Title

Getting the Hippocampus Under Control: The Interplay Between the Hippocampus and Prefrontal Cortex

Author

Williams, Ashley

Publication Date

2023

Peer reviewed|Thesis/dissertation

Getting the Hippocampus Under Control: The Interplay Between the Hippocampus and Prefrontal Cortex

By

ASHLEY WILLIAMS
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

NEUROSCIENCE

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

J. Daniel Ragland, Co-Chair

Charan Ranganath, Co-Chair

Randall O'Reilly

Andrew Yonelinas

Laura Tully

Committee in Charge

2023

Table of Contents

ABSTRACT ..... *iii*

CHAPTER 1: Getting the Hippocampus Under Control: Goal-directed Prefrontal Modulation... *1*

CHAPTER 2: Prefrontal Control of the Hippocampus: A Computational Model..... *37*

CHAPTER 3: Testing the Boundaries: The Testing Effect and Event Segmentation..... *68*

CHAPTER 4: Memory Based Prediction Deficits and Dorsolateral Prefrontal Dysfunction in Schizophrenia *85*

Concluding Remarks *113*

ABSTRACT

The hippocampus has been recognized as critical for several memory processes, but current theories disagree about how to characterize its role. Some theoretical models suggest that the hippocampus forms memories that capture relationships between different experiences ("integration"), whereas others suggest that it forms distinct representations of otherwise similar events ("differentiation"). Different computational models of hippocampal function propose different mechanisms for hippocampal integration and differentiation, and it is unclear whether a single model can explain both processes. In the first chapter of this dissertation, I consider evidence suggesting that the prefrontal cortex might mediate the tradeoff between memory integration and differentiation, flexibly modulating hippocampal functioning depending on task goals. The second chapter describes a proposed proof-of-concept hippocampal-prefrontal model that is able to handle a switch in task goals from integration to differentiation. The third chapter describes possible ways in which items may be relating to each other in a real-world environment, I present a study on the testing effect and event boundaries. We know through extensive literature that retrieval practice, or repeated recall through testing, enhances long-term memory retention. By introducing a manipulation of event boundaries, I show behaviorally, through the use of virtual reality environments, that retrieval practice or restudy of items within an event can cause forgetting of other items within the same event. In the future, this work can be related to the hippocampal and prefrontal findings associated with both the testing effect and event segmentation. Finally, the last chapter introduces an example of a disorder population, namely individuals with schizophrenia, who have deficits in memory processing, possibly related to hippocampal

and prefrontal dysfunction. This chapter showcases a study of temporal sequence processing in schizophrenia, which reveals a deficit in the functioning of both the hippocampus and dorsolateral prefrontal cortex. Together, the work described in this dissertation shows literature from multiple fields of research, empirical studies, and computational modeling of the interplay between the hippocampus and prefrontal cortex.

CHAPTER 1: Getting the Hippocampus Under Control: Goal-directed Prefrontal Modulation

Integration and Differentiation in Hippocampus and Prefrontal Cortex

The hippocampus is known to play a central role in memory processes [1], and computational models and descriptive theories have assigned a broad range of functions to the hippocampus. Many of these functions can be grouped under the umbrella of two key processes: “integration” and “differentiation”. In this review, we will summarize evidence suggesting that integration and differentiation are key functions of the hippocampus. Next, we summarize descriptive and computational models that explain a particular subset of this evidence and consider whether any existing model of hippocampal function is sufficient to explain both integration and differentiation because these goals are computationally incompatible. Next, we summarize evidence linking prefrontal cortex (PFC) to flexible, goal directed modulation of memory encoding and retrieval. These findings suggest that PFC might interact with the hippocampus in a manner allowing memories from single experiences to be adaptively modulated, updated, and recapitulated as necessary to accomplish specific behavioral goals. We close by considering biologically plausible computational mechanisms by which PFC can shape the representational landscape in the hippocampus, enabling the hippocampus to support functions that would be otherwise computationally incompatible.

