et al., 1995) highlight the importance of offline processes—such as reinstatement of recent experience and enhanced interregional communication—for episodic memory. It has been proposed that through hippocampal-neocortical interactions (McClelland et al., 1995; Nadel et al., 2000), memories are reactivated during periods of sleep and awake rest. Such reactivation (or *replay*) is thought to support the strengthening and transfer of memory traces from the hippocampus to neocortical regions for long-term storage, a process termed *consolidation*.

These mechanisms may also support the integration of memories across experiences. Recent theories suggest that hippocampus-mediated replay of event sequences during sleep (Hoffman and McNaughton, 2002; Ji and Wilson, 2007) provides a potential mechanism for constructing networks of related memories that anticipate future decisions and actions (Diekelmann and Born, 2010; Sara, 2010; Lewis and Durrant, 2011)—a process referred to as *prospective consolidation* (Buckner, 2010). Such theories propose that by reactivating memories during sleep, representations are recombined and recoded, resulting in rich networks of related memories that extend beyond initially encoded events. According to this view, stored memories are not veridical representations of events, but rather derived representations formed in anticipation of future use. Sleep-based replay of hippocampal memory traces, therefore, could enhance performance on inference tasks that tap knowledge about the relationship

between overlapping events experienced at different times (Ellenbogen et al., 2007; Werchan and Gómez, 2013). Consistent with this notion, one study (Ellenbogen et al., 2007) using a transitive inference paradigm demonstrated that participants who slept following acquisition of the premise relationships performed better on inferential judgments involving two degrees of separation (B ? E) relative to a comparison group who remained awake.

While these findings provide speculative evidence that sleep may facilitate inference, it is worth noting that the observed improvements were not specific to inferential judgments; that is, the researchers also observed improvements on the novel but non-inferential judgments containing the end elements (A ? F). Therefore, the question of whether sleep enhances memory integration specifically or whether it contributes to improved inferential judgments by simply consolidating memories for individual premise associations is yet to be determined.

In addition to sleep-based mechanisms that might contribute to inference, recent reports have also highlighted the importance of offline processes that occur during periods of awake rest for memories of individual episodes. The mnemonic consequences of reactivation of recent experience has been demonstrated during awake rest using neurophysiological techniques in rodents (Jadhav et al., 2012) and, more recently, in humans using pattern information analysis of fMRI data (Deuker et al., 2013; Staresina et al., 2013). For instance, more delay period reactivation was observed for stimuli that were remembered relative to those that were forgotten in a subsequent test (Staresina et al., 2013). Moreover, studies have shown that the degree of hippocampal-neocortical

47

functional coupling during rest periods following learning relates to later memory for the learned content (Tambini et al., 2010). Yet, how these mechanisms relate to integrated memories that span events remains to be determined (see **Chapter 6**).

The schema literature provides reason to believe that offline neural activity may also affect more complex memory structures, as prior work has demonstrated schema-dependent activation that persists during offline periods following encoding. For example, one study has shown increased hippocampal-MPFC functional coupling during encoding conditions that necessitate schema reorganization and updating; interestingly, this pattern persisted during the post-encoding rest period (van Kesteren et al., 2010a). These findings are consistent with the idea that neural patterns evoked during encoding are reactivated during offline rest periods, potentially reflecting early-phase consolidation mechanisms. While the precise effect of these rest-phase processes for memory integration is yet to be determined, it may be the case that memory reactivation and increased interregional coupling may strengthen connections among related memories, thereby further promoting the formation of schemas and perhaps, integrated memory representations.

2.7. NEGATIVE CONSEQUENCES OF INTEGRATION

While the focus of this dissertation is primarily on the positive effects of memory integration, a few noteworthy studies have highlighted its negative behavioral consequences. For example, integration may lead to false memories (i.e., through overgeneralization) (Warren et al., 2014), and memory misattributions (Hupbach et al., 2007; Jones et al., 2012; Gershman et al., 2013; St. Jacques et al., 2013). Interestingly,

patients with ventral MPFC lesions show reduced false memories relative to healthy control participants for words that were never seen but are thematically related to a studied word list (Warren et al., 2014), consistent with the notion that ventral MPFC constructs generalized memory representations.

Integration may also explain the phenomenon of memory misattribution, in which an episodic experience is incorrectly attributed to a different encoding context than the one in which it occurred (e.g., as measured by intrusions). Misattributions may occur when prior knowledge is reactivated and updated with the current experience to the detriment of memory accuracy. One fMRI study (Gershman et al., 2013) used neural decoding to quantify the neural reinstatement of the context associated with prior memories (List 1) during new learning (List 2). Results showed that greater evidence for reactivation of the List 1 context was associated with more misattributions of List 2 words to List 1. Another study (St. Jacques et al., 2013) showed that when participants reactivated a prior experience during new encoding, ventral MPFC and hippocampal engagement was associated with later memory misattributions, consistent with a role for these regions in linking experiences across time.

2.8. HIPPOCAMPAL REPRESENTATIONS UNDERLYING MEMORY INTEGRATION

Initial research suggests that one way in which the hippocampus supports inferential reasoning is by integrating information across multiple experiences to establish links between related events, either during new learning experiences or offline through replay of related experiences during sleep and rest. However, intriguing questions remain regarding the precise nature of the underlying hippocampal

representations. Several theoretical and computational frameworks have proposed alternate accounts of the properties of memory representations that can support inference.

One hypothesized representational structure supporting inference across experiences is one in which new events are incorporated into existing memory traces to be parsimoniously represented in a single, composite memory representation (**Fig. 9A**). For instance, consider the simplified example of two events that share a common element (AB, BC) as in the associative inference paradigm. When a new event occurs that contains an element overlapping with a previous event (e.g., BC after encoding AB), the overlapping element (B) can trigger pattern completion of the previously encoded memory (AB). According to this hypothesized representational structure, elements from the new, overlapping event (in this case, C) would be encoded into the existing, reactivated memory (AB) to form a single integrated representation that combines the two experiences (ABC). Because these integrated representations directly code the novel relationship between A and C along with the original experiences, this representational format provides a basis for the inferential use of memory, but has a notable cost in that details of the individual experiences may not be preserved (e.g., the knowledge that A and C were presented in two different temporal contexts).

The influential cognitive map theory (Tolman, 1948; O'Keefe and Nadel, 1978)—which first sparked interest in the inferential function of the hippocampus—implicitly assumes such integrated representations. In the context of this theoretical spatial framework, memory traces for newly learned individual events (i.e., recently traveled routes) are combined with memories of previously traveled routes to allow for the

creation of an integrated map of the environment, including information about paths not traveled. As a cognitive map of an environment becomes established, it can be reactivated when an animal enters the same environment at a later point and updated with new experiences in that environment. When familiar routes to a goal are blocked, the cognitive map will enable navigation to the goal via an alternate route because information about this novel (i.e., never before traveled) route is included in a single representational structure of the environment.
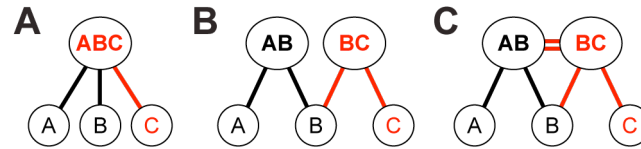
Figure 9. Hippocampal Representations Underlying Inference

Schematic depiction of alternative accounts of hippocampal representation in an associative inference task. Representations of overlapping events (AB, BC) are shown using a simplified two-layer architecture. The bottom layer contains units for each event element; the top layer contains hypothesized patterns of hippocampal representation. (**A**) Single integrated representation for overlapping events. According to this hypothesized structure, new, overlapping event elements (C) are encoded into an existing, reactivated memory (AB) to form a single composite representation for the two related associations. (**B**) Pattern separated representations of individual events. In this view, a new event (BC) with partial overlap to a previous memory (AB) would recruit a distinct hippocampal representation that preserves the details of each individual experience. Links between the common element (B) and each of the individual experiences could be used to mediate inference at encoding or retrieval. (**C**) Relational representation of overlapping events. In this framework, separate representations are maintained for overlapping events (AB, BC) and direct links between those events (at the level of the hippocampus) code their relationship to one another.

In the context of non-spatial inference tasks, there is some evidence to support this hypothesized representational structure. For example, one study showed that successful participants perform as quickly on inferential judgments as on explicitly

trained associations (Shohamy and Wagner, 2008), suggesting similar representations for both directly learned and inferential associations. Moreover, informal assessment suggested that the majority of participants in this study failed to recognize the inferential probes as novel combinations of items, perhaps indicating that some contextual details of original experiences were lost. Returning to the dog-walking example, you may remember that the woman and man are a couple with a dog, but may not remember specific details about how you first encountered them. Future studies may provide a more detailed account of the circumstances under which memory for original experience may become degraded.

The loss of experiential detail is a significant downside to the single, composite representational structure linking elements of discrete events. Other computational perspectives propose a different representational structure for hippocampus, with pattern separation processes preserving distinct individual experiences and recurrent connections between the element and event representations allowing inference across experiences (**Fig. 9B**; Kumaran & McClelland, 2012; McClelland et al., 1995). In our simplified example, this representational structure would predict that a new event partially overlapping with a previous event (i.e., BC) would recruit a different hippocampal representation to make it distinct from the originally experienced event (AB). The two events would be linked through their individual connections to the shared event element (B). Because of the recurrent connections between individual element and event representations (ascribed to ERc and hippocampus, respectively), such a hypothesized structure allows for preservation of event details while also supporting inferential

judgments about the relationship between experiences. For example, when presented with a novel inferential probe (AC), each individual element (A and C) may serve as a partial cue leading to the reactivation of the originally experienced events (AB and BC). Activation of the common item (B) in both cases would lead to successful inference.

Results showing unique hippocampal responses during inferential retrieval (Preston et al., 2004; Zalesak and Heckers, 2009; Zeithamova and Preston, 2010) might reflect the use of such pattern separated inputs to support performance. This representational structure can also explain recruitment of the hippocampus during encoding of overlapping events (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010), which potentially reflects changes in the weights linking common elements to the individually experienced events. It is important to note that even such pattern-separated representations would be expected to change over time and become more generalized. Reactivation of these memory representations during the consolidation process or during offline replay would result in more frequent reactivation of common elements and strengthening of their connections to event representations. In contrast, idiosyncratic elements unique to individual events would be reactivated less frequently and gradually lose their connections to event representations (Lewis and Durrant, 2011). This process would lead to the gradual loss of episodic details in favor of abstracted representations that capture regularities across experiences.

An alternate view that combines elements of both of these frameworks stems from relational memory theory (Cohen and Eichenbaum, 1993). Relational memory theory proposes that the hippocampus maintains representations of individual events while also

directly encoding relationships between separate experiences (Eichenbaum et al., 1999). In my symbolic representation of this theory, different hippocampal units are recruited to represent individual events, but a lateral connection exists at the second level, linking the representations of overlapping events together (**Fig. 9C**). Both pattern separation and pattern completion at the level of the hippocampus would contribute to the formation of such networks of related memories. For example, a new overlapping event (BC) would recruit a hippocampal representation distinct from the originally experienced event (AB). Simultaneously, the overlapping element (B) serves as a partial cue that reactivates the prior event (AB). Based on a Hebbian learning rule, the connection between the two hippocampal memory traces would be strengthened and an explicit link between the overlapping events would be formed. Like the representational structure above, such relational networks would support mnemonic inference while simultaneously preserving memory for individual experiences.

It remains for empirical research to test the predictions stemming from these alternate theories of hippocampal representation. Doing so will afford a more complete understanding of the computational properties of the hippocampus underlying inferential reasoning, and whether or not the same representational format is used across all inference paradigms, training procedures, and individuals.

3. Summary of the Specific Aims

Extensive evidence indicates that MTL and PFC play important roles in inferential judgments that require consideration of the relationship among multiple events. The hippocampus is thought to build flexible memory representations that provide details not only about individual event elements, but also about the relationships between different events. This information may then be combined across time, either as new memories are formed or during the reasoning judgment itself by MPFC and IFG, respectively. This research highlights the function of memory as more than simple recordkeeping. The MTL does not merely enable the retrospective use of memory; rather, hippocampal function is "intrinsically prospective" (Klein et al., 2002), aimed at constructing representations that can be used to successfully negotiate future judgments and actions.

The research in this dissertation builds upon these initial findings to further our understanding of the mechanisms by which MTL and PFC contribute to the generation and stabilization of memory traces that span episodes. I employed the associative inference paradigm, which prior work in my lab has shown to provide a valid model of overlapping event encoding (Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a), in combination with both standard- and high-resolution fMRI (hr-fMRI). In the first study, high-resolution imaging restricted to the MTL region allowed me to test specific predictions made by neurobiological and computational models regarding the contribution of distinct hippocampal subfields to overlapping event encoding. Standard resolution imaging of the whole brain was employed in the second

experiment, which investigates how the MTL works cooperatively with PFC and other neocortical regions during and immediately following encoding of events that relate to prior knowledge. The dissertation culminates in a hr-fMRI study of the whole brain, providing a level of detail and scope not previously possible. In this study, I interrogate how neural representations of individual memory elements change as a function of learning in MTL and PFC subregions.

To investigate the mechanisms by which MTL and PFC aid in the formation of memories that span experiences, this dissertation incorporates both standard univariate analyses as well as a number of advanced imaging analysis methods, including multivoxel pattern analysis (MVPA; Haxby et al., 2001; Norman et al., 2006), neural pattern similarity analysis (NPSA; Kriegeskorte et al., 2008a) and functional connectivity analyses.

3.1. AIM 1: ISOLATE SUBFIELD CONTRIBUTIONS TO INTEGRATIVE ENCODING

The first study in this dissertation used hr-fMRI and NPSA to investigate hippocampal subfield contributions to integrative encoding and inferential reasoning (Schlichting et al., 2014). Participants encoded overlapping AB and BC pairs and were tested on both directly learned and inferential associations during fMRI scanning. The hippocampus is a heterogeneous structure, with a number of distinct subfields thought to support different functions based on their diverse cellular makeup and anatomical connectivity (Manns and Eichenbaum, 2006). One possibility is that CA_3, a region thought to be critical for binding elements within events (Marr, 1971), also plays a role in cross-episode binding. An alternative candidate is area CA_1, a region implicated in

updating memories when new information deviates from existing knowledge (Lisman and Grace, 2005). While prior work has demonstrated hippocampal engagement during overlapping event encoding (Zeithamova and Preston, 2010), this study is the first to test anatomically-inspired hypotheses about subfield contributions to memory integration and inference.

3.2. AIM 2: IDENTIFY MTL-NEOCORTICAL MECHANISMS SUPPORTING INTERACTIONS AMONG MEMORIES

The second experiment investigated how the MTL works with neocortical regions to link distinct episodes (Schlichting and Preston, 2014, under review). Recent empirical studies have shown that the MPFC is critical for successful inference (DeVito et al., 2010b). Furthermore, prior studies suggest that this region may be most involved during and immediately following the encoding of overlapping events (van Kesteren et al., 2012), forming memory-based predictions and enabling updating during the new encoding itself (Shohamy and Wagner, 2008). In this study, participants with strong memories for AB pairs encoded overlapping BC associations in a single exposure during scanning. fMRI data were also be collected during rest periods following encoding. The goal of this experiment was to investigate how MTL-neocortical interactions support memory updating during both online (learning) and offline (rest) periods.

3.3. AIM 3: INVESTIGATE LEARNING-RELATED CHANGES IN NEURAL REPRESENTATIONS

The third experiment used hr-fMRI of the whole brain to investigate the nature of the neural representations in MTL and PFC that underlie inference (Schlichting et al., in

preparation). One perspective suggests that in regions supporting integration, memories for related experiences recruit overlapping populations of neurons (Eichenbaum et al., 1999). Accordingly, while increased neural similarity has recently been demonstrated for items directly experienced together (Schapiro et al., 2012), this view would predict that representations should also become more similar for items indirectly related across events. In contrast, regions supporting inferential reasoning via recombination of individual event memories might develop more *dissimilar* representations for indirectly related items, thus making related memory traces less susceptible to interference (McClelland et al., 1995; O'Reilly and Rudy, 2001). Participants viewed single A, B and C items during fMRI scanning before and after overlapping event encoding to assess how the neural representations of individual items change with experience. The goal of this study was to identify the learning conditions and brain regions that support a) memory integration, as indexed by neural patterns for indirectly related (A and C) items becoming more similar following learning; and b) memory for individual events, as indexed by A and C item representations becoming more dissimilar as a result of learning.

EXPERIMENTS

4. Experiment 1: Hippocampal Subfield Contributions to Memory Integration and Inference[2]

4.1. RATIONALE

As described previously, memory integration may rely on dynamic interactions between old memories and new information during encoding. It has been suggested that related memories are retrieved during new experiences containing overlapping content, which may in turn impact the way the new information is encoded (Bartlett, 1932; Tolman, 1948; O'Keefe and Nadel, 1978; Cohen and Eichenbaum, 1993). Specifically, it has been proposed that encountering a new event that shares content with a previous experience would lead to the reactivation of the existing memory through pattern completion mechanisms supported by the hippocampus (McClelland et al., 1995; Eichenbaum, 2000; O'Reilly and Rudy, 2001). This learning-phase memory retrieval would allow new, externally available information to be encoded in relation to these internally generated (i.e., reactivated) memories, thereby facilitating subsequent inference about the relationships among memories.

Despite the everyday importance of behaviors like inference, the underlying neural substrates are not well understood. One open question is the role of specific hippocampal subfields in encoding experiences containing elements that overlap with prior memories. Hippocampal subfields differ in cellular organization, anatomical

---

[2] This chapter was adapted from Schlichting ML, Zeithamova D, Preston AR (2014) CA1 subfield contributions to memory integration and inference. Hippocampus 24:1248–1260. D Zeithamova and AR Preston supervised the project and contributed to writing the paper.

connectivity and hypothesized mnemonic functionality (Manns and Eichenbaum, 2006). Even in the face of these important known differences, the degree to which specific hippocampal subfields contribute to encoding and retrieval operations that support inference has not been studied in either animals or humans. However, based on their hypothesized functions, one might posit that the CA fields in particular have special significance when experiencing events that overlap with prior knowledge.

Theoretical and computational models (Marr, 1971; McNaughton and Morris, 1987) attribute binding of elements within individual episodes to area CA_3. This is thought to occur via "context fields" that respond preferentially to temporally contiguous stimuli or events (Wallenstein et al., 1998). Simulation experiments demonstrate that the development of CA_3 context fields during the learning phase of an associative inference task are critical for inference performance (Wallenstein et al., 1998). These data suggest that in addition to supporting within-episode binding, this CA_3 mechanism may also enable novel inference across related experiences.

On the other hand, CA_1 has also been implicated in relating information across individual episodes. In particular, CA_1 is hypothesized to serve as a comparator, signaling when new experiences deviate from memory-based expectations (Lisman and Grace, 2005; Chen et al., 2011; Duncan et al., 2012). Differences between reactivated memories and current events may trigger a specialized encoding process, leading to the formation of links between current experience and existing knowledge (Shohamy and Wagner, 2008; Wang and Morris, 2010; van Kesteren et al., 2012). Thus, a primary aim of this study is

to assess the contributions of areas CA_3 and CA_1 to encoding new content that relates to existing memories.

I employed an associative inference task (Preston et al., 2004; Zeithamova and Preston, 2010) in combination with hr-fMRI and neural pattern similarity analysis (NPSA; Kriegeskorte et al., 2008a). During study phases (**Fig. 10A**), participants were first presented with a series of AB object pairs (e.g., clipboard-truck) followed by the corresponding BC object pairs (e.g., truck-binoculars), where the B item (truck) was common to both associations. After each study phase, participants completed a two-alternative forced choice test of both the directly learned (AB, BC) and inference (AC; e.g., clipboard-binoculars) associations (**Fig. 10B**).

I first sought to isolate hippocampal subfields whose encoding activation predicted subsequent inference success, and interrogate whether such activation was unique to overlapping event encoding. Second, I examined whether inference was supported by reactivation of encoding patterns engaged during overlapping event encoding using NPSA. Specifically, I assessed the following prediction: test-phase inference may be facilitated when connections among memories are formed during study of overlapping associations. Moreover, the degree to which the neural patterns evoked during overlapping event encoding are reinstated during inference should relate to behavioral performance (**Fig. 11A**). In the context of the present task, learning overlapping BC associations provides a unique opportunity to engage a memory integration process, whereby prior AB memories are retrieved and connected with current experience (**Fig. 11B**). I reasoned that the same neural signatures might be reinstated

during successful inference judgments. That is, for subfields supporting this specialized, BC-specific learning mechanism, AC test patterns should be more similar to the neural patterns evoked during BC learning than to those evoked during the encoding of initial AB associations (**Fig. 11B,** bidirectional arrows). I also predicted that as inferences made via this study-phase process would be facilitated, the overall difference in pattern similarity may be driven specifically by fast inference judgments.

The results of this study are reported in Schlichting et al., 2014.
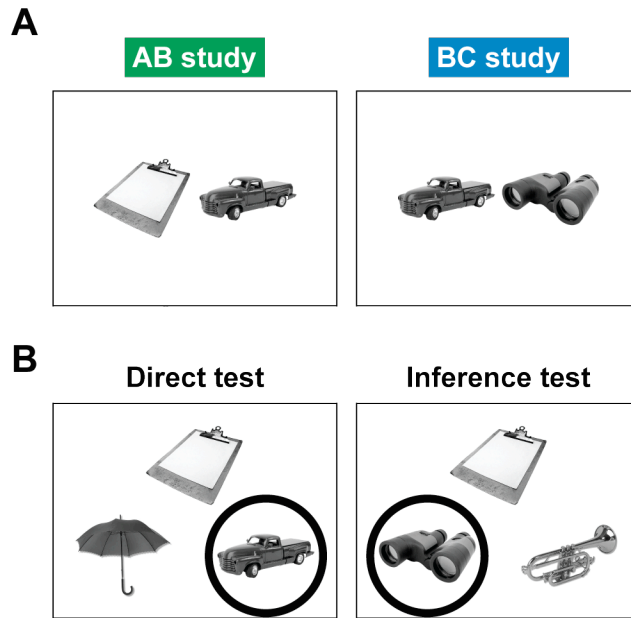
Figure 10. Experiment 1: Memory Task

(**A**) Participants learned overlapping pairs of objects during the study phases. AB (e.g., clipboard-truck) pairs were presented first. BC (e.g., truck-binoculars) pairs were learned later and included familiar items from the AB pairs (i.e., the truck in this example). (**B**) During test phases, participants were presented with three objects. The top item served as the cue; the bottom items were the two choices. A direct test trial is shown on the left, in which the participants were required to select truck when cued with clipboard. In the inference example (right), the participant should choose the binoculars, as both the clipboard and binoculars were paired with the truck during learning. For both direct and inference test trials, familiar items that were members of a

different triad from the same study scan served as foils. Correct choices are circled for illustrative purposes only (not shown to participants).
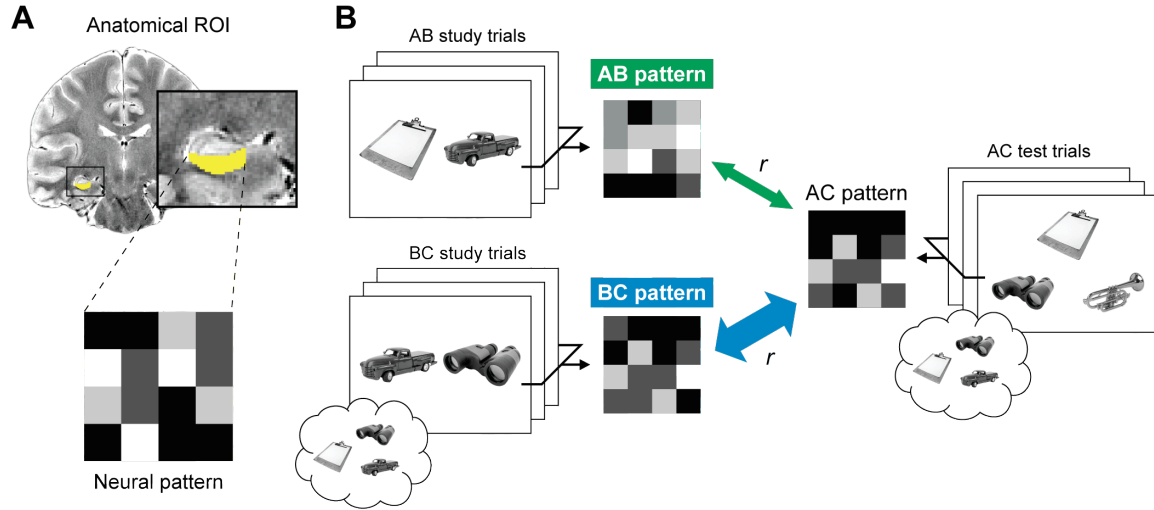
Figure 11. Experiment 1: Neural Pattern Similarity Analysis Rationale

(**A**) Average patterns of activation associated with specific trial types were extracted for each anatomical ROI. Here I depict the cross-participant analysis, which was estimated irrespective of memory performance. Trials were modeled according to event type (AB, BC for study phase; AC for test phase) using a general linear model (GLM). Parameter estimates associated with conditions of interest were then extracted for each voxel within the ROI (example ROI shown in yellow). The intensity in each cell in the grayscale matrix schematic represents the parameter estimate for a single voxel in the brain. The similarity of two patterns is then determined by computing a Pearson correlation between the two matrices. (**B**) Predictions for NPSA when existing memories are retrieved and linked to current experience during overlapping event encoding. Example AB study, BC study, and inference test screens are shown; simplified hypothetical mean patterns of activation associated with each trial type are depicted next to the corresponding condition. The neural representation evoked during BC learning contains A, B, and C information. This pattern can be later reinstated to make the novel AC inference judgment. Reinstatement of study patterns evoked during overlapping event encoding would be reflected by a higher correlation between BC-AC study-test (thick blue arrow) than between AB-AC study-test (thin green arrow).

4.2. HYPOTHESES

In line with prior work (Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a), I predicted that encoding of overlapping events (BC pairs)

would provide a unique opportunity for reactivation and integration of prior memories (AB pairs) with new experiences.

Hypothesis 1: *CA$_1$ engagement during overlapping event encoding will predict subsequent inference*. As CA$_1$ novelty detection mechanisms are thought to trigger a memory integration process, encoding activation in this region will be more predictive of subsequent inference success during BC relative to AB study.

Hypothesis 2: *Regions supporting learning-phase memory integration will be re-engaged during successful inference*. NPSA will be used to isolate regions supporting memory integration. Within participant, there will be greater study-test pattern similarity for BC relative to AB study trial types, particularly when the corresponding AC inference judgments are correct. As integration during encoding should result in facilitated inferences, this difference will be driven by those inference judgments that were made quickly.

Hypothesis 3: *Degree of study-test similarity will predict inference performance across participants*. Greater study-test pattern similarity will be associated with superior inference across participants. This relationship will be observed for BC, but not AB, study trial types.

4.3. METHODS

4.3.1. Participants

Twenty-five right-handed volunteers from the Stanford University community participated in this study. Participants were in good general health and were screened for contraindications to MRI. Consent was obtained in accordance with an experimental

protocol approved by Stanford University and The University of Texas at Austin Institutional Review Boards. Participants received monetary compensation for their involvement in the study. Data from four participants were excluded for the following reasons: failure to achieve above chance accuracy on directly learned associations (3 participants) and loss of anatomical data (1 participant). Data from the remaining 21 participants (12 females, ages 18-31, median = 22 years) were used for all reported analyses.

4.3.2. Behavioral Methods

Materials

Stimuli consisted of 360 grayscale images of common objects organized into 144 overlapping pairs (72 AB pairs and 72 BC pairs, forming a total of 72 ABC triads; **Fig. 10A**) and 72 non-overlapping pairs (XY). Overlapping pairs were those for which two objects (A and C) were each associated with a third overlapping object (B). Non-overlapping XY pairs consisted of two unique objects not paired with any other items. To control for the viewing order and pair type of the object stimuli, participants were assigned to one of six randomization groups.

Procedures

Participants completed an associative inference task (Preston et al., 2004) during fMRI scanning. The task consisted of six alternating study and test phases; both study and test phases were scanned. During study phases, participants intentionally encoded object pairs (AB, BC, XY) and single objects (X) (**Fig. 10A**). Participants saw each pair only once during study, requiring rapid acquisition of associative information. During the

test phases, participants were assessed on their memory for associations learned during the immediately preceding study phase. Memory for both the directly learned (AB, BC, XY) and inference (AC) associations was tested using a two-alternative forced choice procedure (**Fig. 10B**).

Study materials were presented in a mixed fMRI design in which stimuli were blocked by type. Study scans consisted of four cycles during which each of four condition blocks (AB, BC, X, XY) was presented exactly once. Condition blocks were presented such that AB blocks immediately preceded BC blocks and X blocks immediately preceded XY blocks; the order was counterbalanced both within and across participants. Each condition block consisted of 12 study trials lasting a total of 72 s. For each trial, stimuli (a pair of images or a single object) were presented on the screen for 3 seconds with a 1 second response period, during which time participants made a judgment of learning (1, will remember; 2, may remember; 3, will forget). These responses were collected solely to ensure participants' attention during the study task and were not considered in the data analysis. Trial onsets were jittered within each condition block. Between study trials, participants completed a variable number of 2 s baseline trials (range 0-3 baseline trials) in which a single digit ranging from 1 to 8 was presented on the screen; participants indicated with a button press whether the digit was odd or even (Stark and Squire, 2001). Baseline task blocks (12 s) were also presented at the beginning and end of each study phase scan and in between each condition block.

An event-related test scan occurred after each study scan. Tested materials included directly learned object pairs (AB, BC, XY) and inference associations (AC) that

66

were viewed during the immediately preceding study phase. Each test trial lasted 4 s during which time 3 objects were presented on the screen (**Fig. 10B**): a cue object at the top (e.g., object A from triad 1, denoted A_1) with two options at the bottom (the correct stimulus from the same triad, denoted B_1, and a studied stimulus from a different triad, e.g., B_2). Participants indicated which of these two choice objects was associated with the cue object by pressing a button. For inference (AC) judgments, participants were told that the relationship between the cue (e.g., A_1) and the correct choice (C_1) was mediated through their common association with a third item (B_1). For both direct and inference test trials, incorrect choices were familiar objects from a different triad in the same study scan. The order of test trials was pseudo-random such that the inference test trial for a given triad was presented before the corresponding direct test trials (i.e., A_1C_1 was tested prior to testing A_1B_1 and B_1C_1) to prevent additional learning of the direct associations during the test phases. Odd/even digit baseline trials (range 0-5) were intermixed with test trials. For both study and test phases, the order of conditions was determined by a sequencing program to optimize the efficiency of event-related fMRI design (Dale, 1999). Because study and test phases were presented in alternation, participants were aware during learning that they would be tested on both direct and inference associations.

4.3.3. Imaging Methods

MR Data Acquisition

Imaging data were acquired on a 3.0 T GE Signa MRI system (GE Medical Systems). Functional images were acquired using gradient echo spiral in/out pulse sequence (TR = 2 s; TE = 30 ms; 2 shot; flip angle = 61; 128 x 128 matrix; 2 x 2 x 2 mm

voxels; 14 oblique axial slices oriented parallel to the main axis of the hippocampus) (Glover and Law, 2001). While the slices were prescribed to cover hippocampus proper in all participants, I did not achieve full coverage of the parahippocampal gyrus in the majority of participants. Thus, the analyses reported here are restricted to hippocampal subfields. A T2-weighted inplane structural image was acquired in the same prescription as the functional images (TR = 3 s, TE = 68 ms, 512 x 512 matrix, 0.43 x 0.43 mm in-plane resolution). To delineate anatomical regions of interest, I also acquired a high-resolution T2-weighted coronal structural image (TR = 3 s, TE = 68 ms, 512 x 512 matrix, 0.43 x 0.43 mm in-plane resolution, 30 3-mm thick slices). A T1-weighted 3D SPGR structural volume (256 x 256 x 156 matrix, 0.86 x 0.86 x 1 mm) was also collected to facilitate coregistration of the inplane and coronal images. Foam padding was used to minimize head motion.

To obtain a field map for correction of magnetic field heterogeneity, the first volume of each functional scan was collected with an echo time 2 ms longer than all subsequent volumes. For each slice, the map was calculated from the phase of the first two time frames and applied as a first order correction during reconstruction of the functional images. In this way, blurring and geometric distortion were minimized on a per-slice basis. In addition, correction for off-resonance due to breathing was applied on a per-volume basis using phase navigation (Pfeuffer et al., 2002). This initial volume was then discarded as well as the following five volumes of each scan (for a total of six discarded volumes, or 12 s) to allow for T1 stabilization.

fMRI Preprocessing

Data were preprocessed and analyzed using SPM5 (Wellcome Department of Imaging Neuroscience, University College London, London, UK) and custom Matlab (MathWorks, MA) routines. Functional images were corrected to account for differences in slice acquisition times by interpolating the voxel time series using sinc interpolation and resampling the time series using the middle slice as a reference point. Functional images were then realigned to the first volume to correct for motion. The inplane and SPGR images were registered to the mean functional image created during the realignment step. Timeseries data were high-pass filtered with a 128 s cutoff and converted to percent signal.

4.3.4. Data Analysis

Both neural pattern similarity and univariate analyses were performed under the assumptions of the general linear models (GLM). With the exception of group-level GLMs, all analyses were done in the native space of each participant. For all GLMs described below, regressor functions were constructed by modeling stimulus-related activation as a stick function convolved with a canonical hemodynamic response function. Study and test scans were modeled in separate GLMs for all analyses. X items and XY pairs served no purpose for the current study and thus were treated as events of no interest in all models. Additionally, test trials for which no response was generated in the 4 s time window were excluded from all performance-based analyses.

Participant-Level Univariate Analysis

If existing knowledge is retrieved and linked to new experiences during learning, it is necessarily the case that this must occur during the second, overlapping associations (i.e., BC) in the present paradigm. Thus, to examine the role of hippocampal subfields in this process, I searched for regions whose encoding activation is more predictive of subsequent inference success during BC than AB study trials (Zeithamova and Preston, 2010). AB and BC study trials were sorted according to later success on the corresponding inference judgment; study events were then modeled separately according to trial type (AB, BC) and subsequent inference (inference correct, inference incorrect) to generate one model per participant. This resulted in four regressors of interest: AB-inference correct, AB-inference incorrect, BC-inference correct and BC-inference incorrect study trials. I tested the study trial type by inference interaction (BC > AB × inference correct > inference incorrect) to isolate those regions whose activation was more predictive of subsequent inference during BC than AB study trials.

Spatial Normalization

A custom template was generated using Advanced Normalization Tools (ANTS) (http://picsl.upenn.edu/software/ants/; Avants et al., 2011). The T2-weighted inplane (i.e., oblique axial) images from a subset of ten participants with canonical hippocampi were selected for template generation. After template generation, each participant's inplane image was normalized to the group template image as follows. To maximize alignment of hippocampi across participants, a bilateral hippocampal ROI was delineated by hand on each participant's inplane image as well as on the group template. Hippocampal ROIs

were used as labels to guide the spatial normalization from each participant's inplane image to the group inplane template.

Group-Level Univariate Analysis

The contrast images representing the study trial type by inference interaction from the participant-level GLMs were transformed to template space for group-level univariate analyses. Specifically, the transformations calculated for each participant's inplane image during the previously described spatial normalization step were applied to each participant's statistics image. The normalized contrast images were then smoothed with a 2.5 mm FWHM Gaussian kernel and submitted to a group-level GLM. As I anticipated more neural evidence for this specialized encoding process among participants with good relative to poor inference ability, I was interested in those regions for which the interaction term was significantly modulated by AC performance across participants. The group analysis was masked with a generous hippocampus ROI that was created by inflating the group template ROI with a 2 mm radius sphere. An uncorrected voxelwise threshold of $p < 0.025$ was applied to all group statistic images. Correction for multiple comparisons was performed using small volume correction to established the cluster size corresponding to a cluster-corrected threshold of $p < 0.05$. This was carried out using 3dClustSim, part of AFNI (Cox, 1996). 3dClustSim uses a Monte Carlo simulation approach to take into account the size and shape of the volume as well as the smoothness of the data in determining a critical cluster size. Simulations were performed separately for the right and left hippocampus ROIs. Cluster sizes that occurred with a probability of less than 0.05 across 5000 simulations were considered statistically significant. This

procedure yielded a critical cluster size of 13 voxels for both the left and right hippocampus.

Neural Pattern Similarity Analysis

NPSA was used to examine the similarity between patterns of activation evoked during study with those evoked during inferential test trials for each of the five anatomical regions of interest (ROIs): anterior hippocampus, CA_1, a combined dentate gyrus/$CA_{2,3}$ region (DG/$CA_{2,3}$), subiculum, and posterior hippocampus. ROIs were manually segmented on the high-resolution T2-weighted coronal images using established guidelines (Amaral and Insausti, 1990; Insausti et al., 1998; Pruessner et al., 2000, 2002; Zeineh et al., 2000, 2003; Preston et al., 2010), registered to the SPGR and downsampled to functional resolution. Anterior and posterior hippocampal ROIs were defined as those portions of the hippocampus for which subfields could not be reliably delineated. All voxels within a given anatomical region were included in the analysis (i.e., no voxel selection was performed).

Activation patterns were generated at the individual participant level under the assumptions of the GLM. Three sets of GLMs were performed for the purposes of the neural pattern similarity analyses: one based on AC performance (performance-based analysis), one based on response time (RT) on the AC correct triads (RT-based analysis), and one based on event type, irrespective of performance (cross-participant analysis).

For all three NPS analyses, voxelwise parameter estimates were extracted for each participant, condition of interest, and region, resulting in a vector of parameter estimates (one per voxel) associated with each condition in each hippocampal subfield ROI. The

72

resulting study and test patterns were then used to address my hypotheses about the nature of the neural signatures re-engaged during the inference test. In all analyses, Pearson correlations were computed to assess the similarity between study and inference test patterns. Correlation coefficients were then Fisher transformed to more closely conform to the assumptions of normality underlying standard statistical tests.

Performance-Based Analysis

The performance-based analysis was carried out to address how study-test pattern similarity related to inference ability within each participant. I hypothesized that study-test similarity measures would show an interaction between study trial type and inference success, such that BC-AC similarity would be more predictive of performance within participant than would AB-AC similarity. I also anticipated greater study-test pattern similarity for BC than AB study trials, particularly for triads on which the AC inference was correct. I limited this analysis to those triads for which both corresponding direct associations (AB and BC) were correctly remembered, henceforth referred to as "direct correct triads." This was done to investigate the processes involved specifically in inference, i.e., those processes that go above and beyond individual pair encoding.

Study and test trials associated with direct correct triads were split based on inference performance (AC correct or AC incorrect). For the study-phase GLMs, regressors were constructed representing the following conditions of interest (all limited to direct correct triads): AB trials for which AC was later correct; AB trials for which AC was later incorrect; BC trials for which AC was later correct; and BC trials for which AC was later incorrect. AB and BC study trials for which one or both of the corresponding

73

direct test trials were incorrect (i.e., direct incorrect triads) were modeled as events of no interest. For the test-phase GLMs, AC trials were first limited to those for which the corresponding direct pairs were correct and then further split based on AC performance. This resulted in two regressors of interest: AC correct test trials and AC incorrect test trials, both limited to direct correct triads. Direct incorrect AC test trials, AB test trials, and BC test trials were modeled as regressors of no interest.

Voxelwise parameter estimates were extracted for each participant, ROI and condition of interest. The following measures of neural similarity between study and test patterns were then calculated using Pearson correlation and Fisher transformed: AB-AC study-test similarity for AC correct triads; AB-AC study-test similarity for AC incorrect triads; BC-AC study-test similarity for AC correct triads; BC-AC study-test similarity for AC incorrect triads. I then performed a 2 × 2 repeated measures analysis of variance of study-test pattern similarity with study trial type (AB, BC) and inference success (AC correct, AC incorrect) as factors for each region. For regions demonstrating significant interactions, follow-up comparisons were performed using paired t-tests. Pattern similarity measures for this analysis were also compared across ROIs using a 5 × 4 repeated measures ANOVA, with region (anterior hippocampus, CA_1, $DG/CA_{2,3}$, subiculum, posterior hippocampus) and condition (AB-AC correct, AB-AC incorrect, BC-AC correct, BC-AC incorrect) as factors.

RT-Based Analysis

The RT-based models were constructed to further interrogate those regions showing a significant study trial type × inference success interaction from the

74

performance-based analysis described above. I hypothesized that in regions supporting subsequent inference, this interaction would be driven by the fast inferences. That is, I predicted that BC-AC study-test pattern similarity would be higher than AB-AC study-test pattern similarity for fast inference judgments, while there should be no difference for slow inferences. Moreover, BC-AC study-test pattern similarity would be higher for fast than for slow inference judgments, reflecting the speed advantage associated with linking new and old memories during the study phase.

Regressors were constructed for these GLMs as previously described in the performance-based analysis with a single exception: for each participant, study and test trials associated with AC correct triads were further split by median response time (RT) on the critical AC inference judgment. For the study scans, this resulted in GLMs with the following four conditions of interest: AB study trials for which AC was later correct, fast RT; AB study trials for which AC was later correct, slow RT; BC study trials for which AC was later correct, fast RT; BC study trials for which AC was later correct, slow RT (all limited to direct correct triads). AB and BC study trials associated with AC incorrect triads were modeled as regressors of no interest, regardless of RT. GLMs for the test phases included regressors for two conditions of interest: AC correct test trials, fast RT; and AC correct test trials, slow RT. AC incorrect trials were modeled as a separate regressor regardless of RT that was not considered for this analysis. Additional regressors of no interest for study- and test-phase GLMs were identical to those described in the performance-based GLMs above.

After extracting patterns associated with each condition of interest and ROI for every participant, I computed four pattern similarity values: AB-AC study-test similarity for fast AC correct triads; AB-AC study-test similarity for slow AC correct triads; BC-AC study-test similarity for fast AC correct triads; BC-AC study-test similarity for slow AC correct triads. Comparisons of interest were then performed using paired t-tests.