Mnemonic Differentiation

In our everyday lives, it is important to form precise memories to distinguish between events with overlapping content. For example, think of the last two weddings

you attended. They overlapped in content (i.e., a ceremony, cake cutting, etc.), but we can distinguish one event from the other. The hippocampus is thought to form distinct representations of overlapping events ("pattern separation") [2-3] to minimize competition in the face of high interference. Functional magnetic resonance imaging (fMRI) studies in humans investigated pattern separation and retrieval precision by examining hippocampal activity with the Mnemonic Similarity Task [4]. In this task, participants were shown repetitions of studied items (e.g., a rubber duck), highly similar "lure" items (e.g., a rubber duck viewed from a different perspective), and unrelated new items (e.g., an apple). Studies using this paradigm demonstrate that activity in the hippocampus [5], particularly the CA3 and/or dentate gyrus (DG) subfields [6-7], differentiates between repeated stimuli versus similar lures and new items.

These findings are also consistent with electrophysiological studies. In a rodent study, researchers found that the DG is particularly important for pattern separation [8]. Leutgeb and colleagues (2007) put rats in a slowly morphing enclosed environment (e.g., moving from a square enclosure to a circle enclosure by slight modifications). Even with slight changes in the rat's environment, rats showed changes in activity patterns within the DG [9-10]. This finding that the DG is sensitive to very slight changes between environments shows that the DG is capable of separating representations for highly similar events. Additionally, relying on converging evidence [8,10-11], including the work discussed above, Guzowski and colleagues (2004) suggest that the CA3 is also sensitive to small changes in context [9]. Together, this evidence supports the idea that the DG and CA3 subfields of the hippocampus are necessary for successful pattern separation.

In addition to neurophysiological evidence, lesion studies support the hippocampus being involved in differentiation. In a recognition memory task, participants with hippocampal damage were shown common images of items (i.e., easily recognized and named items) [12]. Following a distraction task, they were shown the same images again together with new images. Images that were only slightly different from the original images were also presented, and participants determined if the images were old, new, or similar. Behavioral results showed that those with hippocampal damage had more difficulty determining when images were similar, indicating a deficit in pattern separation [12]. These results are supported by rat lesion studies [13]. However, it is important to note that a review investigating pattern separation in the hippocampus concluded that the ability to differentiate types of stimuli heavily relies on task design [14].

There is also evidence that the role of the hippocampus in mnemonic differentiation is especially prominent when people must repeatedly encode or retrieve highly similar information. A number of fMRI studies have shown that differences in hippocampal representations between two highly overlapping memories can become grossly exaggerated over the course of learning, a phenomenon known as *repulsion* [14–16]. For example, one fMRI study [15] examined hippocampal representations as participants were repeatedly exposed to a series of photos depicting a route between two destinations in a city. Following each of many runs, they were given random photos along the routes and asked with which destination that image was associated. Critically, the routes that were studied were highly overlapping, such that participants needed to focus specifically on the unique aspect of each route in order to select the correct

destination. Early in learning, hippocampal activity patterns for each route were distinct such that similarity between highly overlapping routes was about the same as the similarity between nonoverlapping routes. Following repeated route exposure, however, the representations between overlapping routes further separated – as if they were repelling one another – with the representations of the overlapping routes becoming *less* similar than the representations of nonoverlapping routes. The authors asserted that the hippocampus is responsible for increasing distance between competing representations [15]. Another fMRI study supporting these findings [18] found that items that did not overlap in context (spatial and temporal) showed more similar representations in DG and CA3 than those that did overlap in context, suggesting repulsion. These findings are consistent with the idea that pattern separation in the hippocampus can be dramatically increased when it is necessary to resolve competition between highly similar memories.