Cross-Participant Analysis

The cross-participant GLMs were used to explore individual differences in the relationship between study-test pattern similarity and inference ability. Events were modeled irrespective of performance to avoid introducing bias into the across-participant correlations. Thus, all trials were included in this analysis. AB study trials and BC study trials were the conditions of interest for the study phase (modeled irrespective of subsequent memory). For the test scans, AC trials served as the single regressor of interest; AB and BC test trials were modeled as regressors of no interest.

I then compared the distributed activation patterns associated with AC test to those evoked during AB and BC encoding. These calculations were performed separately for each participant, resulting in AB-AC study-test and BC-AC study-test similarity measures, computed irrespective of performance. These study-test pattern similarity measures (Fisher's z) were then related to AC inference performance (proportion correct) using Pearson correlation. For those ROIs demonstrating significant relationships between pattern similarity and performance, AB-AC and BC-AC correlations with inference performance were compared using the Hotelling-Williams test (Hotelling, 1940; Williams, 1959).

76

Study-Test Lag Calculations

In the present task, it is necessarily the case that BC study pairs were presented closer in time to the AC test trials than were the AB study pairs. To ensure that any observed differences between BC and AB study-test pattern similarity were not simply due to differences in study-test temporal distance, I calculated mean lag in seconds between study and test trials separately for each study trial type (AB, BC) and inference performance (AC correct, AC incorrect) for each participant. As in the performance-based pattern similarity analysis described above, I limited my comparisons to direct correct triads. This resulted in an average lag for AB-AC correct, AB-AC incorrect, BC-AC correct, and BC-AC incorrect trial types, which I compared using both a 2×2 repeated measures ANOVA with study type and inference performance as factors and follow-up paired t-tests.

4.4. RESULTS

4.4.1. Behavioral Results

Performance on both directly learned AB and BC pairs (range: 63.8%-97.1% correct; mean ± SEM = 79.2% ± 2.5% correct; $t_{20} = 11.49$, $p < 0.001$) and inference associations (AC; 43.1%-93.8%; 65.2 ± 3.4%, $t_{20} = 4.48$, $p = 0.002$) was above chance. Memory for AB pairs was significantly higher than memory for BC pairs (AB: 63.9%-97.2%; 81.6% ± 2.5%; BC: 59.4%-96.9%; 76.8% ± 2.7%; $t_{20} = 4.42$, $p = 0.003$). There were large individual differences in accuracy for inference judgments, which allowed me to determine how hippocampal activation related to inference performance across participants.

77

4.4.2. Imaging Results

Univariate Results

I conducted a univariate analysis to investigate the relationship between hippocampal encoding activation and later inference. Importantly, links between related memories can only be formed in the present task during the second, overlapping (i.e., BC) experience. Thus, I reasoned that in subfields supporting this mechanism, trial-by-trial encoding responses should be more predictive of subsequent inference during study of BC than AB pairs (Zeithamova and Preston, 2010). To test the hypothesis that CA_1 supports this process, I interrogated the study trial type by inference interaction (i.e., BC > AB × inference correct > inference incorrect) across the hippocampus. I was specifically interested in those regions for which the interaction term was significantly modulated by AC performance. Significant activation was observed exclusively in right CA_1 (**Fig. 12**), consistent with this region's hypothesized role in detecting novelty and integrating current experience with prior memories. This cluster was located in the hippocampal body.
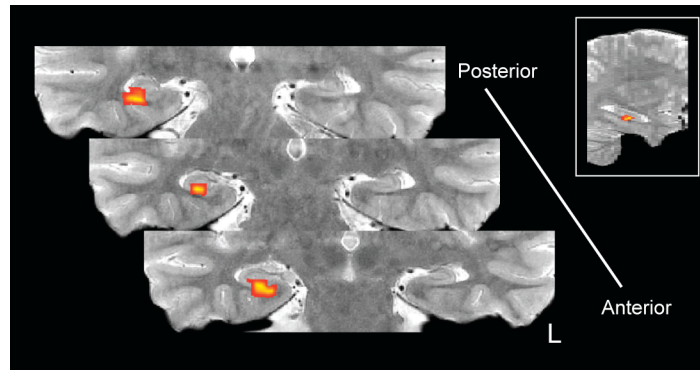
Figure 12. Experiment 1: Univariate Results

Results from the univariate analysis demonstrating relationship between CA_1 processes during overlapping event encoding and subsequent inference success. Right CA_1 was the only hippocampal region to demonstrate a signature consistent with study-phase retrieval of prior memories in service of later inference. This region showed greater subsequent inference effects for BC relative to AB study trials as a function of AC performance. Inset, sagittal view showing location of cluster along anterior-posterior extent of hippocampus. Significant activation was restricted to the hippocampal body. Activation map has been transformed to the space of a single participant's T2 coronal image for visualization purposes. Cluster is significant after correction for multiple comparisons (voxel threshold: $p < 0.025$, uncorrected; cluster size threshold: $p < 0.05$).

Neural Pattern Similarity Results

I employed a multivariate approach to directly compare neural engagement during study and test. I predicted that the degree to which study-phase patterns are reinstated during tests of novel inference would be related to speed or accuracy of the inference decision. The following analyses were designed to test various aspects of this prediction.

Performance-Based Analysis

To examine how patterns of hippocampal activation evoked during AB and BC study were later reinstated during inference judgments, I directly compared BC-AC and AB-AC study-test pattern similarities as a function of inference success. Because BC study trials presented a unique opportunity to retrieve existing, related memories, I

hypothesized that the degree to which BC study patterns were reinstated during inference would uniquely relate to performance, with greater BC-AC study-test similarity for correct relative to incorrect inferences. I anticipated a smaller difference for AB-AC study-test similarity measures. Moreover, I predicted that in cases of correct inference, AC test patterns would be more similar to BC study than to AB study patterns.

Parameter estimates associated with each condition were extracted for every voxel within the anatomically defined hippocampal subfield ROIs, resulting in average patterns of distributed activation for AB study, BC study, and AC inference trials, all for which AC was either correct or incorrect (for a total of six conditions). Importantly, this analysis was limited to only those triads for which the corresponding AB and BC pairs were both remembered. Study-test pattern similarities were then computed for each participant as a function of study trial type and inference performance. Within each hippocampal subfield, I tested the following predictions: (1) that there would be a significant study trial type × inference interaction and (2) that this interaction would be driven by greater BC-AC than AB-AC study-test similarity when AC was correct. Because these statistical tests were performed on all five hippocampal subfields, Bonferroni correction was performed to yield a critical p-value of 0.01 (corrected $p < 0.05$).

Results consistent with memory retrieval during overlapping BC study were found exclusively in CA_1. I observed the predicted study trial type × inference interaction in only this region ($F_{1,20} = 12.26$, $p = 0.002$) (**Fig. 13A**). Moreover, consistent with the second prediction, a follow-up paired t-test revealed that these effects were driven by significantly greater BC-AC than AB-AC study-test pattern similarity for AC correct

triads ($t_{20} = 3.69$, one-tailed $p < 0.001$). Pattern similarity also differentiated correct from incorrect inference performance for BC ($t_{20} = 3.33$, $p = 0.003$) but not AB ($t_{20} = 0.25$, $p = 0.81$) study trial types. Main effects of study trial type ($F_{1,20} = 5.40$, $p = 0.03$) and inference performance ($F_{1,20} = 4.43$, $p = 0.05$) did not reach the corrected significance threshold in this region.

No other ROI showed evidence of an integrative encoding signature (all other study trial type × inference interactions $p > 0.33$; **Figs. 14A-14D**, top charts). However, I did observe a main effect of inference performance that survived Bonferroni correction in anterior hippocampus, subiculum, and posterior hippocampus (all $F_{1,20} > 12.87$, $p < 0.01$), demonstrating greater study-test similarity for correct vs. incorrect trials. There were no significant main effects of study trial type at the corrected threshold (all $F_{1,20} < 6.86$, all $p > 0.01$).

To verify the specificity of the integrative encoding signature to CA_1, I performed a 5 × 4 repeated measures ANOVA with region and condition (AB-AC correct, AB-AC incorrect, BC-AC correct, BC-AC incorrect) as factors. The region × condition interaction was significant ($F_{12,240} = 4.80$, $p < 0.001$), indicating that the relationship between pattern similarity and subsequent inference differed significantly across hippocampal subfields.
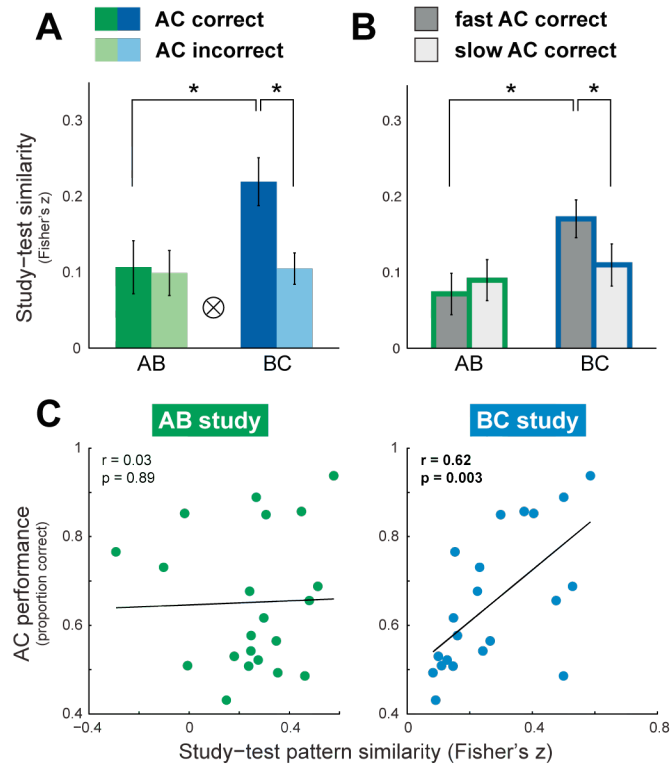
Figure 13. Experiment 1: CA$_1$ Pattern Similarity Results

Results from neural pattern similarity analysis demonstrating an encoding signature specific to overlapping events in bilateral CA$_1$. (**A**) Performance-based analysis. Study-test pattern similarity limited to triads for which both direct memory judgments (AB, BC) were correct. Trials were split based on whether the corresponding inference (AC) judgment was later correct (dark bars) or incorrect (light bars). AB-AC study-test pattern similarity is shown in green; BC-AC study-test pattern similarity is shown in blue. (**B**) RT-based analysis. Study-test pattern similarity for direct correct and AC inference correct triads. Trials were further split based on median reaction time into fast (dark gray bars) and slow (light gray bars) inference judgments. Bars depicting AB-AC study-test pattern similarity are outlined in green; bars for BC-AC pattern similarity are outlined in blue. For both (a) and (b), asterisks (*) denote significant follow-up paired t-tests (p < 0.05); tensor product symbol (⊗) denotes significant interaction (p < 0.01). Error bars denote across-participant SEM. (**C**) Cross-participant analysis. Scatterplots depict continuous relationship between study-test pattern similarity (AB-AC study-test, green; BC-AC study-test, blue) and inference performance. Best-fit lines and statistics on plots were calculated using Pearson correlation. Significant correlation at p < 0.005 denoted with bold type.
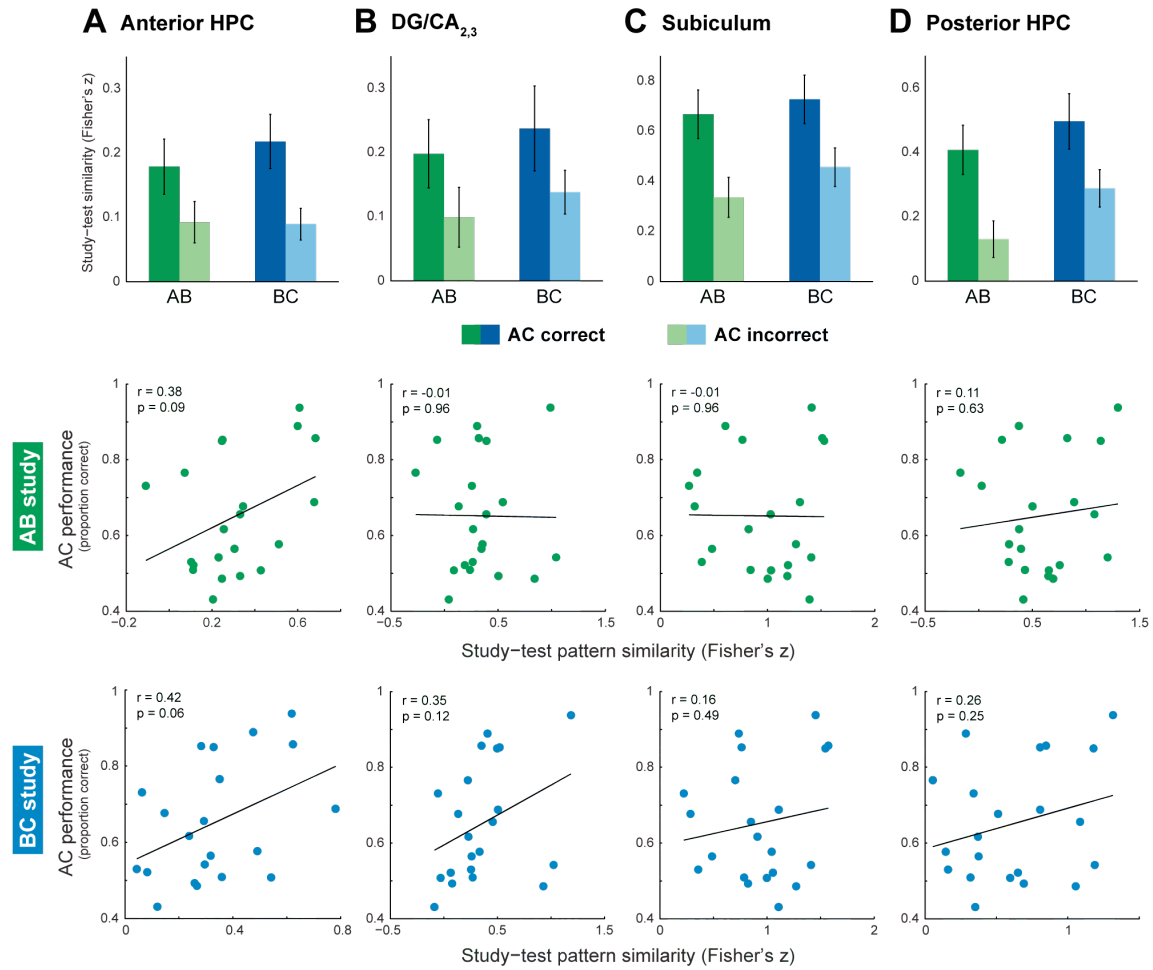
Figure 14. Experiment 1: Other Hippocampal ROI Pattern Similarity Results

Neural pattern similarity results in anterior hippocampus (**A**), DG/CA$_{2,3}$ (**B**), subiculum (**C**), and posterior hippocampus (**D**). None of these regions showed evidence of a specialized BC encoding mechanism. Data are presented as in **Figure 13A** and **C**. Top charts, performance based analysis. Significant main effects of inference performance were observed in anterior hippocampus, subiculum, and posterior hippocampus (all $p < 0.01$; not marked on charts). Bottom scatterplots, cross-participant analysis depicting continuous relationship between study-test pattern similarity and inference performance. All correlations were non-significant at the critical p threshold of 0.005. HPC, hippocampus.

RT-Based Analysis

I reasoned that test-phase reinstatement of neural patterns engaged during encoding would be associated with facilitated inference. Here, I hypothesized that

specifically for fast inferences, AC test patterns would more closely reflect BC than AB study patterns in regions identified above as important for integrative encoding. Moreover, I predicted that evidence for reinstatement of hippocampal patterns evoked during BC study would be greater for fast compared with slow inferences. In this analysis, study and test trials associated with correct inference judgments were median split based on AC test RT. As in the performance-based analysis described above, all study and test trials were limited to only those triads for which AB and BC were both correct. Parameter estimates associated with the conditions of interest were extracted for every voxel in CA_1 for each participant. I then computed study-test pattern similarity as described previously.

Consistent with my predictions, I observed greater BC-AC than AB-AC study-test pattern similarity for fast ($t_{20} = 2.58$, one-tailed p = 0.009) but not slow ($t_{20} = 0.68$, p = 0.50; **Fig. 13B**) inferences. I also found higher BC-AC study-test pattern similarity for fast relative to slow AC correct triads ($t_{20} = 2.13$, one-tailed p = 0.02). This difference was not observed for AB study measures ($t_{20} = -0.67$, p = 0.51). Together, these results suggest that the observed study trial type × inference interaction was driven specifically by the fast inferences, i.e., those likely facilitated by retrieval during learning. Notably, this effect was found only for BC study trials, which provide a unique opportunity for memory retrieval and integration.

Cross-Participant Analysis

I also anticipated that the degree to which hippocampal patterns evoked during BC learning were reactivated during inference test trials would relate to individual

84

differences in inference ability. Specifically, BC-AC study-test similarity should be positively related to inference performance across participants, while there should be no relationship between AB-AC study-test similarity and inference success. To test this prediction, parameter estimates were extracted for each event type (AB study, BC study, AC inference test trials). I did not limit this analysis to particular trials in the experiment (e.g., based on subsequent memory); all study and test trials were included to avoid introducing bias into the across-participant correlations. As I performed two correlations in each of five hippocampal subfield ROIs, my Bonferroni-corrected critical p-value was 0.005.

I related BC-AC and AB-AC study-test pattern similarity in hippocampal subfields to inference performance across participants using Pearson correlation. Again, only CA_1 showed results consistent with test-phase reinstatement of overlapping encoding patterns (**Fig. 13C**). I observed a significant positive relationship between BC-AC study-test pattern similarity and AC performance across participants in this region (r = 0.62, p = 0.003). In contrast, there was no relationship between AB-AC study-test pattern similarity and AC performance (r = 0.03, p = 0.89). A follow-up direct comparison between these two correlation values revealed that they were significantly different from one another (Hotelling-Williams test; t_{20} = 3.73, p < 0.01). No significant relationship was observed between study-test pattern similarity (BC-AC or AB-AC) and inference performance in any other hippocampal subfield (all r < 0.43, p > 0.05) (**Figs. 14A-D**, bottom scatterplots).

Study-Test Lag Results

In the current task design, AB study trials are by definition presented before BC study trials. Therefore, the temporal lag between AB study trials and AC test trials is necessarily longer than the lag between BC study trials and AC test trials. To address the possibility that observed pattern similarity differences might simply reflect discrepancies in intervening time between study and test among my conditions of interest, I performed a control study-test lag analysis. I calculated the average lag in seconds between study and test trials for each participant and condition included in the performance-based analysis (not considering the lag between the study run and the test run). The results demonstrated that pattern similarity did not track with this temporal distance measure. Specifically, despite the large differences in pattern similarity as a function of performance (see **Figs. 13A** and **14A-14D**, top charts), study-test lags did not differ between AC correct and AC incorrect trials ($F_{1,20} = 0.61$, $p = 0.44$). In contrast, large differences in average study-test lag for BC and AB study trial types, dictated by the task design, did not substantially impact the similarity measures in any subfield.

4.5. DISCUSSION

Prevailing views suggest that novel inference relies on the ability to link distinct events across time (Cohen and Eichenbaum, 1993). One process that might support this ability is retrieval of related content during new learning. It has been speculated that overlapping events elicit associative novelty signals (Shohamy and Wagner, 2008; van Kesteren et al., 2012) in CA_1 (Wang and Morris, 2010; Duncan et al., 2012; Larkin et al., 2014), which in turn enable new experiences to be encoded into existing memory traces.

This data provide converging evidence from both univariate and multivariate approaches in support of this idea. I found evidence of a specialized encoding mechanism engaged during learning of overlapping events that supports later inference exclusively in right CA_1. Moreover, I used neural pattern similarity analysis to demonstrate that CA_1 study patterns were reinstated during the inference test, particularly when inference judgments were fast and accurate.

In the present paradigm, events that overlap with prior memories (i.e., BC pairs) provide the unique opportunity to combine current and previous experiences. Thus, I reasoned that associative novelty signals would be possible only during BC study trials. Moreover, such signatures may be associated with superior inference performance on a trial-by-trial basis. The present findings are consistent with the hypothesized role of CA_1 in associative novelty detection. That is, CA_1 was the only region to demonstrate a greater subsequent inference effect during encoding of new events that partially overlapped with prior experiences (BC pairs) relative to the initially experienced events themselves (AB pairs). I underscore that this signature was observed only in the right hemisphere. Interestingly, this converges with a previous standard-resolution study using this paradigm (Zeithamova and Preston, 2010), in which this unique encoding signature related to subsequent inference was also specific to right hippocampus. Right-lateralized hippocampal effects have also been reported in studies of prospection (Addis et al., 2007, 2011; Weiler et al., 2010a, 2010b; Martin et al., 2011), which require similar cognitive processes—recombining prior experiences—in service of imagining future scenarios.

Using neural pattern similarity analysis, I also showed that CA_1 is similarly engaged during overlapping event encoding and successful subsequent inferences. This relationship held across participants, with greater BC, but not AB, study-test pattern similarity relating to superior inference performance. Moreover, this was driven by those inference decisions that were made quickly. Interestingly, these results were also found exclusively in CA_1, consistent with hypotheses regarding this area's special role in comparing prior memories with current experience.

On a mechanistic level, the present neural pattern similarity analysis findings might reflect test-phase reinstatement of memory representations evoked during BC study trials (Zeithamova et al., 2012a). In other words, A, B, and C information is represented in the brain during BC learning and later reactivated during inference to answer the novel AC questions. This interpretation comes with the caveat that the present findings do not provide evidence for reinstatement of specific items at any point in time, as all pattern similarity measures were pooled across multiple trials. Accordingly, a more likely interpretation of the present findings is that similarity of BC encoding patterns with AC inference within CA_1 reflects processes common to BC study and AC test that do not occur during initial AB study. Two likely candidates for such processes are retrieval (i.e., recalling task-relevant memories) and memory-based comparison (i.e., signaling deviations between stored memories and current experience). Critically, if the present results do reflect engagement of such processes, they are processes associated with important behavioral benefits in inference speed and accuracy. Thus, regardless of whether these findings reflect reinstatement of integrated memories or engagement of

common cognitive processes, the present data demonstrate that CA_1 plays a unique and specialized role in encoding mechanisms that support the extension of memory beyond direct experience.

Existing research suggests that there are at least two general mechanisms that may work in a complementary fashion to support flexible memory behaviors such as inference (Zeithamova et al., 2012b). First, information acquired at different times may be combined to make novel inferences during the inferential judgment itself. Under this hypothesized mechanism, individual pattern-separated memory representations are retrieved, manipulated and recombined to address novel questions. This intuitive explanation has received support from a number of fMRI studies (Acuna et al., 2002a; Heckers et al., 2004; Preston et al., 2004; Zalesak and Heckers, 2009; Zeithamova and Preston, 2010), generally on the basis of greater activation in medial temporal and prefrontal regions during inferential judgments requiring consideration of multiple memories relative to memory judgments about single events. In the present paradigm, this framework might suggest that AB memories are retrieved and separately strengthened during BC study, thus leading to better inference.

However, recent work suggests that in some cases, effortful retrieval-based inference can be bypassed through engagement of a different mechanism that operates solely during encoding. In this process, alternatively known as retrieval-mediated learning (Holland, 1981; Hall, 1996; Iordanova et al., 2011; Zeithamova et al., 2012a) or integrative encoding (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010), it has been suggested that existing memory networks are updated with new information while

that new content is being experienced. By initially encoding new information into existing memories, this process is thought to result in hippocampal memory traces that bridge temporally disparate events. Importantly, integrated memories code the relationships among elements in our environment—even those relationships we have not experienced firsthand. Such integrated representations would thus support inferences directly, requiring no further recombination of knowledge during the inference judgment itself. In the present study, an integrative encoding framework would predict that retrieval of AB information during BC study would lead to the formation of an integrated ABC memory representation, thus enabling direct extraction of the novel AC association.

The present data are compatible with both of these accounts. However, I argue that taken in the context of prior work, the results presented here more likely reflect engagement of an integrative encoding mechanism. Recent evidence implicates hippocampal encoding processes in the integration of related memories during learning (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010; Zeithamova et al., 2012a). For example, one study demonstrated that hippocampal activation specifically during overlapping event encoding was predictive of inference success (Zeithamova and Preston, 2010). Moreover, another study (Shohamy and Wagner, 2008) provided behavioral evidence consistent with the idea that memories are integrated during learning, enabling direct extraction of inferential information at test. The authors showed that in successful participants, inference judgments were just as fast as retrieval of directly learned associations. These findings suggest that inferential trials required minimal extra processing during test, presumably because the link between indirectly related items was

already established during encoding. While these studies suggest an integrative encoding process supported by the hippocampus, no study to date has localized this signature to a particular hippocampal subfield. Thus, through the combined use of univariate and multivariate analysis approaches, the present work represents the first empirical evidence highlighting the role of CA_1 in memory integration and inference.

Notably, these results are not readily explained by either memory strength or temporal context accounts. It is true that BC study trials are closer in time to AC test trials than are AB study trials, which might influence the neural pattern similarity analysis in a number of ways. One possible consequence of this design is that BC memories may be stronger (i.e., easier to retrieve) because they were encoded more recently than were AB memories. However, I found significantly better memory for AB than BC associations, suggesting that BC memories are actually *weaker* by comparison. This finding rules out a memory strength account. A second possible consequence relates to differences in temporal context. That is, neural patterns evoked during AC test may more closely reflect BC than AB study patterns simply because BC information was encoded in a more similar temporal context. To this point, it is important to note that study and test phase data were collected in separate scans, thus minimizing the effects of low-level factors such as high within-scan temporal autocorrelation. Nevertheless, it is conceivable that differences in trial timing could give rise to greater BC-AC than AB-AC study-test similarity due to higher-level factors like *cognitive* context. To rule out this possibility, I performed a control study-test lag analysis. I demonstrate that the present pattern similarity results do not track with the temporal distance between study and test,

rendering an account based purely on lag improbable. While these factors may certainly contribute to my pattern similarity measures, they cannot fully explain my findings. Thus, I believe that the pattern similarity data reported here reflect the true engagement of a common neural state during BC study and AC test.

Interestingly, the memory integration signatures observed in CA_1 in the present study converge with studies of rodent hippocampus. First, recent work has implicated CA_1 processes in both detecting novelty in the environment as well as enabling memory updating via increases in plasticity (Larkin et al., 2014). This finding is consistent with an interpretation of the present results as reflecting CA_1-mediated integration of new content into existing memories during study. Additional work has highlighted the importance of CA_1 for nodal coding, in which shared content is represented by the same population of neurons across distinct events (Wood et al., 2000; Singer et al., 2010; McKenzie et al., 2013). For instance, one study demonstrated that while the firing patterns of some CA_1 cells reflected individual episode representations, others fired similarly across different types of episodes that shared content (i.e., a common spatial location) (Wood et al., 2000). Another study showed that in environments with overlapping elements (related locations), a subset of CA_1 and CA_3 neurons responded similarly to the related locations both within and across environments (Singer et al., 2010). Together, these studies suggest that hippocampal neurons in the CA fields can develop generalized firing patterns that encode similarities across episodes. Consistent with this idea, recent work demonstrated that some CA_1 neurons respond similarly to multiple spatial sequence locations learned in a single environment, further evidence that this region may code for "nodes," or

commonalities across experiences (McKenzie et al., 2013). My results are consistent with such a nodal coding scheme, in which integrated memories representing shared B elements across AB and BC learning episodes are formed in CA_1 during learning. Moreover, the data presented here build upon the existing animal literature to demonstrate the potential behavioral significance of nodal coding—specifically, that integrated memories may support novel judgments, allowing for rapid and appropriate action in the absence of direct experience.

The present results suggest that area CA_1 plays a specialized role during the encoding of overlapping information. Utilizing new methodologies—hr-fMRI in combination with neural pattern similarity analysis—I provide a direct comparison of the neural states engaged during study of overlapping events and subsequent inference test. These methods allowed me to take advantage of distributed patterns of hippocampal activation reflecting important content and process, thereby suggesting a mechanistic account of the hippocampal subfield contributions to overlapping event encoding. More broadly, these data are consistent with the notion from the animal literature that CA_1 signals deviations of current events from memory-based expectations, allowing for the construction of memory representations that code the relationships among multiple experiences. The formation of such integrated memory representations supports flexible judgments like novel inference, reflecting the extension of memory beyond direct observation. Moreover, the present results reinforce the idea that encoding and retrieval do not occur in isolation, but rather are highly interactive processes.

5. Experiment 2A: Memory Reactivation During Rest Supports

Upcoming Learning of Related Content[3]

5.1. RATIONALE

The importance of offline processes for memory has been underscored by both empirical (Tambini et al., 2010; Jadhav et al., 2012; Deuker et al., 2013; Staresina et al., 2013) and theoretical (Marr, 1970; McClelland et al., 1995) work. The functional significance of early consolidation signatures for memory—such as reactivation of recent experience—has been demonstrated during awake rest using neurophysiological techniques in rodents (Jadhav et al., 2012) and, more recently, in humans using pattern information analysis of fMRI data (Deuker et al., 2013; Staresina et al., 2013). For instance, more delay period reactivation was observed for stimuli that were remembered relative to those that were forgotten in a subsequent test (Staresina et al., 2013). Moreover, studies have shown that the degree of hippocampal-neocortical functional coupling during rest periods following learning relates to later memory for the learned content (Tambini et al., 2010).

This existing body of work demonstrates that rest-phase neural signatures relate to memory for prior experiences. Yet, one important quality of memory is that it is inherently prospective (Buckner, 2010); that is, memories are formed for maximal utility in future situations. While research shows that rest-phase reactivation impacts memory for the reactivated content itself (Deuker et al., 2013; Staresina et al., 2013), how this

---

[3] This chapter was adapted from Schlichting ML, Preston AR (2014) Memory reactivation during rest supports upcoming learning of related content. Proc Natl Acad Sci U S A 111:15845–15850. AR Preston supervised the project and contributed to writing the paper.

mechanism might be prospectively advantageous remains unknown. In the present study, I turn my attention to this question: How does spontaneous reactivation of established memories and enhanced hippocampal-neocortical connectivity during rest affect learning during subsequent related episodes?

A number of theories underscore the highly interactive nature of episodic memories (van Kesteren et al., 2012; Preston and Eichenbaum, 2013). One prominent view, *interference theory,* highlights that existing knowledge may impair learning of related content. A host of studies confirm this intuition; that is, people often have worse memory for information that is related to their existing memories relative to unrelated information, a phenomenon termed proactive interference (Whitely, 1927; Underwood, 1949; Postman, 1962). However, this impairment is not universally observed, even in the classic literature; on the contrary, prior knowledge can also be beneficial to new learning under some circumstances (Bransford and Johnson, 1972). For example, one study showed a memory advantage for new responses paired with well-learned old stimuli (i.e., stimuli previously learned with a different response), a phenomenon known as associative facilitation (Underwood, 1949). Such facilitation may also extend to novel judgments that require the simultaneous consideration of multiple memories (e.g., inferences).

While these data and others (Tse et al., 2007) suggest that strong prior knowledge may facilitate new learning, the neural mechanisms supporting such associative facilitation are not well understood. One possible explanation stems from *integrative encoding*, which describes how new memories are created in relation to existing knowledge (Shohamy and Wagner, 2008; Schlichting and Preston, 2015).

Mechanistically, it has been proposed that when newly encountered content overlaps with one's stored memory representations, the neural patterns associated with that pre-existing knowledge may be reactivated in the brain during new learning (Kuhl et al., 2010; Wimmer and Shohamy, 2012; Zeithamova et al., 2012a). New episodes may then be encoded in the context of these internally generated representations, connecting these related memories. A recent fMRI study suggests that reactivation of existing knowledge during encoding of new, overlapping events may strengthen pre-existing memory traces, making the prior knowledge itself less susceptible to interference (Kuhl et al., 2010). Reactivation during learning has also been shown to support novel judgments that span experiences (Zeithamova et al., 2012a), consistent with the notion that this mechanism enables the linking of related memories (Schlichting et al., 2014). However, the potential impact of encoding-phase reactivation on the new learning itself has not been addressed. That is, although reactivation has been shown to strengthen both established memories and the connections among discrete experiences, it is as yet undetermined whether this process also facilitates memory formation for the new, related events through integration.

I propose that the degree to which memory processes are engaged during offline periods influences whether prior knowledge interferes with or facilitates new encoding. Importantly, interference theory and integrative encoding make opposing predictions for the impact of rest-phase processes on subsequent learning of related events. Both perspectives might predict that memories are strengthened during offline periods; and that stronger memories are more likely to be reactivated during learning of new, related events. However, these perspectives diverge in their predictions for the consequences of

that reactivation on new learning. While interference theory would suggest that rest-phase strengthening of the initially acquired information might lead to more "competition" and thus worse memory for new, related content (Ratcliff et al., 1990), integrative encoding predicts the opposite. Because stronger memories are more readily reinstated, they are also more likely to be "updated" with new information during subsequent experiences. For this reason, more engagement of rest-phase memory processing might facilitate both the later encoding of related events and novel judgments that span episodes. I sought to adjudicate between these perspectives by investigating the impact of offline reactivation and functional coupling on subsequent encoding of distinct but related experiences.

I employed an associative inference task (Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a; Schlichting et al., 2014) that mirrors classic interference paradigms (Underwood, 1949; Postman, 1962), in which adult participants with prior knowledge encoded new, overlapping pairs. I first trained participants on a set of face-object (AB) associations across four study-test repetitions (**Fig. 15**). I then collected fMRI data while participants engaged in passive rest and encoding of both new overlapping (BC) object-object pairs and non-overlapping (i.e., unrelated; XY) object-object pairs in a single exposure. Importantly, the order of BC and XY learning was counterbalanced across participants. After scanning, participants completed a cued recall test for studied associations (BC and XY) and a surprise test of inferential (AC) relationships. The AC inference test required participants to recall the A_{face} item that was indirectly related to the C_{object} cue through their common association with B_{object}, indexing each individual's ability to combine remembered associations across episodes. This

97

paradigm enables investigation of the neural mechanisms that modulate how existing memories (AB) impact future learning (BC) and inference (AC), thus improving our fundamental understanding of the interactive nature of real-world memory.

The results of this study are reported in Schlichting and Preston, 2014.
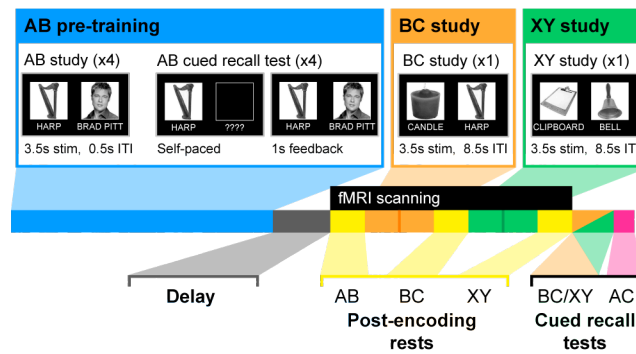
Figure 15. Experiment 2A: Memory Task

Participants encoded 60 initial associations (AB pairs; HARP—BRAD PITT) in four alternating study-test repetitions during the pre-training phase (blue). A stimuli were famous faces to facilitate memory and allow for a cued recall test. After a delay of approximately 5 to 20 minutes, participants were scanned during passive rest (yellow, left). Participants then studied new overlapping (BC pairs; CANDLE—HARP; orange) and non-overlapping (XY pairs; CLIPBOARD—BELL; green) associations (60 each) in a single exposure during a total of four slow event-related scans (two each BC and XY, denoted by dark dividing lines). BC and XY study blocks were each followed by post-encoding rest scans (yellow, middle and right); order of BC vs. XY encoding scans was counterbalanced across participants. After scanning, memory for BC, XY (intermixed; orange/green) was tested via cued recall, followed by a surprise AC inference test (CANDLE—BRAD PITT; pink).

5.2. HYPOTHESES

Interference theory predicts that more reactivation of the AB memory and greater hippocampal-neocortical connectivity during rest following AB learning should be associated with stronger AB memories and will thus lead to impaired BC learning. Conversely, integrative encoding predicts that the same rest-phase mechanisms should be

associated with *better* BC learning. I hypothesized that following initial learning, offline mechanisms—specifically, reactivation of prior knowledge and enhanced interregional communication—would strengthen memories representing the reactivated (AB) content. This process will in turn support later learning of related information (BC), as stronger initial memories are more readily retrieved and available for integration during encoding.

Hypothesis 1: *Rest-phase processes following initial learning will support upcoming encoding of related content.* Integrative encoding posits that existing memories must be reinstated during new learning experiences to enable connections to be formed across episodes. As stronger memories are easier to reinstate during learning, rest-phase strengthening will facilitate subsequent encoding of related information. More evidence for reactivation of prior knowledge and enhanced hippocampal-neocortical connectivity during post-AB rest will be associated with superior learning of overlapping BC pairs when controlling for performance on non-overlapping XY pairs.

Hypothesis 2: *Related knowledge will be reinstated to support learning during new experiences.* AB pairs consisted of a face paired with an object; thus, face (stimulus A) information will be reinstated during encoding of overlapping BC object-object associations to support memory. This will be manifested as engagement of face-sensitive areas that is predictive of subsequent memory for BC, but not XY pairs.

Hypothesis 3: *Evidence for offline mechanisms following initial learning will predict neural engagement during subsequent encoding.* Across participants, more post-AB reactivation and connectivity will be associated with greater fMRI activation during the new learning itself.

5.3. METHODS

5.3.1. Participants

Forty-eight right-handed volunteers (27 females; ages 20-33, mean ± SEM = 24.6 ± 0.5 years) participated in the experiment. Consent was obtained in accordance with an experimental protocol approved by the Institutional Review Board at the University of Texas at Austin. Participants received monetary compensation for their involvement in the study. Data from a total of 13 participants were excluded for the following reasons: hardware malfunction (N = 5), handedness concerns (N = 1), and low memory performance (N = 7). Low memory performance was defined as either (a) failure to subsequently recall more than 10% of BC and XY associations studied in the scanner (N = 6) or (b) failure to reach near-perfect performance on initial AB associations (< 80% cued recall accuracy; N = 1). Data from the remaining 35 participants were included in all analyses (21 females; ages 20-30, 24.1 ± 0.5 years).

5.3.2. Behavioral Methods

Materials

Stimuli consisted of 60 grayscale images of famous faces (30 male and 30 female) and 240 grayscale images of common objects organized into 60 triads (denoted ABC) and 60 pairs (denoted XY) (**Fig. 15**). All images were presented with verbal labels. ABC triads consisted of one face and two objects and were presented as overlapping AB and BC pairs, with the B item shared between pairs. That is, AB pairs always consisted of one famous face (A) paired with one object (B); the same object (B) was then paired with a different object (C) to form a BC pair. Non-overlapping XY pairs consisted of two

objects. All items were unique to their triad or pair, such that a single face or object image was a member of only one ABC triad or one XY pair. Four randomization groups were created to control for the organization of images into triads and pairs and viewing order. Objects were randomly assigned to item type (B, C, X or Y), which determined both whether it belonged to a triad or pair and during which phase(s) it served as a recall cue (see below). An equal number of BC pairs associated with male and female faces (A) were presented within each of two BC encoding scans; no other constraints in item assignment or trial order were imposed. As described below, the order of BC encoding versus XY encoding was counterbalanced across participants.

Procedures

Memory Task

Participants completed a modified version of the associative inference task (**Fig. 15**) (Preston et al., 2004; Zeithamova and Preston, 2010; Zeithamova et al., 2012a; Schlichting et al., 2014). Prior to scanning, participants were trained to near-perfect performance on all 60 AB (face-object) pairs (**Fig. 15**, blue). The goal of the pre-training phase was to create established memories for the AB pairs, such that overlapping BC information could then be encoded in relation to strong existing memories. The AB pre-training phase consisted of 4 study-test alternations. During the study phase, participants viewed each AB pair once (**Fig. 15**, blue; 3.5 s stimulus, 0.5 s ITI). A items (faces) were always shown on the right; B items (objects) were always on the left. Participants were encouraged to construct a visual or verbal story linking the items to aid memory but were

not required to make any explicit response. Each study phase in the pre-training portion of the experiment lasted 4 minutes.

Following each study phase, participants completed a self-paced cued recall test (**Fig. 15**, blue). The B item (object) was presented on the left side of the screen next to an empty box. Participants were asked to say aloud the name of the face that was paired with it. After either a verbal response had been produced or the trial was "passed," participants viewed a feedback display in which the correct image appeared in place of the empty box. Including the feedback displays, each pair was viewed a total of 8 times during the pre-training phase.

After completing the initial AB pair pre-training, participants were transferred to the scanner after a delay (for more details see **Section 5.4.3. *Delay and Encoding Order Analyses***). Importantly, at no time were participants made aware by the experimenter of the relationship between the pre-training phase and subsequent study and test tasks. That is, participants were not told that they would be learning overlapping associations in the scanner or that their memory for AB associations would later be probed via the AC inference test. Thus, while participants became aware of the overlap between the AB and BC pairs after the fact (all 35 reported awareness of this structure in a post-experiment questionnaire), there was no such expectation established prior to scanning. Once in the scanner, fMRI data was collected during 6 minutes of passive rest (**Fig. 15**, yellow). During this time, a white fixation cross was displayed on a black screen. Participants were instructed to think about whatever they liked while remaining awake and alert with their eyes open.