Memory Integration

A separate line of work has shown that the hippocampus can also be critical for memory tasks that require integration of information from different events. For example, statistical learning tasks involve gradual learning of associations between temporally contiguous events. In one fMRI study of statistical learning [19], participants were shown a stream of fractal images that included pairs of fractals that were shown in the same temporal order (e.g., A always followed by B). Both prior to and after encoding, the fractals were viewed in random order so that researchers could characterize the representational changes before and after the manipulation. Although participants were not explicitly instructed about the sequentially pairing of the two fractals,

representational similarity within the hippocampus was increased for fractals that were paired relative to those that were not contiguously presented with one another [19]. Using a different statistical learning paradigm, Schapiro and colleagues [20] observed impaired learning of associations between sequences of images or tones in a patient with bilateral hippocampal lesions.

In studies using an associative inference task, participants are trained on paired items and are asked to infer relationships across pairs. For example, one could learn an "AB" pair, where A is an image of a face and B is an image of a house, and a "BC" pair, where B is the same image of a house and C is an image of a different face. During the inference test, participants are given face A and are asked to choose either face C or face D, for example. The answer, if the correct relationship has been established (i.e., a shared B house), is face C [18-21]. fMRI studies using these paradigms demonstrate that hippocampus is involved in this integrative process [19-20, 22]. Ultimately, these paradigms are used to explain how we integrate (and differentiate) similar content and how we use that information to draw conclusions.

The importance of the hippocampus for memory *integration* tasks like associative inference (AB+BC=AC) is interesting because the hippocampus is also thought to be necessary for differentiating between overlapping associations (AB≠ AC) [26]. Interestingly, some fMRI studies have observed evidence of representational separation and integration in the same paradigm, just in different subregions of the hippocampus [6,8,21]. Schlichting and colleagues [24] showed that, during an associative inference task, the posterior hippocampus represents individual associations, whereas anterior hippocampal representations generalize across overlapping AB and AC associations .

Dimsdale-Zucker [18] and Bakker [6], whose differentiation findings were previously described, showed that either increased similarity within CA1 across items that shared a context [18] or increased activity in CA1 [6] was associated with pattern completion. In a goal-based navigation task using VR environments, participants learned and were tested on a variety of local environments, then entered an MRI where they viewed locations in those environments. The following day, participants were tested on global environments comprised of the previously learned local environments. Results revealed, through analyses of patterns of activation in the hippocampus and entorhinal cortex (EC), that both integration and differentiation are important in constructing local and global environments and are, therefore, necessary for goal-based navigation [27]. Together, these studies suggest that CA1 is primarily involved in integration while DG/CA3 are involved in differentiation.

Computational Modeling of Integration and Differentiation

Several computational models, some of which are explained further in Table 1, have been proposed to explain the role of the hippocampus in differentiating between similar memories. Norman and O'Reilly [26-27], for instance, used the Complementary Learning Systems (CLS) model [30] to demonstrate how hippocampal differentiation of overlapping memories could enable recollection of past events in the face of significant interference (see [26] for related simulations). The CLS model is based on the known architecture of the hippocampus (see Box 1), and central to this model is the fact that DG has sparse coding due to a high degree of inhibition. Because of this inhibition, two events with many overlapping elements (as with an AB-AC task) are represented by distinct populations of units in DG (i.e., pattern separation). In a standard AB-AC task,

participants learn a list of paired items (“AB”) followed by learning of a different list of paired items, where item A from the AB list is now paired with an item C (“AC”). Results from studies using this paradigm revealed that learning of the AC list interfered with performance on the previously learned AB list [29-30]. By differentiating overlapping representations, a computational model can recover information about the two events without suffering from interference [26]. Ultimately, these simulations show how hippocampal pattern separation can enable the hippocampus to learn in the face of interference and make very fine-grained mnemonic discriminations.

Interestingly, the same CLS model architecture used to explain differentiation has also been used to model hippocampal memory integration. In the Recurrency and Episodic Memory Results in Generalization (REMERGE) Model [23]. Kumaran and colleagues used a principle known as “big loop recurrence” – referring to the idea that a retrieved hippocampal representation can be fed back into the hippocampus to cue retrieval of a related memory representation. Consider our earlier example: if people separately learn to associate face A and face C with house B, then they can infer an association between A and C. In REMERGE, each association (AB and BC) is encoded separately by the hippocampus. During the inference phase, when the hippocampus receives input about face A in the superficial layers of EC, it is then able to recover the representation of house B in the deep (output) layers of EC. Information about house B is then recirculated into the superficial layers of EC. During a second pass, the face A/house B pair is transferred from the superficial layers of EC into the hippocampus, which allows retrieval of the face C memory into the deep layers of EC. The key point from the REMERGE model is that implementation of big loop recurrence can potentially

explain hippocampal memory integration, but this principle would seem to be counterproductive for memory tasks that require strong differentiation.

To address this issue, Schapiro and colleagues proposed [32] another variant of the CLS model that attempts to explain that integration and differentiation could be happening in different hippocampal pathways [34-35]. In this model, integration occurs through the monosynaptic pathway (MSP)—i.e., the direct connection between the EC and CA1—whereas differentiation occurs through the trisynaptic pathway (TSP), described in Box 1. To test these pathways, the model learned tasks that mimicked statistical learning and episodic memory. Results showed that the TSP was able to support rapid learning of individual items, but due to high pattern separation, it was not sufficient to support statistical learning of associations between items. In contrast, the MSP was able to support slow learning of statistical associations in CA1. To mimic an underdeveloped TSP, the TSP pathway was "lesioned" (i.e., pathway projection strength set at 0), and the ability to do statistical learning was not impaired, further proving that MSP was involved in integration while TSP was important for differentiation [33]. Although this model can account for both encoding of single items and associations, it is not clear that the model can succeed at integration while maintaining the capacity to differentiate between highly similar items (as in the mnemonic discrimination task) or retain previously learned associations while learning new overlapping associations (as in AB-AC learning). Given the differences between the MSP and TSP observed in Schapiro et al.'s simulations, it is possible that a mechanism could be incorporated to flexibly prioritize pattern separation in the TSP for mnemonic differentiation in the face of high interference and the MSP when memory integration is

necessary. As we describe later, this would require a mechanism to flexibly select or prioritize processing in one of these pathways to perform depending on task demands [32].

Another framework to understand hippocampal integration and differentiation is the Non-Monotonic Plasticity Hypothesis (NMPH) [35]. Rather than focusing on unpacking circuit level interactions in the hippocampus, the model focuses on a learning rule that could be implemented in the hippocampus. According to the NMPH, when a memory retrieval cue triggers moderate co-activation of overlapping associations, differentiation occurs due to weakening of synaptic connections, but strong co-activation between overlapping associations leads to integration [23,33]. The NMPH can also potentially account for the phenomenon of repulsion. Specifically, when overlapping memories are repeatedly reactivated, the NMPH learning rule prunes the parts of the representation that are maximally overlapping. As a result, the competing memories become more differentiated from one another than they are from other items that do not overlap much [37]. The key idea in the NMPH is that there is a “sweet spot” for repulsion: if two representations are just different enough, then repeated retrieval will exaggerate the differences between them. The NMPH can explain a broad range of data, but it is somewhat inflexible in that it does not readily explain how the same stimuli can be integrated or differentiated under different task conditions. As we pointed out earlier, tasks like associative inference share a lot in common with tasks that require differentiation between overlapping associations, so it is unclear how representational overlap alone could be sufficient to explain the range of data described above.