Following the initial post-AB rest scan, participants were scanned during encoding of overlapping BC (**Fig. 15**, orange) and non-overlapping XY pairs (**Fig. 15**, green). Pairs were segregated by type into separate scans. There were a total of four slow event-related scans (2 BC scans and 2 XY scans; 3.5 seconds stimulus, 8.5 seconds ITI). Participants were encouraged to construct a visual or verbal story while they encoded the new associations; no explicit responses were required. Each pair was presented just once, requiring rapid acquisition of associative information. C and Y objects were on the left; B and X objects were on the right. BC study scans always occurred in immediate succession, as did XY study scans. Encoding order of BC and XY scans was counterbalanced across participants. That is, for half of the participants, all BC pairs (scans 1 and 2) were learned before XY pairs (scans 3 and 4); for the other half, the order was reversed. Each study scan was 6 minutes long. Post-encoding rest scans were acquired immediately following BC (e.g., after study scan 2) and XY (e.g., after study scan 4) learning (**Fig. 15**, yellow). These scans were identical to the post-AB encoding rest scan described previously.

After the final rest scan, participants were taken out of the scanner to complete a cued recall test on BC and XY pairs (**Fig. 15**, orange/green). C and Y items, presented on the left, served as probes. BC and XY test trials were randomly intermixed. Following completion of the BC/XY test, the structure of the inferential (AC) associations was explained to participants. That is, participants were told that A and C items both paired with the same B item were indirectly related. Only two participants reported that they anticipated this inference test, even though all participants became aware of the overlap

between the AB and BC associations during the BC study phase. They then completed a cued recall test on these surprise inference associations. The same items (C) served as probes, but this time participants were asked to name the indirectly related item (A, always a face). No feedback was provided during post-scanning BC/XY or AC inference tests to prevent additional learning of the directly learned (BC/XY) associations.

Participants had the opportunity to practice the memory task before beginning the experiment. The practice included only non-overlapping face-object associations using different stimuli from the main the experiment.

Visual Localizer Task

After the memory task, participants completed a blocked design functional localizer during fMRI scanning to obtain neural patterns associated with viewing different types of visual stimuli. Participants viewed blocks of faces, objects, and scrambled objects while performing a 1-back task. For each image, they pressed one of two buttons on a keypad to indicate whether the picture was new or a repeat of the immediately preceding picture. Responses were collected solely to ensure attention to the task and were not considered as part of the analysis. The images used in the localizer task were different from those used in the memory task. There were four blocks of each stimulus type per run, plus additional interleaved blocks of passive fixation. Blocks were 18 seconds long, yielding a total run length of 5 minutes. Three localizer scans were collected. Participants had the opportunity to practice the visual localizer task before beginning the experiment.

5.3.3. Imaging Methods

MR Data Acquisition

Imaging data were acquired on a 3.0T GE Signa MRI system (GE Medical Systems). All functional data were collected in 33, 3-mm thick oblique axial slices using an EPI sequence (TR = 2000 ms, TE = 30.5 ms, flip angle = 73; 64 x 64 matrix, 3.75 x 3.75 mm in-plane resolution, bottom-up interleaved acquisition, 0.6 mm gap). T2-weighted structural images were acquired in the same prescription as the functional images for the memory (TR = 3200 ms, TE = 68 ms, 512 x 512 matrix, 0.46 x 0.46 mm in-plane resolution) and visual localizer (TR = 3200 ms, TE = 68 ms, 256 x 256 matrix, 0.94 x 0.94 mm in-plane resolution) tasks. A T1-weighted 3D SPGR structural volume (256 x 256 x 172 matrix, 1 x 1 x 1.3 mm voxels) was also collected to facilitate image coregistration and for automated parcellation using Freesurfer (http://surfer.nmr.mgh.harvard.edu/) (Desikan et al., 2006).

fMRI Preprocessing

Functional data were preprocessed using FSL version 5.0.2 (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl). The first 4 volumes of all functional scans were discarded to allow for T1 stabilization. Motion correction was performed within each functional scan using MCFLIRT by aligning all images in the run to the middle volume in the timeseries. Coregistration of functional data across runs was performed by calculating and applying the affine transformation from each run to a reference run using FLIRT, part of FSL. The 3D SPGR structural volume was registered to the functional reference run using the EPI registration utility (part of FLIRT) and then resampled to

functional space. Brain extraction was performed on all structural and functional images using BET. With the exception of the group-level general linear models (GLMs), all analyses were done in the native functional space of each participant.

5.3.4. Data Analysis

Behavioral Data

Cued recall responses were hand scored. Responses were scored as correct if the participant produced the correct label or, for famous faces, if they provided a unique and accurate description of the person (e.g., by naming a film in which the actor was featured). I used this liberal criterion for recall because I found that participants would often recall a stimulus in great detail despite an inability to remember the specific verbal label. This type of recall performance was true for virtually only the A_{face} stimuli, and was particularly common early in the initial AB pair pre-training. For example, instead of recalling Daniel Radcliffe, the participant may state "the guy who plays Harry Potter." This criterion has been employed in prior studies using a similar stimulus set and paradigm (Kuhl et al., 2011). A proportion correct was calculated for each participant, pair type (for AB, BC, XY and AC) and repetition (for AB only).

To investigate the relationship between memory for the initial AB pairs and performance on overlapping BC pairs and AC inferences, I also sorted triads according when the AB pair was initially learned. I hypothesized that memory for those AB pairs learned early in the pre-training phase (e.g., that were correctly recalled on the first and all subsequent test blocks) may have stronger memory traces than for those learned late in pre-training (e.g., correctly recalled only on the final test block). An integrative encoding

perspective predicts that stronger AB memories would support BC learning and AC inference. Thus, we would expect higher BC and AC performance when the corresponding AB pairs were learned early relative to when they were learned late. Interestingly, interference theory might make the opposite prediction, with better BC and AC performance for later-learned AB pairs, as these weaker memories are less likely to interfere.

Accordingly, I sorted BC and AC test trials according to the AB test block on which the corresponding AB pair was correctly recalled for the first time. Due to the small number of participants (N = 16) who had any pairs that were first recalled on the fourth and final test block, I collapsed across pairs whose first recall occurred during blocks 3 and 4 for this analysis (referred to hereafter as Block 1 AB pairs, Block 2 AB pairs, and Block 3/4 AB pairs). To be considered in this analysis, the AB pairs also had to be correctly recalled on all subsequent test blocks. That is, a Block 1 AB pair was correctly recalled on blocks 1-4; a Block 2 AB pair was correctly recalled on blocks 2-4; and a Block 3/4 AB pair was either correctly recalled either on blocks 3 and 4 or only on block 4. I then assessed BC and AC performance as a function of AB acquisition time using a repeated measures ANOVA. All but three participants who had no Block 3/4 AB pairs were included in this analysis (N = 32).

Regions of Interest

Anatomical Region of Interest Definition

The hippocampus and MTL cortex (ERc, PRc, and PHc) were delineated by hand on the 1 mm MNI template brain and reverse-normalized to each individual's functional

space using Advanced Normalization Tools (ANTS) (Avants et al., 2011). Specifically, a non-linear transformation was calculated from the MNI template brain to each participant's 3D SPGR volume. This warp was then concatenated with the SPGR to functional space transformation calculated using FLIRT. After applying the transformation using ANTS, the anatomical MTL ROIs were aligned to each participant's functional data.

Functional Region of Interest Definition

Functional data from the localizer task were used to define face-sensitive voxels within the fusiform gyrus (i.e., fusiform face area, FFA). Analysis of fMRI data from the localizer task was carried out using FEAT (fMRI Expert Analysis Tool) version 6.00, part of FSL. The following pre-statistics processing was applied: grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 64 s); and spatial smoothing (5 mm FWHM). FILM prewhitening was used. Stimulus presentation blocks were modeled as events of 18 s duration, with one regressor for each stimulus type (face, object, scrambled object, passive fixation). Stimulus regressors were convolved with the canonical (double-gamma) HRF. Motion parameters calculated during the motion correction step and their temporal derivatives were added as additional confound regressors. Two measures of framewise data quality were also calculated to identify volumes that may be adversely impacted by motion artifacts: framewise displacement (FD) and DVARS (Power et al., 2012). FD measures the overall change in head position from one time point to the next, and is calculated by summing the absolute

values of the derivatives of the six motion parameters calculated during the realignment step. DVARS measures the overall change in image intensity from one time point to the next. This index is calculated as the root mean square of the derivatives of the timecourses across all voxels in the brain. Both FD and DVARS were added to the model as regressors of no interest (Schonberg et al., 2014). Additional regressors were created for each time point in which motion exceeded a threshold of both 0.5 mm for FD and 0.5% change in BOLD signal for DVARS (plus one time point before and two time points after) (Power et al., 2012). Temporal filtering was then applied to the model.

After modeling functional data within each run, the resulting statistical images were combined across localizer runs for each participant using fixed effects. As data were already coregistered across runs, no additional registration or spatial normalization was necessary. Face-selective regions were defined for each participant as those voxels responding more to faces than objects and scrambled objects. The procedure for FFA definition was as follows: I created 14 mm spheres centered at each participant's peak voxel (i.e., the maximum z-statistic from the face > objects + scrambled objects contrast image) located within the posterior half of their Freesurfer-defined fusiform gyrus. This sphere was then masked with fusiform gyrus to restrict FFA to gray matter voxels. This method was used to ensure FFAs of approximately the same size across participants. This procedure was carried out separately for the left and right hemispheres; lateralized ROIs were then summed to create a bilateral FFA (ROI size range: 205-336 voxels, mean ± SEM = 288.7 ± 5.7 voxels). As ROI definition took place in the native functional space of each participant, no realignment or resampling was necessary.

Post-AB Pattern Classification Analysis

MVPA was carried out using sparse multinomial logistic regression (SMLR) with a penalty term of 10 implemented in PyMVPA (Hanke et al., 2009). The bilateral FFA mask described above was used for the main pattern classification analysis. In a follow up analysis, I also interrogated reactivation within the posterior half of the fusiform gyrus defined on each participant (as above). By looking at this larger activation pattern, I was able to increase my statistical power. For each participant, a classifier was trained to differentiate patterns of activation associated with face, object, scrambled object, and passive fixation blocks using data from the visual localizer task. Following training, the classifier trained on localizer data from FFA was able to predict the visual content associated with new, unlabeled activation patterns from the same participant's brain with high accuracy (range: 70.4%-96.1%, mean ± SEM: 84.4% ± 1.1% correct). Similar results were found for the posterior fusiform gyrus ROI (74.8%-95.6%, 87.4% ± 0.7% correct).

The goal of the MVPA approach in the present study was to detect reactivation of previously encoded content (i.e., face information) during offline rest periods. To test the hypothesis that face information would be reactivated following the AB pair pre-training phase, I applied the trained classifier to each volume of the post-AB rest scan, while the participant was viewing a fixation cross. The classifier estimates were extracted for each volume, yielding reactivation timecourses specific to each stimulus category for each participant.

A sliding window analysis was used to determine how the relationship between reactivation and BC memory performance changed across the duration of the post-AB rest scan. For each 60-TR window in the rest scan, a reactivation index was calculated for each participant as the average face evidence across the window. While I focus primarily on face evidence (**Figs. 17** and **18**, top left), this measure was also calculated for the remaining stimulus categories, for which I did not see any relationship to behavior (see **Fig. 18**, top right). This index was then related to BC performance across participants using partial correlation (controlling for XY performance; see below for further justification as to why partial correlation was used for the primary analyses). This resulted in a partial correlation value (r) that was Fisher transformed (z) and assigned to the middle TR of the window. I defined 95% confidence intervals on the partial correlation statistic for each window using a bootstrapping procedure as follows: for each of 1,000 iterations, a group of 35 participants was drawn randomly with replacement from my sample. I then repeated the pattern classification analysis with this simulated group of participants, yielding a partial correlation statistic (Fisher's z) for each analysis. After repeating this procedure 1,000 times, partial correlation statistics were sorted and confidence intervals were drawn at the 25th (the upper 2.5%) and 975th (lower 97.5%) z values. The window was then shifted by one TR and the entire procedure was repeated. This analysis was also performed separately for individual relationships with BC and XY performance using Pearson's correlation (**Fig. 18**, bottom row). As the relationship between reactivation and performance was unique to the beginning of the scan, the results

111

reported below focus on how reactivation relates to behavior during the first 60-TR window (**Fig. 17**).

I used partial correlation to index the degree to which prior memories facilitate versus interfere with the acquisition of new, related knowledge. Because general associative memory ability (i.e., XY performance) was highly related to both BC memory and AC inference across participants, I needed to statistically control for these differences to answer my central question—how prior knowledge specifically impacts overlapping encoding, *relative* to one's general associative encoding ability. Mathematically, this is accomplished by performing a correlation on the residuals after regressing both reactivation and BC or AC performance on the controlling variable, XY performance. This analysis approach mirrors other studies that control for various factors such as age (Feldman et al., 2012; Suzuki et al., 2013), sex (Suzuki et al., 2013), general cognitive ability (Chiu et al., 2006; Zeithamova et al., 2012a), or neural measures (Supekar et al., 2010; Barron et al., 2013).

Individual relationships with BC, AC, and XY performance were also assessed in the first 60-TR window using Pearson's correlation. I also compared the separate relationships of reactivation with BC and AC with that of XY using the Hotelling-Williams procedure (Hotelling, 1940; Williams, 1959), a statistical test that accounts for shared variance among related measures.

I also performed several control analyses that leveraged the post-XY encoding rest period to investigate the specificity of this reactivation-performance relationship to the post-AB rest. These analyses were performed to assess the hypothesis that the

observed measures relate specifically to memory integration; and are modulated by experience (i.e., they are not stable within an individual over time). For these analyses, I treated the post-XY rest as a baseline period. I believe this serves as an appropriate control in the present study for two reasons. First, of the three rest scans acquired in the present study (post-AB, post-BC and post-XY encoding), the post-XY encoding rest is the most removed from the demands of overlapping encoding (e.g., memory integration or interference resolution). For this reason, it is likely that this period would reflect the most recent experience: encoding of object-object (XY) associations that did not overlap with the critical A_{face} information. Second, while post-XY encoding rest may not reflect "true" baseline activity in the traditional sense, it is a stringent and appropriate control for studying the phenomenon of interest. That is, as I expect post-XY encoding rest to reflect persistence of signatures associated with simple associative encoding, any additional effects observed during post-AB encoding rest are all the more likely to relate *specifically* to overlapping encoding. Importantly, as the present neural measures are defined specifically to index processing of face-related memories, I would not expect the degree of FFA reactivation during the post-XY period to relate to XY performance.

Post-AB Functional Connectivity Analysis

Functional connectivity was examined using two approaches: first, a timeseries correlation within anatomically and functionally-defined ROIs; and second, a voxelwise regression approach using FFA as a seed region.

Timeseries Correlation Analysis

For each participant, functional data from the post-AB rest scan were high-pass filtered with a cutoff of 0.009 Hz, which has been employed in previous studies examining rest-phase connectivity (Fox et al., 2005; Tambini et al., 2010). For each FFA and MTL ROI, the first eigenvariate of the signal across all voxels in the mask was extracted across the full 6 minutes (180 TRs) of the post-AB rest period. For each participant, the FFA timeseries was then correlated with the timeseries from each MTL ROI. This procedure resulted in four correlation values (r) per participant (FFA-hippocampus, FFA-ERc, FFA-PRc, FFA-PHc), representing the degree of functional connectivity between FFA and each of the MTL ROIs during the post-AB rest period. For each pair of ROIs, connectivity was related to BC and AC performance across participants using partial correlation (controlling for XY performance). Individual relationships with BC, AC, and XY performance were also assessed using Pearson's correlation.

I also repeated this analysis after regressing out potential sources of noise from the resting state data. Specifically, I extracted the first eigenvariate of the signal across the duration of the rest scan in anatomically defined white matter and lateral ventricle ROIs. The signal from these two ROIs and their temporal derivatives were used to construct a GLM along with: motion parameters, FD, DVARS, and their temporal derivatives; and timepoints with excessive motion (as described above). The high-pass filtered data from the previously described step was then regressed on these noise sources, and the first eigenvariate was extracted from the resulting data (i.e., the residual

114

timeseries). As in the above-described analysis, each participant's FFA timeseries was then correlated with their MTL timeseries, resulting in four indices representing functional connectivity between FFA and the four MTL regions. These measures were then related to performance using partial correlation and Pearson's correlation.

Seed-Based Analysis

This analysis interrogated activation within MTL (inclusive of hippocampus, ERc, PRc and PHc) for voxels that showed rest-phase connectivity with FFA that related to subsequent performance. I performed this analysis to account for the potential heterogeneity in response profiles within the MTL, thereby complementing the above ROI approach above. Each participant's high-pass filtered functional data were spatially smoothed (5 mm FWHM). The FFA timeseries and its first temporal derivative were entered into a GLM, along with motion-related confound regressors (motion parameters, FD, DVARS, and their temporal derivatives; and timepoints with excessive motion). The resulting parameter estimate image for each participant reflected the degree to which activation in each MTL voxel tracked with FFA activation across the post-AB rest scan. I then warped these images to the 1 mm MNI template using ANTS (Avants et al., 2011) and combined them across participants using a group-level GLM as follows.

I constructed the group-level model with both BC and XY performance as covariates. I was specifically interested in those MTL voxels whose connectivity with FFA was modulated by individual differences in BC performance (after covarying out XY performance), as well as those voxels that tracked more with BC than XY performance (BC > XY contrast). Voxelwise statistics were calculated within MTL using

permutation tests implemented in FSL. I first applied an uncorrected voxelwise threshold of $p < 0.025$ to the group statistics images. I then corrected for multiple comparisons within the MTL using a small volume correction procedure to determine the cluster size corresponding to a cluster-corrected threshold of $p < 0.05$. This calculation was carried out using 3dClustSim, part of AFNI (Cox, 1996). 3dClustSim performs Monte Carlo simulations that take into account the size and shape of the ROI as well as the smoothness of the data in determining a critical cluster size. Cluster sizes that occurred with a probability of less than 0.05 across 2000 simulations were considered statistically significant. This procedure yielded a critical cluster size of 326 (1 mm^3) voxels for the MTL ROI (see **Figs. 20B** and **21A**).

Multiple Regression Analysis

Multiple linear regression analyses were performed to further assess the degree to which reactivation and connectivity indices were independently related to performance. Two regression models were run: one with BC performance as the dependent variable, and one with AC performance as the dependent variable. Reactivation, connectivity, and XY performance measures for each participant were entered into the regression as predictors. Participants were treated as a random effect.

Univariate Analysis

Analysis of fMRI data from the memory task was carried out using FEAT (fMRI Expert Analysis Tool) version 6.00, part of FSL. As with the localizer data, the following pre-statistics processing was applied: grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; high-pass temporal filtering (Gaussian-weighted

116

least-squares straight line fitting, with sigma = 64 s); and spatial smoothing (5 mm FWHM). FILM prewhitening was used. Encoding trials were sorted based on subsequent memory in the cued recall test to create four conditions: BC encoding trials that were later correct, BC encoding trials that were later incorrect, XY encoding trials that were later correct, and XY encoding trials that were later incorrect. Stimulus presentations were modeled as events with 3.5 s durations, with one regressor for each of the four conditions. The model was convolved with the canonical (double-gamma) HRF. Motion parameters calculated during the motion correction step and their temporal derivatives were added as additional regressors of no interest. As described above, FD, DVARS, and individual regressors for time points exceeding FD and DVARS thresholds were added to the model to additionally account for motion effects (Power et al., 2012; Schonberg et al., 2014). Temporal filtering was applied to the model.

After modeling functional data within each run, the resulting statistics images were warped to the 1 mm MNI template brain using ANTS (Avants et al., 2011). The warped images were combined across encoding runs for each participant using fixed effects, and then across the group using mixed effects. Correction for multiple comparisons were carried out on group-level voxel-wise statistical images according to cluster-based Gaussian random field theory (Worsley et al., 2002), with a cluster-forming threshold of $z > 2.3$ and a whole-brain corrected cluster significance level of $p < 0.05$.

In addition to a group mean model, two models with covariates were created to investigate how individual differences in rest-phase reactivation and functional connectivity relate to subsequent neural activation during learning of BC and XY pairs.

117

For each of these two models, a second regressor was created containing FFA reactivation and FFA-hippocampal connectivity indices, respectively, for every participant. I then examined regions whose activity was modulated by these covariates. Correction for multiple comparisons was performed as described previously.

Delay and Encoding Order Analyses

Participants experienced a variable delay between AB pre-training and the post-AB encoding rest scan (median delay length = 22 minutes, S.E.M. = 1.76, range = 15-58). Moreover, as encoding order was counterbalanced across participants, there was also a difference in time between the initial learning of the AB pairs and subsequent encoding over the overlapping BC pairs. To rule out these potential confounds, I performed two analyses. First, I assessed the continuous relationship between delay duration and FFA reactivation; and delay duration and FFA-hippocampal connectivity measures using Pearson's correlation. I also performed one-way analyses of covariance (ANCOVA) to interrogate whether the observed relationships between neural measures and performance differed significantly as a function of encoding order. The neural measure (FFA reactivation or FFA-hippocampal connectivity) served as the predictor variable; the behavioral measure served as the response. Encoding order was the grouping variable.

5.4. RESULTS

5.4.1. Behavioral Results

As intended, AB pairs were well learned by the fourth test block (mean ± SEM: 97.3% ± 0.9% correct recall; **Fig. 16A**). I defined proactive interference as performance on overlapping BC relative to non-overlapping XY pairs (i.e., XY-BC accuracy), with

higher values indicating more interference of AB pair knowledge on new BC encoding. Importantly, BC (11.7%-86.7%, 41.5% ± 3.3% correct) and XY (10%-78.3%, 42.4% ± 3.4% correct) were matched in terms of both content type and number of presentations, allowing me to directly compare performance in these two conditions.
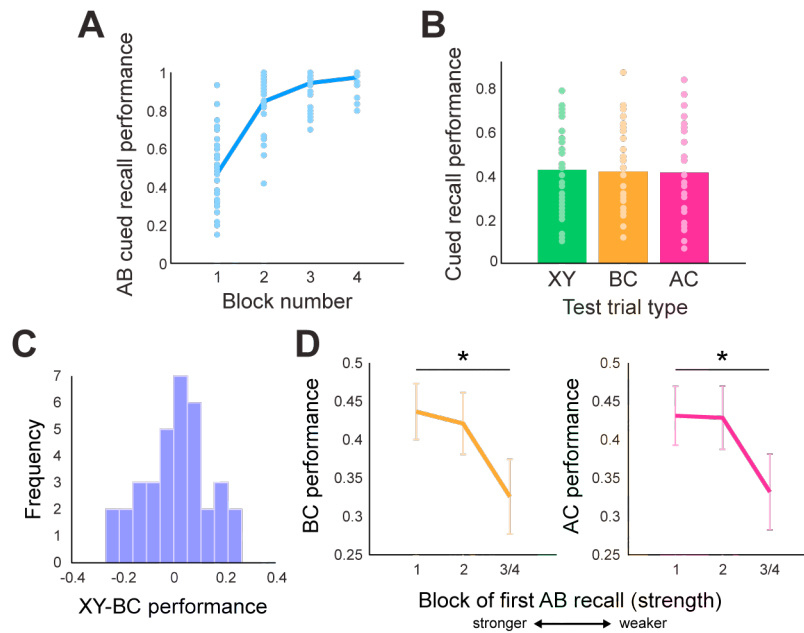
Figure 16. Experiment 2A: Behavioral Results

(**A**) AB memory performance as proportion correct on each test block. Line represents the group mean; points show individual participants. (**B**) Performance for non-overlapping XY pairs (green), overlapping BC pairs (orange), and AC inferences (pink). Bar heights represent group means; points show individual participants. (**C**) Histogram depicting distribution of behavioral interference in my sample (displayed as the difference between XY and BC performance). Positive values indicate better XY than BC performance, evidencing interference of prior AB knowledge on new encoding; negative values indicate better BC than XY performance, indicating AB knowledge-related facilitation. (**D**) BC performance (orange; left) and AC performance (pink; right) as a function of AB memory strength. AB pairs were sorted according to whether they were correctly recalled initially on the first, second or third/fourth pre-training test block (x-axis). BC and AC performance was then computed separately for these groups of triads (y-axis). There was a significant linear effect of block of first AB recall on both BC and AC performance; participants demonstrated better memory for those AB pairs learned earlier in the pre-training phase, suggesting that strong AB memories facilitate new learning. Significant linear effects denoted by asterisks.

119

Interestingly, I observed neither proactive interference nor facilitation across the group (XY vs. BC performance: $t_{34} = 0.40$, $p = 0.693$; **Fig. 16B**); rather, I found that the degree of proactive interference was highly variable across individuals (mean of XY-BC performance = 0.9%; S.E.M. = 2.2%; range = -26.7%-26.7%; **Fig. 16C**). Approximately half of the participants showed interference (N = 16 had numerically better XY than BC memory), while half showed facilitation (N = 15 had numerically better BC than XY memory). The remaining four participants had equal memory for BC and XY pairs. This variability enabled me to investigate how rest-phase processes following initial AB learning modulate encoding of overlapping BC relative to control XY pairs.

I also found that performance on AC inferences (6.7%-83.3%; 41% ± 3.5% correct)—which notably, required retrieval of the initially learned A_{face} item—paralleled BC memory, further demonstrating the strong nature of the AB memories at the end of the experiment. Both across participants ($r_{33} = 0.98$) and on a triad-by-triad basis within participant, performance on BC pairs and AC inferences was highly correlated. Importantly, C items served as cues for both BC and AC test trials. Thus, using the example from **Fig. 15**, the participant was first presented with CANDLE and asked to produce HARP as a test of BC memory. When the participant was again presented with CANDLE during the AC inference test, they should recall BRAD PITT. Because of this structure, AC inference can serve as an approximation of AB performance at the end of the memory task. I computed AC performance for only those triads for which the corresponding BC pair was recalled. Participants correctly recalled the vast majority of AC inferences when BC was correct (mean accuracy = 88.6%; median = 91.7%; S.E.M.

= 2.1%; range = 57.1%-100%). This corresponded to on average just over two incorrect AC inferences when the overlapping BC was recalled (mean number of incorrect AC trials = 2.1; median = 2; S.E.M. = 0.3; range = 0-8). These high levels of performance suggest that AB pairs were not forgotten over the course of the experiment.

I also investigated how the strength of the initially acquired AB pairs impacted later BC learning and AC inference. I found that both within and across individuals, superior AB memory was associated with better performance on overlapping BC pairs and AC inference judgments. I found a significant linear effect of AB memory strength on BC performance ($F_{1,31}$ = 5.19, p = 0.030; **Fig. 16D**, left chart). This suggests that, consistent with an integrative encoding perspective, stronger initial memories are associated with superior encoding of overlapping content. Performing the same analysis for AC performance revealed a similar relationship ($F_{1,31}$ = 3.99, p = 0.055; **Fig. 16D**, right chart), further suggesting that strong prior knowledge facilitates memory integration for subsequent flexible use.

5.4.2. Imaging Results

Due to the volume of complementary and control analyses performed, the reader should note that only especially pertinent findings are mentioned in the main results text. Full details are provided in subsections entitled **Control Analyses** within each set of results.

Post-AB Pattern Classification Results

I examined the impact of neural engagement during the post-AB rest period on encoding of related BC information and AC inferences. Here, I focus on reactivation of

face information in face-sensitive regions of visual cortex (e.g., fusiform face area, FFA). To measure spontaneous reactivation during the rest period, I trained a pattern classifier to distinguish between different types of visual content on the basis of activation patterns in each participant's functionally defined FFA. Importantly, the classifier was trained on independent visual localizer data. The trained classifier was then applied to each volume of the post-AB rest period (**Fig. 17**, top).
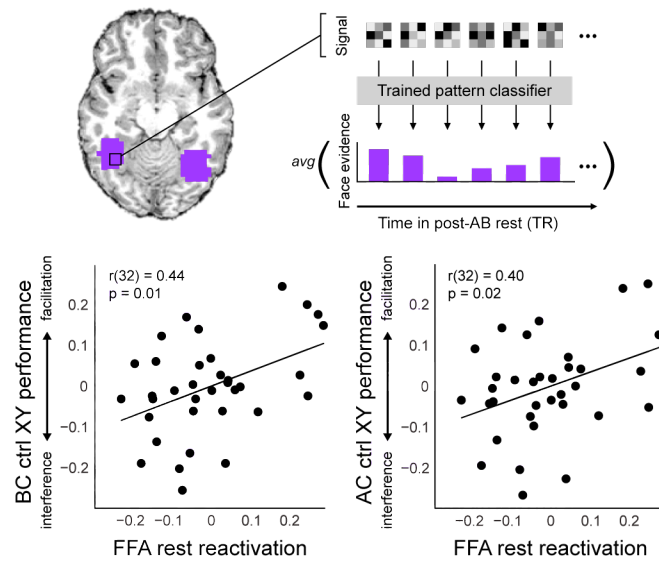
Figure 17. Experiment 2A: Post-AB MVPA Results

Reactivation following initial learning predicts subsequent encoding of related content. Top, depiction of rest-phase pattern classification analysis. A pattern classifier was trained to discriminate FFA (purple) activation patterns associated with faces, objects, scrambled objects, and fixation baseline (not depicted). The trained classifier was then applied to each time point of the rest data (grayscale matrices). Bottom, relationship between FFA reactivation and BC (left) and AC (right) performance, plotted as residuals after regressing each on XY performance.

I first examined the relationship between reactivation and performance over time using a 60-volume (2-minute) window swept across the rest scan. For each window, I

calculated a reactivation index (defined as the mean classifier evidence for faces) for each participant. I then related this reactivation index to BC learning and AC inference using two approaches. As my primary approach, I quantified the degree of facilitation for BC encoding and AC inference by performing across-participant partial correlations. Specifically, I interrogated the relationships between (a) reactivation and BC memory performance and (b) reactivation and AC inference performance, after statistically controlling for the effects of XY performance (our metric of general associative memory ability). This analysis was performed to index the unique relationship between memory reactivation and later encoding of related information. As a secondary approach, I also investigated the individual relationships between reactivation and performance on related BC pairs, AC inferences, and unrelated XY pairs using Pearson's correlation. However, I note that due to the high correlation between XY pair memory and performance on both BC pairs ($r_{33} = 0.80$, $p < 1 \times 10^{-8}$) and AC inferences ($r_{33} = 0.81$, $p < 1 \times 10^{-8}$), these relationships are heavily influenced by general associative memory and thus do not specifically reflect the impact of prior knowledge on subsequent encoding of related information.

I found a significant relationship between face reactivation and BC performance controlling for XY that was unique to the beginning of the rest period (partial correlation during first 2-minute window; $r_{32} = 0.44$, $p = 0.010$; **Fig. 17**, bottom left; **Fig. 18**, top left) and that was not observed for other classes of visual content (**Fig. 18**, top right). Reactivation during the first 2-minute window also tracked AC performance after controlling for XY memory ($r_{32} = 0.40$, $p = 0.019$; **Fig. 17**, bottom right). That is,

participants who showed more face reactivation following initial AB learning also showed superior memory for related BC associations and AC inferences after controlling for general associative memory ability. This finding can also be conceptualized as a *negative* association between face reactivation and proactive interference, i.e., less reactivation was observed for participants who showed more proactive interference. Importantly, neither the degree of reactivation nor its relationship to performance was significantly impacted by differences in lag duration or encoding order across participants (see ***Delay and Encoding Order Results***).

Moreover, the relationship between reactivation and performance was specific to the post-AB rest scan. Reactivation during the post-XY rest period did not relate to XY performance; or to BC or AC performance after controlling for general associative memory.

Next, to further validate these findings, I repeated the same analysis using an expanded region of interest (ROI) encompassing the entire posterior fusiform gyrus. I found significant relationships between reactivation in posterior fusiform and BC performance ($r_{32} = 0.38$, $p = 0.028$) as well as AC inference ($r_{32} = 0.37$, $p = 0.032$) when controlling for general associative memory. Moreover, I observed significant individual correlations of BC performance and AC inference with reactivation that were not observed in the smaller FFA ROI (**Fig. 19**). Importantly, reactivation during the post-AB scan was not related to XY performance for either ROI (**Fig. 19**).
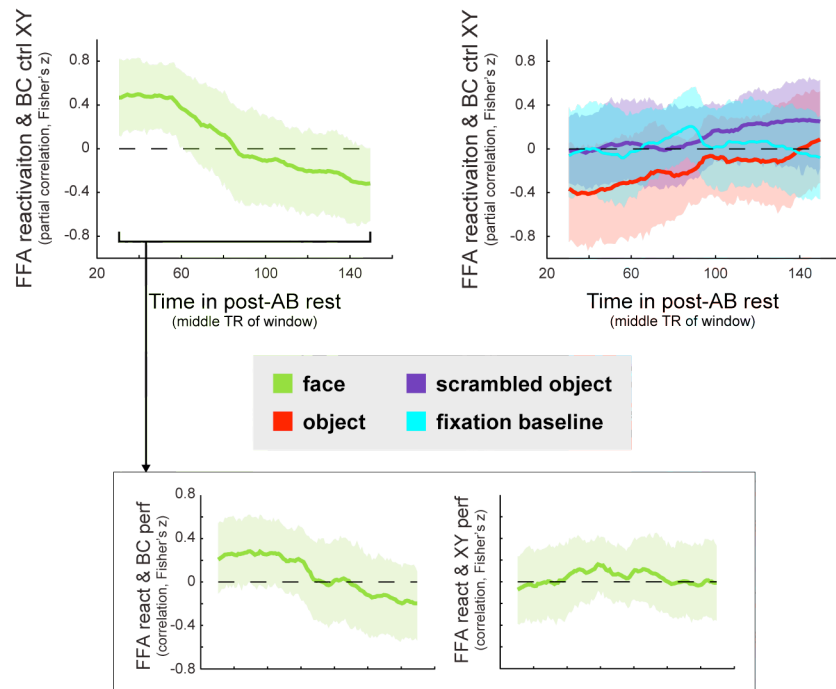
Figure 18. Experiment 2A: Post-AB MVPA Sliding Window Results

Relationship between FFA reactivation and memory over time. Top row, relationship between FFA reactivation and BC performance controlling for XY performance over time. For each 60-TR window, I calculated the partial correlation between mean classifier evidence for all four stimulus categories (face, left; object, scrambled object, fixation baseline, right) in FFA and BC performance, controlling for XY performance. Partial correlation values (r) were Fisher transformed (z) and assigned to the middle TR of the window. Data are plotted as partial correlation values (y-axis) over time (middle TR; x-axis). Face evidence (left; green) was significantly related to BC performance (controlling for XY performance) at the beginning of the rest scan; this relationship decreased over time. Right panel, relationship between classifier evidence for remaining stimulus categories (right; object, red; scrambled object, purple; fixation baseline, cyan) and performance. There was no significant relationship between evidence for any of these stimulus types and BC performance controlling for XY across the duration of the rest scan. Bottom row, individual relationships between FFA reactivation and BC and XY performance over time. I calculated separate Pearson's correlations between mean classifier evidence for faces in FFA and BC performance (left) and XY performance (right) for each window. As above, correlation values (r) were Fisher transformed an assigned to the middle TR. Data are plotted as correlation values (y-axis) over time (middle TR; x-axis). The significant partial correlation (top left, early in scan) was driven by a positive relationship with BC performance early in the scan and no relationship to XY performance. For all plots, shaded region represents 95% bootstrapped confidence intervals.

A Fusiform Face Area (FFA)

B Posterior fusiform gyrus
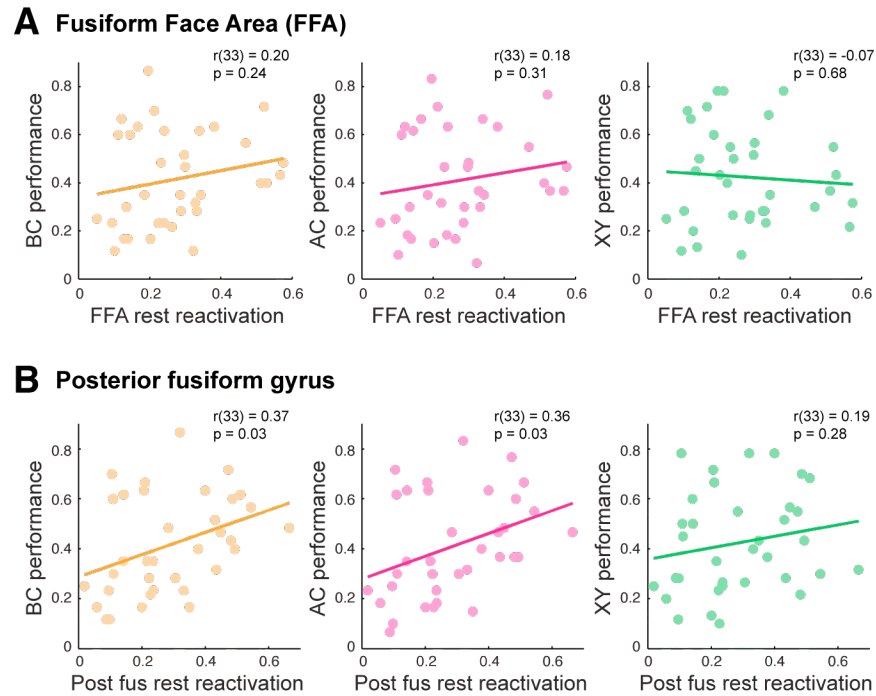
Figure 19. Experiment 2A: Reactivation First Order Correlations with Performance

Individual relationships between reactivation and performance for FFA (**A**) and posterior fusiform gyrus (**B**) ROIs. (**A**) Relationships between FFA reactivation (x-axes) and BC (left; orange), AC (middle; pink), and XY (right; green) performance (y-axes). BC and AC performance were positively (though nonsignificantly) associated with reactivation; XY was negatively (though nonsignificantly) associated with reactivation. (**B**) Relationships between posterior fusiform reactivation and performance. Data are displayed as in (**A**). Both BC and AC performance showed significantly positive relationships with reactivation, such that greater reactivation was associated with superior performance. There was also a numerically positive but nonsignificant relationship with XY. For both (**A**) and (**B**), results were similar after subtracting baseline reactivation levels derived from the post-XY encoding rest period (not depicted). All statistics reflect Pearson's correlations.

I also performed a control analysis to determine whether these findings could be attributed to individual differences in baseline levels of face reactivation. Because the post-XY encoding rest scan was the most removed from face-related encoding, I reasoned that this scan should be the least likely to contain face-related neural signatures that would support memory. Thus, I subtracted each participant's post-XY encoding

126

reactivation index from their post-AB reactivation index (Tambini et al., 2010). The resulting difference scores reflecting the degree to which post-AB reactivation deviated from baseline (as indexed by the post-XY scan) were related to performance as described above. For both FFA and posterior fusiform gyrus, I observed significant relationships between post-AB reactivation with BC and AC performance controlling for general associative memory using partial correlation. Moreover, in the posterior fusiform gyrus ROI, individual relationships between reactivation and performance were significant for BC and AC, but not XY.

Control Analyses

The partial correlation between reactivation and memory facilitation during the post-AB rest was driven by a numerically positive relationship between reactivation and BC performance ($r_{33} = 0.20$, $p = 0.239$); and a numerically negative relationship between reactivation and XY performance ($r_{33} = -0.07$, $p = 0.681$; **Figs. 18**, bottom row and **19**). While neither relationship was statistically significant on its own, they were significantly different from one another (Hotelling-Williams test; $t_{32} = 2.74$, $p = 0.010$). Tracking BC performance, the relationship between reactivation and AC performance was also positive, though it too did not reach statistical significance ($r_{33} = 0.18$, $p = 0.310$; **Fig. 19**). The relationship between reactivation and AC performance also differed significantly from XY (Hotelling-Williams test; $t_{32} = 2.50$, $p = 0.018$). Importantly, I present these data with the caveat that the individual relationships are difficult to interpret, as they include effects attributable to general associative memory ability.

Using the entire posterior fusiform gyrus ROI, the partial correlations between reactivation and BC performance ($r_{32} = 0.38$, $p = 0.028$) and reactivation and AC performance ($r_{32} = 0.37$, $p = 0.032$) remained statistically significant. Moreover, the individual relationships (**Fig. 19**) between reactivation and BC performance ($r_{33} = 0.37$, $p = 0.028$) and reactivation and AC performance ($r_{33} = 0.36$, $p = 0.032$) were significant in this larger ROI, while the relationship with XY performance remained nonsignificant ($r_{33} = 0.19$, $p = 0.282$). The difference in the reactivation-performance relationships for both BC and XY (Hotelling-Williams test; $t_{32} = 1.81$, $p = 0.079$) and AC and XY (Hotelling-Williams test; $t_{32} = 1.78$, $p = 0.085$) trended toward significance in this expanded region.

I also interrogated the relationship between FFA reactivation during post-XY encoding rest measures and memory performance. One alternative account of the present findings is that the observed neural measures simply reflect general associative memory ability and are unrelated to memory integration in particular. To assess this possibility, I related FFA reactivation during the post-XY period to general associative memory performance (XY) across participants. I found that face reactivation in FFA following XY encoding was not related to general associative memory (XY) performance ($r_{33} = 0.11$, $p = 0.52$), rendering an account based purely on general associative memory signatures improbable. A second alternative account is that stable individual differences in reactivation are associated with superior memory integration. If this were indeed the case, one would expect that reactivation measures from an unrelated rest period (e.g., post-XY encoding rest) should predict BC performance after controlling for XY. This was not the case in these data; I observed no relationship between post-XY reactivation

and BC (r_{32} = -0.06, p = 0.747) or AC (r_{32} = -0.09, p = 0.587) performance after controlling for XY performance using partial correlation. Thus, these control analyses are consistent with the notion that enhanced reactivation during rest confer a specific behavioral advantage on subsequent learning of related content.