Control of Memory Functioning by the Prefrontal Cortex

One key assumption of the models described above is that they attempt to explain differentiation and/or integration solely through computations in the hippocampus. As we have seen, however, differentiation and integration are observed in highly similar tasks that differ primarily in terms of the intended goal—i.e., to focus on distinctive attributes of individual items or associations or to focus on shared elements. Ultimately, it is unclear whether a single brain area with a particular set of cellular and circuit-level anatomy can support both extreme differentiation (e.g., discrimination between two nearly identical rubber ducks [12] in the mnemonic differentiation task, or creating hyper-differentiated representations of competing associations [15]) while still supporting extreme integration (e.g., integration of overlapping associations in associative inference tasks [23] or integration of temporally contiguous items in statistical learning paradigms [21]). It also seems implausible that such complex memory tasks would be supported by the hippocampus in isolation. Given the high degree of extrinsic connectivity of the hippocampal-entorhinal system [38], it is worth considering whether the tradeoff between differentiation and integration might depend on interactions between the hippocampus and other brain regions.

In particular, the prefrontal cortex (PFC) might play a critical role in hippocampal differentiation and integration by modulating hippocampal representations to fit task demands. A wide range of research suggests that different areas of the PFC and anterior cingulate cortex (ACC) [see Box 2] play a critical role in using representations of goals and rules to support performance on complex memory tasks [43,45] involving differentiation or integration [41]. For instance, prefrontal regions are critical for mitigating interference in AB-AC paired associate learning tasks [42] and in tasks that

require fine grained mnemonic distinctions between studied items and similar lures [40–43]. In humans, inactivation of the ventrolateral PFC via transcranial magnetic stimulation (TMS) impaired mnemonic discrimination between previously learned objects (e.g., a hammer) and unseen lure items (e.g., a hammer from a slightly different angle) [45]. Frontal lesions in rats and humans can sometimes increase susceptibility to falsely remember even items that are unrelated to study items [44–46] suggesting that differentiation can break down with frontal dysfunction. In addition, one study found that inactivation of the medial PFC in rodents impaired mnemonic differentiation [46]. This empirical evidence suggests the necessity of a PFC to successfully differentiate between similar items.

Several lines of research also suggest that frontal areas are critical for memory integration, as in the transitive inference task (e.g., learn that if A<B and B<C, then A<C). For instance, in an fMRI study investigating the hippocampus and prefrontal cortex, researchers found that activation of the rostrolateral PFC was required for relational integration instead of the hippocampus [50]. In a rodent study, mice with damage to the medial PFC had impaired transitive inference abilities [51]. Similarly, those with VMPFC damage had difficulties with transitive inference [52-53]. Other studies have shown that the VMPFC plays a role in associative inference (i.e., AB-BC task). One study found that individuals with VMPFC damage were able to successfully learn AB and BC pairs, but they were unable to infer a relationship between A and C [54]. In an fMRI AB-BC task, the ability to have successful associative inference was related to the functional coupling of the hippocampus and VMPFC [53]. This suggests that integration is able to occur when the hippocampus and PFC work together.

Given that PFC regions seem to support flexible, goal-directed processing, it is reasonable to think that the PFC might support memory integration and differentiation via modulation of activity in the hippocampus. This idea is consistent with theoretical models suggesting that, during retrieval, the PFC interacts with the MTL by cueing the correct contextual information to complete the task [55]. During retrieval, the PFC directs the MTL to attend to the type of information necessary for the current task (i.e., integration or differentiation) to succeed. There have also been lesion studies demonstrating that PFC-hippocampal interactions are necessary for integration and differentiation [51–53]. Given a confluence of data suggesting a role for PFC in mnemonically challenging memory integration and differentiation tasks [44–49], there is good reason to believe that performance on these tasks might depend on interactions between the PFC and hippocampus.

Computational Mechanisms for Prefrontal Control of Hippocampal Function

There are multiple mechanisms by which the PFC could plausibly play a significant role in controlling hippocampal function [43,45,50-51] (Box 2). Here, we propose three possibilities: 1) Input Modulation, 2) Output Gating, and 3) Direct Modulation (Figure 2). Input modulation involves the PFC interacting with regions that provide input to the hippocampus. The perirhinal cortex (PRC), parahippocampal cortex (PHC), and presubiculum account for much of the direct input to the hippocampal system [61], and these areas receive direct inputs from both the ventrolateral PFC [62] and medial [63] PFC.