I also subtracted post-XY reactivation from post-AB reactivation for each participant (Tambini et al., 2010); these difference scores were then related to performance as described previously. The goal of this approach was to account for differences in baseline connectivity across participants. This analysis replicated my prior findings for both FFA and anatomically defined posterior fusiform gyrus ROIs. Specifically, reactivation in FFA was related to BC memory (r_{32} = 0.33, p = 0.054) and inference (r_{32} = 0.34, p = 0.050) after controlling for general associative memory performance. Moreover, the correlations between reactivation and BC (Hotelling-Williams test; t_{32} = 2.14, p = 0.040) and AC (Hotelling-Williams test; t_{32} = 2.18, p = 0.037) performance each differed significantly from the relationship between reactivation and XY performance. In posterior fusiform gyrus, reactivation also maintained its significant relationships with BC memory (r_{32} = 0.39, p = 0.023) and inference (r_{32} = 0.42, p = 0.013) after controlling for XY performance. In addition, the individual relationships between reactivation and BC performance (r_{33} = 0.39, p = 0.022) and AC inference (r_{33} = 0.40, p = 0.017)—but not XY performance (r_{33} = 0.20, p = 0.259)—were significant. The difference between BC and XY correlations with reactivation approached significance (Hotelling-Williams test; t_{32} = 1.87, p = 0.071); its relationship with AC was significantly different from XY (Hotelling-Williams test; t_{32} = 2.13, p = 0.041).

Post-AB Functional Connectivity Results

Next, I sought to determine how FFA connectivity with medial temporal lobe (MTL) regions predicted subsequent learning of object-object pairs. I employed two approaches: first, a timeseries correlation approach within anatomically and functionally defined ROIs; and second, a voxelwise regression approach using FFA as a seed region. Both analyses were performed with consideration of the entire rest scan, as prior reports have shown the importance of sufficiently long timeseries for extracting stable measures of functional connectivity (Van Dijk et al., 2010).

Timeseries Correlation Approach

ROIs included functionally defined FFA and anatomically defined hippocampus, perirhinal, entorhinal and parahippocampal cortices (all bilateral). I extracted the first eigenvariate across all voxels in each ROI from the high-pass filtered post-AB rest data. I then correlated the FFA timeseries with the timeseries from each of the four MTL ROIs (**Fig. 20A**, top). As in the rest-phase reactivation analysis, functional coupling during post-AB rest was related to BC performance after controlling for XY performance (r_{32} = 0.37, p = 0.033; **Fig. 20A**, bottom left). A similar relationship was observed with AC performance (r_{32} = 0.34, p = 0.049; **Fig. 20A**, bottom right). That is, participants showing enhanced FFA-hippocampal functional coupling following AB encoding also showed an advantage specific to learning of the overlapping BC associations and inferring the AC relationships. Connectivity between FFA and all other MTL regions showed no significant relationship to BC or AC performance after controlling for XY (all $|r_{32}| < 0.18$, p > 0.312). Moreover, neither the degree of connectivity itself nor its relationship to

performance was modulated by lag duration or encoding order (see ***Delay and Encoding***

Order Results).
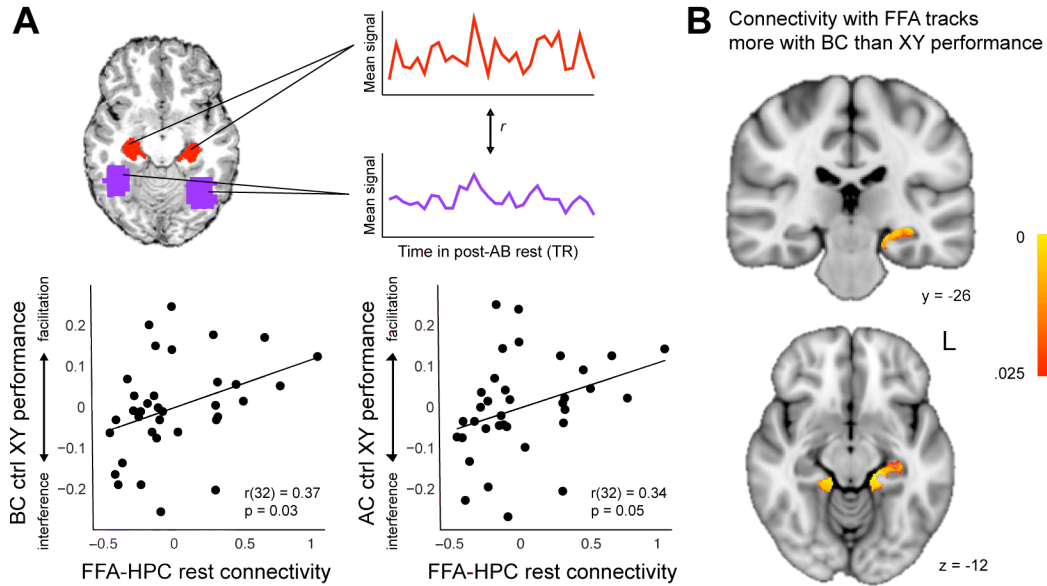
Figure 20. Experiment 2A: Functional Connectivity Results

FFA-hippocampal connectivity following initial learning predicts subsequent encoding of related content. (**A**) Top, depiction of timeseries correlation analysis. For each participant, FFA (purple) and hippocampal (red) timecourses from the post-AB rest scan were correlated to quantify the degree to which FFA and hippocampus exhibit similar activations over time (top right). Bottom, relationship between FFA-hippocampal connectivity and BC (left) and AC (right) performance, controlling for XY performance. Data are displayed as in **Fig. 17**. (**B**) Hippocampus showed connectivity with FFA during rest that was significantly more predictive of BC than XY performance (displayed on the 1 mm MNI template brain). Color bar indicates uncorrected voxelwise p-value. Coordinates are in mm. HPC, hippocampus.

Importantly, the relationship between connectivity and performance was also specific to the post-AB rest; there was no correlation between FFA-hippocampal connectivity during the post-XY rest period and XY performance; or with BC or AC performance controlling for XY performance. When various nuisance sources (signal

from white matter and ventricular ROIs and motion-related regressors) were regressed out from the post-AB rest data, the pattern of results was similar but slightly weaker (partial correlations with BC: $r_{32} = 0.30$, $p = 0.088$ and AC: $r_{32} = 0.30$, $p = 0.089$ controlling for XY).

As with the reactivation results, I performed a control analysis to account for individual differences in baseline connectivity by subtracting the degree of post-XY FFA-hippocampal connectivity from each participant's post-AB connectivity measure of interest. I found that my results held, with significant correlations between connectivity and BC learning (as measured by both individual and partial correlations), as well as AC inference. There was no relationship between FFA-hippocampal connectivity during post-AB rest and XY performance.

Control Analyses

Relationships to BC and XY; and AC and XY were compared using the Hotelling-Williams test (Hotelling, 1940; Williams, 1959). The individual relationships to BC ($r_{33} = 0.12$, $p = 0.500$), AC ($r = 0.09$, $p = 0.59$), and XY ($r_{33} = -0.13$, $p = 0.467$) performance did not reach statistical significance. However, the connectivity-performance relationships for BC and XY performance were significantly different from one another (Hotelling-Williams test; $t_{32} = 2.36$, $p = 0.024$); as was the difference between relationships with AC and XY performance (Hotelling-Williams test; $t_{32} = 2.18$, $p = 0.037$).

I also calculated connectivity measures for the post-XY encoding rest period for the purposes of the control analyses. Connectivity was related to XY performance using

Pearson's correlation; and to BC and AC performance controlling for XY using partial correlation. I found no relationship between FFA-hippocampal connectivity following XY encoding and XY memory performance (r_{33} = -0.22, p = 0.20), suggesting that this signature does not simply reflect general associative memory ability. There was also no relationship between connectivity and BC (r_{32} = -0.11, p = 0.538) or AC (r_{32} = -0.11, p = 0.534) performance after controlling for XY.

To control for possible individual differences in baseline connectivity, post-XY connectivity was also subtracted from post-AB connectivity for each participant (Tambini et al., 2010). The resulting difference scores were then related to performance as above. The difference scores showed a significant relationship with BC (r_{32} = 0.40, p = 0.018) and AC (r_{32} = 0.38, p = 0.025) performance after controlling for general associative memory, as well as significant individual relationships with BC (r_{33} = 0.35, p = 0.040) and AC (r_{33} = 0.33, p = 0.052) performance. There was no association between FFA-hippocampal connectivity and XY performance (r_{33} = 0.13, p = 0.444). BC and XY (Hotelling-Williams test; t_{32} = 2.10, p = 0.043); and AC and XY (Hotelling-Williams test; t_{32} = 1.98, p = 0.056) relationships with FFA-hippocampal connectivity also differed from one another.

Seed-Based Approach

I also employed a more sensitive seed-based regression approach to identify specific MTL voxels for which connectivity with FFA tracked subsequent BC performance. I regressed each participant's MTL data on their FFA timeseries from the post-AB encoding rest scan, resulting in a statistics image representing the degree of

correspondence between each MTL voxel and FFA activation over time. These results were combined across participants in a group level general linear model. I found two regions for which FFA connectivity tracked more with BC than XY performance: one in left (MNI coordinates (mm) x, y, z = -16, -30, -19) and one in right (14, -33, -11) hippocampus, extending into PHC (**Fig. 20B**). An overlapping cluster (-18, -29, -13) in the left hemisphere was found to predict BC performance (**Fig. 21A**). Similar results were obtained after repeating the seed-based analysis on the residuals from the nuisance regressed analysis described above (**Fig. 21B**).

Multiple Regression Results

To investigate the degree to which reactivation and connectivity independently explained variance in subsequent learning, I next performed two multiple linear regression analyses with indices of reactivation, connectivity, and XY performance as independent variables; and BC and AC performance, respectively, as the dependent variables. The model fit was significant ($F_{3,31}$ = 38.90, p < 0.0001), accounting for 77.0% of the variance in BC performance (adjusted R^2). Moreover, all three independent variables showed a significant positive relationship to BC performance (reactivation: β = 0.33, p = 0.0004; connectivity: β = 0.30, p = 0.001; XY performance: β = 0.87, p < 0.0001; all statistics reflect standardized β), demonstrating the unique contributions of rest-phase reactivation and FFA-hippocampal connectivity to subsequent learning of related experiences. Similar results were found in the regression model predicting AC performance ($F_{3,31}$ = 36.67, p < 0.0001; adjusted R^2 = 0.759; reactivation: β = 0.30, p = 0.002; connectivity: β = 0.27, p = 0.004; XY performance: β = 0.87, p < 0.0001).

A Connectivity with FFA tracks with BC performance

y = -28   z = -12   L

B Connectivity with FFA tracks more with BC than XY performance

y = -26   z = -12   L
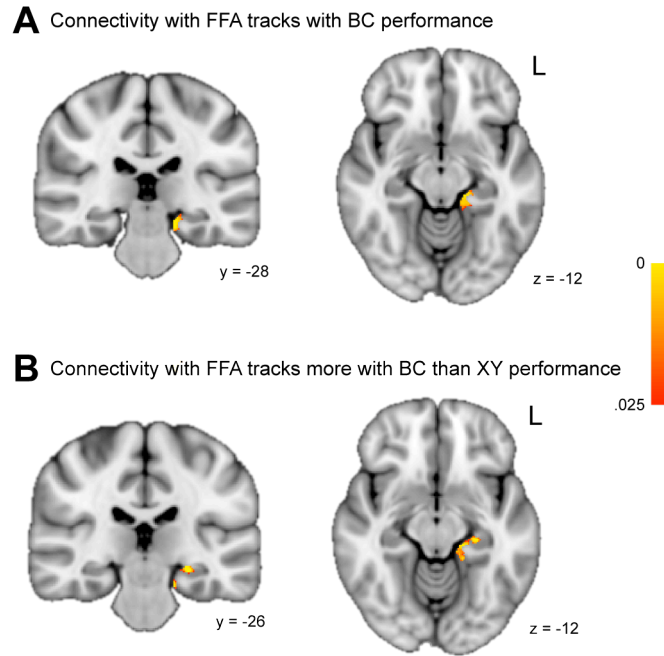
Figure 21. Experiment 2A: Control Seed-Based Connectivity Analyses

(**A**) MTL regions for which connectivity with FFA during rest significantly predicts BC performance. Left hippocampus [-18, -29, -13] was the only region for which greater post-AB rest-phase connectivity with FFA was associated with superior BC performance. No regions showed a positive relationship with XY performance. (**B**) MTL regions for which connectivity with FFA during rest is significantly more predictive of BC than XY performance. These results depict analysis performed on rest data following nuisance regression. This contrast revealed a cluster in left hippocampus [-17, -30, -20] that overlapped with the results from the high-pass filtered rest data (see **Fig. 20B**). For both panels, clusters are significant after correction for multiple comparisons within the MTL. Color bar indicates uncorrected voxelwise p-value. Coordinates are in mm.

Univariate Results

I next investigated neural engagement during encoding of new object-object associations. I was specifically interested in regions demonstrating a subsequent memory effect for BC (i.e., more engagement during study of BC pairs that were subsequently remembered vs. forgotten) but not XY pairs. As the above results showed that certain rest-phase processes can facilitate BC encoding, I hypothesized that (1) face-sensitive

regions of visual cortex would be engaged during BC trials, indicative of reinstatement of previously learned A_{face} stimuli; and that (2) such engagement would support encoding of the new BC object-object pairs. A whole-brain analysis confirmed my predictions, revealing a significant subsequent recall by condition interaction (correct > incorrect × BC > XY) in left fusiform gyrus (-20, -75, -15; **Fig. 22A**).

I then considered how individual differences in rest-phase reactivation and FFA-hippocampal connectivity related to neural engagement during subsequent learning. I created two GLMs at the group level that included each participant's FFA reactivation and FFA-hippocampal connectivity indices, respectively, as covariates. I hypothesized that greater reactivation and connectivity during rest would be associated with more reactivation of A_{face} stimuli during encoding. No region showed a significant relationship between the interaction term and FFA reactivation. I did, however, find activation in medial parietal and occipital cortex (centered on -5, -39, 22), including fusiform gyrus, for which the interaction term tracked positively with the degree of FFA-hippocampal connectivity following AB encoding (**Fig. 22B**).
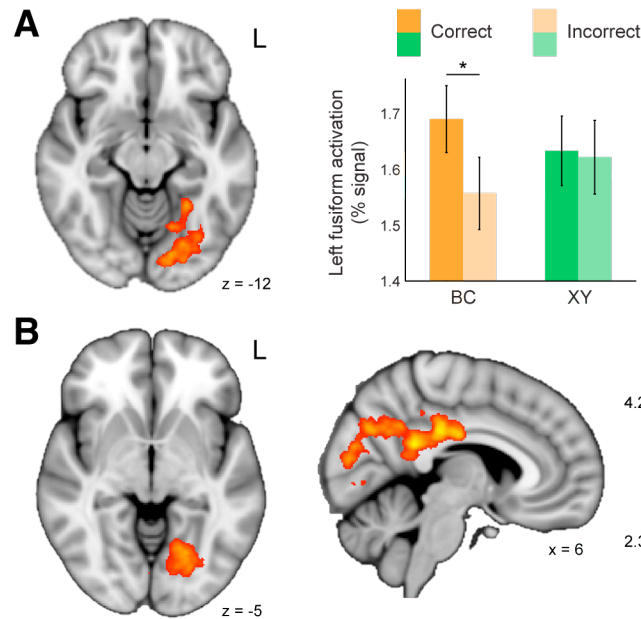
Figure 22. Experiment 2A: Univariate Results

Regions showing a significant subsequent recall by condition interaction during encoding. (**A**) Left, left fusiform gyrus was the only region to show greater subsequent recall effects for BC relative to XY pairs. Right, encoding activation in left fusiform gyrus relates to memory for overlapping BC but not XY pairs. Estimates were converted to percent signal and are shown separately for BC pairs subsequently remembered (dark orange); BC pairs subsequently forgotten (light orange); XY pairs subsequently remembered (dark green); and XY pairs subsequently forgotten (light green). The significant interaction in this region was driven by a strong subsequent memory effect for BC pairs (paired t-test; $t_{34} = 6.23$, $p = 4.4 \times 10^{-7}$; dark vs. light orange bars) but not XY pairs ($t_{34} = 0.49$, $p = 0.625$; dark vs. light green bars). Asterisk indicates significant difference at $p < 0.05$. (**B**) The interaction term was significantly modulated by the degree of FFA-hippocampal connectivity during the post-AB rest scan in portions of medial parietal and occipital cortex, including the fusiform gyrus and posterior cingulate cortex. Color bar indicates z-score. Coordinates are in mm.

Delay and Encoding Order Results

There was no relationship across the group between duration of the delay and the post-AB neural measures of interest (FFA-hippocampal connectivity: $r_{33} = 0.19$, $p = 0.27$), suggesting that the length of the delay did not significantly influence the degree to which these processes were engaged.

Encoding order did not significantly impact the relationships between connectivity (main effect and interaction; all $F_{1,31} < 1.02$, all $p > 0.32$) or reactivation (main effect and interaction; all $F_{1,31} < 0.32$, all $p > 0.57$) and BC or AC performance controlling for XY performance. Individual relationships between reactivation and performance (BC: both $F_{1,31} < 0.49$, $p > 0.49$; AC: both $F_{1,31} < 0.64$, $p > 0.43$; XY: both $F_{1,31} < 1.40$, $p > 0.24$); and FFA-hippocampal connectivity and performance (BC: both $F_{1,31} < 0.58$, $p > 0.45$; AC: both $F_{1,31} < 0.32$, $p > 0.58$; XY: both $F_{1,31} < 0.14$, $p > 0.72$) were also not modulated by encoding order.

5.5. DISCUSSION

I employed a combination of methods—multivariate pattern classification, functional connectivity approaches, and task-based univariate analyses—to provide empirical evidence that offline neural processes may mediate the relationship between prior knowledge and new encoding. These findings converge to suggest that rest-phase reactivation may benefit future learning by promoting subsequent integration during encoding. Interestingly, this benefit was observed even despite a long delay of approximately one hour between overlapping event encoding and test. Moreover, I suggest that the rest-phase reactivation observed in the present study occurred spontaneously. As participants were unaware of the overlap between the pre-training phase (AB learning) and the scanned portion of the experiment (BC and XY learning) at the time of post-AB rest, it is unlikely that my results reflect intentional rehearsal of AB associations during that period.

While the benefits conferred by offline processes on prior memories have been shown previously, the present work is the first to demonstrate how such benefits might also be prospectively advantageous. That is, rest-phase reactivation and connectivity serve to make our memories better suited for new learning in future situations, providing a foundational knowledge base upon which new experiences can be encoded. Importantly, because my classifier was trained on data from an independent localizer task consisting of a different stimulus set and task, these results suggest that reinstatement of episodic content (i.e., face information from learned AB face-object pairs)—rather than reinstatement of a learning-related state or context—supports the formation of memories for related information.

However, I note that like the vast majority of studies on this topic, these data do not provide evidence for processing of specific AB memories during the post-AB rest period. Thus, one possible alternative explanation of my findings is that participants reactivated task-irrelevant face information; and that doing so supported their later ability to encode BC content, as AB memories became *less* likely to interfere with new learning. However, I feel this possibility is unlikely given the present results. For example, I found that greater reactivation and connectivity were associated with superior AC inference performance, which requires retrieval of the A_{face} learned during pre-training. Moreover, both BC memory and AC inference were better for the AB pair memories acquired earlier in the pre-training phase, suggesting that strong AB knowledge promotes BC encoding. Thus, I suggest that the most likely interpretation of the data presented here is that AB memories were spontaneously reinstated during the post-AB rest period. Through this

139

process, they became stronger or more readily accessible (Jadhav et al., 2012; Deuker et al., 2013; Staresina et al., 2013), and therefore easier to reactivate and integrate during BC study.

I would also note that associative facilitation was not observed across the entire group of participants; rather I observed large individual differences in the degree to which participant showed facilitation versus interference. In fact, approximately half of my participants did not show evidence of facilitation at all, but rather showed interference (**Fig. 16C**). Thus, I would not conclude that proactive interference does not occur, but rather that there exists a tradeoff between interference and facilitation; and that this tradeoff may be mediated by offline processes. In other words, participants with a greater degree of post-AB reactivation and connectivity may show associative facilitation through an integrative encoding mechanism, while participants showing a lesser degree may tend to exhibit interference.

More broadly, these data suggest that one factor that may determine whether prior knowledge facilitates or interferes with the acquisition of new information is the strength of the initial memory, with strong prior knowledge being predominantly facilitative. This may occur not only during overt encoding or rehearsal, but also spontaneously during periods of passive rest. Notably, prior work suggests that low to moderate levels of memory reactivation may weaken traces, while high reactivation serves to strengthen memories (Detre et al., 2013). In the present study, lesser offline reactivation may thus be associated with weaker AB memories. Such weak traces may fail to be reactivated at all or may be weakly reactivated during learning, perhaps being forgotten when integration

140

fails (Anderson and McCulloch, 1999) or interfering with new encoding. In contrast, greater reactivation during rest strengthens AB memories, which can then later support BC learning through learning-phase retrieval and integration. Future work may characterize differences in item-level reactivation within participants to address how strengthening of individual memories impacts the balance between facilitation and interference.

One interesting aspect of the present data is that the observed relationships between reactivation and both BC learning and AC inference were specific to early in the post-AB rest scan. Importantly, this finding cannot be explained by differences in the amount of time between the encoding and rest scan across participants; delay duration did not predict the degree of face reactivation. Thus, interpreted in the context of the converging rest- and task-based neural measures provided here, I believe this reactivation measure serves as a valid index of neural processes that mediate the interactions between prior memories, new learning, and subsequent inference. While a mechanistic explanation for the temporal dependence of this signature is unclear, I believe the present measure of reactivation provides additional insight into how processing of prior memories during offline periods shape later learning experiences.

These results converge across multiple measures to demonstrate the important relationship between post-encoding reactivation, functional connectivity, and episodic memory, consistent with a host of findings from rodent (Jadhav et al., 2012) and human (Tambini et al., 2010; Deuker et al., 2013; Staresina et al., 2013) studies. In addition, the present data provide new evidence that the mnemonic advantage conferred by offline

processes extends beyond the initial memories themselves to influence the subsequent encoding of related content. These findings suggest a specific mechanism through which offline reactivation and hippocampal-neocortical connectivity lead to the strengthening of memory traces, thereby supporting later learning of related content via integrative encoding (Preston and Eichenbaum, 2013). Consistent with this interpretation, I found greater engagement of face-sensitive regions (i.e., fusiform gyrus) during encoding of object-object pairs that related to prior face knowledge as a function of FFA-hippocampal connectivity at rest. This extends prior work demonstrating the benefits of learning-phase reactivation for the reactivated memories themselves (Kuhl et al., 2010) and for linking experiences across time (Zeithamova et al., 2012a). I suggest that memory strengthening during rest facilitates retrieval of related content during subsequent learning experiences, thereby supporting new encoding by enabling linking of related memories (Schlichting and Preston, 2015).

6. Experiment 2B: MTL-MPFC Interactions During Learning and Post-Encoding Rest

6.1. RATIONALE

Recent theory has implicated hippocampal-MPFC interactions in integrating new content into strong prior knowledge, with MPFC guiding hippocampal encoding and retrieval (van Kesteren et al., 2012; Preston and Eichenbaum, 2013; Schlichting and Preston, 2015). I hypothesize that this circuit may be similarly engaged during offline periods to promote integration across episodes (Lewis and Durrant, 2011). While empirical research has broadly demonstrated involvement of hippocampus and MPFC in memory integration (Tse et al., 2007, 2011; van Kesteren et al., 2010a; Zeithamova et al., 2012a), existing reports are conflicting. While some have shown hippocampal-MPFC connectivity increases during events associated with strong prior knowledge (Tse et al., 2011; Zeithamova et al., 2012a), others have observed decreases (van Kesteren et al., 2010a), leaving open questions as to when and how this circuit supports encoding.

Here, I seek to fill this gap in understanding by providing a controlled test of theories regarding hippocampal-MPFC interactions during integration (van Kesteren et al., 2012; Preston and Eichenbaum, 2013). I propose that integration through hippocampal-MPFC interactions reduces interference among related memories, facilitating encoding of overlapping experiences. I manipulate the overlap between prior knowledge and new learning on a memory-by-memory basis, allowing for the examination of hippocampal-MPFC interactions mediating updating of individual events. Moreover, the hetereogeneity of MPFC has been underappreciated in prior research on

this topic; differences in structural connectivity across MPFC (Ongür and Price, 2000) suggest that its posterior aspects might be particularly important for integration (Nieuwenhuis and Takashima, 2011). I formally test this idea by interrogating neural signatures within subregions of MPFC.

I investigated hippocampal and MPFC contributions to integration both across and within individuals, comparing on- and offline neural engagement during periods reflecting memory updating versus simple associative encoding. Behavioral work has shown sleep-related increases in performance on novel judgments spanning experiences (Ellenbogen et al., 2007; Coutanche et al., 2013), consistent with the idea that sleep facilitates integration (Buckner, 2010). Mechanistically, hippocampus-driven memory reactivation during sleep may enable the formation of connections among co-activated representations, promoting integration in neocortical regions like MPFC (Lewis and Durrant, 2011). Inspired by such theories, I propose that similar reorganizational processes may occur during rest periods immediately following learning. Furthermore, as behaviors tapping the hippocampal-MPFC circuit should depend on the integrity of the underlying structural connections, I hypothesize that more coherent white matter tracts connecting these regions will be associated with superior integration.

Data from some portions of this experiment were described in a previous report (Schlichting and Preston, 2014) examining how processes prior to encoding influence new learning (see **Chapter 5**). Here, I focus on how the hippocampal-MPFC circuit is engaged during and immediately following new learning experiences. The results of this study are reported in Schlichting and Preston, under review.

6.2. HYPOTHESES

Comparing neural engagement during BC vs. XY encoding and post-encoding rest periods enabled me to investigate specific mechanisms engaged when new learning relates to prior experience (**Fig. 23**). That is, while BC and XY pairs consisted of the same stimulus type (object-object pairs), they differed in their overlap with prior knowledge. For this reason, neural signatures that differentiate BC from XY should reflect processes engaged in the context of overlapping events.

Hypothesis 1: *Learning-phase connectivity will be increased during overlapping relative to non-overlapping associations*. Greater hippocampal-MPFC connectivity will be observed during encoding of overlapping BC relative to non-overlapping XY pairs. Encoding-phase connectivity will relate to memory performance on a trial-by-trial basis more for BC than XY pairs.

Hypothesis 2: *Memory updating signatures will persist during the post-encoding period*. Hippocampal-MPFC functional coupling will be enhanced following encoding of overlapping BC pairs relative to encoding of non-overlapping XY pairs, reflecting updating of prior memories with new experience. These measures will relate to behavioral evidence of memory integration.

Hypothesis 3: *Hippocampal-MPFC structural connectivity will predict integration*. White matter integrity of structures connecting hippocampus to MPFC will relate to behavioral evidence of memory integration.
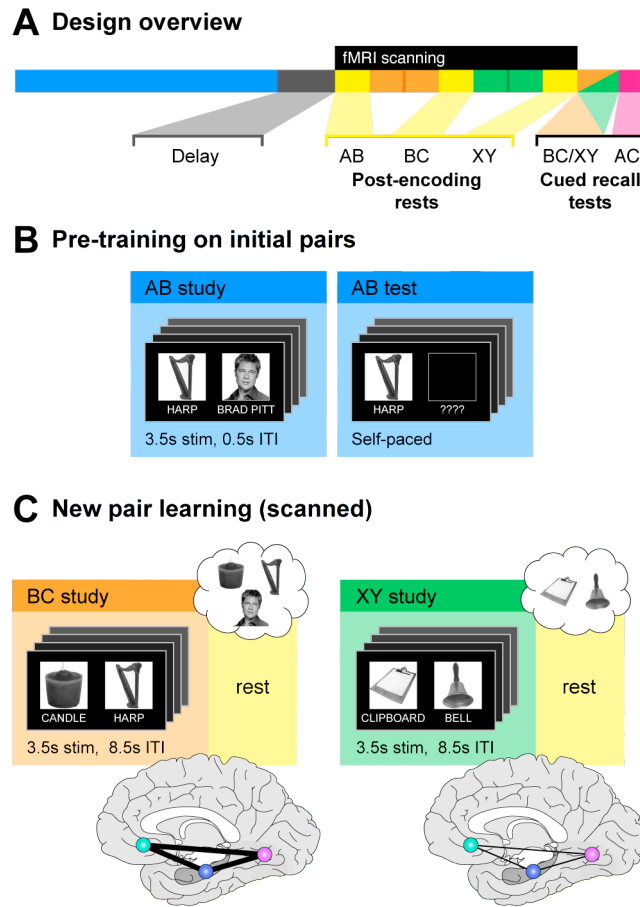
Figure 23. Experiment 2B: Design Overview and Predictions

Experimental design and theoretical predictions. **(A)** Schematic depiction of the experiment timeline. **(B)** Participants learned a series of AB face-object associations across four alternating study-test repetitions during the pre-training phase. Memory performance was near ceiling by the final test block, demonstrating strong memories for the AB pairs. **(C)** Following the pre-training phase, participants were transferred to the fMRI scanner for study of overlapping BC and non-overlapping XY object-object associations. BC and XY study blocks were each followed by a rest period (yellow). The order of BC and XY study was counterbalanced across participants. Left, overlapping BC associations (orange) included one object (here, HARP) that had previously been paired with a face during the pre-training phase (panel A, blue). Such overlap allows for prior AB memories to be reactivated and updated with the new BC information. I hypothesize that this process engages a network comprising hippocampus (indigo), medial prefrontal cortex (MPFC, teal), and content-sensitive visual regions (e.g., fusiform face area [FFA], magenta). Right, non-overlapping XY associations (green) included two new objects that had not been previously learned during the pre-training phase. Brain schematics represent the predicted functional connectivity enhancement among hippocampus, MPFC and FFA for overlapping BC (thick black lines) relative to XY (thin black lines) study and post-encoding rest periods.

146

6.3. METHODS

6.3.1. Participants

Participants were as reported previously (Schlichting and Preston, 2014; see **Chapter 5**). Data from 35 participants were included in all functional analyses (21 females; ages 20-30, 24.1 ± 0.5 years). An additional 10 participants were excluded from the diffusion tensor imaging (DTI) analysis due to data acquisition error. Thus, 25 participants were included in the structural connectivity analysis (15 females; ages 20-30, 24.2 ± 0.7 years).

6.3.2. Behavioral Methods

Materials and Procedures were identical to those described in **Chapter 5**.

6.3.3. Imaging Methods

MR Data Acquisition

Imaging data were acquired on a 3.0T GE Signa MRI system (GE Medical Systems) as described above (see **Chapter 5**). Diffusion-weighted data were also acquired to characterize white matter structure (TR = 12000 ms, TE = 87.1 ms, 25 diffusion directions, maximum b-value = 1000, 128 x 128 matrix, 0.94 x 0.94 in-plane resolution, 41 straight axial slices, 3-mm thickness, no gap).

fMRI Preprocessing

Functional data were preprocessed using FSL version 5.0.2 (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl). The first 4 volumes of all functional scans were discarded to allow for T1 stabilization. Motion correction was performed within each functional scan using MCFLIRT by aligning all images in the run to the middle volume

in the timeseries. Coregistration of functional data across runs was performed by calculating and applying the affine transformation from each run to a reference run using FLIRT, part of FSL. The 3D SPGR structural volume was registered to the functional reference run using the EPI registration utility (part of FLIRT) and then resampled to functional space. Brain extraction was performed on all structural and functional images using BET.

6.3.4. Data Analysis

Regions of Interest

Anatomical Region of Interest Definition

The hippocampus was delineated by hand on the 1 mm MNI template brain and reverse-normalized to each individual's functional space using Advanced Normalization Tools (ANTS) (Avants et al., 2011). Specifically, a non-linear transformation was calculated from the MNI template brain to each participant's 3D SPGR volume. This warp was then concatenated with the SPGR to functional space transformation calculated using FLIRT. After applying the transformation using ANTS, the anatomical hippocampus ROI was aligned to each participant's functional data.

MPFC ROIs were generated for each participant using output from Freesurfer (Desikan et al., 2006) run on each individual's SPGR. Because previous studies on similar topics (Kumaran et al., 2009; van Kesteren et al., 2010b, 2014; Sweegers et al., 2013) have identified a range of activation foci throughout the medial surface of PFC, I combined medial orbitofrontal and rostroanterior cingulate to create an MPFC ROI for

each participant. ROIs were then aligned to each individual's functional data using transformations applied in FLIRT, as described above.

Functional Region of Interest Definition

Procedures used to define functional FFA ROIs were identical to those described in **Chapter 5**.

Univariate Analysis

Analysis of fMRI data from the memory task was carried out using FEAT (fMRI Expert Analysis Tool) version 6.00, part of FSL. Grand-mean intensity normalization, high-pass temporal, filtering, spatial smoothing, and pre-whitening were carried out as described above.

I interrogated the relationship between encoding activation and subsequent memory using a general linear model (GLM). The GLM was designed to isolate effects specific to the encoding of new associations that overlapped with prior knowledge. Accordingly, encoding trials were sorted based on both subsequent memory and prior knowledge condition to create four conditions: BC encoding trials that were later correct, BC encoding trials that were later incorrect, XY encoding trials that were later correct, and XY encoding trials that were later incorrect. Contrasts of interest were (1) BC correct > XY correct and (2) BC correct – BC incorrect > XY correct – XY incorrect (i.e., the subsequent recall × prior knowledge interaction).

Stimulus presentations were modeled as events with 3.5 s durations, with one regressor for each condition of interest. The model was convolved with the canonical (double-gamma) HRF. Temporal derivatives were included. Motion parameters

calculated during the motion correction step and their temporal derivatives were added as additional regressors of no interest. As described above (see **Chapter 5**), FD, DVARS, and individual regressors for time points exceeding FD and DVARS thresholds were added to the models to additionally account for motion effects (Power et al., 2012; Schonberg et al., 2014). Temporal filtering was applied to the model.

After modeling functional data within each run, the resulting statistics images were warped to the MNI template brain resampled to functional resolution (3.75 x 3.75 x 3.6 mm) using ANTS (Avants et al., 2011). The warped images were combined across encoding runs for each participant using fixed effects, and then across the group using mixed effects. For both contrasts of interest, correction for multiple comparisons was carried out at both the whole-brain level and using small volume correction within my *a priori* hippocampus and MPFC ROIs. Correction for multiple comparisons at the whole brain level was carried out on group-level voxel-wise statistical images according to cluster-based Gaussian random field theory (Worsley et al., 2002), with a cluster-forming threshold of $z > 2.3$ and a whole-brain corrected cluster significance level of $p < 0.05$. I also interrogated activation within *a priori* hippocampus and MPFC ROIs for integration-specific effects using small volume correction implemented in 3dClustSim, part of AFNI (Cox, 1996). Group statistics images were first masked with each ROI. I then applied a primary voxelwise threshold of $p < 0.05$ to the masked images to identify those voxels within my ROIs surpassing this initial p-value threshold. I corrected for multiple comparisons within the ROIs by determining the cluster extent corresponding to a cluster-corrected threshold of $p < 0.05$ using Monte Carlo simulations. Cluster sizes that

150

occurred with a probability of less than 0.05 across 2000 simulations were considered statistically significant.

Task-Phase Functional Connectivity

Functional connectivity during task was examined using a psychological-physiological interactions (PPI; Friston et al., 1997; O'Reilly et al., 2012) approach carried out in FEAT version 6.00. Grand-mean intensity normalization, high-pass temporal, filtering, spatial smoothing, and pre-whitening were carried out as described previously.

The PPI analyses were performed to isolate regions whose functional connectivity with hippocampus and MPFC respectively were significantly modulated by the interaction of subsequent recall by prior knowledge condition. Encoding trials were sorted according to subsequent memory, and a correct – incorrect regressor was constructed for each run. Notably, because BC and XY encoding trials occurred in different scans, this regressor represented the correct – incorrect difference for either BC or XY study trials, depending on the run. This served as the psychological regressor. The first eigenvariate of the filtered timeseries (derived from the univariate analyses described previously) was extracted from anatomically defined hippocampus and MPFC and functionally defined FFA ROIs and entered as the physiological regressors for the three PPI analyses. The PPI regressor served as the regressor of interest and was generated as the interaction between the psychological and physiological regressors. An additional task regressor was added to account for variance associated with both correct and incorrect trials.

For all three PPI models, stimulus presentations were modeled as events with 3.5 s durations. Task-related regressors and their temporal derivatives were convolved with the canonical (double-gamma) HRF and filtered. Physiological and PPI regressors were not convolved with the HRF or filtered, as these regressors were derived from neural signal that had previously undergone temporal filtering. Motion parameters calculated during the motion correction step and their temporal derivatives were added as additional regressors of no interest. As described above, FD, DVARS, and individual regressors for time points exceeding FD and DVARS thresholds were added to the models to additionally account for motion effects (Power et al., 2012; Schonberg et al., 2014).

After modeling functional data within each run, the statistics images associated with the PPI regressor were warped to the MNI template brain resampled to functional resolution using ANTS (Avants et al., 2011). The warped PPI images were contrasted for BC and XY study runs for each participant using fixed effects, yielding a subsequent recall × prior knowledge interaction contrast (i.e., BC correct – BC incorrect > XY correct – XY incorrect). These statistics images were then combined across the group using mixed effects. Correction for multiple comparisons was performed within *a priori* hippocampus and MPFC ROIs using small volume correction implemented in 3dClustSim as described above.

As the anatomical MPFC ROI spans a large and likely functionally diverse region (Ongür and Price, 2000; Price and Drevets, 2009; Roy et al., 2012), I also performed the PPI analysis described above using the MPFC clusters defined in the univariate

interaction contrast (depicted in **Fig. 24B**) as the seed. This analysis was otherwise identical to above.

Rest-Phase Functional Connectivity

Functional connectivity during rest was interrogated using a voxelwise regression approach with anatomically defined hippocampus and MPFC and functionally defined FFA as seed regions. This analysis approach is similar to that I have employed previously (Schlichting and Preston, 2014; **Chapter 5**). I interrogated neural activation for clusters of voxels (1) that showed enhanced rest-phase connectivity following overlapping (post-BC) relative to non-overlapping (post-XY) encoding conditions and (2) for which the degree of enhancement was related to subsequent performance. Importantly, I would expect the post-XY rest period to reflect neural signatures related to simple associative encoding (i.e., of object-object pairs), Thus, using this period as a baseline provides a stringent control for isolating those processes related to memory integration above and beyond those engaged during learning more generally.

I first regressed out potential sources of noise from the resting state data. Specifically, I extracted the first eigenvariate of the signal across the duration of the each rest scan in anatomically defined white matter and lateral ventricle ROIs. The signal from these two ROIs and their temporal derivatives were used to construct a GLM along with: motion parameters, FD, DVARS, and their temporal derivatives; and timepoints with excessive motion (as described previously). Rest data were high-pass filtered with a cutoff of 0.009 Hz (Fox et al., 2005; Tambini et al., 2010) and regressed on these noise sources. The resulting residual timeseries data for each participant were spatially

smoothed (5 mm FWHM). As motion-related nuisance signals had already been removed, each model included only the seed ROI timeseries and its temporal derivative. The resulting parameter estimate image for each participant reflected the degree to which activation in each voxel tracked with activation in the seed ROI across each rest scan. I then warped these images to the MNI template resampled to functional dimensions using ANTS (Avants et al., 2011). Images were contrasted within participant (i.e., each participant's post-XY connectivity statistics image was subtracted from their post-BC connectivity image) and combined them across participants using a group-level GLM as follows.

I constructed the group-level model with both BC and XY performance as covariates. To identify those voxels showing enhanced connectivity with the seed region during post-BC relative to post-XY encoding (irrespective of memory performance), I combined the difference images across participants using a one-sample t-test. I also isolated voxels for which the degree of connectivity enhancements tracked more with BC than XY performance (i.e., contrast of covariates, BC performance > XY performance; see Schlichting and Preston, 2014). Voxelwise statistics were calculated using permutation tests implemented in FSL. Correction for multiple comparisons was then performed within *a priori* hippocampus and MPFC ROIs using small volume correction implemented in 3dClustSim (Cox, 1996) as described above. I also assessed the relationship between connectivity and AC inference performance within the performance-related regions identified above. Contrast values quantifying the degree of connectivity enhancement with the seed region following overlapping BC encoding were

extracted and related to AC performance using partial correlation (controlling for XY performance).

As the above analysis searches for regions demonstrating a somewhat complex association with behavior (the difference in post-BC and post-XY connectivity measures relating more to BC than XY performance), it is difficult to know the precise underlying pattern that gives rise to this result. To further describe the results of the above analysis, I extracted post-BC and post-XY connectivity measures separately from significant regions identified above (i.e., showing either an overall connectivity enhancement or an enhancement-performance relationship). These measures were compared across two groups of individuals demonstrating different behavioral signatures: those showing behavior consistent with a facilitative effect of prior knowledge on new learning (BC performance > XY performance; N = 15), and those evidencing proactive interference (BC performance < XY performance; N = 16). This analysis excluded individuals who had identical BC and XY performance (N = 4). Effects were quantified using 2×2 mixed ANOVAs with rest scan (post-BC or post-XY) as the within-subjects factor and behavioral signature (facilitation or interference) as the between subjects factor t-tests. The reader should note that these regions were identified precisely because they showed a significant post-BC enhancement or relationship between enhancement and behavior. Thus, it is not the relationship itself but the nature of that relationship among individuals showing these two behavioral signatures that I sought to isolate.

Due to the large and potentially heterogeneous nature of the anatomical MPFC ROI (Ongür and Price, 2000; Price and Drevets, 2009; Roy et al., 2012), I repeated the

main rest-based connectivity analysis described above using the MPFC clusters defined in the interaction contrast (depicted in **Fig. 24B**) as seed regions.

Additionally, to alleviate the possibility that rest-phase connectivity might be attributed to lingering effects of the task itself, I repeated the analysis omitting the first 30 TRs (60 seconds) of the rest scan. This analysis lengthened the delay between study and rest connectivity analyses to roughly 3-4 minutes (in addition to the approximate 2-3 minute delay between scans for image reconstruction and communication with the participant).