The second potential mechanism is Output Gating, where the ACC works as a mediator between the PFC and EC. A specific way in which ACC and PFC could

interact with the hippocampus is as follows: for integration, the representations of similar information become more similar through the big loop (see Box 1). In this case, the ACC is not needed as there is no conflict between the two similar representations. However, if these integrated representations are not helpful to the task at hand (e.g., mnemonic discrimination), then the ACC registers the conflict and proceeds to drive control processes in the PFC that cause the hippocampus to shift to a differentiation mode of encoding.

The last pathway is through Direct Modulation. This pathway involves input from the PFC to hippocampus through the nucleus reuniens in the thalamus, which may suppress pattern completion by suppressing unwanted hippocampal activity [59]. A study investigating fear memory generalization shows that inactivation of the nucleus reuniens (or prefrontal inputs into the nucleus reuniens) showed increased generalization, whereas activation showed a decreased ability to generalize across memories [57]. This showcases the necessity of the nucleus reuniens for tasks where generalization does not help participants reach task goals and suggests a role in differentiation vs integration.

Note that the three proposed mechanisms are not mutually exclusive, and it is reasonable to think that all three mechanisms are engaged to control memory encoding and retrieval depending on task demands. Through these pathways, we propose that during integration, the PFC enhances shared features during retrieval leading to updated representations that are *more* similar, and during differentiation, the PFC inhibits shared features leading to an updated representations that are *less* similar (Figure 3).

Concluding Remarks and Future Perspectives

Ultimately, adaptive human behavior depends on flexible deployment of learned information to accomplish goals. At some times, this might require differentiation between similar experiences and at others it might require integration of bits of information acquired at different moments. Computational models have been effective at demonstrating how the unique architecture of the hippocampus could meet the demands of differentiation or integration in the context of specific task demands. To create a single model that is capable of a wide variety of tasks with different goals, it may be necessary for computational models of memory to incorporate contributions of the PFC to flexibly monitor and control hippocampal functions depending on task goals.

If, as we propose, the PFC plays a role in modulating hippocampal function in the service of task demands, this would have far-reaching implications for our understanding of hippocampal function. Numerous theories have been proposed to explain how the hippocampus contributes to memory, but this approach might be myopic. The nature of hippocampal computations and their apparent memory functions might be radically altered based on task demands, goals, and an organism's understanding of the current situation. Accordingly, it may be time to embrace complexity and consider a more flexible view of memory, in which the hippocampus is part of a broader circuit that supports the use of memory to guide goal-directed behavior in the present and planning for the future [25,65].

BOX 1: Medial Temporal Lobe Circuitry

Most computational models of hippocampal function focus on how information flows between the EC and hippocampal subfields (Figure 1). Information from different neocortical hub areas converges in layers II and III of EC, which in turn, directs inputs to the hippocampal subfields via two pathways. Neurons in EC II convey input to the dentate gyrus (DG), which inputs into the CA3, before reaching the CA1 and re-entering the EC via layer V. These connections comprise the “trisynaptic pathway” (TSP), which is central to the computational specializations of the hippocampal EC-hippocampal circuit. Due to sparse coding within DG, there is a greater ability within this region to encapsulate small changes in representations, which allows for differentiation between similar inputs, ultimately leading to pattern separation. In contrast, neurons in EC III input directly into the CA1, which is known as the “monosynaptic pathway” (MSP) [38, 65]. The deep and superficial layers in this system have bidirectional connections, resulting in an interconnected loop of information transfer [30], also known as the “big loop”. The movement of information from the deep layers into the superficial layers of the EC is often called “big loop recurrence” [23, 66], which supports memory integration. Ultimately, the hippocampus is receiving information from various regions, through EC, to bind together information before entering higher-order cortical regions [38-41].