Control Analyses

As encoding order was counterbalanced across participants, different individuals experienced differences in time (Schlichting and Preston, 2014; **Chapter 5**) and mnemonic demand between initial AB pair pre-training and overlapping BC study. While one group encoded overlapping BC pairs following AB learning, the other encoded non-overlapping XY pairs in between learning AB and BC. One might predict that these differences in encoding order may impact my neural measures of functional activation or interregional connectivity during study and/or post-encoding rest. For instance, one possibility is that the present findings are driven primarily by the group that encoded BC before XY, as the degree to which AB knowledge is brought to bear during BC learning might decrease as AB experience becomes more remote (i.e., temporal proximity of the related experiences may influence these neural mechanisms). Accordingly, I performed additional control analyses to assess the effects of encoding order on my neural measures of interest.

Effects of Encoding Order at Task

I interrogated the encoding phase to determine whether the degree of task-related functional activation or connectivity differed significantly as a function of encoding order. I first focused on clusters showing a significant subsequent recall by prior knowledge condition interaction in univariate activation across the group. I performed a 2 × 2 × 2 mixed ANOVA with subsequent recall and trial type as within-subjects factors and encoding order as the between-subjects factor. Univariate activation served as the dependent measure. I carried out this analysis for all three regions identified in the main analysis as showing a significant interaction effect (two MPFC clusters and one hippocampal cluster; see **Fig. 24B**). I corrected for multiple comparisons in the three ROIs using Bonferroni correction, which yielded a critical p-value of 0.017. Encoding order did not significantly affect the observed interaction between subsequent recall and prior knowledge condition in any region (main effects and 3-way interactions; all $F_{1,33}$ < 2.49, all p > 0.125).

I also determined whether task-phase functional connectivity assessed using PPI differed significantly across counterbalancing groups. I performed a 2 × 2 mixed ANOVA with trial type as the within-subjects factor and encoding order as the between-subjects factor. I note that subsequently correct versus incorrect trials were already contrasted within run; thus, the dependent measures in this analysis reflected the connectivity difference for subsequently correct relative to incorrect trials, separately for BC and XY conditions. This analysis was performed for all six regions identified previously in the main PPI analyses for FFA, hippocampus, and MPFC seeds (including

157

the 18-voxel MPFC cluster from the analysis seeded with FFA, for a total of four hippocampal clusters and two MPFC clusters; see **Fig. 25**). Bonferroni correction across the six ROIs yielded a critical p-value of 0.008. I observed no significant effect of encoding order on connectivity measures in any of these regions (main effects and interactions; all $F_{1,33} < 3.27$, all $p > 0.08$).

Effects of Encoding Order at Rest

Next, I investigated the relationship between functional connectivity during rest and encoding order. I first performed a one-way ANOVA to assess whether the degree of connectivity enhancement following BC learning differed as a function of encoding order. Neural measures (degree of connectivity enhancement for post-BC vs. post-XY rest) were grouped by encoding order. There was no significant effect of encoding order on the degree of connectivity enhancement ($F_{1,33} = 1.39$, $p = 0.248$).

I performed one-way analyses of covariance (ANCOVA) to determine whether the observed relationships between functional connectivity during rest and performance were related to encoding order. The neural measure served as the independent measure; behavior was the dependent measure. Encoding order was the grouping variable. This analysis was performed for all eight regions identified in the main analyses as relating to performance when FFA, hippocampus, or MPFC served as the seed region (including the 18-voxel MPFC cluster from the analysis seeded with FFA, for a total of five hippocampal clusters and three MPFC clusters; see **Figs. 26** and **28**). I corrected for multiple comparisons in the eight ROIs using Bonferroni correction, yielding a critical p-value of 0.006. There was no significant effect of encoding order on the observed

relationship between connectivity and BC performance controlling for XY performance at my corrected threshold for any region (main effects and interactions; all $F_{1,33} < 5.64$, all $p > 0.024$). I note that the one region in which the interaction effect would be considered significant at a more lenient statistical threshold ($F_{1,33} = 5.64$, $p = 0.024$) was the cluster in left anterior hippocampus when FFA served as the seed. Interestingly, this interaction was driven by a stronger relationship between connectivity and performance among participants who encoded XY before BC. This is the opposite of what would be predicted by a temporal proximity account, suggesting that the present findings are unlikely to reflect merely lingering effects of AB encoding.

Structural Connectivity

DTI data analysis was carried out using tools from FDT (FMRIB's Diffusion Toolbox; Behrens et al., 2007) version 3.0, and Tract-Based Spatial Statistics (TBSS; Smith et al., 2006), both part of FSL. I was interested in characterizing how white matter structure in pathways connecting hippocampus and MPFC related to memory for critical BC associations relative to control XY associations. I hypothesized that as the hippocampal-MPFC circuit supports the updating of existing memories with new information, greater white matter integrity in these tracts should relate to superior overlapping BC pair memory, after accounting for overall differences in memory ability (i.e., performance on non-overlapping XY pairs).

To test this hypothesis, I first isolated tracts connecting anatomical hippocampus and MPFC ROIs as follows. Registration was carried out within FDT using FLIRT and FNIRT for structural and standard space transformations, respectively. DTI data were

first corrected for eddy currents, and then a diffusion tensor model was fit at each voxel of the DTI data for each participant. I generated probability distributions of diffusion parameters at each voxel in the brain for each participant using Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques (BEDPOSTX). I then performed tractography between hippocampus and MPFC ROIs using PROBTRACKX, which uses the voxelwise probability distributions to generate a distribution of pathways connecting specified regions. Probabilistic tractography was carried out for each individual in native DTI space. Each individual's anatomical ROIs were registered to their DTI data using FLIRT transformations computed previously. Tractography was run twice for each participant (number of samples = 5000, curvature threshold = 0.2, maximum number of steps = 2000, step length = 0.5 mm): once from hippocampus to MPFC (waypoint or inclusion mask); and once from MPFC to hippocampus. In using this approach, I identified tracts that pass through both ROIs for each individual.

Each individual's hippocampal-MPFC and MPFC-hippocampal connectivity distribution maps were thresholded to exclude voxels through which fewer than 5% (250) of all sampled pathways passed. They were then added together to create a single bidirectional white matter ROI for each participant. These white matter ROIs were warped to MNI space using FNIRT transformations computed previously in FDT. ROIs were summed across participants to create a single, group ROI in standard space encompassing all tracts connecting hippocampus and MPFC. The overlap of white matter ROIs across participants is depicted in **Fig. 29A**.

160

Fractional anisotropy (FA) was used as the metric of white matter integrity within this ROI. FA quantifies the degree to which water molecules diffuse in a directional (i.e., anisotropic) manner, and tends to be high when white matter fibers are oriented similarly (i.e., when coherence is high). All subjects' FA data were aligned to common space using FNIRT carried out in TBSS. A mean FA image was then generated and thinned to create a mean FA skeleton, which represents the centers of all tracts common to the group. Each subject's aligned FA data was then skeletonized, i.e., projected onto this group skeleton. Mean FA values were extracted from each individual's skeletonized data across all voxels within the group white matter ROI. FA was related to BC memory and AC inference performance across participants controlling for XY memory using partial correlation (for a similar approach see Schlichting and Preston, 2014 and **Chapter 5**).

One possibility is that individual differences in overall white matter structure could explain performance differences across participants. Thus, to assess the specificity of this relationship to hippocampal-MPFC pathways, I also identified a control white matter ROI, the corticospinal tract, using the Johns Hopkins University (JHU) white-matter tractography atlas (Hua et al., 2008). FA values were then derived from the skeletonized FA data and related to performance measures using partial correlation as described above.

Multiple Regression

Multiple linear regression analyses were performed to further assess the degree to which rest-phase functional and structural connectivity measures were independently related to performance. Mean FA values extracted from the hippocampal-MPFC tract of

interest served as the measure of structural connectivity. For functional connectivity, I averaged across the connectivity measures from all four clusters (two hippocampal, two MPFC; depicted in **Figs. 26A**, right and **26B**) identified as significantly relating to performance in the rest-phase analyses seeded with hippocampus or MPFC. Two regression models were run: one with BC performance as the dependent variable, and one with AC performance as the dependent variable. Structural connectivity, functional connectivity, and XY performance measures for each participant were entered into the regression as predictors. Participants were treated as a random effect.

6.4. RESULTS

6.4.1. Behavioral Results

Behavioral performance was as described previously (Schlichting and Preston, 2014; **Chapter 5**). Briefly, as intended, AB pairs were well learned by the fourth test block (mean ± SEM: 97.3% ± 0.9% correct recall). Importantly, as BC (11.7%-86.7%, 41.5% ± 3.3% correct) and XY (10%-78.3%, 42.4% ± 3.4% correct) were matched on all dimensions except prior knowledge, I was able to directly compare neural engagement and performance in these two conditions (**Fig. 23C**). Interestingly, I observed no difference between BC and XY performance across the group (t_{34} = 0.40, p = 0.693); rather, relative performance on these two conditions was highly variable across individuals (see Schlichting and Preston, 2014 and **Chapter 5** for a more in-depth discussion of the factors modulating the interfering versus facilitative effects of prior knowledge on new learning). This variability enabled me to investigate how neural processes engaged during task and post-encoding rest periods related to performance for

overlapping BC relative to control non-overlapping XY pairs. I also found that performance on AC inferences (6.7%-83.3%; 41% ± 3.5% correct) paralleled BC memory (Schlichting and Preston, 2014), demonstrating the strong nature of the AB memories at the end of the experiment.

6.4.2. Imaging Results

Univariate Results

The above framework suggests that hippocampus and MPFC will be recruited during learning, specifically during episodes in which interference is reduced through memory updating. I predicted greater activation in these regions during encoding of overlapping BC relative to non-overlapping XY associations. Importantly, I focused on only those associations that were subsequently remembered in both prior knowledge conditions. I searched for this BC correct > XY correct effect both at the whole brain level and within hippocampus and MPFC regions of interest (ROIs). Two clusters survived correction for multiple comparisons across the whole brain: one in posterior cingulate (-3.75, -39.75, 21.6) and one extending from precuneus on the medial surface to superior parietal lobule (-7.5, -66, 39.6) (**Fig. 24A**). Small volume correction revealed significant clusters in both left hippocampus (-33.75, -13.4, -21.6) and in MPFC (-3.75, 35.25, -21.6) (**Fig. 24A**).
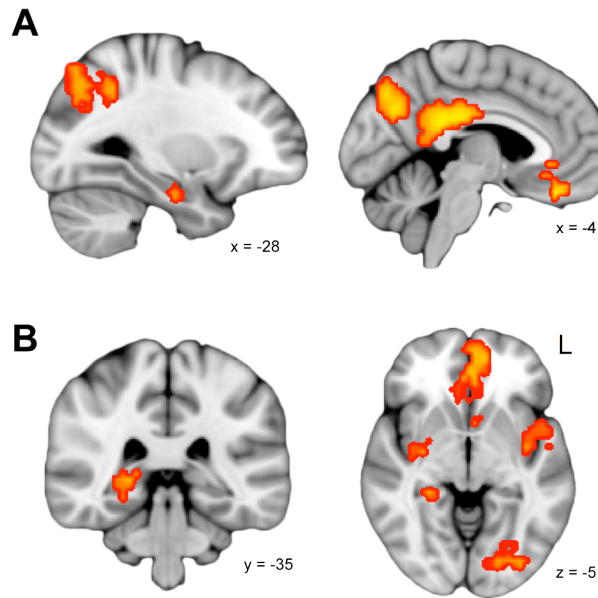
Figure 24. Experiment 2B: Univariate Results

Activation during overlapping encoding predicts memory updating. Maps include clusters significant at the whole-brain level as well as those that survived correction within *a priori* hippocampus and MPFC ROIs. **(A)** Regions in parietal cortex and posterior cingulate survived correction for multiple comparisons across the whole brain. Clusters in both hippocampus (left) and MPFC (right) showed significantly greater activation during encoding of BC relative to XY pairs (subsequently correct trials only). Clusters are significant after small volume correction within hippocampus and MPFC, respectively. For all figures, activations were resampled to the 1mm MNI template for visualization. **(B)** Clusters in left fusiform and bilateral insula were significant at the whole brain level. Regions in hippocampus (left) and MPFC (right) showed a significantly greater subsequent recall effect for BC relative to XY pairs during encoding. Clusters are significant after small volume correction within hippocampus and MPFC, respectively. Coordinates are in mm.

I next assessed the link between encoding activation and subsequent behavior. My key prediction is that hippocampal-MPFC engagement and connectivity will be enhanced when new information is successfully integrated into memory. Thus, engagement should be greater for those overlapping BC associations that are later remembered relative to those that are forgotten. Moreover, because BC study events provide the unique opportunity for memory updating via integration, I predicted hippocampus and MPFC

164

would show a larger subsequent memory effect for BC than XY associations (i.e., a significant interaction between prior knowledge condition and subsequent memory). At the whole brain level, significant clusters were observed in right (41.25, -2.25, 28.75) and left insula (-37.5, -6, 0) as well as left fusiform gyrus (-15, -81, -10.75) (**Fig. 24B**). I note that due to differences in the spatial normalization approach used between the current study and the previous experiment (Schlichting and Preston, 2014; **Chapter 5**), the insula regions were not identified in the prior report of this contrast. Using small volume correction, I found one additional significant cluster in right hippocampus (22.5, -36, 0) and two in MPFC (anterior: -11.25, 50.25, 0; posterior: -22.5, 1.5, -14.4) that showed the predicted interaction (**Fig. 24B**). Control analyses verified that the observed interaction between subsequent recall and prior knowledge condition did not differ by order of BC versus XY encoding in any region. As my key hypotheses relate to integration-specific neural processes and behavior, all subsequent analyses focus on the prior knowledge condition by subsequent memory interaction.

Task-Phase Functional Connectivity Results

While the activation results described previously suggest that both hippocampus and MPFC are engaged during integration, I also interrogated whether their degree of functional coupling would predict subsequent memory for overlapping but not non-overlapping pairs. To examine how functional coupling during new encoding related to the presence or absence of existing knowledge, I performed two psychological-physiological interaction (PPI) analyses with anatomically defined hippocampus and MPFC, respectively, as seed regions. Using this approach, I was able to test my

hypothesis that hippocampal-MPFC connectivity would be modulated on a trial-by-trial basis as a function of subsequent memory and prior knowledge condition. I predicted greater functional coupling during BC study trials that were later correct relative to those that were incorrect. Moreover, as this effect should be specific to the overlapping study trials, I looked for hippocampal and MPFC regions that demonstrated this connectivity difference more for BC relative to XY trials.

Consistent with my predictions, I found evidence for recruitment of the hippocampal-MPFC circuit during successful encoding of overlapping information. I observed significant activation within MPFC (0, 24, -21.6) when hippocampus served as the seed (**Fig. 25A**); and conversely, two significant clusters in hippocampus (left: -30, -32.25, -7.2; right: 37.5, -24.75, -7.2) when MPFC was the seed (**Fig. 25B**). These effects did not differ significantly by encoding order. Because my anatomical MPFC ROI spanned a large region, I also performed this analysis using the regions identified in the interaction contrast as seeds (i.e., MPFC clusters in **Fig. 24B**). This analysis yielded similar results (two hippocampal clusters; left: -33.75, -32.25, -7.2; right: 33.75, -17.25, -14.4).

Connectivity with FFA

A number of existing studies have shown that prior memories may be reactivated during encoding of overlapping experiences (Hupbach et al., 2007; Kuhl et al., 2010; Jones et al., 2012; Zeithamova et al., 2012a; Schlichting and Preston, 2014). Such reactivation has been linked to better memory for the reactivated content itself (Kuhl et al., 2010), superior learning of the overlapping content (Schlichting and Preston, 2014),

166

and an enhanced ability to make novel judgments that span the two events (Zeithamova et al., 2012a). Accordingly, I hypothesized that greater functional coupling of hippocampus and MPFC with face-sensitive visual regions should also be associated with better subsequent memory specifically for the overlapping associations due to the related A stimuli being faces.
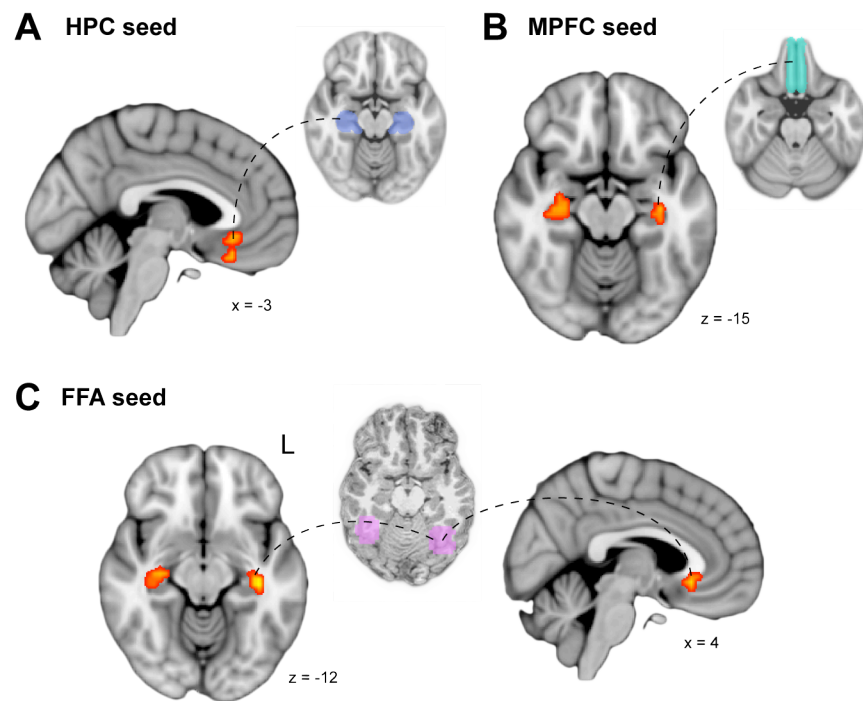
Figure 25. Experiment 2B: Task-Phase Functional Connectivity Results

Functional connectivity during overlapping encoding predicts memory updating. Regions for which functional connectivity with MPFC (**A**), hippocampus (**B**), and FFA (**C**) showed a significantly greater subsequent recall effect for BC relative to XY pairs during encoding. For (**A**) and (**B**), clusters are significant after small volume correction within hippocampus and MPFC, respectively. For (**C**), while the clusters in hippocampus (left) did reach my small volume corrected cluster size threshold, the cluster in MPFC (right) was one voxel shy of statistical significance (18 voxels). Coordinates are in mm. HPC, hippocampus.

To test this prediction, I performed a PPI analysis with each individual's functionally defined bilateral FFA as the seed region. I found significant clusters in bilateral hippocampus (left: -37.5, -21, -14.4; right: 33.75, -17.25, -14.4; **Fig. 25C**, left). I also observed a cluster in MPFC that was one voxel shy of the critical cluster size threshold (cluster size = 18 voxels; 7.5, 31.5, -3.6) (**Fig. 25C**, right). Results did not differ significantly by encoding order.

Rest-Phase Functional Connectivity Results

Existing work suggests that hippocampus and MPFC interact during sleep to integrate and generalize across discrete experiences (Lewis and Durrant, 2011). Extending these ideas to awake rest periods, I hypothesized that greater hippocampal-MPFC connectivity following overlapping encoding would be associated with more memory updating. I interrogated neural engagement following encoding of the overlapping compared with the non-overlapping associations. I hypothesized that I would observe (1) integration-related enhancements in hippocampal-MPFC connectivity during post-BC encoding rest relative to post-XY and (2) that the degree of enhancement would relate to behavioral evidence for memory integration across participants.

I further interrogated the clusters identified as showing the predicted patterns to determine how connectivity during post-BC and post-XY rest scans related to different behavioral signatures. I did this by dividing my group into individuals showing behavior consistent with a facilitative (BC performance > XY performance) versus interfering (BC performance < XY performance) effect of prior AB knowledge on BC learning, with XY pairs serving as the baseline.

Using a regression approach, I first searched for voxels in hippocampus or MPFC whose connectivity with the seed region (MPFC or hippocampus, respectively) was greater following BC than XY encoding. There were no significant clusters in hippocampus when MPFC served as the seed. However, when seeding with hippocampus, I observed a significant region of MPFC that showed the predicted integration-related enhancement (3.75, 16.5, -10.8; **Fig. 26A**, left), consistent with the notion that overlapping events trigger a memory updating mechanism. This effect was driven by enhanced connectivity for BC versus XY in the group demonstrating behavioral facilitation (**Fig. 27A**, top).

I also hypothesized that the degree to which participants showed enhanced connectivity following encoding of overlapping relative to non-overlapping associations would correlate with their behavioral evidence for memory integration. That is, if the observed functional connectivity truly reflects integration-specific processes, one would expect that these connectivity enhancements would support performance only for overlapping content. Thus, I identified MPFC and hippocampal voxels whose connectivity was more predictive of BC than XY performance. I found two significant clusters in MPFC (anterior: 3.75, 50.25, -14.4; posterior: -18.75, 5.25, -10.8) when hippocampus served as the seed (**Fig. 26A**, right). Posterior MPFC showed a pattern consistent with memory updating only in the facilitation group, with greater connectivity in that group after BC than XY encoding (**Fig. 27A**, bottom left). Anterior MPFC demonstrated a similar connectivity enhancement following BC learning for the facilitation group, but showed the opposite pattern for the interference group; that is,

greater connectivity following XY than BC encoding (**Fig. 27A**, bottom right). I also identified two clusters in hippocampus (left: -26.25, -28.5, -7.2; right: 26.25, -28.5, -7.2) when MPFC served as the seed (**Fig. 26B**). These regions both showed patterns consistent with integration for the facilitation group (**Fig. 27B**), albeit weakly in right hippocampus (**Fig. 27B**, right).
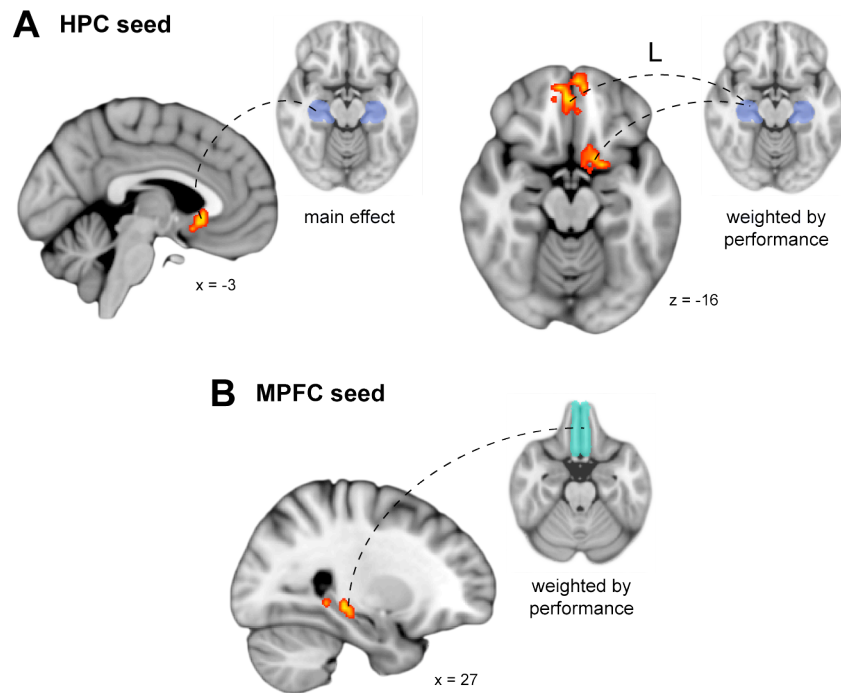
Figure 26. Experiment 2B: Rest-Phase Hippocampal-MPFC Functional Connectivity.

Hippocampal-MPFC functional connectivity enhancements during post-encoding rest tracked integration performance. (**A**) Left, cluster in MPFC showing greater connectivity with hippocampus following BC than XY encoding. Right, two clusters in MPFC for which the degree of functional connectivity enhancement with hippocampus during post-BC encoding rest period tracked more with BC than XY performance across participants. (**B**) Hippocampal clusters for which enhancement tracked more with BC than XY performance. For both panels, clusters are significant after small volume correction within MPFC and hippocampus, respectively. HPC, hippocampus.
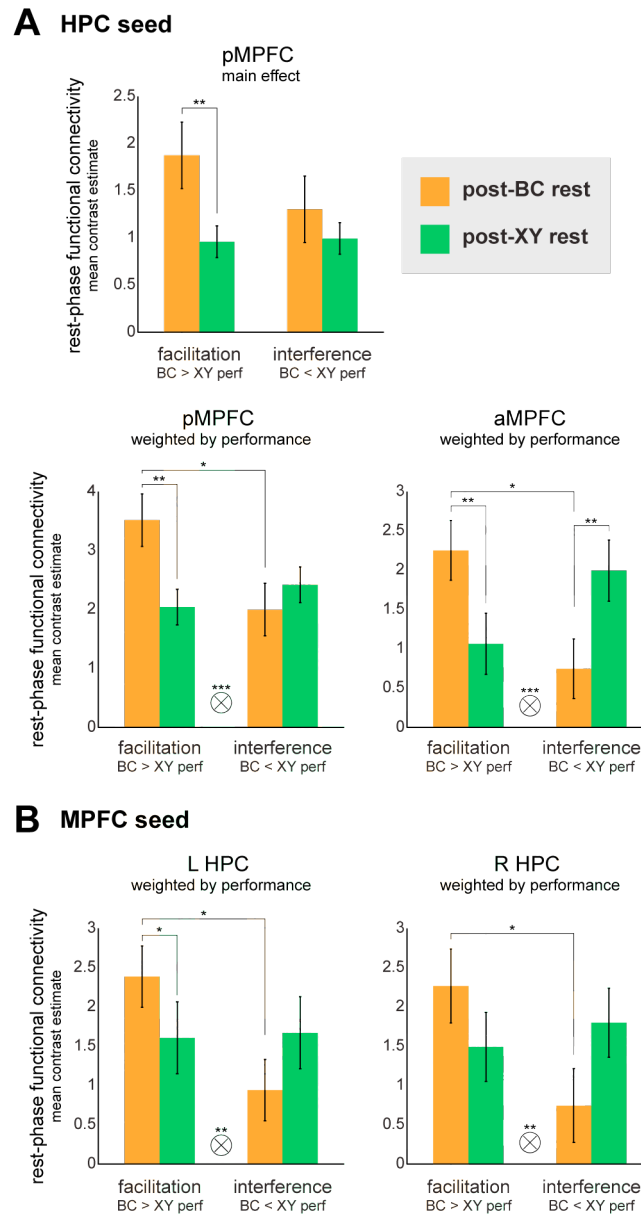
Figure 27. Experiment 2B: Rest-Phase Functional Connectivity by Behavioral Signature.

Rest-phase functional connectivity as a function of rest scan and behavioral signature in clusters demonstrating significant enhancement effects (see **Fig. 26**). Regions were defined using (**A**) hippocampus and (**B**) MPFC seeds. Connectivity values from the post-BC (orange) and post-XY (green) rest scans were extracted from these clusters and averaged across participants showing facilitation (left bar pairs) and those showing interference (right bar pairs). pMPFC, posterior MPFC; aMPFC, anterior MPFC. Tensor product symbol denotes rest scan by behavioral signature interaction. Asterisks indicate significant level, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. HPC, hippocampus.

Control analyses revealed that these effects did not differ significantly as a function of encoding order. To limit the possible effects of the immediately preceding study on rest-phase connectivity, I also performed the above analyses excluding the first minute of the rest scan. This analysis yielded similar results (hippocampal seed, main effect MPFC cluster: 7.5, 27.75, -14.4; performance-related MPFC cluster 11.25, 39, -10.8; MPFC seed, performance-related right hippocampal cluster: 26.25, -32.25, 0), suggesting that these findings are unlikely to be the result of continuing study-phase engagement.

I also assessed the across-participant relationship between connectivity enhancements in these regions and AC inference performance. As expected, all regions exhibited significant positive associations with inference performance (partial correlations controlling for XY performance; all $r_{32} > 0.49$, all $p < 0.003$). Moreover, seeding with MPFC regions defined from the univariate interaction contrast instead of the large anatomical MPFC ROI yielded similar clusters in hippocampus (left: -15, -39.75, 0; right: 33.75, -32.25, -7.2). These results are consistent with the prediction that enhanced hippocampal-MPFC functional coupling following overlapping encoding supports new learning via the integration of that newly learned content into prior related memories.

Connectivity with FFA

A number of studies in both rodents and humans have provided evidence that mnemonic content is processed in the brain during offline periods (e.g., rest and sleep) (Tambini et al., 2010; Jadhav et al., 2012; Deuker et al., 2013; Staresina et al., 2013). For instance, content-specific increases in hippocampal-neocortical connectivity have been

172

demonstrated following associative encoding (Tambini et al., 2010). Such connectivity enhancements were also related to performance, suggesting that offline processing strengthens memory for recent experiences. I reasoned that immediately following overlapping encoding, hippocampus and MPFC might show greater connectivity with neocortical regions sensitive to the overlapping content (here, A_{faces}), reflecting persistence of the memory updating process. That is, although participants most recently encoded object-object associations in both BC and XY study phases, I predicted that enhanced functional coupling with face-sensitive regions (FFA) during post-BC relative to post-XY would be associated with superior performance.

Accordingly, I performed a voxelwise regression analysis with FFA serving as the seed region. I found three significant clusters within hippocampus for which the degree of connectivity enhancement during the post-BC encoding rest period related more to BC than XY performance (**Fig. 28**, left): one in right (30, -21, -14.4) and two in left (anterior: -11.25, -9.75, -21.6; posterior: -15, -36, 0) hippocampus. Moreover, there was an 18-voxel cluster in MPFC that was one voxel short of statistical significance at my corrected threshold (7.5, 1.5, -10.8; **Fig. 28**, right). Connectivity within these hippocampal and MPFC clusters were also significantly related to AC inference performance (partial correlations controlling for XY performance; all $r_{32} > 0.38$, all $p < 0.026$). These findings did not differ by encoding order. Similar results were obtained after omitting the first minute of the rest scan (right hippocampus: 18.75, -24.75, -14.4; left hippocampus: -22.5, -39.75, 3.6; MPFC: 3.75, 5.25, -10.8).

FFA seed

L

y = -23   z = -21
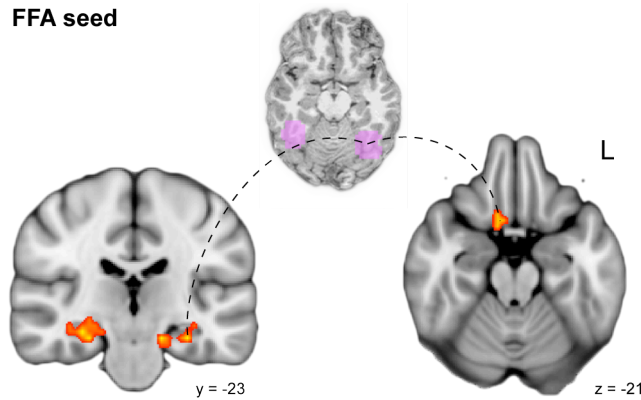
Figure 28. Experiment 2B: Rest-Phase Functional Connectivity with FFA.

Functional connectivity enhancements with FFA during post-BC encoding rest period tracked more with BC than XY performance across participants in hippocampus (left) and MPFC (right). While the clusters in hippocampus (left) did reach the small volume corrected cluster size threshold, the cluster in MPFC (right) was one voxel shy of statistical significance (18 voxels). Coordinates are in mm.

Structural Connectivity Results

The preceding findings indicate that hippocampal-MPFC activation and functional coupling predicts subsequent memory updating, both during learning on a trial-by-trial basis and during offline periods across participants. I hypothesized that the structural integrity of the white matter tracts connecting these regions would also predict individual differences in the ability to learn overlapping relative to non-overlapping associations. Here, I define white matter integrity as fractional anisotropy (FA), a commonly used measure quantifying the degree to which water molecules diffuse in a directional manner. High FA suggests high white matter integrity or tract coherence. There was no relationship between FA in the control (corticospinal) tract and behavior (i.e., partial correlation with BC and AC performance controlling for XY; BC: $r_{22} = 0.29$, $p = 0.169$; AC: $r_{22} = 0.28$, $p = 0.178$). However, I did find a positive relationship between

174

FA in tracts connecting hippocampus and MPFC (**Fig. 29A**) and both BC pair memory

and AC inference, after statistically controlling for individual differences in general

memory ability (i.e., partial correlation with BC and AC performance controlling for XY;

BC: $r_{22} = 0.42$, p = 0.039, **Fig. 29B**; AC: $r_{22} = 0.45$, p = 0.028).
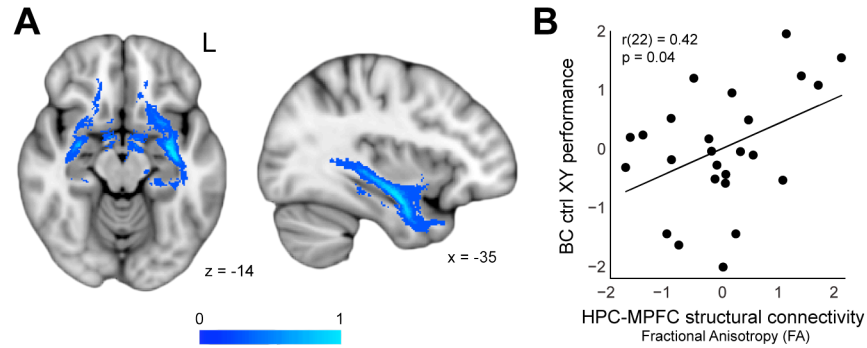
Figure 29. Experiment 2B: Structural Connectivity Results

Hippocampal-MPFC structural connectivity predicts memory updating across participants. **(A)** Group map depicting overlap across participants of tracts connecting hippocampus and MPFC. Tracts were determined for each individual using probabilistic tractography, warped to the MNI template and combined across participants. Colorbar indicates the proportion of participants with tracts in each voxel. **(B)** Mean fractional anisotropy (FA) values were extracted for each participant within the group ROI depicted in panel A. Scatterplot shows across-participant relationship between FA within hippocampal-MPFC white matter tracts and BC performance, plotted as residuals after regressing both on XY performance. HPC, hippocampus.

Multiple Regression Results

To investigate the degree to which hippocampal-MPFC structural and functional

connectivity measures independently explained variance in subsequent learning, I next

performed two multiple linear regression analyses with indices of structural connectivity,

functional connectivity, and XY performance as independent variables; and BC and AC

performance, respectively, as the dependent variables. The model fit was significant ($F_{3,21}$

= 64.379, p < 1 x 10^{-9}), accounting for 88.8% of the variance in BC performance (adjusted R^2). Moreover, all three independent variables showed a significant positive relationship to BC performance (structural connectivity: β = 0.17, p = 0.022; functional connectivity: β = 0.50, p < 1 x 10^{-6}; XY performance: β = 0.92, p < 1 x 10^{-10}; all statistics reflect standardized β), demonstrating the unique contributions of structural and functional hippocampal-MPFC connectivity to subsequent learning of related experiences. Similar results were found in the regression model predicting AC performance ($F_{3,21}$ = 46.547, p < 1 x 10^{-8}; adjusted R^2 = 0.851; structural connectivity: β = 0.20, p = 0.021; functional connectivity: β = 0.48, p < 1 x 10^{-4}; XY performance: β = 0.90, p < 1 x 10^{-9}).

6.5. DISCUSSION

The present study provides convergent evidence from functional and structural MRI methodologies that hippocampal-MPFC interactions promote integration of new content into existing knowledge, both at the level of individual memories and across participants. This work is consistent with the idea that MPFC resolves competition among memories through integration (Preston and Eichenbaum, 2013) and promotes generalization across episodes (Schnider, 2003; Gilboa et al., 2006; Koscik and Tranel, 2012; Warren et al., 2014). I show that recruitment of the hippocampal-MPFC circuit benefits new learning, providing a neurobiological account of the age-old observation that knowledge can promote the formation of new, related memories (Bartlett, 1932). Integration-related signatures were observed in posterior MPFC, consistent with this region's proposed role in integrating across inputs (Nieuwenhuis and Takashima, 2011).

Structural measures also predicted performance, with greater tract coherence between hippocampus and MPFC being associated with superior memory integration. I further extend prior human and animal work by showing that content-specific cortical modules interacted with this circuit in service of integration. Specifically, I found that FFA was functionally connected to the hippocampal-MPFC circuit when new content related to previously learned face-object associations.

Enhanced interregional coupling when new learning related to prior knowledge was observed not only during intentional study, but also during post-encoding rest periods. Despite the fact that immediately preceding experiences consisted of object-object associations in both cases, functional connectivity and its relationship to memory performance differentiated post-BC from post-XY encoding rest periods. I found that enhanced hippocampal functional coupling with posterior MPFC following overlapping encoding was associated with behavioral evidence for superior integration. This work extends theories on the sleep-based mechanisms thought to support memory reorganization and integration (Lewis and Durrant, 2011) by providing an empirical test during awake rest. I also observed enhanced hippocampal and MPFC functional coupling with FFA during this period, suggesting processing of integrated traces incorporating more remotely encoded but related content (A_{face} items). Taken together, these results suggest a behavioral benefit to offline processing of integrated memories on subsequent recall of the newly learned information.

Theories propose that hippocampus and MPFC form a highly interactive and dynamic circuit impacting both encoding and retrieval. In particular, MPFC is thought to

form memory models (Schacter et al., 2012; St. Jacques et al., 2013) that bias hippocampal retrieval toward task-relevant memories (Kroes and Fernández, 2012; van Kesteren et al., 2012; Preston and Eichenbaum, 2013). When new content overlaps with existing knowledge, MPFC memory models may thus be activated, guiding hippocampal retrieval of relevant knowledge. Hippocampal encoding mechanisms would then bind current experience to the reactivated content to form integrated memory traces (Preston and Eichenbaum, 2013; Schlichting and Preston, 2015), updating memory models (van Kesteren et al., 2012) through projections to MPFC.

The present findings build upon these theories to provide an understanding of the roles played by subregions of hippocampus (Giovanello et al., 2009; Poppenk et al., 2013) and MPFC (Ongür and Price, 2000; Nieuwenhuis and Takashima, 2011; Roy et al., 2012) in memory integration. For instance, univariate analyses revealed that activation in posterior hippocampus predicted subsequent memory for overlapping associations. As previous work has implicated posterior hippocampus in representing specific event elements (Komorowski et al., 2009; Liang et al., 2012; Poppenk et al., 2013; Preston and Eichenbaum, 2013), one possibility is that posterior hippocampus drives retrieval of mnemonic details to support new BC encoding.

Notably, in contrast to the univariate activation observed in posterior hippocampus, task-phase connectivity with MPFC was observed more anteriorly. In addition to having direct anatomical projections to MPFC (Barbas and Blatt, 1995), anterior hippocampus has been implicated in forming generalized representations that span events (Komorowski et al., 2009; Liang et al., 2012; Poppenk et al., 2013; Preston

and Eichenbaum, 2013). Thus, one possible interpretation of this finding is that detailed memories reactivated by posterior hippocampus are communicated to anterior hippocampus, which in turn integrates and transfers them to MPFC. Offline processing of integrated traces may occur through coordinated hippocampal-MPFC interactions during rest, with reactivation of episodic detail (posterior hippocampus) and integration across episodes (anterior hippocampus) occurring simultaneously following overlapping encoding.

In the present study, I observed effects primarily in posterior aspects of MPFC (i.e., subgenual MPFC/anterior cingulate cortex [ACC]) during both encoding and rest. These findings are consistent with work suggesting that this subregion in particular carries out the integrative functions of MPFC (Nieuwenhuis and Takashima, 2011; Roy et al., 2012). Subgenual MPFC exhibits a widespread pattern of anatomical connectivity (Barbas and Blatt, 1995; Cavada et al., 2000; Ongür and Price, 2000), allowing it to integrate across limbic inputs during learning. Damage to this region results in reduced false memory formation (Warren et al., 2014) and poor schema representation (Ghosh et al., 2014), suggesting its involvement in integrating across sources of information. Engagement of subgenual MPFC also increases with consolidation (Nieuwenhuis and Takashima, 2011), perhaps tracking the increasing importance of abstracted neocortical representations as experiences become more remote. The present work extends these ideas to suggest that enhanced engagement and connectivity of subgenual MPFC with hippocampus promotes integration of overlapping events into prior memories. These

results may thus reflect both learning-phase biasing of hippocampal retrieval mechanisms toward relevant memories and rest-phase processing of integrated memories.

I also observed performance-related connectivity of anterior aspects of MPFC (i.e., rostromedial PFC) with hippocampus during rest. Interestingly, this region showed a different pattern of connectivity than did subgenual MPFC. In anterior MPFC, connectivity was enhanced during post-BC relative to post-XY rest for individuals who showed superior BC relative to XY learning, while the opposite pattern was observed for individuals showing better memory for XY than BC pairs (**Fig. 27A**, bottom right). One speculative interpretation of this finding is that while subgenual MPFC integrates overlapping content, anterior MPFC performs a more general mnemonic function. This region has been implicated in episodic simulation and future thinking (Okuda et al., 2003; Addis et al., 2009), which require retrieval and restructuring of episodic details. Thus, one possibility is that during post-encoding rest, anterior MPFC guides reinstatement of both overlapping and non-overlapping associations. Enhanced rest-phase anterior MPFC-hippocampal connectivity may therefore be associated with superior memory for the preceding pairs, regardless of their relationship to prior knowledge.