BOX 2: Subregions of the Prefrontal Cortex

The prefrontal cortex (PFC) occupies about 1/3 of the human neocortex, and it consists of multiple subregions. In broad terms, researchers have differentiated these subfields along the rostro-caudal, dorsal-ventral, and medial-lateral axes. Anatomical research has suggested that areas in the ventromedial PFC ([70]) and the Subgenual Anterior Cingulate Cortex (ACC) are positioned to play the most significant role in modulation of hippocampal function. These areas have direct connections with EC layer 5 [59], which receives direct output from CA1, and they are positioned to modulate activity in CA1 via a disynaptic pathway through the nucleus reuniens of the thalamus. The Anterior Cingulate Cortex (ACC) is known to be involved in conflict monitoring processes and works with the PFC to resolve conflict [63–65]. When tasks involved competition between multiple answers, ACC was found to be highly active, which is in line with the idea that ACC is necessary when there is conflict [71], [72]. The ACC has also been linked to accessing information from *and* updating established schemas in cortical regions [74], which suggests the ability of the ACC to work with the PFC in updating information to better match the task at hand. Further, research suggests that the co-activation of medial prefrontal regions and the ACC is the key to controlling what type of information is needed for successful retrieval and, importantly, inhibiting the information that is not needed [73].

In contrast, the anatomy of the frontopolar cortex (area 10) has not been well characterized. However, the medial frontopolar cortex might work alongside ventromedial PFC and ACC in modulating hippocampal function because of its role in monitoring competing goals and switching between tasks [75]–[77]. There are also

connections between the PFC and other medial temporal regions that input into the EC and, ultimately, the hippocampus. Anterograde and retrograde tracings studies show that orbital PFC is connected to the perirhinal cortex, which is a part of the anterior temporal system involved in object perception, while the medial PFC is connected to the parahippocampal cortex, which is associated with the posterior medial system involved in retrieval of contextual information [63], [78]. Additionally, there is a bi-directional relationship between the parahippocampal cortex and the dorsolateral PFC [79], which is involved in strategic planning and resolving competition between competing representations [80], [81]. Together, these regions of the PFC, along both the medial-lateral and dorsal-ventral axes, work together to monitor competing representations, recruit necessary regions, and bias neural processing to achieve task goals.

Table 1: Computational Models

Model	Model Parameters
CLS Model [28], [29]	The CLS framework posits that the hippocampus is specialized for encoding specific experiences (hippocampus), in contrast to the neocortex, which generalizes across multiple experiences (see also [2]). Successive variants of this model have incorporated increasing degrees of detail regarding the anatomy and physiology of the hippocampal system. The model was originally proposed to explain how people can rapidly learn new information and still retain memories of previously learned events with overlapping elements. The CLS model is able to reduce interference between memories because it has a relatively fast learning rate and because of pattern separation in the DG. Pattern separation occurs because overlapping inputs from the EC are assigned to sparse, nonoverlapping representations in the DG. Simulations of this model have focused on the importance of mnemonic differentiation to overcome interference. [28]
Recurrency and Episodic Memory Results in	The REMERGE model by Kumaran and McClelland was proposed in order to demonstrate how the original idea behind the CLS model—that the hippocampus separates memories—

Generalization (REMERGE) Model [82]	can be aligned with evidence for hippocampal involvement in rapid generalization. Using associative inference, Kumaran and colleagues were able to show that “big loop recurrence” (see Box 1) happens through a recirculation of information between the superficial and deep layers of the entorhinal cortex. To better characterize integration tasks, the authors implement a bi-directional, layered network. Through a combination of localist coding and inhibitory competition, the conjunctive layer works to limit interference and allow for generalization. Importantly, activity within this layer increases along with the memory strength of the features, which they associate with increased activity in the hippocampus. The feature layer represents the external input into the model and has a recurrent, bi-directional relationship with the conjunctive layer [82]. Together, these layers work together to generalize across item pairs, as with associative inference.
C-Horse [33]	Schapiro and colleagues used the architecture of the CLS model to explain how the tension between mnemonic differentiation and integration can be resolved through the interaction of two intra-hippocampal circuits, the trisynaptic (TSP) and monosynaptic (MSP) pathways (see Box 1). As described in earlier simulations of the CLS model, the high degree of pattern separation in the TSP (via sparse coding in the DG and CA3) enables rapid