Notably, the present findings contrast with a prior study that reported decreased functional coupling for participants with strong prior knowledge (van Kesteren et al., 2010a). I propose that there are at least three factors that might account for these discrepancies. First, while the present study trained both initial and overlapping associations within a single day, the previous study (van Kesteren et al., 2010a) imposed a 24-hour delay between initial learning and overlapping encoding. This difference could

have significant implications for memory processes given the demonstrated importance of sleep for integrating and generalizing across experiences (Buckner, 2010; Lewis and Durrant, 2011). Second, the degree of compatibility between existing knowledge and new experiences—thought to impact integration demands (van Kesteren et al., 2012)—is difficult to compare across paradigms. Prior work (van Kesteren et al., 2010a) has theorized that integration demands may be higher when encoding in the context of incompatible related memories, while this paradigm may encourage resolving incompatibilities via integration precisely when strong prior knowledge exists (Zeithamova et al., 2012a; Preston and Eichenbaum, 2013; Bein et al., 2014). Third, previous work (van Kesteren et al., 2010a) used a large, heterogeneous MPFC ROI to assess connectivity with hippocampus, thus making it impossible to determine which subregion might have driven their results. Here, I use a voxelwise approach to demonstrate that integration-related connectivity enhancements were found only in posterior MPFC.

I also interrogated white matter structure to find that fractional anisotropy in hippocampal-MPFC white matter pathways related to memory integration behavior. Prior work (Gerraty et al., 2014) has shown individual differences in resting-state functional connectivity relating to memory integration processes that support generalization. In apparent contrast to the present findings, that study found that intrinsic (i.e., not task-evoked) hippocampal-MPFC functional connectivity was negatively correlated with behavioral evidence for memory integration (Gerraty et al., 2014). However, hippocampus and MPFC also showed opposite connectivity-performance relationships

181

with the default mode network (DMN), with low hippocampal-DMN and high MPFC-DMN connectivity being associated with superior integration. Thus, it is unclear whether the negative relationship between intrinsic hippocampal-MPFC connectivity and behavior implies more or less reliance on these structures during integration itself. These results complement this prior report in suggesting that even stable individual differences in structure and function of hippocampal-MPFC circuitry relate to complex memory behaviors. Importantly, these results rely on an unrelated measure of structural connectivity, highlighting the conceptual convergence across methodologies.

In summary, I demonstrate that fluctuations in hippocampal-MPFC connectivity track integration demands on a trial-by-trial basis. These findings also provide novel insight into the timecourse of integration, suggesting the importance of post-encoding rest periods for offline reorganization of overlapping memories. Evidence of rest-phase integration was specific to posterior MPFC, consistent with the notion of dissociable mnemonic functions across MFPC. Moreover, I demonstrate how underlying hippocampal-MPFC structure relates to integration ability, providing insight into why some individuals are better able to integrate knowledge than others.

7. Experiment 3: Neural Representations Underlying Inference

7.1. RATIONALE

One of the most difficult problems facing the memory system is simultaneously representing important commonalities across experiences while still maintaining the details of individual events. For instance, while in some cases it might be advantageous to generalize across many trips to the supermarket (e.g., when inferring where to find the milk in a new store), doing so will not help you remember where you parked your car during today's grocery run. From an intuitive perspective, it seems clear that our brains must represent both—storing separated traces that keep even highly overlapping memories distinct and integrated traces that combine across memories. Here, I will first discuss how the tension between these opposing goals has been described in terms of two hippocampal operations: pattern separation and pattern completion (O'Reilly and Rudy, 2001). I then turn to the aims of the present study, which include (1) assessing the learning conditions promoting integration over separation and (2) determining the specific regions supporting each.

On the one hand, the hippocampus has traditionally been conceptualized as a fast-learning system whose primary purpose it is to rapidly encode pattern-separated memories for individual events. Under this model, even related events are coded by largely non-overlapping neuronal populations (McClelland et al., 1995; O'Reilly and Rudy, 2001; Norman and O'Reilly, 2003). Such a bias in the hippocampus to pattern separate is thought to support retrieval of specific episodic details after just a single experience and protect from interference among related memories (O'Reilly and Rudy,

2001). The complementary learning systems (CLS) framework contrasts hippocampal function with that of a slow-learning neocortical system, which generalizes across memories over time (McClelland et al., 1995; O'Reilly and Rudy, 2001; Norman and O'Reilly, 2003). Thus, according to the CLS model, hippocampus and neocortex will store complementary separated and integrated memory traces that allow for retrieval of specific event details and generalizations, respectively. This theory has received a host of empirical support from studies showing increasing reliance on neocortical over hippocampal memory representations over time (i.e., as a result of consolidated; Scoville and Milner, 1957; Zola-Morgan and Squire, 1990; Kim and Fanselow, 1992; Squire, 1992; Rempel-Clower et al., 1996; Bontempi et al., 1999; Haist et al., 2001; Squire et al., 2004; Douville et al., 2005; Bayley et al., 2006; Takashima et al., 2006, 2007; Gais et al., 2007; Sterpenich et al., 2009; Smith and Squire, 2009) as recollections become more schematized (Winocur et al., 2010).

However, alternate theoretical accounts (Eichenbaum, 1999) suggest that under some circumstances, the hippocampus may forgo its default pattern-separating mode to instead pattern complete. In particular, when two experiences are substantially overlapping, the presence of shared features may bias the hippocampus to retrieve, or pattern complete to, the related memory instead of laying down a new, pattern-separated trace (O'Reilly and Rudy, 2001). Pattern completion during new memory formation would result in the recruitment of overlapping populations of hippocampal neurons to encode the related events (Eichenbaum, 1999; O'Reilly and Rudy, 2001). This representational scheme, termed nodal coding, is one mechanism by which memories for

184

related experiences might become linked to capture regularities across events (Eichenbaum et al., 1999). Empirical work has demonstrated the existence of such overlapping codes when experiences share features (Eichenbaum et al., 1999; Wood et al., 2000; Singer et al., 2010; McKenzie et al., 2013, 2014) and have linked integration-related neural signatures to a host of flexible behaviors such as novel inference (Zeithamova and Preston, 2010; Zeithamova et al., 2012a; for reviews see Zeithamova et al., 2012b; Schlichting and Preston, 2015).

A goal of this experiment was to test the predictions stemming from separation and integration perspectives to better describe the neural mechanisms supporting memory flexibility. I sought to tease apart these neural signatures by teaching participants overlapping AB and BC associations, where the B item was shared between pairs. Participants later completed a surprise inference test, in which they were asked to link indirectly related A and C items through their common association with B. I then interrogated the learning-related changes to the neural representations of A and C memory elements supporting inference for evidence of separation and integration.

Critically, these two representational schemes make opposing predictions for how indirectly related A and C memory elements will be coded. Pattern separation predicts that neural representations for A and C items will become more dissimilar after learning, as AB and BC memories will be coded by non-overlapping neuronal populations. In contrast, integration via nodal coding predicts increases in the similarity of A and C representations. While increased neural similarity in the human hippocampus has been observed for immediately adjacent items in a temporal sequence (Schapiro et al., 2012),

185

similarity increases for indirectly related memory elements remain to be demonstrated. Importantly, both coding schemes could support novel inference judgments in which participants linked A and C items (O'Reilly and Rudy, 2001; Kumaran, 2012; Kumaran and McClelland, 2012; Zeithamova et al., 2012b). In the case of separation, inferences may be made via a logical recombination process (Zeithamova et al., 2012b) or through recurrent connections (Kumaran and McClelland, 2012). In contrast, integration would allow for direct extraction of the inferential A-C relationship from the integrated memory trace (Zeithamova et al., 2012b; Schlichting and Preston, 2015). Importantly, these coding schemes are not mutually exclusive; rather, separation and integration may alternately support successful inference on an event-by-event basis. In this study, I sought to determine the conditions that bias the memory system towards pattern completion and integration over pattern separation.

One factor that may influence the neural representations recruited for a given pair of overlapping events is the manner in which the events are experienced. Existing theories make opposing predictions for how overlapping events encoded in a blocked versus intermixed manner may support inference. For instance, schema theory (van Kesteren et al., 2012) would predict that blocked learning will result in better inference, as BC information is learned only after AB memories are well established (hereafter referred to as the AB strength → integration prediction). Such strong AB representations would promote pattern completion during BC events, thereby allowing for the formation of overlapping traces (O'Reilly and Rudy, 2001). Consistent with this hypothesis, existing empirical and theoretical work has suggested that strong initial memories—as is

the case for blocked learning in the present study—promotes integration across related episodes (Bartlett, 1932; Heckers et al., 2004; Preston et al., 2004; Greene et al., 2006; van Kesteren et al., 2013a, 2010a, 2010b, 2012; Zeithamova et al., 2012a; Ghosh and Gilboa, 2013; Preston and Eichenbaum, 2013; Schlichting and Preston, 2014).

However, signatures consistent with hippocampal integration have also been observed after just a single experience with the initial AB association (Zeithamova and Preston, 2010; Schlichting et al., 2014) or when overlapping events are presented in an intermixed fashion (Shohamy and Wagner, 2008; Howard et al., 2009), suggesting that strong initial memories are not a prerequisite for integration. One possible explanation for these findings stems from temporal context theories (Estes, 1955; Howard and Kahana, 2002), which predict that encoding AB and BC in a more similar temporal context leads to more highly overlapping representations (i.e., more integration). In hippocampal terms, this might be due to a greater tendency to pattern complete to more recently experienced events, a prediction that also receives support from models of hippocampal function (O'Reilly and Rudy, 2001). In this framework, intermixed learning would be expected to yield learning-related increases in neural pattern similarity for A and C items (hereafter, AB recency → integration).

In addition to distinct separation and integration signatures across events within a given brain region, different regions may also preferentially exhibit one type of coding strategy over the other. For example, recent work suggests dissociable functions across the hippocampal anterior-posterior axis (converging with rodent work; Moser and Moser, 1998), with anterior generalizing across events and posterior coding for event detail

(Giovanello et al., 2009; Demaster and Ghetti, 2013; Poppenk et al., 2013). These findings suggest that while posterior hippocampus might pattern separate overlapping events, anterior hippocampus would be more likely to integrate.

Outside of the hippocampus, both medial and lateral aspects of PFC have been implicated in overlapping encoding and successful inference (Schlichting and Preston, under review; Zeithamova and Preston, 2010; Zeithamova et al., 2012a). Accordingly, I also interrogated activation in MPFC and IFG in the context of these perspectives to determine the coding schemes recruited by each. One theoretical framework (Preston and Eichenbaum, 2013) suggests that while MPFC generalizes across events, IFG represents individual memories for later recombination. In particular, schema theory (van Kesteren et al., 2012) suggest that hippocampus interacts with MPFC (van Kesteren et al., 2010a) to acquire new information in the context of existing knowledge (Tse et al., 2007, 2011), which can later be flexible applied to make novel judgments (Kumaran et al., 2009). In this framework, MPFC would be engaged during overlapping encoding, reflecting the updating of existing memories with new information and the formation of integrated memory representations. Consistent with this idea, MFPC engagement (Zeithamova et al., 2012a) and connectivity with hippocampus (Schlichting and Preston, under review; Zeithamova et al., 2012a) during encoding has been related to neural and behavioral signatures of memory integration. Thus, I predicted that MPFC would show neural evidence for integration of indirectly related A and C items in the present study.

In contrast, IFG is thought to protect from interference by maintaining distinct representations of similar events (Paller and Wagner, 2002; Badre and Wagner, 2007;

Öztekin et al., 2009; Jenkins and Ranganath, 2010; Ranganath, 2010; Kuhl et al., 2012a). IFG is recruited during episodic retrieval tasks that require recollection of prior event details (Mitchell and Johnson, 2009), particularly when distinctive features of individual events must be isolated (Raposo et al., 2009). Moreover, encoding activation in this region is associated with successful memory for individual events (Wagner et al., 1998; Uncapher and Rugg, 2005; Straube et al., 2009; Kuhl et al., 2012b), but does not predict subsequent inference performance (Zeithamova and Preston, 2010). These findings provide support for the hypothesis that although IFG plays an important role in memory, its involvement is not specific to integration. During retrieval, however, IFG activation does predict inference success (Zeithamova and Preston, 2010), consistent with this region's hypothesized role in recombining individual event memories during novel judgments. Accordingly, I predicted that IFG would separate A and C items.

While the existing literature suggests the importance of hippocampus, MPFC, and IFG in successful inference, prior work has universally failed to demonstrate the mechanism by which each contributes. In this experiment, I aimed to move beyond describing simply *where* this signature occurs to instead show *how* these regions represent overlapping information. Participants viewed rapid presentations of single A, B, and C items during hr-fMRI scanning both before and after learning of overlapping AB and BC associations. I assessed learning-related changes in the neural representations of individual A and C items using neural pattern similarity analysis (NPSA; **Fig. 30**). As each participant learned half of the pairs in a blocked and half in an intermixed manner, I

189

was additionally able to investigate how changes in neural representations were modulated by learning condition.

The results of this study are reported in Schlichting et al., in preparation.
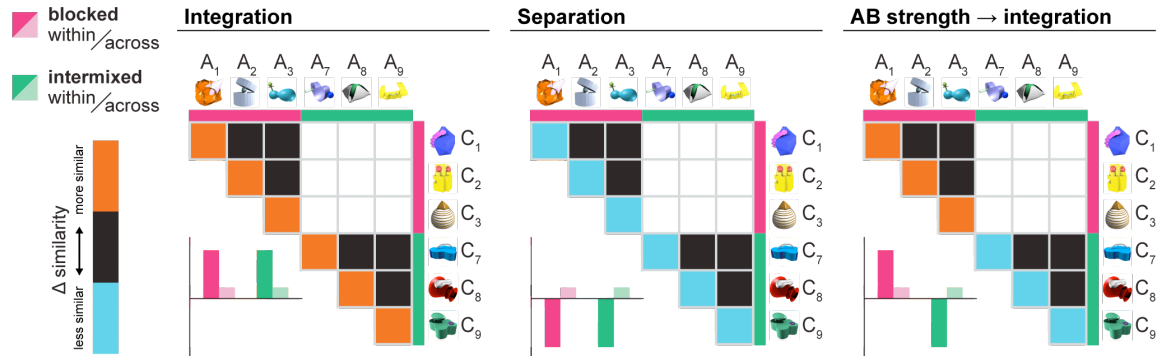
Figure 30. Experiment 3: Predictions for Neural Pattern Similarity Analysis

Schematic depiction of neural pattern similarity analysis and predictions for a subset of six triads. Triads 1-3 were studied in a blocked manner (pink); triads 7-9 were intermixed (green). For all matrices, the brightness in each cell represents the change in (Δ) neural pattern similarity between an A item (horizontal) and C item (vertical) from pre- to post-study. Black cells indicate no change in similarity from pre- to post-study; orange cells indicate learning-related increases in similarity; blue cells indicate learning-related decreases in similarity. White cells represent comparisons across learning condition (e.g., intermixed item A7 with blocked item C1) and were excluded from the all analyses. Matrices show predicted item similarities for regions showing integration for both blocked and intermixed learning (left), separation for both blocked and intermixed learning (middle), and an interaction with learning condition (right). For simplicity, only the AB strength $\rightarrow$ integration prediction is depicted. Inset bar graphs show predicted average similarities across all within- (darker bars) and across-triad (lighter bars) comparisons. Integration and separation were operationalized as significantly greater (integration) or less (separation) within- than across-triad Δ neural pattern similarity.

7.2. HYPOTHESES

Hypothesis 1: *Neural pattern similarity in anterior hippocampus and MPFC will increase for indirectly related items as a result of learning*. As hippocampus and MPFC are thought to support memory integration, there will be increased pattern similarity from

pre- to post-study for A and C items from the same triad relative to items from different triads.

Hypothesis 2: *Posterior hippocampus and IFG will exhibit pattern separation, with decreased pattern similarity for indirectly related items following learning*. IFG and posterior hippocampus may support inference via retrieval of specific events and logical recombination; thus, these regions may maintain separate memory traces for the individual related episodes. Accordingly, neural patterns associated with A and C items will become less similar in these regions as a result of learning.

Hypothesis 3: *Blocked learning will be associated with evidence for integration in anterior hippocampus and MPFC*. In accordance with schema theory, blocked learning will be associated with evidence for integration in anterior hippocampus and MPFC.

7.3. METHODS

7.3.1. Participants

Thirty right-handed volunteers (15 females; ages 18-27; mean ± SEM = 21.7 ± 0.5 years) participated in the experiment. Consent was obtained in accordance with an experimental protocol approved by the Institutional Review Board at the University of Texas at Austin. Participants received monetary compensation for their involvement in the study. Data from a total of four participants were excluded for hardware malfunction (N = 1), failure to complete the experiment due to illness (N = 1), instruction error (N = 1), and low memory performance (N = 1). Low memory performance was defined as failure to reach above 80% correct on the directly learned (AB, BC) associations as

assessed during the post-scan memory test. Data from the remaining 26 participants were included in all analyses (14 females; ages 18-27; 21.6 ± 0.5 years).

7.3.2. Behavioral Methods

Materials

Stimuli consisted of 36 multicolored novel objects (a subset of which were adapted from Hsu et al., 2014) created using Blender, an open source 3D animation suite (www.blender.org). Novel objects were made to appear physically feasible but distinct from real-world objects. I chose novel rather than real-world objects to avoid stimuli with pre-existing knowledge representations and associations.

Procedures

Novel objects were arranged into 12 ABC triads. ABC triads were presented to participants as overlapping AB and BC pairs, with the B item shared between pairs (**Fig. 31A**). That is, AB pairs consisted of two novel objects, A and B; the B object was then later paired with a new novel object C to form a BC pair. Overlapping AB and BC pairs were divided into two learning conditions comprising 6 ABC triads each: blocked and intermixed. In the blocked learning condition, all AB pair presentations occurred prior to the presentation of any BC pairs (**Fig. 31A**, pink). In the intermixed learning condition, AB and BC pairs were presented in alternation; that is, for a given ABC triad, the participant first saw AB, then BC, then AB, then BC, and so on (**Fig. 31A**, green). For both conditions, triads were presented in a pseudo-random order, with the constraint that two pairs from the same triad (multiple presentations of a single AB or BC; or an AB and its corresponding BC) were not presented in immediate succession. The assignment of

stimuli to conditions and the order of learning conditions (i.e., whether blocked or intermixed learning occurred first) were counterbalanced across participants.
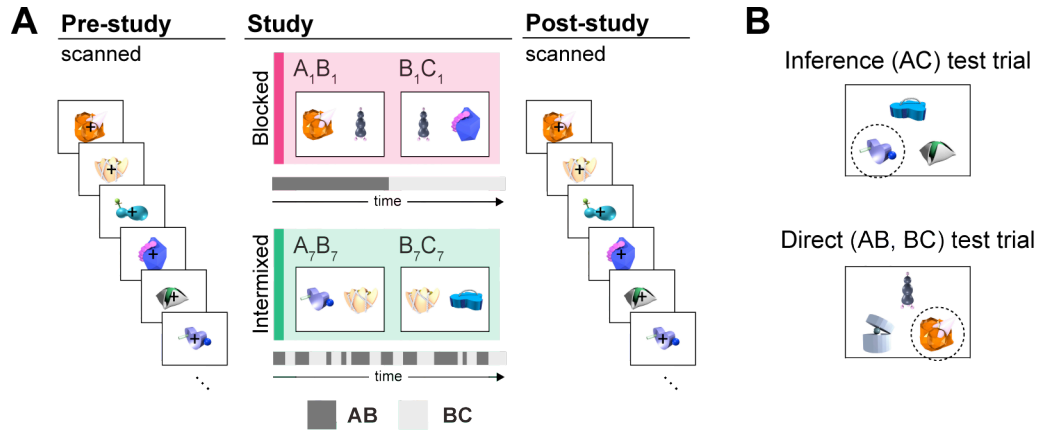
Figure 31. Experiment 3: Memory & Exposure Tasks

Experimental paradigm. (**A**) During the study phase (middle), participants intentionally encoded pairs of novel objects. Half of the pairs were presented in a blocked fashion (top, pink); half were intermixed (bottom, green). Identical stimulus exposure phases occurred immediately before (left) and after (right) the study task during fMRI scanning. Pre- and post-study exposure phases were used to obtain estimates of the neural patterns evoked by specific stimuli learned during the study phase. Trial timing and order were matched between pre- and post-study to avoid introducing unequal bias into the neural pattern estimates from the two phases. (**B**) After scanning, participants completed a two-alternative forced choice test for inference (top, tested first) and directly learned (bottom, tested second) associations. Participants selected which of the two choice stimuli (bottom of screen) was associated with the cue stimulus (top of screen). Correct answers are circled for display only (not shown to participants).

During a study phase (not scanned; **Fig. 31A**), novel object pairs were presented on the screen for 3.5 seconds with an ITI of 0.5 seconds (**Fig. 31A**). Each of the 24 AB and BC pairs was presented 12 times. The left/right position of stimuli on the screen was randomized across presentations. Participants were asked to try their best to remember the pairs by creating a visual or verbal story (e.g., *try to picture these objects interacting in some way*); no behavioral response was required during the study phase. Importantly,

participants were not made aware of the overlap between AB and BC pairs prior to beginning the experiment; that is, no instructions were given as to how they should remember the overlapping associations. The study was broken up into two parts (blocked and intermixed learning) of 9.6 minutes each, with the order counterbalanced across participants. Participants were given the opportunity to take a short break between study portions if they so wished.

Participants were exposed to single items both immediately prior to and following study during fMRI scanning (**Fig. 31A**). The goal of the pre- and post-study exposure phases was to assess the effect of experience on the neural representations of individual A, B and C items. During this task, participants viewed novel objects on the screen in isolation while performing an orthogonal change detection task (Kriegeskorte et al., 2008b). Trials were 4 s in duration. At the beginning of the trial, a single novel object appeared on screen for 300 ms. At a random time during this 300 ms interval, the superimposed white fixation cross turned either blue or green. Subjects were instructed to press a button with their index finger if it turned blue, and a button with their middle finger if it turned green. Subjects had until the end of the 4 s trial to make their response, but were asked to respond as quickly as possible without sacrificing accuracy. These responses were collected solely to ensure attention to the stimuli and were not considered in the data analysis.

During each functional run, each of the 36 items was presented in isolation exactly twice. Items were presented in a random order, with the additional constraint that items from the same triad (e.g., A, B, and C from triad 1) were presented with no fewer

than two other items in between them. This was done to ensure that the parameter estimate for the item of interest was not contaminated by lingering activation associated with another item in the same triad due to the sluggish hemodynamic response. All items were presented once before any item appeared for a second time (i.e., both halves of each run contained exactly one presentation of all 36 items). The 72 item trials were randomly intermixed with 24 null fixation trials (4 s long) (Kriegeskorte et al., 2008b), yielding a total run length of 6.4 minutes. There were 4 pre-study exposure scans and 4 post-study exposure scans. The ordering of trials and scans was identical between the pre- and post-study exposure scans. This was done to ensure that pre- and post-study activation measures were not differentially impacted by stimulus presentation order (Mumford et al., 2014).

Following scanning, it was explained to participants that A and C items could be indirectly related through their common association with a single item, B. After ensuring that participants understood the inferential test, a two alternative forced choice test over the inferential (AC) and directly learned (AB, BC) associations was administered (**Fig. 31B**). All inferential associations were tested before AB and BC associations to prevent additional learning of the direct associations. As with the study phase, test trials (AC, AB, or BC) from the same triad were not tested in immediate succession; the order of trials was otherwise random. A test screen consisted of three items: a cue item at the top of the screen, and two options on the bottom of the screen. For AC and BC test trials, C items served as cues; B items served as cues for AB test trials. Participants were instructed to select which of the bottom two items was associated with the cue with a key press. The

test was self-paced. Incorrect options (i.e., foils) were always familiar items that were members of another triad in the same condition (blocked or intermixed learning). Proportion correct was computed separately for each learning condition (blocked, intermixed) and test trial (AC, AB, BC) type and averaged across participants. Differences in performance across conditions were assessed using a 3×2 repeated measures ANOVA with test trial type (AC, AB, BC) and learning condition (blocked, intermixed) as within-subjects factors.

Prior to scanning, participants had the opportunity to practice study, test and exposure tasks. The practice study and test stimuli were novel objects that were not included in the main experiment. The practice pairs were not overlapping, so as to not encourage any strategy in particular prior to beginning the experiment. The practice exposure task consisted of a single presentation of each of the 36 items used in the main experiment. This was done to minimize stimulus novelty effects in the scanner.

7.3.3. Imaging Methods

MR Data Acquisition

Imaging data were acquired on a 3.0T Siemens Skyra MRI system. Functional data were collected in 72 oblique axial slices using an EPI sequence, oriented approximately 20° off the AC-PC axis (TR = 2000 ms, TE = 31 ms, flip angle = 73; 128 x 128 x 72 matrix, 1.7 mm isotropic voxels, multiband acceleration factor = 3, GRAPPA factor = 2). Two field maps were collected (TR = 589 ms, TE = 5 ms/7.46 ms, flip angle = 5 degrees; matrix size = 128 x 128 x 60; 1.5 x 1.5 x 2 mm voxels) to allow for correction of magnetic field distortions. One was collected before the first functional run

(before the first pre-study exposure) and one before the fourth functional run (before the first post-study exposure). Two oblique coronal T2-weighted structural images in the same prescription were acquired perpendicular to the main axis of the hippocampus (TR = 13150 ms, TE = 82 ms, 512 x 60 x 512 matrix, 0.4 x 0.4 mm in-plane resolution, 1.5 mm thru-plane resolution, 60 slices, no gap). A T2-weighted structural image in the same prescription as the functional images (i.e., coplanar image; TR = 15780 ms, TE = 82 ms, 512 x 512 matrix, 0.4 x 0.4 mm in-plane resolution) and a T1-weighted 3D MPRAGE volume (256 x 256 x 192 matrix, 1 mm isotropic voxels) were also collected to facilitate image coregistration, intracranial volume (ICV) estimation using Freesurfer (Desikan et al., 2006), and spatial normalization to the MNI template brain.

fMRI Preprocessing

Data were preprocessed and analyzed using FSL version 5.0 (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl) and Advanced Normalization Tools (ANTS) (Avants et al., 2011). Motion correction was applied to each functional run using MCFLIRT, part of FSL. Functional runs were then registered to a single functional reference run (here, the fourth run) by applying affine transformations calculated in ANTS to each functional timeseries. Each participant's coplanar, MPRAGE and mean coronal structural images were registered to their functional data using affine transformations implemented in ANTS. The coplanar image was registered to the functional reference run following fieldmap-based unwarping of the functional data (see below). Transformations were computed for the MPRAGE to the coplanar image and the mean coronal to the MPRAGE. Appropriate transformations were concatenated and

applied such that all structural images were moved to functional space and resampled to functional dimensions. For each registration to functional space (coplanar, MPRAGE and mean coronal) resampling occurred only once. Non-brain structures were removed from the coplanar, MPRAGE and functional images using BET, part of FSL. Brain extraction for the mean coronal image was performed using the brain mask from the MPRAGE. With the exception of group-level statistics, all analyses were carried out in the functional space of each participant.

Pre-statistics processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00 part of FSL. The following processing was applied: fieldmap-based EPI unwarping using PRELUDE+FUGUE (Jenkinson, 2003), grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50 s); and spatial smoothing using a Gaussian kernel of 4 mm full width at half maximum (FWHM). The first field map was used to unwarp the pre-study exposure scans; the second field map was used to unwarp the post-study scans. This was done to provide equally optimal unwarping for both scanned phases of the experiment, as any movements during the intervening 20-minute study phase may have caused small changes in the magnetic field.

7.3.4. Data Analysis

Regions of Interest

Anatomical ROIs were used to restrict neural pattern similarity searchlight analyses. Hippocampus was manually demarcated on a custom template generated from

the mean coronal images; MPFC and IFG ROIs were created on the MNI template brain. These procedures are described in detail below.

A custom coronal template was generated using ANTS (Avants et al., 2011). The T2-weighted mean coronal images from a subset of ten participants with canonical hippocampi were selected for template generation. A bilateral hippocampus ROI was delineated by hand on the coronal template using established guidelines (Amaral and Insausti, 1990). Hippocampus was further segmented into head, body and tail subregions for volume analysis using anatomical landmarks as follows. The posterior boundary of the hippocampal head was the last slice on which the uncal apex was visible (Weiss et al., 2005; Poppenk and Moscovitch, 2011). The anterior boundary of the hippocampal tail was the first slice on which the fornix became visibly separated from the hippocampus (Watson et al., 1992). Nonlinear transformations were calculated to normalize each participant's mean coronal image to the coronal template. The inverse of these transformations were then concatenated with the coronal to functional (affine) transformation and applied to the hippocampal ROIs. This process resulted in bilateral hippocampal head, hippocampal body, hippocampal tail, and overall hippocampal ROIs in the native functional space of each participant.

Prefrontal ROIs were defined on the MNI template brain. MPFC was delineated by hand to approximate all cytoarchitectonic subdivisions thought to be part of the medial prefrontal network, which is most related to limbic structures such as the hippocampus (Ongür et al., 2003; Price and Drevets, 2009). IFG was created by summing subdivisions pars opercularis, pars triangularis, and pars orbitalis derived from Freesurfer (Desikan et

al., 2006). ROIs were then inflated to allow for variability in neocortical anatomy across participants and reverse normalized to each participant's native functional space.

Estimation of Item-Level Neural Patterns

Item-level neural patterns were generated under the assumptions of the general linear model (GLM) using a modified LS-S approach, which has been described previously (Mumford et al., 2012). Parameter estimate images associated with each of the 36 novel objects were extracted for each run and each participant using custom Python routines. Item presentations were modeled as 0.3 s events and convolved with the double gamma hemodynamic response function. The two presentations of each item were modeled together in a single regressor. Motion parameters calculated during the motion correction step and their temporal derivatives were added as additional confound regressors. Two measures of framewise data quality were also calculated to identify volumes that may be adversely impacted by motion artifacts: framewise displacement (FD) and DVARS (Power et al., 2012). FD measures the overall change in head position from one time point to the next, and is calculated by summing the absolute values of the derivatives of the six motion parameters calculated during the realignment step. DVARS measures the overall change in image intensity from one time point to the next. This index is calculated as the root mean square of the derivatives of the timecourses across all voxels in the brain. Both FD and DVARS were added to the model as regressors of no interest (Schlichting and Preston, 2014; Schonberg et al., 2014). Additional regressors were created for each time point in which motion exceeded a threshold of both 0.5 mm for FD and 0.5% change in BOLD signal for DVARS (plus one time point before and

two time points after) (Power et al., 2012). Temporal filtering was applied to the model. This process resulted in one statistics image for each of 36 items in each of eight runs (for a total of 288 images per participant).

Neural Pattern Similarity Searchlight

Searchlight neural pattern similarity analysis (NPSA) was carried out in the native space of each participant using the PyMVPA toolbox (Hanke et al., 2009) and custom Python routines. This approach was used to identify voxels showing one of the three hypothesized learning-related signatures in A-C neural pattern similarity: (1) integration for both learning conditions (**Fig. 30**, left); (2) separation for both learning conditions (**Fig. 30**, middle); (3) interaction, with AB strength → integration (**Fig. 30**, right); and (4) interaction, with AB recency → integration. For the AB strength → integration interaction, I searched for voxels showing integration for blocked triads and separation for intermixed triads (i.e., blocked > intermixed × within > across triad similarity). The AB recency → integration analysis searched for the opposite pattern (i.e., intermixed > blocked × within > across triad similarity). All analyses were limited to those A and C items for which the AC inference judgment was correct during the test.

I searched for each of these four effects using searchlights restricted to voxels in three *a priori* anatomical ROIs: bilateral hippocampus, MPFC, and IFG. For completeness, I also ran searchlight analyses across the whole brain, unrestricted to any particular region. ROIs were defined on custom coronal (hippocampus) and MNI (MPFC, IFG) template brains as described previously and reverse normalized to each participant's functional space to carry out the searchlight analyses. Group ROIs were used to ensure

that the spatially normalized statistical maps (see below) would be overlapping, thereby enabling comparisons across the group.

A spherical searchlight (radius = 3 voxels; volume = 123 voxels, except along ROI boundaries) was swept across each anatomical ROI. For each sphere (i.e., centered on every voxel in the ROI), a statistic of interest was calculated from the item-level pairwise comparisons of activation patterns as follows. Pairs of item-level activation patterns from pre- and post-study exposure phases were compared separately using Pearson's correlation, transformed to Fisher's z, and subtracted, resulting in a post-pre similarity score for each pair of items (hereafter, change in [Δ] similarity). Changes in pairwise similarity for comparisons of interest were then averaged to yield mean within- (i.e., the average of all within-triad [e.g., A1-C1] comparisons) and across-triad (i.e., the average of all across-triad [e.g., A1-C2] comparisons) Δ similarities for blocked and intermixed learning. Importantly, for both within- and across-triad Δ similarity calculations, comparisons were limited to pairs of activation patterns extracted from different scanning runs. This was done to ensure independence of the activation patterns going into Δ similarity calculations, thus preserving the false positive rate (Mumford et al., 2014). Across-triad comparisons were additionally limited to items from the same learning condition (e.g., a blocked A item would never be compared to an intermixed C item; **Fig. 30**, white cells). (The reader should note that within-triad comparisons are necessarily limited to items from the same learning condition.)

For each sphere in the searchlight, contrasts representing the three effects of interest were computed (**Fig. 30**, inset barcharts) using the mean Δ similarities as

follows: (1) integration, blocked within – blocked across + intermixed within – intermixed across; (2) separation, blocked across – blocked within + intermixed across – intermixed within; (3) AB strength → integration interaction, blocked within – blocked across + intermixed across – intermixed within; (4) AB recency → integration interaction, intermixed within – intermixed across + blocked across – blocked within. Contrasts were converted to p-values by comparing the observed contrast value to a bootstrapped null distribution. I generated null distributions by shuffling within- and across-triad Δ similarities (within learning condition) and re-computing the statistic of interest for each of 1,000 iterations. This p-value was assigned to the center voxel of the current sphere; the sphere was then shifted and the entire procedure repeated. Conducting the searchlight NPSA for each participant resulted in four p-value maps for each of the three anatomical ROIs as well as four for the whole brain.

Each participant's voxelwise p-values were converted to z-statistics and the resulting maps were warped to the MNI template (resampled to the functional dimensions of the present study, 1.7 mm isotropic) by applying nonlinear transformations computed previously using ANTS (Avants et al., 2011). Z-statistics were then combined across the group by using nonparametric one-sample t-tests implemented in Randomise (Winkler et al., 2014), part of FSL. For the whole brain analysis, statistics images were first masked to exclude white matter. I then applied a primary voxelwise threshold of p < 0.01 (uncorrected) to all group statistics images to identify those voxels surpassing this initial p-value threshold. Significant cluster sizes within each anatomical ROI (hippocampus, MPFC, IFG) and across the whole brain were determined using Monte Carlo simulations

implemented in 3dClustSim, part of AFNI (Cox, 1996). Cluster sizes that occurred with a probability of less than 0.05 across 2000 simulations were considered statistically significant. This was performed separately for each ROI and for the whole brain, restricted to gray matter.

Follow-up Functional ROI Analyses

Because the NPSA searchlights described above interrogated the brain for main effects and interactions across learning conditions, one possibility is that any observed effects might be driven by significant within- versus across-triad differences in one learning condition but not the other. To assess this possibility, I performed a follow-up analysis in which I assessed the within- versus across-triad Δ similarities for blocked and intermixed learning conditions separately within clusters identified in the main searchlight analyses as showing one of the three effects of interest.

Searchlights were run for each of the following learning-changes in A-C neural pattern similarity: (1) integration for blocked triads; (2) integration for intermixed triads; (3) separation for blocked triads; (4) separation for intermixed triads. I searched for these patterns within hippocampus, MPFC and IFG. With the exception of the contrast values calculated within each sphere, searchlight analyses at the individual participant level were identical to those described previously. Contrasts were calculated in the following manner: (1) integration for blocked triads, blocked within – blocked across; (2) integration for intermixed triads, intermixed within – intermixed across; (3) separation for blocked triads, blocked across – blocked within; (4) separation for intermixed triads, intermixed across – intermixed within. Bootstrapping was performed as described

previously, yielding four p-value maps for each of the three ROIs for each participant. Each participant's voxelwise p-value maps were converted to z-statistics and the resulting images were warped to the 1.7 mm isotropic MNI template using ANTS (Avants et al., 2011). Average z-statistics across each cluster identified in the main searchlight analyses were extracted and compared with zero using a bootstrapping approach. Participants were resampled with replacement and the average z-statistic across the simulated group was computed for each ROI on each of 100,000 iterations. P-values were determined as the proportion of iterations on which the group average was less than zero.

Hippocampal Volume-Δ Similarity Analysis

I next interrogated the relationship between measures of hippocampal structure and coding strategy. Specifically, I related the volume of hippocampal subregions (head, body, tail) to the neural evidence for integration and separation from the pattern similarity analyses described above. I tested the hypothesis that hippocampal head volume would relate to the degree of integration, particularly for blocked triads.

I extracted bilateral hippocampal head, body, and tail volumes for each participant, as well as ICV measures derived from Freesurfer (Desikan et al., 2006). To account for differences in overall head size (approximated by ICV) across participants, hippocampal volumes were adjusted using an established analysis of covariance approach (Raz et al., 2005). As my neural similarity measures, I extracted average z-statistics representing integration for blocked and separation for intermixed conditions for each participant across the whole hippocampal ROI. I then z-scored these measures and related hippocampal subregion volumes to neural similarity using robust regression. Volume

served as the independent variable and neural similarity served as the dependent variable.

I also performed two robust multiple regressions, with hippocampal head, body and tail

volumes as independent variables and neural evidence for integration in the blocked

condition and separation in the intermixed condition, respectively, as the dependent

measures.

Control Analyses

As encoding order (i.e., whether blocked or intermixed learning occurred first)

was counterbalanced across participants, individuals had substantial differences in

learning experience that might impact their behavioral memory performance and/or

neural coding. Accordingly, I performed control analyses to assess the effects of

encoding order on behavioral and neural measures of interest.

Effects of Encoding Order on Behavior

I first interrogated whether behavioral performance was significantly modulated

by encoding order. I performed a 3 × 2 × 2 mixed ANOVA with test trial type (AC, AB,

BC) and learning condition (blocked, intermixed) as within-subjects factors and encoding

order as the between subjects factor. Memory performance (proportion correct) served as

the dependent measure.

Effects of Encoding Order on Δ Neural Pattern Similarity

Average z-statistics were extracted for each participant across every cluster

identified in the main searchlight analyses within our *a priori* anatomical ROIs. These z-

statistics represented the degree to which each participant exhibited the effect of interest

(i.e., the effect that was significant in that particular cluster). I tested whether these values

differed as a function of encoding order using a two-sample t-test. Correction for multiple comparisons was performed using Bonferroni correction, yielding a critical p-value of 0.005 (0.05/11 clusters).

Effects of Encoding Order on Hippocampal Volume-Δ Neural Pattern Similarity Relationship

I also performed one-way analyses of covariance (ANCOVA) to interrogate whether the observed relationships between hippocampal head volume and neural similarity measures differed significantly as a function of encoding order. Hippocampal volumes (head, body, and tail) served as the predictor variables; neural similarity (integration in the blocked condition, and separation in the intermixed condition) served as the response. Encoding order was the grouping variable. Correction for multiple comparisons was performed using Bonferroni correction, yielding a critical p-value of 0.017 (0.05/3 subregions).

7.4. RESULTS

7.4.1. Behavioral Results

Behavioral performance for both directly learned AB (collapsed across learning condition; range: 83.3-100%; mean ± standard error of the mean (SEM) = 95.8 ± 1.3%; $t_{25} = 34.52$, $p < 1 \times 10^{-21}$) and BC (91.7-100%; 97.4 ± 0.8%; $t_{25} = 61.67$, $p < 1 \times 10^{-28}$) pairs was well above chance, suggesting that the paradigm provided sufficient training on the premise associations. AC inference performance (58.3-100%; 90.4 ± 2.3%; $t_{25} = 17.24$, $p < 1 \times 10^{-14}$) was also above chance. Investigating performance as a function of learning condition (**Fig. 32**), I observed only a significant main effect of test trial type

207

($F_{2,50} = 6.65$, p = 0.003), driven by lower AC inference than direct pair (AB, BC) memory

performance. Interestingly, there was no significant main effect of learning condition

($F_{1,50} = 0.43$, p = 0.518) nor a test trial type × learning condition interaction ($F_{2,50} = 0.96$, p

= 0.390). Because there were no behavioral differences across learning conditions, I note

that any neural differences would not be attributable to differences in performance.

Importantly, encoding order did not significantly affect behavior (main effect of order

and two- and three-way interactions; all F < 1.91, all p > 0.180).
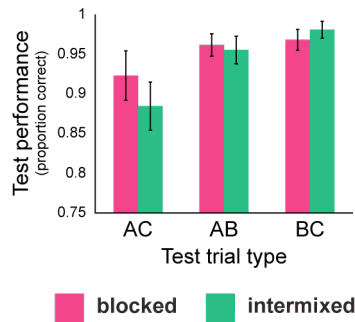
Figure 32. Experiment 3: Behavioral Results

Behavioral performance as a function of test trial and learning condition. Left bar pair, AC inference performance for blocked (pink; 33.3-100%, 92.3 ± 3.1%) and intermixed (green; 50-100%, 88.5 ± 3.0%) triads. Middle bar pair, AB performance (blocked: 83.3-100%, 96.2 ± 1.4%; intermixed: 66.7-100%, 95.5 ± 1.7%). Right bar pair, BC performance (blocked: 83.3-100%, 96.8 ± 1.3%; intermixed: 83.3-100%, 98.1 ± 1.1%). Bar heights represent group means; error bars denote SEM.

7.4.2. Imaging Results

Hippocampal Pattern Similarity Searchlight Results

I first interrogated anatomical hippocampus for voxel clusters showing learning-

related representational changes for indirectly related A and C items. I observed a large

cluster in right posterior hippocampus (**Fig. 33A**, blue cluster; MNI template coordinates, rounded to the nearest millimeter [mm]: x, y, z = 35, -37, -6) that showed a significant main effect of separation, consistent with the proposed role of hippocampus in maintaining orthogonalized representations for overlapping events (McClelland et al., 1995). Follow-up analyses demonstrated significant separation (i.e., larger decreases in neural pattern similarity for within- than across-triad) in this cluster for both blocked and intermixed triads when considered separately (both p < 0.005). In contrast, two regions in anterior hippocampus (left: -18, -16, -23; right: 30, -11, -28) showed a significant AB strength → integration interaction (**Fig. 33A**, green clusters). In the left cluster, this effect was driven by a significant separation effect for intermixed triads (p < 0.0001), while the integration effect for blocked triads was at trend level (p = 0.081). Effects of both integration for blocked triads and separation for intermixed triads were significant in the right hippocampal cluster (both p < 0.047). No regions showed a significant main effect of integration or an AB recency → integration interaction. Encoding order did not significantly impact neural similarity measures in any of the three hippocampal clusters (Bonferroni-corrected α threshold for significance < 0.005; all t_{24} < 2.06, all p > 0.051).
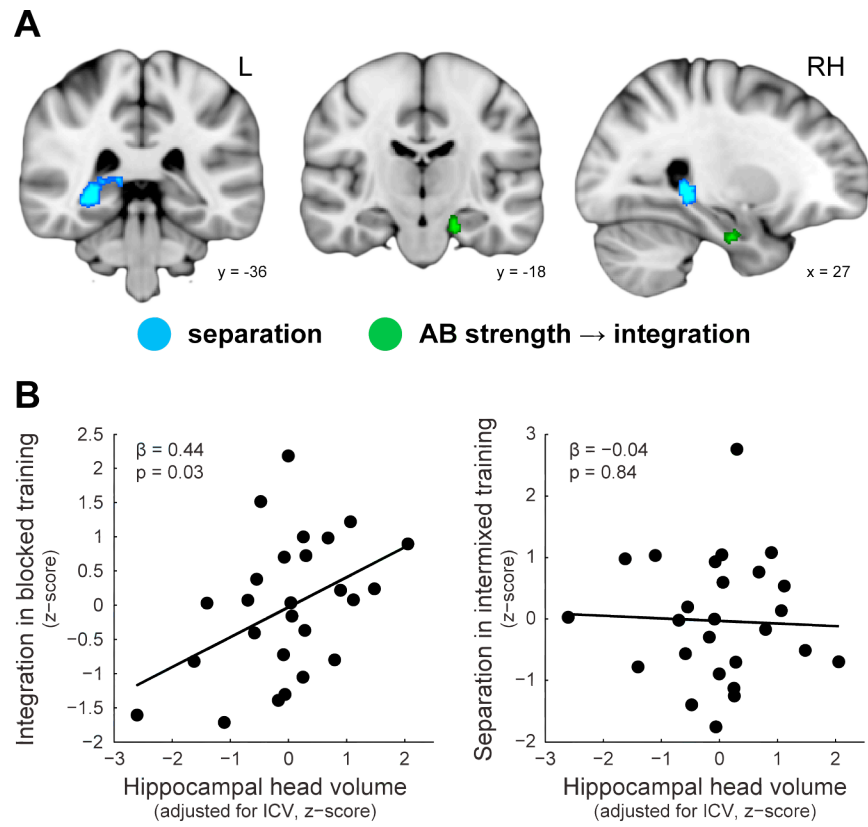
Figure 33. Experiment 3: Hippocampal Pattern Similarity Searchlight Results

(A) Hippocampal regions showing a significant main effect of separation (blue) or an AB strength → integration interaction (green). Clusters are significant after correction for multiple comparisons within anatomical hippocampus. Coordinates are in millimeters. (B) Across-participant relationship between hippocampal head volume (x-axes) and evidence for integration in the blocked learning condition (y-axis, left scatterplot) and separation in the intermixed learning condition (y-axis, right scatterplot). Best-fit lines and statistics were calculated using robust regression.

Hippocampal Volume-Δ Similarity Results

I next assessed whether the degree to which Δ neural pattern similarities reflected integration related to hippocampal structural measures. I predicted that more neural evidence for integration—particularly in the blocked learning condition—would relate to across participants to volume of the hippocampal head. This region shares the strongest

anatomical connections with MPFC (Barbas and Blatt, 1995) and has previously been implicated in forming generalized representations that span events (Nieuwenhuis and Takashima, 2011; Roy et al., 2012; Ghosh et al., 2014; Warren et al., 2014), making it a good candidate region for explaining variability in the coding strategies employed across individuals.

Hippocampal head volumes were positively related to the degree to which evidence for integration was observed in the blocked learning condition ($\beta = 0.44$, $p = 0.032$; **Fig. 33B**, left), such that larger hippocampal heads were associated with more evidence for integration. There was no such relationship for separation in the intermixed condition ($\beta = -0.04$, $p = 0.839$; **Fig. 33B**, right). Neither hippocampal body nor tail volumes were significantly related to the degree of integration in the blocked condition (both $|\beta| < 0.16$, both $p > 0.466$) or the degree of separation in the intermixed condition (both $|\beta| < 0.21$, both $p > 0.343$).

To further assess the specificity of this effect to the hippocampal head, I next performed two robust multiple regression analyses. When hippocampal head, body and tail volumes were simultaneously considered as independent variables, only hippocampal head volume was a significant predictor of integration in the blocked condition ($\beta = 0.51$, $p = 0.035$; for body and tail volumes, both $|\beta| < 0.24$, $p > 0.293$). These findings highlight the relatively greater contribution of the hippocampal head in integration. There was no relationship between volume of any subregion volume and the degree of separation in the intermixed condition (all $|\beta| < 0.07$, all $p > 0.776$) using robust multiple regression.

Encoding order did not significantly impact the relationships between hippocampal head volumes and integration in the blocked condition (main effect and interaction; both $F_{1,22} < 1.80$, both $p > 0.194$) or separation in the intermixed condition (both $F_{1,22} < 0.49$, both $p > 0.490$). This was also true for hippocampal body (main effects and interactions; all $F_{1,22} < 2.97$, all $p > 0.099$) and tail (main effects and interactions; all $F_{1,22} < 2.52$, all $p > 0.127$) volumes.

Prefrontal Pattern Similarity Searchlight Results

Within MPFC, neural pattern similarity searchlights revealed significant clusters for three of the four possible effects—main effects of integration, separation, and an AB strength → integration interaction, highlighting the functional hetereogeneity of this region. A cluster in anterior MPFC exhibited a main effect of separation (**Fig. 34**, blue; -3, 58, -26). Follow up analyses revealed significant separation effects in this cluster for both blocked and intermixed learning conditions (both $p < 0.03$). A slightly more posterior cluster demonstrated a main effect of integration (**Fig. 34**, orange; -6, 37, -19), with significant effects present in both learning conditions when considered separately (both $p < 0.002$). In the most posterior aspects of MPFC (i.e., subgenual cortex), two clusters demonstrated an AB strength → integration interaction with learning condition (**Fig. 34**, green; 13, 13, -21; -8, 10, -26), with integration for blocked and separation for intermixed triads. In both regions, effects of integration and separation were significant for blocked and intermixed learning conditions, respectively (all $p < 0.033$). No regions showed an AB recency → integration interaction. Learning-related Δ neural pattern

similarities were not significantly related to encoding order in any cluster (Bonferroni-corrected α threshold for significance < 0.005; all $|t_{24}| < 2.63$, all $p > 0.014$).

Figure 34. Experiment 3: MPFC Pattern Similarity Searchlight Results

MPFC regions showing a significant main effect of integration (orange), a main effect of separation (blue), or an interaction with learning condition (green). Clusters are significant after correction for multiple comparisons within anatomical MPFC. Coordinates are in millimeters.

The searchlight restricted to IFG revealed clusters showing main effects of separation and an AB strength → integration interaction with learning condition. Main effects of separation (**Fig. 35**, blue) were observed in an anterior left region (pars triangularis subregion; -49, 29, 5) and a more posterior right region (pars opercularis subregions; 50, 8, -1). Significant separation was observed in both blocked and intermixed learning conditions for both regions (all $p < 0.020$). Neighboring regions in both the left and right hemispheres showed a significant interaction with learning condition (**Fig. 35**, green; left: -27, 29, -4, right: 38, 13, 1; in the right hemisphere, this cluster extended into insula). In both clusters, integration effects were significant for blocked and separation effects significant for intermixed triad types (all $p < 0.039$). No

regions showed a main effect of integration or an AB recency → integration interaction. Encoding order did not significant modulate Δ neural pattern similarities in any region (Bonferroni-corrected α threshold for significance < 0.005; all $|t_{24}| < 2.14$, all $p > 0.043$).
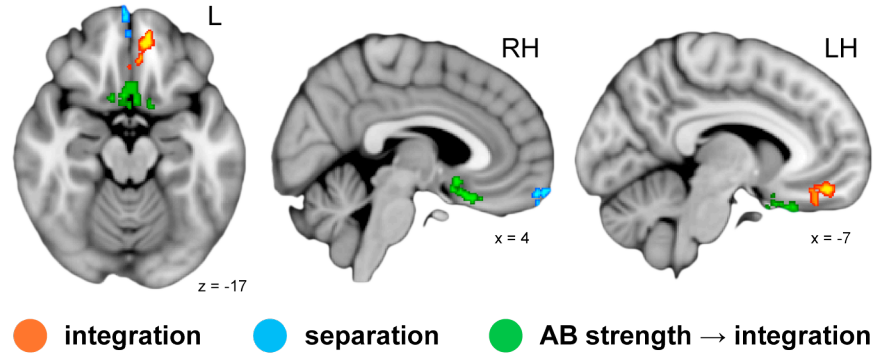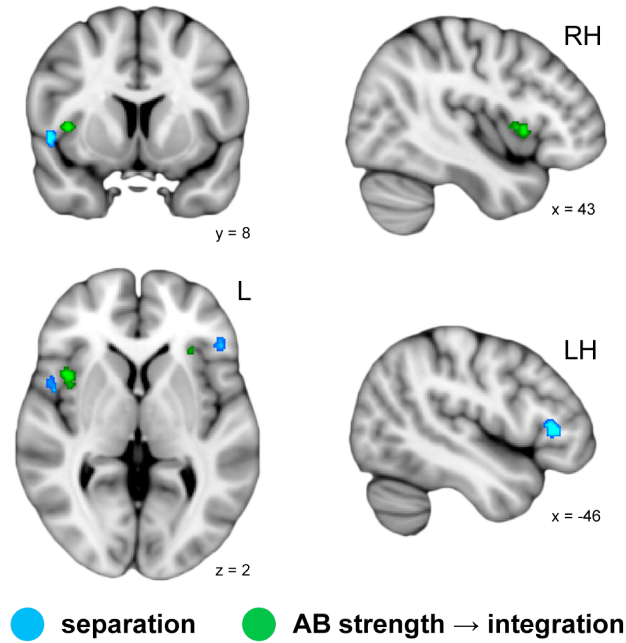

Figure 35. Experiment 3: IFG Pattern Similarity Searchlight Results

IFG regions showing a significant main effect of separation (blue) or an AB strength → integration with learning condition (green). Clusters are significant after correction for multiple comparisons within anatomical IFG. Coordinates are in millimeters.

Whole Brain Pattern Similarity Searchlight Results

Neural pattern similarity searchlights unrestricted to any particular brain region revealed a number of clusters showing significant effects of integration, separation, and both interactions with learning condition (**Table 1**). Just two regions showed a main effect of integration, both in prefrontal cortex. In contrast, numerous regions showed a

main effect of separation, including ventral occipito-temporal areas, temporal pole, and insula. Still others showed interactions with learning condition. Regions including midbrain, PFC and higher order visual regions showed evidence for integration in the blocked condition and separation for intermixed. In contrast, precuneus, middle temporal gyrus, and dorsal PFC showed the opposite pattern.

7.5. DISCUSSION

The present study combined hr-fMRI with neural pattern similarity analysis to investigate the neural representations underlying inference. I showed that the brain simultaneously maintains both integrated and separated representations of overlapping events, providing empirical evidence for the assumptions underlying much memory theory (Bartlett, 1932; Marr, 1970; McClelland et al., 1995; O'Reilly and Rudy, 2001). These distinct representational coding schemes were dissociable across the brain and within subregions of hippocampus, MPFC and IFG, highlighting the functional heterogeneity of these structures. Specifically, within my *a priori* defined ROIs, I showed that posterior hippocampus, bilateral IFG, and anterior MPFC evidenced separation of overlapping events, with A and C items becoming less similar to one another after learning. In contrast, mid-MPFC demonstrated integration of A and C items. In other regions—anterior hippocampus, posterior MPFC, and more medial aspects of bilateral IFG—these representational changes were modulated by the manner in which the overlapping events occurred, with blocked learning yielding integration and intermixed learning yielding separation, consistent with the AB strength → integration prediction.

Region	Laterality	N vox[a]	t_{25}	X^b	Y^b	Z^b
Integration						
Middle frontal gyrus	R	79	3.06	26	5	47
Frontal pole	L	72	2.99	-15	63	6
Separation						
Fusiform/lingual gyrus	B	484	2.97	23	-90	-24
Inferior temporal/fusiform gyrus	R	342	3.33	47	-50	-30
Temporal pole	R	320	2.4	25	10	-48
Pons	B	320	3.84	-11	-37	-45
Hippocampus, posterior	R	261	2.97	25	-30	-7
Midbrain/cerebellum	R	261	3.2	23	-31	-33
Temporal pole/planum porale	R	240	3.33	49	6	-13
Insula/pars opercularis	L	223	3.2	-34	-11	-7
Lateral occipital cortex	R	204	3.37	50	-71	3
Supramarginal gyrus (IPL)	L	161	3.44	-54	-35	33
Precentral gyrus	B	132	2.9	1	-31	52
Midbrain/thalamus	L	123	3.44	-5	-35	6
Superior temporal gyrus	L	111	2.63	-61	-19	-1
Precuneus	L	106	3.63	-5	-54	37
Midbrain/parahippocampal cortex	L	96	3	-10	-25	-19
Nucleus accumbens	B	90	4.34	4	3	-6
Lingual gyrus	L	87	3.92	-17	-44	-7
Intracalcarine sulcus	R	76	4.11	16	-73	3
Superior frontal gyrus	R	69	3.37	8	20	66
Medial frontal pole	R	69	2.77	4	67	-18
Anterior cingulate gyrus	R	67	3.55	13	34	18
AB strength → Integration (blocked > intermixed × integration > separation)						
Midbrain	R	328	4	6	-21	-23
Lingual gyrus	R	229	3.33	16	-50	-13
Insula	L	205	3.33	-44	-13	-4
Subgenual cortex	B	181	2.42	-6	20	-26
Planum temporale	L	124	2.88	-41	-38	16
Middle frontal gyrus	L	98	3.45	-34	32	35
Putamen	R	77	3.38	25	12	-9
Lingual gyrus	B	70	3.58	-5	-76	-11
AB recency → Integration (intermixed > blocked × integration > separation)						
Pre/Postcentral gyrus	R	92	3.13	40	-23	67
Precuneus	R	83	3.22	9	-56	16
Superior frontal gyrus	L	72	3.01	-23	46	37
Precuneus	R	72	3.77	9	-61	37
Middle temporal gyrus	R	70	2.99	56	-2	-36

[a]1.7 mm isotropic voxels
[b]MNI coordinates, rounded to the nearest mm

Table 1. Experiment 3: Whole Brain NPSA Searchlight Results

The dissociation of integration and separation signatures along the hippocampal axis converges with previous research. Specifically, the fact that posterior hippocampus shows separation across both learning conditions is consistent with existing views on the role of this subregion in individuating events from other, often highly similar, events memory (Komorowski et al., 2009; Brown et al., 2010; Liang et al., 2012; Poppenk et al., 2013; Preston and Eichenbaum, 2013). For instance, recent empirical work in humans (Hsieh et al., 2014) demonstrated lower neural pattern similarity in posterior hippocampus between items in overlapping sequences, suggesting that this region disambiguates overlapping events. The present study builds upon the existing literature to show that indirectly related memory elements are coded as distinct in this region.

In contrast, anterior hippocampus (particularly on the right) showed evidence for integration in the blocked learning condition. These results add to a body of work suggesting the role of this region in forming generalized representations (Komorowski et al., 2009; Liang et al., 2012; Poppenk et al., 2013; Preston and Eichenbaum, 2013), processing relational information (Schacter and Wagner, 1999; Davachi et al., 2003; Jackson and Schacter, 2004; Kirwan and Stark, 2004; Chua et al., 2007), and combining across episodes (Preston et al., 2004; Addis et al., 2007, 2011; Addis and Schacter, 2008; Barron et al., 2013). Interestingly, I also found a positive relationship between neural evidence for integration in the blocked condition and anterior hippocampal (head) volumes, such that individuals with larger hippocampal heads exhibited more integration. This result may appear at odds with prior work, which has demonstrated superior memory ability for individuals with smaller hippocampal heads (Maguire et al., 2000,

2006; Poppenk and Moscovitch, 2011; Demaster et al., 2013). One possible explanation that might reconcile these findings is that a tendency to integrate is in fact harmful for performance on standard memory tasks. Specifically, if a memory task requires retrieval of episodic details, it may be more advantageous to separate rather than integrate, as integrated representations may code the commonalities across experiences while losing the specifics. Thus, one prediction that stems from this finding is that larger hippocampal heads (i.e., more integration) would be advantageous in the context of a task tapping generalizations across episodes, while smaller hippocampal heads would be more beneficial when a pattern separation scheme—supporting retrieval of specific episodic details—is required.

The specificity of both the integration signature itself and its relationship to hippocampal head volume to the blocked learning condition is consistent with schema theory and modeling frameworks (O'Reilly and Rudy, 2001), which suggest that more robust pattern completion of strong prior memories will lead to a greater likelihood of overlapping representations. More broadly, these findings corroborate a host of work suggesting that strong initial memories can serve as a foundation into which new information is encoded (Bartlett, 1932; van Kesteren et al., 2010a, 2013a, 2013b, 2014; Schlichting and Preston, 2014; van Buuren et al., 2014) and further suggest that such mechanisms might depend specifically on anterior hippocampus, perhaps in its interactions with MPFC.

MPFC is often treated in the fMRI literature as a homogeneous region; yet, it is becoming increasingly clear that this region, like the hippocampus, comprises functional

subregions that differ in important ways—perhaps also along the anterior-posterior axis. In the present study, anterior aspects of MPFC (rostromedial PFC) exhibited separation of overlapping memory representations. Taken in the context of my prior work (Schlichting et al., in preparation; **Chapter 6**) this result suggests a more general mnemonic function of anterior MPFC, perhaps in processing memories irrespective of their relationships to one another. In contrast, I observed evidence for integration in posterior MPFC, converging with previous studies (Schlichting and Preston, under review; Nieuwenhuis and Takashima, 2011; Roy et al., 2012; Ghosh and Gilboa, 2013; Ghosh et al., 2014; Warren et al., 2014; **Chapter 6**). In mid-MPFC, this was true irrespective of learning condition. In contrast, the most posterior (i.e., subgenual) portion of MPFC exhibited this pattern for blocked but not intermixed learning conditions, mirroring the results in anterior hippocampus. While the neural measures reported here are not based on an index of functional connectivity per se, one possible interpretation of this parallel is that anterior hippocampus and subgenual MPFC work in concert to form and store integrated representations. This interpretation is consistent with prior theoretical (Nieuwenhuis and Takashima, 2011; Preston and Eichenbaum, 2013) and empirical research (Schlichting and Preston, under review), including work demonstrating the direct anatomical connection between subgenual MPFC and anterior hippocampus (Barbas and Blatt, 1995).

Within IFG, I observed regions bilaterally that showed separation and an AB strength → integration interaction. In the left hemisphere, these regions were both in mid-IFG, corresponding roughly to pars triangularis/Brodmann area (BA) 45 (Badre and

Wagner, 2007; Levy and Wagner, 2011). This region has been widely implicated in resolving competition among similar alternatives and resolving interference (Paller and Wagner, 2002; Badre and Wagner, 2007; Öztekin et al., 2009; Jenkins and Ranganath, 2010; Ranganath, 2010; Levy and Wagner, 2011; Kuhl et al., 2012a), consistent with a pattern separation scheme. I also observed a similar pattern in the right hemisphere in slightly more posterior regions (posterior-IFG or pars opercularis/BA 44; Badre and Wagner, 2007; Levy and Wagner, 2011). Although right IFG has most often been implicated in inhibiting motor responses (Aron et al., 2014), some evidence suggests that right posterior-IFG is also sensitive to relational integration and interference resolution demands (Cho et al., 2010). Thus, one possible interpretation of the present results is that both right and left IFG maintain separate representations to resolve interference among related memories; however, how this function relates to the more widely demonstrated roles of right IFG in particular (Aron et al., 2014) remains to be studied. I also demonstrated that under some conditions—in particular, in the context of strong AB memories—IFG may perform integration. These data suggest that for prefrontal regions, as in hippocampus, the specifics of the learning experience can influence neural representation.

At the whole brain level, a number of regions exhibited each of the four potential effects (i.e., integration, separation, AB strength $\rightarrow$ integration and AB recency $\rightarrow$ integration), highlighting how complementary neural representations might be stored in different brain regions to simultaneously accomplish both specificity and generalizability of memories. I observed a widespread network of regions across the brain that showed

evidence for separation; in contrast, just two regions exhibited integration irrespective of learning conditions, both in PFC. While this may appear to contradict the CLS framework, which proposes integration in neocortical sites, one possibility is that pattern separation is generally preferred for the neocortex as well as the hippocampus on short timescales like that of the present experiment. The CLS framework suggests that the cortex learns slowly over many experiences; thus, it may take more experience with overlapping associations or more time for such an integration bias to emerge in cortex, as memory traces become strengthened and consolidated.

Interestingly, midbrain showed both pattern separation across conditions and an AB strength → integration interaction. Prior work has also implicated the midbrain in encoding overlapping associations (Shohamy and Wagner, 2008). In one study (Shohamy and Wagner, 2008), researchers observed that greater test-phase activation in both hippocampus and midbrain was related to superior ability to generalize across overlapping experiences. Moreover, changes in learning-phase engagement of these structures were coupled, consistent with the idea that dopaminergic midbrain inputs modulate hippocampal mechanisms (Lisman and Grace, 2005; Shohamy and Wagner, 2008). In particular, such inputs may mediate the switch between hippocampal encoding and retrieval (Hasselmo and Schnell, 1994; Hasselmo et al., 1995; Hasselmo and McClelland, 1999) by detecting deviations in the environment from memory-based expectations (Strange and Dolan, 2001; Yamaguchi et al., 2004; Köhler et al., 2005; Lisman and Grace, 2005; Duncan et al., 2012). In the present study, midbrain signatures may similarly reflect a mismatch response regulating the switch between pattern

221

separation (encoding) and pattern completion (retrieval). Interestingly, midbrain exhibited the same pattern similarity signatures as did hippocampus, consistent with the notion that hippocampal processes are modulated by midbrain during the encoding of overlapping events.

Recent work has also used a pattern similarity approach to investigate learning-related changes in neural representations following overlapping encoding (Milivojevic et al., 2015). The authors of that study interpreted their data as suggesting integration of related and separation of unrelated events (in their case, scenes that either did or did not fit into a narrative) in MPFC and posterior hippocampus. In contrast, they concluded that anterior hippocampus exhibited pattern separation on the basis of its univariate response. These conclusions may be seen as contradicting those put forth by the present study. However, I argue that there are important differences in the paradigms employed that may explain these apparent discrepancies and suggesting caution in accepting the authors' interpretation of the prior findings.

First, their study (Milivojevic et al., 2015) did not allow for comparisons between scenes that were part of the same versus different narratives, as it was confounded with temporal proximity in their experimental design. For this reason, it is impossible to determine whether the reported effects reflect integration of specific scenes within a narrative or, rather, a common process engaged for only the scenes that were a part of any narrative. For instance, an equally plausible explanation of their data is that the increased neural similarity for related scenes reflects a retrieval process—for instance, recalling verbal labels used to describe the narratives. Such a process would not be engaged during

the unrelated scenes, which may indeed elicit a novelty response rather than retrieval. This would explain both their anterior hippocampal findings and the observed neural similarity decreases between related and unrelated scenes. Second, I note that their task explicitly encouraged an integration strategy, as participants were told to determine the relationships among the scenes in a narrative; this aspect of their design coupled with the slow trial structure would allow for—or even encourage—explicit retrieval strategies like the one described above.

I believe the present study addresses these shortcomings to allow for a cleaner interpretation in the context of existing theory. First, I quantified integration and separation as the learning-related increases or decreases in similarity for items from the same *relative to those from different triads*. This rules out a process account in the present data. Second, fMRI scanning took place while participants viewed brief presentations of individual items (e.g., item A_1) rather than multi-item events (e.g., $A_1B_1^-$ in the present study, analogous to the scenes above). Importantly, this allowed for estimation of item-specific neural patterns. Moreover, the present task design discouraged the use of intentional strategies, both during the learning phase and during the pre- and post-study exposure scans. During learning, participants were unaware that they would later need to link indirectly related A and C items. This design choice was made to decrease the likelihood that participants would attempt to integrate in preparation for an upcoming test. Moreover, as the neural data were acquired while participants viewed rapid presentations of individual items and made an unrelated visual decision, my neural representations are unlikely to reflect the active retrieval or suppression of related items.

223

Rather, I argue that the changes reported here reflect shifts in the neural representations of the items themselves, which is a claim that cannot be made by previous reports (Milivojevic et al., 2015).

Here, I provide novel support for the notion that the same memory elements may be represented in strikingly different manner across brain regions. These results provide an empirical account for the intuitive notion that memory representations do not come in a single form; rather, a given experience may have multiple representations, each advantageous in a different scenario. Moreover, consistent with computational frameworks, the specific type of learning experience influences representations in hippocampus and PFC. I demonstrate dissociations across subregions of hippocampus, MPFC, and IFG, underscoring the importance of considering the functional heterogeneity of these regions in human neuroimaging research.

These findings suggest a theoretical framework in which posterior hippocampus supports separation of related events in connection with lateral and rostromedial PFC, while anterior hippocampus and subgenual MFPC integrate across experiences, especially when initial memories are strong. In contrast, experiencing related events in close temporal proximity may lend itself to a pattern separation scheme throughout this network, as weaker overlapping memories are especially prone to interference. Moreover, the results presented here provide a promising avenue for future computational research, which may incorporate neural pattern similarity analyses into formal tests of model frameworks.

CONCLUSIONS[4]

New learning events are not isolated from the past. Encoding in the real world occurs on the backbone of existing knowledge, with new learning events triggering retrieval of prior episodes. Retrieved memories then impact how the new information is encoded into memory. Several influential memory theories—cognitive map (Tolman, 1948; O'Keefe and Nadel, 1978), relational memory (Cohen and Eichenbaum, 1993), and schema (Bartlett, 1932) theories—propose that new events are encoded in the context of existing knowledge, resulting in integrated memory networks that relate experiences. As influential as these theories are, neurobiological research has almost universally ignored this critical aspect of memory. Instead, the vast majority of studies have focused on functions of MTL and PFC that link information presented within single episodes. This dissertation aims to provide an initial understanding of how MTL and PFC interact to integrate information acquired across distinct events; and how such derived knowledge can be exploited to support novel behaviors.

The field's growing interest in understanding these complex, real-world aspects of episodic memory has been realized thanks to the introduction of elegant behavioral paradigms and advanced analysis methods for neural data (Hupbach et al., 2007; Tse et al., 2007, 2011; Zeithamova and Preston, 2010; Kuhl et al., 2011, 2012a; Zeithamova et al., 2012a; Gershman et al., 2013; McKenzie et al., 2014). This dissertation incorporates hr-fMRI and pattern information analyses (Polyn et al., 2005; Kriegeskorte et al., 2008a)

---

[4] Portions of this section were adapted from Schlichting ML, Preston AR (2015) Memory integration: neural mechanisms and implications for behavior. Curr Opin Behav Sci 1:1–8. AR Preston contributed to the writing of this review paper.

to precisely characterize the neural systems, computations, and coding strategies that underlie memory integration and inference. This research provides novel insight into how MTL regions interact with medial and lateral PFC to promote the formation of memory traces allowing for flexible memory behaviors. In particular, focusing on MTL at the level of subregions using hr-fMRI bridges the gap between animal and human research, allowing for the formation of novel, testable predictions that are relatable across species. Taken together, these studies greatly inform how the brain uses prior experience to make sense of new events, a topic of interest to much memory theory and research.

This dissertation tested key but essentially unstudied predictions that memory integration both facilitates new learning and enables novel inference about related events by allowing memory to extend beyond direct experience. Broadly speaking, the experiments detailed here provide converging support for the idea that memories may become integrated when new learning experiences share content (e.g., a person, place, or thing) with an existing memory trace. During the new experience, hippocampal pattern completion reactivates the previously stored, overlapping memory (Zeithamova et al., 2012b; Preston and Eichenbaum, 2013). Experiment 2A (Schlichting and Preston, 2014) showed evidence for such reactivation, with engagement of content-specific visual regions predicting later memory for new associations that were related to prior knowledge. This finding adds to a body of work evidencing neural reactivation of prior memories during overlapping learning experiences (Kuhl et al., 2012a; Zeithamova et al., 2012a; Gershman et al., 2013). Moreover, I built upon the existing literature to demonstrate that both the degree of neural reactivation during learning and behavioral

measures of integration are mediated by offline processes engaged prior to encoding (Schlichting and Preston, 2014), consistent with the notion that stronger initial memories lead to better integration due to facilitated reactivation (O'Reilly and Rudy, 2001).

With the related content reactivated in the brain, hippocampal area CA_1 is thought to compare prior memories with incoming information from the environment. CA_1 may signal the presence of novelty (i.e., when new experiences violate memory-based predictions) and facilitate new encoding by increasing the plasticity of neighboring CA_3 neurons (Larkin et al., 2014). In Experiment 1, I showed that activation in human CA_1 during the encoding of events that overlap with prior experiences relates to a behavioral measure of memory integration (Schlichting et al., 2014), consistent with the idea that CA_1 triggers integration. As demonstrated in animal work (McKenzie et al., 2013, 2014), the resulting integrated memories are highly structured, with shared elements coded similarly across experiences. Such a coding scheme could then be exploited to extract commonalities across episodes and support a host of behaviors. Experiment 3 provided a novel demonstration of this phenomenon in the human brain, isolating evidence for such integrated codes to anterior hippocampus. This finding builds upon prior work (Poppenk et al., 2013) that implicates this subregion in coding commonalities across experiences. Moreover, I found that whether such codes are recruited depends on the specifics of the learning experience, converging with Experiment 2A in suggesting that stronger initial memories promote integration.

MPFC may influence memory integration by biasing reactivation toward behaviorally relevant memories (Kroes and Fernández, 2012; van Kesteren et al., 2012;

Preston and Eichenbaum, 2013). Across a number of domains, MPFC is thought to represent mental models that guide behavior (Roy et al., 2012; Wilson et al., 2014). While its specific role in memory is only starting to be uncovered, some suggest that MPFC forms mental models based on mnemonic content (i.e., memory models) (Schacter et al., 2012; St. Jacques et al., 2013), which may include features such as behavioral relevance and appropriate response (Kroes and Fernández, 2012). These memory models may be activated when incoming information relates to existing knowledge, with MPFC selecting specific task-relevant memories for reactivation (Kroes and Fernández, 2012; van Kesteren et al., 2012; Wilson et al., 2014). Following integration in hippocampus, memory models may be updated with new content as needed through direct hippocampal inputs to MPFC (van Kesteren et al., 2012). Converging with these ideas, Experiment 2B showed enhanced posterior MPFC engagement and connectivity with hippocampus that predicted memory integration. I found similar results when interrogating structural connectivity, suggesting that individual differences in hippocampal-MPFC white matter integrity might additionally explain tendencies to integrate across participants. Through this process, MPFC may come to represent integrated memories that have been abstracted away from individual episodes (i.e., schema) over time (van Kesteren et al., 2012; Richards et al., 2014). As shown in Experiment 3, posterior MPFC indeed shows evidence of integration mirroring that of anterior hippocampus, particularly when initial memories were strong.

A number of studies suggest that memory integration persists into post-encoding rest (van Kesteren et al., 2010a) and sleep (Coutanche et al., 2013), with offline

228

consolidation processes facilitating generalization across episodes. Specifically, hippocampus-driven reactivation during slow-wave sleep is thought to transform memories, allowing connections to be formed among representations co-activated in neocortex (Lewis and Durrant, 2011). This process is thought to promote both the integration of new information into existing memories and abstraction across episodes in neocortical regions, particularly MPFC (Lewis and Durrant, 2011). In Experiment 2B, I extended these ideas to awake rest periods, showing that connectivity enhancements during learning of overlapping events persisted into post-encoding rest periods and related to memory integration behaviors. These results are consistent with the notion that offline periods are important for the strengthening of integrated representations.

While this dissertation focuses primarily on how memory integration supports performance on a single task, this phenomenon indeed has widespread implications for behavior and could support performance on a range of tasks. Below, I review recent work highlighting these benefits across a number of cognitive domains.

Spatial Navigation

Perhaps the most familiar and widely studied form of memory integration stems from Tolman's seminal work on cognitive maps (Tolman, 1948). Tolman proposed that navigation relies on the coherent representation of spatial layouts, which can flexibly give rise to new inferences about the relative locations of landmarks in the environment (Tolman, 1948). Recent work in humans has demonstrated a relationship between hippocampal volumes and the ability to infer novel spatial relationships among a set of trained landmarks (Schinazi et al., 2013), consistent with the idea that the hippocampus

constructs integrated spatial maps. A behavioral study further found sleep-related increases in spatial relational inference (Coutanche et al., 2013), indicating that early phase consolidation processes may facilitate the construction of cognitive maps.

Moreover, work in rodents demonstrates that the firing patterns of hippocampal CA_1 neurons predict animals' future routes (Pfeiffer and Foster, 2013). These trajectories can represent even novel paths (Gupta et al., 2010; Pfeiffer and Foster, 2013), suggesting that the hippocampus—perhaps guided by MPFC (de Bruin et al., 1994)—may support flexible navigation by simulating and evaluating possible trajectories in the context of current goals.

Decision Making

Integrated memories may also influence non-mnemonic decision making. For example, one recent fMRI study (Wimmer and Shohamy, 2012) suggests that the hippocampus supports the transfer of monetary value across related experiences through additional recruitment of reward regions. The researchers showed greater reactivation of prior related knowledge during encoding of new reward information for stimuli that showed more evidence of subsequent preference shifts, a behavioral index of value transfer. Hippocampal-striatal functional coupling was also associated with value-related preference changes (Wimmer and Shohamy, 2012), suggesting that hippocampus may interact with domain-specific regions (e.g., striatum in value learning tasks) in service of integration.

Consistent with a domain-general role for hippocampus in memory integration, rodent work (Blanquat et al., 2013) has found that the hippocampus was necessary for

updating a known goal location with new value information. These updated memories may then be transferred to neocortex, as MPFC was necessary for retaining the updated knowledge to support performance on the next day (Blanquat et al., 2013). Thus, integrated memories incorporating value information may be maintained as memory models in MPFC that will later bias behavior. I note that this role for MPFC is likely also domain-general given its documented involvement in a number of tasks lacking an explicit value component.

Schemas

Recent attention has focused on the behavioral benefits conferred by memory schemas. For instance, research in rodents has shown that prior knowledge of a spatial layout (i.e., a spatial schema) can both facilitate acquisition of new related memories and speed their consolidation (Tse et al., 2007, 2011). Echoing these results, a number of human studies have reported behavioral benefits in learning and memory when new information can be incorporated into an existing schema (Kumaran, 2013; van Kesteren et al., 2013a, 2014). Application of a schema to a new scenario has also been shown to recruit hippocampus (Kumaran et al., 2009; de Hoz and Martin, 2014). For example, one fMRI study (Kumaran et al., 2009) found that while engagement of and connectivity between hippocampus and ventral MPFC was enhanced during generation of a task schema, the application of schemas to guide behavior in a novel but similarly structured task selectively recruited hippocampus.

Rodent (Tse et al., 2011) and human (van Kesteren et al., 2010a, 2013a, 2014) work further suggests that MPFC may be activated along with hippocampus during

learning of schema-related information. Recent empirical data indicate that one factor that may influence the relative engagement of MTL and MPFC is the degree of consistency between new information and existing schema. Specifically, one study (van Kesteren et al., 2013a) demonstrated that MPFC engagement was more predictive of subsequent memory for information congruent with existing schemas, perhaps reflecting direct encoding of new content into prior knowledge. (The reader should note that this idea contrasts with standard views of consolidation, which propose that hippocampal memories are transferred to neocortex after long time periods; however, recent work suggests the possibility of neocortical encoding of new information independent of the hippocampus [Sharon et al., 2011; see however Smith et al., 2014; Warren and Duff, 2014]). In contrast, MTL engagement was more predictive of successful encoding of incongruent information.

One theory (van Kesteren et al., 2012) of schemas suggest that with increasing congruency, MPFC becomes increasingly able to bias reactivation toward related memories. Increasing congruency would also be associated with decreasing novelty, which may result in decreased reliance on hippocampal integration triggered by area CA_1. In such cases, MPFC memory models may guide reactivation and be updated directly, thus bypassing hippocampal involvement. In contrast, when an existing memory model is weak or nonexistent, MPFC would play no role in guiding memory retrieval. In this case, new content would be encoded by hippocampus. Across multiple related experiences (i.e., when forming a new schema), MPFC may come online (Zeithamova et al., 2012a), reflecting the emergence of guided reactivation and the abstraction across experiences.

However, in many cases, new events are likely to be neither entirely novel nor identical replications of prior experience. These events will instead share a moderate level of congruency with existing memory models, and would thus be expected to involve both MPFC and hippocampus.

Creativity and Imagination

Memory integration may also underlie the ability to recombine prior memories to construct new ideas and imagine future scenarios (Schacter et al., 2012). Consistent with this notion, recent work (Duff et al., 2013) has demonstrated that hippocampal damage results in impaired performance on creativity tasks in which participants generate novel responses on the basis of existing knowledge. MPFC may also support performance in such tasks; one recent fMRI study (Takeuchi et al., 2012) showed that individual differences in resting state functional connectivity of MPFC with posterior cingulate cortex predicted creativity.

Hippocampus and MPFC are also engaged during imagination (Martin et al., 2011; Barron et al., 2013), particularly when imagined scenarios are rich in episodic detail. One human fMRI study showed enhanced connectivity between hippocampus and MPFC during imagination of future scenarios that were later remembered (Martin et al., 2011), consistent with the notion that these regions are important for creating and maintaining integrated memories—even those representing imagined events. Another study (Barron et al., 2013) required participants to construct mental representations of novel foods from two familiar ingredients. Using an fMRI adaptation paradigm, researchers found that imagining novel foods engaged the same neuronal populations as

did the ingredients in both anterior hippocampus and MPFC, reflecting retrieval and recombination of prior memories during mental construction. The ingredient items themselves also came to recruit overlapping neuronal populations, perhaps reflecting integration of the simultaneously reactivated memories. Interestingly, the degree of representational overlap of the ingredients in hippocampus and MPFC tracked across participants with subjective value of the imagined foods, suggesting that integration may be enhanced according to behavioral relevance (here, for high value items).

The findings reviewed here collectively suggest the importance of a hippocampal-MPFC circuit for linking related experiences across a number of cognitive domains. Memory integration has broad implications for behavior, supporting abilities ranging from navigating our environment to imagining the future. While recent years have brought a surge of attention to this area of study, I believe this is just the beginning of a rich scientific enterprise. How do integrated memories give rise to semantic representations? How do memory integration and behavioral flexibility change across the lifespan (Brod et al., 2013), and how can these ideas be applied to promote learning in an educational setting? The research presented in this dissertation offers an empirical and conceptual framework for investigating these and other critical questions that remain the subject of future investigation.

REFERENCES

Acuna BD, Eliassen JC, Donoghue JP, Sanes JN (2002a) Frontal and parietal lobe activation during transitive inference in humans. Cereb Cortex 12:1312–1321.

Acuna BD, Sanes JN, Donoghue JP (2002b) Cognitive mechanisms of transitive inference. Exp Brain Res 146:1–10.

Addis DR, Cheng T, Roberts RP, Schacter DL (2011) Hippocampal contributions to the episodic simulation of specific and general future events. Hippocampus 21:1045–1052.

Addis DR, Pan L, Vu M-A, Laiser N, Schacter DL (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.

Addis DR, Schacter DL (2008) Constructive episodic simulation: temporal distance and detail of past and future events modulate hippocampal engagement. Hippocampus 237:227–237.

Addis DR, Schacter DL (2012) The hippocampus and imagining the future: where do we stand? Front Hum Neurosci 5:173.

Addis DR, Wong AT, Schacter DL (2007) Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45:1363–1377.

Amaral DG, Insausti R (1990) Hippocampal formation. In: The Human Nervous System (Paxinos G, ed), pp 711–755. San Diego: Academic Press.

Amaral DG, Insausti R, Cowan WM (1987) The entorhinal cortex of the monkey: I. Cytoarchitectonic organization. J Comp Neurol 264:326–355.

Anderson MC, McCulloch KC (1999) Integration as a general boundary condition on retrieval-induced forgetting. J Exp Psychol Learn Mem Cogn 25:608–629.

Aron AR, Robbins TW, Poldrack RA (2014) Inhibition and the right inferior frontal cortex: One decade on. Trends Cogn Sci 18:177–185.

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54:2033–2044.

Badre D, Poldrack RA, Paré-Blagoev EJ, Insler RZ, Wagner AD (2005) Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. Neuron 47:907–918.

Badre D, Wagner AD (2005) Frontal lobe mechanisms that resolve proactive interference. Cereb Cortex 15:2003–2012.

Badre D, Wagner AD (2007) Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 45:2883–2901.

Bakker A, Kirwan CB, Miller M, Stark CEL (2008) Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus. Science 319:1640–1642.

Barbas H, Blatt GJ (1995) Topographically Specific Hippocampal Projections Target Functionally Distinct Prefrontal Areas in the Rhesus Monkey. Hippocampus 5:511–533.

Barrett LF (2006) Solving the emotion paradox: categorization and the experience of emotion. Pers Soc Psychol Rev 10:20–46.

Barron HC, Dolan RJ, Behrens TEJ (2013) Online evaluation of novel choices by simultaneous representation of multiple memories. Nat Neurosci 16:1492–1498.

Bartlett F (1932) Remembering: A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press.

Bayley PJ, Hopkins RO, Squire LR (2006) The fate of old memories after medial temporal lobe damage. J Neurosci 26:13311–13317.

Behrens TEJ, Berg HJ, Jbabdi S, Rushworth MFS, Woolrich MW (2007) Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34:144–155.

Bein O, Reggev N, Maril A (2014) Prior knowledge influences on hippocampus and medial prefrontal cortex interactions in subsequent memory. Neuropsychologia 64:320–330.

Blanquat PDS, Hok V, Save E, Poucet B, Chaillan FA (2013) Differential role of the dorsal hippocampus, ventro-intermediate hippocampus, and medial prefrontal cortex in updating the value of a spatial goal. Hippocampus 23:342–351.

Bontempi B, Laurent-Demir C, Destrade C, Jaffard R (1999) Time-dependent reorganization of brain circuitry underlying long-term memory storage. Nature 400:671–675.

Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS (2009) How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. Neuron 62:733–743.

Bransford JD, Johnson MK (1972) Contextual prerequisites for understanding: Some investigations of comprehension and recall. J Verbal Learning Verbal Behav 11:717–726.

Brod G, Werkle-Bergner M, Shing YL (2013) The Influence of Prior Knowledge on Memory: A Developmental Cognitive Neuroscience Perspective. Front Behav Neurosci 7:139.

Brown TI, Ross RS, Keller JB, Hasselmo ME, Stern CE (2010) Which way was I going? Contextual retrieval supports the disambiguation of well learned overlapping navigational routes. J Neurosci 30:7414–7422.

Buckmaster CA, Eichenbaum H, Amaral DG, Suzuki WA, Rapp PR (2004) Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. J Neurosci 24:9811–9825.

Buckner RL (2010) The role of the hippocampus in prediction and imagination. Annu Rev Psychol 61:27–48.

Bunge SA, Helskog EH, Wendelken C (2009) Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. Neuroimage 46:338–342.

Bunge SA, Wendelken C, Badre D, Wagner AD (2005) Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. Cereb Cortex 15:239–249.

Bunsey M, Eichenbaum H (1996) Conservation of hippocampal memory function in rats and humans. Nature 379:255–257.

Burgess PW, Shallice T (1996) Confabulation and the control of recollection. Memory 4:359–411.

Cacioppo JT, Gardner WL (1999) Emotion. Annu Rev Psychol 50:191–214.

Cacioppo JT, Uchino BN, Crites SL, Snydersmith MA, Smith G, Berntson GG, Lang PJ (1992) Relationship Between Facial Expressiveness and Sympathetic Activation in Emotion: A Critical Review, With Emphasis on Modeling Underlying Mechanisms and Individual Differences. J Pers Soc Psychol 62:110–128.

Cavada C, Compañy T, Tejedor J, Cruz-Rizzolo RJ, Reinoso-Suárez F (2000) The anatomical connections of the macaque monkey orbitofrontal cortex. A review. Cereb Cortex 10:220–242.

Chen J, Olsen RK, Preston AR, Glover GH, Wagner AD (2011) Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. Learn Mem 18:523–528.

Chib VS, Rangel A, Shimojo S, O'Doherty JP (2009) Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. J Neurosci 29:12315–12320.

Chiu C-YP, Schmithorst VJ, Brown RD, Holland SK, Dunn S (2006) Making memories: a cross-sectional investigation of episodic memory encoding in childhood using FMRI. Dev Neuropsychol 29:321–340.

Cho S, Moody TD, Fernandino L, Mumford JA, Poldrack RA, Cannon TD, Knowlton BJ, Holyoak KJ (2010) Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. Cereb Cortex 20:524–533.

Christoff K, Prabhakaran V, Dorfman J (2001) Rostrolateral prefrontal cortex involvement in relational integration during reasoning. Neuroimage 14:1136–1149.

Chua EF, Schacter DL, Rand-Giovannetti E, Sperling RA (2007) Evidence for a specific role of the anterior hippocampal region in successful associative encoding. Hippocampus 17:1071–1080.

Cohen NJ, Eichenbaum H (1993) Memory, amnesia, and the hippocampal system. Cambridge, MA: The MIT Press.

Coutanche MN, Gianessi CA, Chanales AJH, Willison KW, Thompson-Schill SL (2013) The role of sleep in forming a memory representation of a two-dimensional space. Hippocampus 23:1189–1197.

Coutureau E, Killcross AS, Good M, Marshall VJ, Ward-Robinson J, Honey RC (2002) Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. J Exp Psychol Anim Behav Process 28:388–396.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114.

Davachi L (2006) Item, context and relational episodic encoding in humans. Curr Opin Neurobiol 16:693–700.

Davachi L, Mitchell JP, Wagner AD (2003) Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. Proc Natl Acad Sci U S A 100:2157–2162.

De Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I (2005) Cognitive modulation of olfactory processing. Neuron 46:671–679.

De Bruin JPC, Sànchez-Santed F, Heinsbroek RPW, Donker A, Postmes P (1994) A behavioural analysis of rats with damage to the medial prefrontal cortex using the morris water maze: Evidence for behavioural flexibility, but not for impaired spatial navigation. Brain Res 652:323–333.

De Hoz L, Martin SJ (2014) Double dissociation between the contributions of the septal and temporal hippocampus to spatial learning: The role of prior experience. Hippocampus 24:990–1005.

Demaster DM, Ghetti S (2013) Developmental differences in hippocampal and cortical contributions to episodic retrieval. Cortex 49:1482–1493.

Demaster DM, Pathman T, Lee JK, Ghetti S (2013) Structural development of the hippocampus and episodic memory: developmental differences along the anterior/posterior axis. Cereb Cortex 24:3036–3045.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980.

Detre GJ, Natarajan A, Gershman SJ, Norman KA (2013) Moderate levels of activation lead to forgetting in the think / no-think paradigm. Neuropsychologia 51:2371–2388.

Deuker L, Olligs J, Fell J, Kranz TA, Mormann F, Montag C, Reuter M, Elger CE, Axmacher N (2013) Memory consolidation by replay of stimulus-specific neural activity. J Neurosci 33:19373–19383.

DeVito LM, Kanter BR, Eichenbaum H (2010a) The hippocampus contributes to memory expression during transitive inference in mice. Hippocampus 20:208–217.

DeVito LM, Lykken C, Kanter BR, Eichenbaum H (2010b) Prefrontal cortex: role in acquisition of overlapping associations and transitive inference. Learn Mem 17:161–167.

Diana RA, Yonelinas AP, Ranganath C (2007) Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn Sci 11:379–386.

Diekelmann S, Born J (2010) The memory function of sleep. Nat Rev Neurosci 11:114–126.

Douville K, Woodard JL, Seidenberg M, Miller SK, Leveroni CL, Nielson KA, Franczak M, Antuono P, Rao SM (2005) Medial temporal lobe activity for recognition of recent and remote famous names: An event-related fMRI study. Neuropsychologia 43:693–703.

Duff MC, Kurczek J, Rubin R, Cohen NJ, Tranel D (2013) Hippocampal amnesia disrupts creative thinking. Hippocampus 23:1143–1149.

Duncan K, Ketz N, Inati SJ, Davachi L (2012) Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. Hippocampus 22:389–398.

Dusek JA, Eichenbaum H (1997) The hippocampus and memory for orderly stimulus relations. Proc Natl Acad Sci U S A 94:7109–7114.

Duvernoy HM (1998) The Human Hippocampus Functional Anatomy, Vascularization and Serial Sections with MRI. New York: Springer.

Eichenbaum H (1999) The hippocampus and mechanisms of declarative memory. Behav Brain Res 103:123–133.

Eichenbaum H (2000) A cortical-hippocampal system for declarative memory. Nat Rev Neurosci 1:41–50.

Eichenbaum H, Cohen NJ (2001) From Conditioning to Conscious Recollection: Memory Systems of the Brain. New York, NY: Oxford University Press.

Eichenbaum H, Dudchenko PA, Wood E, Shapiro M, Tanila H (1999) The hippocampus, memory, and place cells: is it spatial memory or a memory space? Neuron 23:209–226.

Eichenbaum H, Otto T, Cohen NJ (1992) The hippocampus—what does it do? Behav Neural Biol 57:2–36.

Eichenbaum H, Schoenbaum G, Young B, Bunsey M (1996) Functional organization of the hippocampal memory system. Proc Natl Acad Sci U S A 93:13500–13507.

Eichenbaum H, Yonelinas AP, Ranganath C (2007) The medial temporal lobe and recognition memory. Annu Rev Neurosci 30:123–152.

Eldridge LL, Engel SA, Zeineh MM, Bookheimer SY, Knowlton BJ (2005) A dissociation of encoding and retrieval processes in the human hippocampus. J Neurosci 25:3280–3286.

Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP (2007) Human relational memory requires time and sleep. Proc Natl Acad Sci U S A 104:7723–7728.

Estes WK (1955) Statistical Theory of Spontaneous Recovery and Regression. Psychol Rev 62:145–154.

Etkin A, Egner T, Kalisch R (2011) Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn Sci 15:85–93.

Feldman HM, Lee ES, Yeatman JD, Yeom KW (2012) Language and reading skills in school-aged children and adolescents born preterm are associated with white matter properties on diffusion tensor imaging. Neuropsychologia 50:3348–3362.

Ferino F, Thierry AM, Glowinski J (1987) Anatomical and electrophysiological evidence for a direct projection from Ammon's horn to the medial prefrontal cortex in the rat. Exp Brain Res 65:421–426.

Fletcher PC, Henson RN (2001) Frontal lobes and human memory: insights from functional neuroimaging. Brain 124:849–881.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102:9673–9678.

Frank MJ, Rudy JW, Levy WB, O'Reilly RC (2005) When logic fails: implicit transitive inference in humans. Mem Cognit 33:742–750.

Frank MJ, Rudy JW, O'Reilly RC (2003) Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. Hippocampus 13:341–354.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229.

Gabrieli JDE, Brewer J, Desmond J, Glover GH (1997) Separate Neural Bases of Two Fundamental Memory Processes in the Human Medial Temporal Lobe. Science 276:264–266.

Gais S, Albouy G, Boly M, Dang-Vu TT, Darsaud A, Desseilles M, Rauchs G, Schabus M, Sterpenich V, Vandewalle G, Maquet P, Peigneux P (2007) Sleep transforms the cerebral trace of declarative memories. Proc Natl Acad Sci U S A 104:18778–18783.

Gerraty RT, Davidow JY, Wimmer GE, Kahn I, Shohamy D (2014) Transfer of learning relates to intrinsic connectivity between hippocampus, ventromedial prefrontal cortex, and large-scale networks. J Neurosci 34:11297–11303.

Gershman SJ, Schapiro AC, Hupbach A, Norman KA (2013) Neural context reinstatement predicts memory misattribution. J Neurosci 33:8590–8595.

Ghosh VE, Gilboa A (2013) What is a memory schema? A historical perspective on current neuroscience literature. Neuropsychologia:1–11.

Ghosh VE, Moscovitch M, Melo Colella B, Gilboa A (2014) Schema Representation in Patients with Ventromedial PFC Lesions. J Neurosci 34:12057–12070.

Gilboa A, Alain C, Stuss DT, Melo B, Miller S, Moscovitch M (2006) Mechanisms of spontaneous confabulations: a strategic retrieval account. Brain 129:1399–1414.

Giovanello KS, Schnyer DM, Verfaellie M (2009) Distinct hippocampal regions make unique contributions to relational memory. Hippocampus 19:111–117.

Glover GH, Law CS (2001) Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 46:515–522.

Goel V (2007) Anatomy of deductive reasoning. Trends Cogn Sci 11:435–441.

Goel V, Dolan RJ (2001) Functional neuroanatomy of three-term relational reasoning. Neuropsychologia 39:901–909.

Greene AJ, Gross WL, Elsinger CL, Rao SM (2006) An FMRI analysis of the human hippocampus: inference, context, and task awareness. J Cogn Neurosci 18:1156–1173.

Gupta AS, van der Meer MAA, Touretzky DS, Redish AD (2010) Hippocampal replay is not a simple function of experience. Neuron 65:695–705.

Haist F, Gore JB, Mao H (2001) Consolidation of human memory over decades revealed by functional magnetic resonance imaging. Nat Neurosci 4:1139–1145.

Hall G (1996) Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. Anim Learn Behav 24:233–255.

Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S (2009) PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7:37–53.

Hannula DE, Ranganath C (2008) Medial temporal lobe activity predicts successful relational memory binding. J Neurosci 28:116–124.

Hannula DE, Tranel D, Cohen NJ (2006) The long and the short of it: relational memory impairments in amnesia, even at short lags. J Neurosci 26:8352–8359.

Hare TA, Camerer CF, Rangel A (2009) Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. Science 324:646–648.

Hasselmo ME, McClelland JL (1999) Neural models of memory. Curr Opin Neurobiol 9:184–188.

Hasselmo ME, Schnell E (1994) Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: computational modeling and brain slice physiology. J Neurosci 14:3898–3914.

Hasselmo ME, Schnell E, Barkai E (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. J Neurosci 15:5249–5262.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science 293:2425–2430.

Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D (2004) Hippocampal activation during transitive inference in humans. Hippocampus 14:153–162.

Henke K, Reber TP, Duss SB (2013) Integrating events across levels of consciousness. Front Behav Neurosci 7:68.

Holland PC (1981) Acquisition of representation-mediated conditioned food aversions. Learn Motiv 12:1–18.

Hotelling H (1940) The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. Ann Math Stat 11:271–283.

Howard MW, Jing B, Rao VA, Provyn JP, Datey AV (2009) Bridging the Gap: Transitive Associations Between Items Presented in Similar Temporal Contexts. J Exp Psychol Learn Mem Cogn 35:391–407.

Howard MW, Kahana MJ (2002) A Distributed Representation of Temporal Context. J Math Psychol 46:269–299.

Hsieh L-T, Gruber MJ, Jenkins LJ, Ranganath C (2014) Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. Neuron 81:1165–1178.

Hsu NS, Schlichting ML, Thompson-Schill SL (2014) Feature Diagnosticity Affects Representations of Novel and Familiar Objects. J Cogn Neurosci 26:2735–2749.

Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PCM, Mori S (2008) Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. Neuroimage 39:336–347.

Hupbach A (2011) The specific outcomes of reactivation-induced memory changes depend on the degree of competition between old and new information. Front Behav Neurosci 5:33.

Hupbach A, Gomez R, Hardt O, Nadel L (2007) Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. Learn Mem 14:47–53.

Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkänen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. Am J Neuroradiol 19:659–671.

Iordanova MD, Good M, Honey RC (2011) Retrieval-mediated learning involving episodes requires synaptic plasticity in the hippocampus. J Neurosci 31:7156–7162.

Iordanova MD, Killcross AS, Honey RC (2007) Role of the medial prefrontal cortex in acquired distinctiveness and equivalence of cues. Behav Neurosci 121:1431–1436.

Jackson O, Schacter DL (2004) Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. Neuroimage 21:456–462.

Jadhav SP, Kemere C, German PW, Frank LM (2012) Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory. Science 336:1454–1457.

Jenkins LJ, Ranganath C (2010) Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. J Neurosci 30:15558–15565.

Jenkinson M (2003) Fast, automated, N-dimensional phase-unwrapping algorithm. Magn Reson Med 49:193–197.

Jones B, Bukoski E, Nadel L, Fellous J-M (2012) Remaking memories: Reconsolidation updates positively motivated spatial memory in rats. Learn Mem 19:91–98.

Karlsson MP, Frank LM (2009) Awake replay of remote experiences in the hippocampus. Nat Neurosci 12:913–918.

Kim JJ, Fanselow MS (1992) Modality-Specific Retrograde Amnesia of Fear. Science 256:675–677.

Kirwan CB, Stark CEL (2004) Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. Hippocampus 14:919–930.

Klein SB, Loftus J, Kihlstrom JF (2002) Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. Soc Cogn 20:353–379.

Köhler S, Danckert S, Gati JS, Menon RS (2005) Novelty responses to relational and non-relational information in the hippocampus and the parahippocampal region: A comparison based on event-related fMRI. Hippocampus 15:763–774.

Komorowski RW, Manns JR, Eichenbaum H (2009) Robust conjunctive item-place coding by hippocampal neurons parallels learning what happens where. J Neurosci 29:9918–9929.

Koscik TR, Tranel D (2012) The human ventromedial prefrontal cortex is critical for transitive inference. J Cogn Neurosci 24:1191–1204.

Kriegeskorte N, Mur M, Bandettini PA (2008a) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141.

Kroes MCW, Fernández G (2012) Dynamic neural systems enable adaptive, flexible memories. Neurosci Biobehav Rev 36:1646–1666.

Kuhl BA, Bainbridge WA, Chun MM (2012a) Neural reactivation reveals mechanisms for updating memory. J Neurosci 32:3453–3461.

Kuhl BA, Rissman J, Chun MM, Wagner AD (2011) Fidelity of neural reactivation reveals competition between memories. Proc Natl Acad Sci U S A 108:5903–5908.

Kuhl BA, Rissman J, Wagner AD (2012b) Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. Neuropsychologia 50:458–469.

Kuhl BA, Shah AT, DuBrow S, Wagner AD (2010) Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nat Neurosci 13:501–506.

Kumaran D (2012) What representations and computations underpin the contribution of the hippocampus to generalization and inference? Front Hum Neurosci 6:157.

Kumaran D (2013) Schema-driven facilitation of new hierarchy learning in the transitive inference paradigm. Learn Mem 20:388–394.

Kumaran D, McClelland JL (2012) Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. Psychol Rev 119:573–616.

Kumaran D, Melo HL (2013) Transitivity performance, relational hierarchy knowledge and awareness: Results of an instructional framing manipulation. Hippocampus 23:1259–1268.

Kumaran D, Summerfield JJ, Hassabis D, Maguire EA (2009) Tracking the emergence of conceptual knowledge during human decision making. Neuron 63:889–901.

Larkin MC, Lykken C, Tye LD, Wickelgren JG, Frank LM (2014) Hippocampal output area CA1 broadcasts a generalized novelty signal during an object-place recognition task. Hippocampus 24:773–783.

Lavenex P, Amaral DG (2000) Hippocampal-neocortical interaction: a hierarchy of associativity. Hippocampus 10:420–430.

Lee I, Rao G, Knierim JJ (2004) A double dissociation between hippocampal subfields: differential time course of CA3 and CA1 place cells for processing changed environments. Neuron 42:803–815.

Levy BJ, Wagner AD (2011) Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. Ann N Y Acad Sci 1224:40–62.

Lewis PA, Durrant SJ (2011) Overlapping memory replay during sleep builds cognitive schemata. Trends Cogn Sci 15:343–351.

Liang JC, Wagner AD, Preston AR (2012) Content representation in the human medial temporal lobe. Cereb Cortex 23:80–96.

Lisman JE, Grace AA (2005) The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46:703–713.

Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RSJ, Frith CD (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci U S A 97:4398–4403.

Maguire EA, Woollett K, Spiers HJ (2006) London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. Hippocampus 16:1091–1101.

Manns J, Eichenbaum H (2006) Evolution of declarative memory. Hippocampus 16:795–808.

Marr D (1970) A theory for cerebral neocortex. Proc R Soc Lond B Biol Sci 176:161–234.

Marr D (1971) Simple Memory: A Theory for Archicortex. Philos Trans R Soc London 262:23–81.

Martin VC, Schacter DL, Corballis MC, Addis DR (2011) A role for the hippocampus in encoding simulations of future events. Proc Natl Acad Sci U S A 108:13858–13863.

McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:419–457.

McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate Neural Systems Value Immediate and Delayed Monetary Rewards. Science 306:503–507.

McKenzie S, Eichenbaum H (2011) Consolidation and reconsolidation: two lives of memories? Neuron 71:224–233.

McKenzie S, Frank AJ, Kinsky NR, Porter B, Rivière PD, Eichenbaum H (2014) Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. Neuron 83:202–215.

McKenzie S, Robinson NTM, Herrera L, Churchill JC, Eichenbaum H (2013) Learning causes reorganization of neuronal firing patterns to represent related experiences within a hippocampal schema. J Neurosci 33:10243–10256.

McNaughton BL, Morris RGM (1987) Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends Neurosci 10:408–415.

Milivojevic B, Vicente-Grabovetsky A, Doeller CF (2015) Insight Reconfigures Hippocampal-Prefrontal Article Insight Reconfigures Hippocampal-Prefrontal Memories. Curr Biol:1–10.

Mitchell KJ, Johnson MK (2009) Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? Psychol Bull 135:638–677.

Mizumori SJ, McNaughton BL, Barnes CA, Fox KB (1989) Preserved spatial coding in hippocampal CA1 pyramidal cells during reversible suppression of CA3c output: evidence for pattern completion in hippocampus. J Neurosci 9:3915–3928.

Moser MB, Moser EI (1998) Functional differentiation in the hippocampus. Hippocampus 8:608–619.

Moses SN, Villate C, Ryan JD (2006) An investigation of learning strategy supporting transitive inference performance in humans compared to other species. Neuropsychologia 44:1370–1387.

Mumford JA, Davis T, Poldrack RA (2014) The impact of study design on pattern estimation for single-trial multivariate pattern analysis. Neuroimage 103:130–138.

Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643.

Munnelly A, Dymond S (2014) Relational memory generalization and integration in a transitive inference task with and without instructed awareness. Neurobiol Learn Mem 109:169–177.

Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, Ferris S, Golomb J, Schnirman G, Schwartz R (2003) Dissociating hippocampal versus basal ganglia contributions to learning and transfer. J Cogn Neurosci 15:185–193.

Nadel L, Hardt O (2011) Update on memory systems and processes. Neuropsychopharmacology 36:251–273.

Nadel L, Hupbach A, Gomez R, Newman-Smith K (2012) Memory formation, consolidation and transformation. Neurosci Biobehav Rev 36:1640–1645.

Nadel L, Samsonovich A, Ryan L, Moscovitch M (2000) Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. Hippocampus 10:352–368.

Nagode JC, Pardo JV (2002) Human hippocampal activation during transitive inference. Neuroreport 13:939–944.

Navawongse R, Eichenbaum H (2013) Distinct pathways for rule-based retrieval and spatial mapping of memory representations in hippocampal neurons. J Neurosci 33:1002–1013.

Nieuwenhuis ILC, Takashima A (2011) The role of the ventromedial prefrontal cortex in memory consolidation. Behav Brain Res 218:325–334.

Norman KA, O'Reilly RC (2003) Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev 110:611–646.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430.

O'Keefe J, Nadel L (1978) The Hippocampus as a Cognitive Map. London: Clarendon.

O'Reilly JX, Woolrich MW, Behrens TEJ, Smith SM, Johansen-Berg H (2012) Tools of the trade: psychophysiological interactions and functional connectivity. Soc Cogn Affect Neurosci 7:604–609.

O'Reilly RC, Norman KA (2002) Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. Trends Cogn Sci 6:505–510.

O'Reilly RC, Rudy JW (2001) Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. Psychol Rev 108:311–345.

Okuda J, Fujii T, Ohtake H, Tsukiura T, Tanji K, Suzuki K, Kawashima R, Fukuda H, Itoh M, Yamadori A (2003) Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. Neuroimage 19:1369–1380.

Ongür D, Ferry AT, Price JL (2003) Architectonic subdivision of the human orbital and medial prefrontal cortex. J Comp Neurol 460:425–449.

Ongür D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cereb Cortex 10:206–219.

Öztekin I, Curtis C, McElree B (2009) The medial temporal lobe and the left inferior prefrontal cortex jointly support interference resolution in verbal working memory. J Cogn Neurosci 21:1967–1979.

Paller KA, Wagner AD (2002) Observing the transformation of experience into memory. Trends Cogn Sci 6:93–102.

Pavlov I (1927) Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. London: Oxford University Press.

Pfeiffer BE, Foster DJ (2013) Hippocampal place-cell sequences depict future paths to remembered goals. Nature 497:74–81.

Pfeuffer J, Van de Moortele P-F, Ugurbil K, Hu X, Glover GH (2002) Correction of physiologically induced global off-resonance effects in dynamic echo-planar and spiral functional imaging. Magn Reson Med 47:344–353.

Plassmann H, O'Doherty J, Shiv B, Rangel A (2008) Marketing actions can modulate neural representations of experienced pleasantness. Proc Natl Acad Sci U S A 105:1050–1054.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-Specific Cortical Activity Precedes Retrieval During Memory Search. Science 310:1963–1966.

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. Trends Cogn Sci 17:230–240.

Poppenk J, Moscovitch M (2011) A Hippocampal Marker of Recollection Memory Ability among Healthy Young Adults: Contributions of Posterior and Anterior Segments. Neuron 72:931–937.

Postman L (1962) Transfer of training as a function of experimental paradigm and degree of first-list learning. J Verbal Learning Verbal Behav 1:109–118.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.

Preston AR, Bornstein AM, Hutchinson JB, Gaare ME, Glover GH, Wagner AD (2010) High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. J Cogn Neurosci 22:156–173.

Preston AR, Eichenbaum H (2013) Interplay of hippocampus and prefrontal cortex in memory. Curr Biol 23:R764–R773.

Preston AR, Shrager Y, Dudukovic N, Gabrieli JDE (2004) Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus 14:148–152.

Preston AR, Wagner AD (2007) The Medial Temporal Lobe and Memory. In: Neurobiology of Learning and Memory (Kesner RP, Martinez JL, eds), pp 305–337. Academic Press.

Price JL, Drevets WC (2009) Neurocircuitry of Mood Disorders. Neuropsychopharmacology 35:192–216.

Pruessner JC, Köhler S, Crane J, Pruessner M, Lord C, Byrne A, Kabani N, Collins DL, Evans AC (2002) Volumetry of entorhinal and parahippocampal cortex from high-resolution MR images: Considering the variability of the collateral sulcus. Cereb Cortex 12:1342–1353.

Pruessner JC, Li LM, Serles W, Pruessner M, Collins DL, Kabani N, Lupien S, Evans AC (2000) Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. Cereb Cortex 10:433–442.

Ranganath C (2010) Binding Items and Contexts: The Cognitive Neuroscience of Episodic Memory. Curr Dir Psychol Sci 19:131–137.

Raposo A, Han S, Dobbins IG (2009) Ventrolateral prefrontal cortex and self-initiated semantic elaboration during memory retrieval. Neuropsychologia 47:2261–2271.

Ratcliff R, Clark SE, Shiffrin RM (1990) List-strength effect: I. Data and discussion. J Exp Psychol Learn Mem Cogn 16:163–178.

Raz N, Lindenberger U, Rodrigue KM, Kennedy KM, Head D, Williamson A, Dahle C, Gerstorf D, Acker JD (2005) Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. Cereb Cortex 15:1676–1689.

Rempel-Clower NL, Barbas H (2000) The laminar pattern of connections between prefrontal and anterior temporal cortices in the Rhesus monkey is related to cortical structure and function. Cereb Cortex 10:851–865.

Rempel-Clower NL, Zola SM, Squire LR, Amaral DG (1996) Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. J Neurosci 16:5233–5255.

Richards BA, Xia F, Santoro A, Husse J, Woodin MA, Josselyn SA, Frankland PW (2014) Patterns across multiple memories are identified over time. Nat Neurosci 17:981–986.

Roberts AC, Tomic DL, Parkinson CH, Roeling TA, Cutter DJ, Robbins TW, Everitt BJ (2007) Forebrain Connectivity of the Prefrontal Cortex in the Marmoset Monkey (Callithrix jacchus): An Anterograde and Retrograde Tract-Tracing Study. J Comp Neurol 502:86–112.

Robin N, Holyoak KJ (1995) Relational complexity and the functions of the prefrontal cortex. In: The Cognitive Neurosciences (Gazzaniga MS, ed), pp 987–997. Cambridge, MA: MIT Press.

Rosene DL, Van Hoesen GW (1977) Hippocampal Efferents Reach Widespread Areas of Cerebral Cortex and Amygdala in the Rhesus Monkey. Science 198:315–317.

Roy M, Shohamy D, Wager TD (2012) Ventromedial prefrontal-subcortical systems and the generation of affective meaning. Trends Cogn Sci 16:147–156.

Sara SJ (2010) Reactivation, retrieval, replay and reconsolidation in and out of sleep: connecting the dots. Front Behav Neurosci 4:185.

Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK (2012) The future of memory: remembering, imagining, and the brain. Neuron 76:677–694.

Schacter DL, Harbluk JL, McLachlan DR (1984) Retrieval without recollection: An experimental analysis of source amnesia. J Verbal Learning Verbal Behav 23:593–611.

Schacter DL, Wagner AD (1999) Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. Hippocampus 9:7–24.

Schapiro AC, Kustner L V, Turk-Browne NB (2012) Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. Curr Biol:1622–1627.

Schiller D, Phelps EA (2011) Does reconsolidation occur in humans? Front Behav Neurosci 5:24.

Schinazi VR, Nardi D, Newcombe NS, Shipley TF, Epstein RA (2013) Hippocampal size predicts rapid learning of a cognitive map in humans. Hippocampus 23:515–528.

Schlichting ML, Mumford JA, Preston AR (in preparation) Learning-related changes in item representations reveal dissociable integration and separation signatures in hippocampus and prefrontal cortex.

Schlichting ML, Preston AR (2014) Memory reactivation during rest supports upcoming learning of related content. Proc Natl Acad Sci U S A 111:15845–15850.

Schlichting ML, Preston AR (2015) Memory integration: neural mechanisms and implications for behavior. Curr Opin Behav Sci 1:1–8.

Schlichting ML, Preston AR (under review) Hippocampal-medial prefrontal circuit supports memory updating during learning and post-encoding rest.

Schlichting ML, Zeithamova D, Preston AR (2014) CA1 subfield contributions to memory integration and inference. Hippocampus 24:1248–1260.

Schnider A (2003) Spontaneous confabulation and the adaptation of thought to ongoing reality. Nat Rev Neurosci 4:662–671.

Schonberg T, Bakkour A, Hover AM, Mumford JA, Nagar L, Perez J, Poldrack RA (2014) Changing value through cued approach: an automatic mechanism of behavior change. Nat Neurosci 17:625–630.

Scoville WB (1954) The Limbic Lobe in Man. J Neurosurg 11:64–66.

Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry 20:11–21.

Sharon T, Moscovitch M, Gilboa A (2011) Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. Proc Natl Acad Sci U S A 108:1146–1151.

Shimamura AP, Janowsky JS, Squire LR (1990) Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. Neuropsychologia 28:803–813.

Shohamy D, Wagner AD (2008) Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. Neuron 60:378–389.

Simons JS, Spiers HJ (2003) Prefrontal and medial temporal lobe interactions in long-term memory. Nat Rev Neurosci 4:637–648.

Simons JS, Verfaellie M, Galton CJ, Miller BL, Hodges JR, Graham KS (2002) Recollection-based memory in frontotemporal dementia: implications for theories of long-term memory. Brain 125:2523–2536.

Singer AC, Karlsson MP, Nathe AR, Carr MF, Frank LM (2010) Experience-dependent development of coordinated hippocampal spatial activity representing the similarity of related locations. J Neurosci 30:11586–11604.

Smith CN, Squire LR (2009) Medial temporal lobe activity during retrieval of semantic memory is related to the age of the memory. J Neurosci 29:930–938.

Smith CN, Urgolites ZJ, Hopkins RO, Squire LR (2014) Comparison of explicit and incidental learning strategies in memory-impaired patients. Proc Natl Acad Sci U S A 111:475–479.

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31:1487–1505.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol Rev 99:195–231.

Squire LR, Stark CEL, Clark RE (2004) The medial temporal lobe. Annu Rev Neurosci 27:279–306.

St. Jacques PL, Olm C, Schacter DL (2013) Neural mechanisms of reactivation-induced updating that enhance and distort memory. Proc Natl Acad Sci U S A 110:19671–19678.

Staresina BP, Alink A, Kriegeskorte N, Henson RN (2013) Awake reactivation predicts memory in humans. Proc Natl Acad Sci U S A 110:21159–21164.

Stark CEL, Squire LR (2001) When zero is not zero: The problem of ambiguous baseline conditions in fMRI. Proc Natl Acad Sci U S A 98:12760–12766.

Sterpenich V, Albouy G, Darsaud A, Schmidt C, Vandewalle G, Dang Vu TT, Desseilles M, Phillips C, Degueldre C, Balteau E, Collette F, Luxen A, Maquet P (2009) Sleep promotes the neural reorganization of remote emotional memory. J Neurosci 29:5143–5152.

Strange BA, Dolan RJ (2001) Adaptive anterior hippocampal responses to oddball stimuli. Hippocampus 11:690–698.

Straube B, Green A, Weis S, Chatterjee A, Kircher T (2009) Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. J Cogn Neurosci 21:821–836.

Supekar K, Uddin LQ, Prater K, Amin H, Greicius MD, Menon V (2010) Development of functional and structural connectivity within the default mode network in young children. Neuroimage 52:290–301.

Suzuki H, Botteron KN, Luby JL, Belden AC, Gaffrey MS, Babb CM, Nishino T, Miller MI, Ratnanather JT, Barch DM (2013) Structural-functional correlations between hippocampal volume and cortico-limbic emotional responses in depressed children. Cogn Affect Behav Neurosci 13:135–151.

Suzuki WA, Amaral DG (1990) Cortical inputs to the CA1 field of the monkey hippocampus originate from the perirhinal and parahippocampal cortex but not from area TE. Neurosci Lett 115:43–48.

Suzuki WA, Amaral DG (1994) Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. J Comp Neurol 350:497–533.

Sweegers CCG, Takashima A, Fernández G, Talamini LM (2013) Neural mechanisms supporting the extraction of general knowledge across episodic memories. Neuroimage 87:138–146.

Takashima A, Nieuwenhuis ILC, Rijpkema M, Petersson KM, Jensen O, Fernández G (2007) Memory trace stabilization leads to large-scale changes in the retrieval network: a functional MRI study on associative memory. Learn Mem 14:472–479.

Takashima A, Petersson KM, Rutters F, Tendolkar I, Jensen O, Zwarts MJ, McNaughton BL, Fernández G (2006) Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. Proc Natl Acad Sci U S A 103:756–761.

Takeuchi H, Taki Y, Hashizume H, Sassa Y, Nagase T, Nouchi R, Kawashima R (2012) The association between resting functional connectivity and creativity. Cereb Cortex 22:2921–2929.

Tambini A, Ketz N, Davachi L (2010) Enhanced brain correlations during rest are related to memory for recent experiences. Neuron 65:280–290.

Thompson-Schill SL, Bedny M, Goldberg RF (2005) The frontal lobes and the regulation of mental activity. Curr Opin Neurobiol 15:219–224.

Tolman EC (1948) Cognitive maps in rats and men. Psychol Rev 55:189–208.

Treves A, Rolls ET (1994) Computational analysis of the role of the hippocampus in memory. Hippocampus 4:374–391.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RGM (2007) Schemas and Memory Consolidation. Science 316:76–82.

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris RGM (2011) Schema-Dependent Gene Activation and Memory Encoding in Neocortex. Science 333:891–895.

Uncapher MR, Rugg MD (2005) Encoding and the durability of episodic memory: a functional magnetic resonance imaging study. J Neurosci 25:7260–7267.

Underwood BJ (1949) Proactive inhibition as a function of time and degree of prior learning. J Exp Psychol 39:24–34.

Van Buuren M, Kroes MCW, Wagner IC, Genzel L, Morris RGM, Fernandez G (2014) Initial Investigation of the Effects of an Experimentally Learned Schema on Spatial Associative Memory in Humans. J Neurosci 34:16662–16670.

Van der Jeugd A, Goddyn H, Laeremans A, Arckens L, D'Hooge R, Verguts T (2009) Hippocampal involvement in the acquisition of relational associations, but not in the expression of a transitive inference task in mice. Behav Neurosci 123:109–114.

Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL (2010) Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J Neurophysiol 103:297–321.

Van Hoesen GW, Pandya DN, Butters N (1975) Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. II. Frontal lobe afferents. Brain Res 95:25–38.

Van Kesteren MTR, Beul SF, Takashima A, Henson RN, Ruiter DJ, Fernández G (2013a) Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. Neuropsychologia 51:2352–2359.

Van Kesteren MTR, Fernández G, Norris DG, Hermans EJ (2010a) Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. Proc Natl Acad Sci U S A 107:7550–7555.

Van Kesteren MTR, Rijpkema M, Ruiter DJ, Fernandez G (2010b) Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. J Neurosci 30:15888–15894.

Van Kesteren MTR, Rijpkema M, Ruiter DJ, Fernández G (2013b) Consolidation differentially modulates schema effects on memory for items and associations. PLoS One 8:e56155.

Van Kesteren MTR, Rijpkema M, Ruiter DJ, Morris RGM, Fernández G (2014) Building on prior knowledge: Schema-dependent encoding processes relate to academic performance. J Cogn Neurosci 26:2250–2261.

Van Kesteren MTR, Ruiter DJ, Fernández G, Henson RN (2012) How schema and novelty augment memory formation. Trends Neurosci 35:211–219.

Vertes RP, Hoover WB, Szigeti-Buck K, Leranth C (2007) Nucleus reuniens of the midline thalamus: Link between the medial prefrontal cortex and the hippocampus. Brain Res Bull 71:601–609.

Wager TD, Barrett LF, Bliss-Moreau E, Lindquist KA, Duncan S, Kober H, Joseph J, Davidson M, Mize J (2008) The Neuroimaging of Emotion. In: Handbook of Emotions.

Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL (1998) Building Memories: Remembering and Forgetting of Verbal Experiences as Predicted by Brain Activity. Science 281:1188–1191.

Wallenstein GV, Eichenbaum H, Hasselmo ME (1998) The hippocampus as an associator of discontiguous events. Trends Neurosci 21:317–323.

Wang S-H, Morris RGM (2010) Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. Annu Rev Psychol 61:49–79.

Wang S-H, Tse D, Morris RGM (2012) Anterior cingulate cortex in schema assimilation and expression. Learn Mem 19:315–318.

Warren DE, Duff MC (2014) Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. Hippocampus 24:920–933.

Warren DE, Jones SH, Duff MC, Tranel D (2014) False recall is reduced by damage to the ventromedial prefrontal cortex: Implications for understanding the neural correlates of schematic memory. J Neurosci 34:7677–7682.

Watson C, Andermann F, Gloor P, Jones-Gotman M, Peters T, Evans A, Olivier A, Melanson D, Leroux G (1992) Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. Neurology 42:1743–1750.

Weiler JA, Suchan B, Daum I (2010a) When the future becomes the past: Differences in brain activation patterns for episodic memory and episodic future thinking. Behav Brain Res 212:196–203.

Weiler JA, Suchan B, Daum I (2010b) Foreseeing the future: occurrence probability of imagined future events modulates hippocampal activation. Hippocampus 20:685–690.

Weiss AP, Dewitt I, Goff D, Ditman T, Heckers S (2005) Anterior and posterior hippocampal volumes in schizophrenia. Schizophr Res 73:103–112.

Wendelken C, Bunge SA (2010) Transitive inference: distinct contributions of rostrolateral prefrontal cortex and the hippocampus. J Cogn Neurosci 22:837–847.

Werchan DM, Gómez RL (2013) Generalizing memories over time: Sleep and reinforcement facilitate transitive inference. Neurobiol Learn Mem 100:70–76.

Whitely PL (1927) The dependence of learning and recall upon prior intellectual activities. J Exp Psychol 10:489–508.

Williams EJ (1959) The comparison of regression variables. J R Stat Soc Ser B 21:396–399.

Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81:267–279.

Wimmer GE, Shohamy D (2012) Preference By Association: How Memory Mechanisms in the Hippocampus Bias Decisions. Science 338:270–273.

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference for the general linear model. Neuroimage 92:381–397.

Winocur G, Moscovitch M, Bontempi B (2010) Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal-neocortical interactions. Neuropsychologia 48:2339–2356.

Witter M, Van Hoesen GW, Amaral DG (1989) Topographical Organization of the Entorhinal Projection to the Dentate Gyrus of the Monkey. J Neurosci 9:216–228.

Witter MP, Amaral DG (1991) Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. J Comp Neurol 307:437–459.

Wolosin SM, Zeithamova D, Preston AR (2012) Reward modulation of hippocampal subfield activation during successful associative encoding and retrieval. J Cogn Neurosci 24:1532–1547.

Wood ER, Dudchenko PA, Eichenbaum H (1999) The global record of memory in hippocampal neuronal activity. Nature 397:613–616.

Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H (2000) Hippocampal neurons encode information about different types of memory episodes occurring in the same location. Neuron 27:623–633.

Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, Evans AC (2002) A general statistical analysis for fMRI data. Neuroimage 15:1–15.

Xu W, Südhof T (2013) A Neural Circuit for Memory Specificity and Generalization. Science 339:1290–1295.

Yamaguchi S, Hale LA, D'Esposito M, Knight RT (2004) Rapid prefrontal-hippocampal habituation to novel events. J Neurosci 24:5356–5363.

Zalesak M, Heckers S (2009) The role of the hippocampus in transitive inference. Psychiatry Res Neuroimaging 172:24–30.

Zeineh MM, Engel SA, Bookheimer SY (2000) Application of cortical unfolding techniques to functional MRI of the human hippocampal region. Neuroimage 11:668–683.

Zeineh MM, Engel SA, Thompson PM, Bookheimer SY (2003) Dynamics of the Hippocampus During Encoding and Retrieval of Face-Name Pairs. Science 299:577–580.

Zeithamova D, Dominick AL, Preston AR (2012a) Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron 75:168–179.

Zeithamova D, Preston AR (2010) Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. J Neurosci 30:14676–14684.

Zeithamova D, Schlichting ML, Preston AR (2012b) The hippocampus and inferential reasoning: Building memories to navigate future decisions. Front Hum Neurosci 6:70.

Zola-Morgan SM, Squire LR (1990) The Primate Hippocampal Formation: Evidence for a Time-Limited Role in Memory Storage. Science 250:288–290.