	encoding of specific memories, whereas the distributed, overlapping representations in the MSP enable generalization across overlapping memories. This work suggests that the hippocampal complex can use either pathway depending on task goals [32-33].
Non-Monotonic Plasticity Hypothesis (NMPH) [35]	The Non-Monotonic Plasticity Hypothesis (NMPH) does not model hippocampal function per se, but instead proposes a learning rule that may be implemented in the hippocampus to adjudicate between differentiation and integration. The model starts with the premise that activation of one memory can lead to co-activation of features of other memories, which can lead to updating of the competing memories. When there is low co-activation between memories, there is no learning-related change in the strength of the overlapping memories. However, when there is moderate co-activation between memories, such as when two memories are associated with the same cue, there is a weakening of synaptic connections between those memories, causing differentiation. When co-activation is high, there is a strengthening of the synaptic connections, causing integration [35], [36]. Thus, the NMPH proposes that the learning rule will determine whether competing memories will be differentiated or integrated based on the degree of representational overlap.

FIGURES

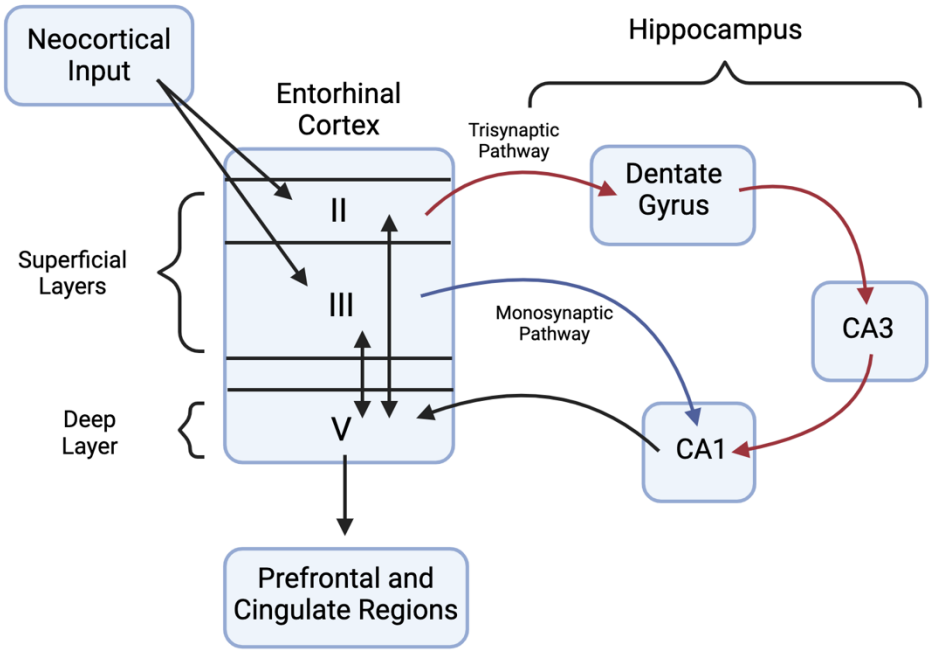

FIGURE 1: Medial Temporal Lobe Circuitry. Neocortical input into EC II and III is then input into the hippocampus via the Monosynpatic and Trisynaptic Pathways The blue arrow represents the Monosynaptic Pathway (EC III -> CA1 -> EC V). The red arrows represent the Trisynaptic Pathway (EC II -> DG -> CA3 -> EC V).

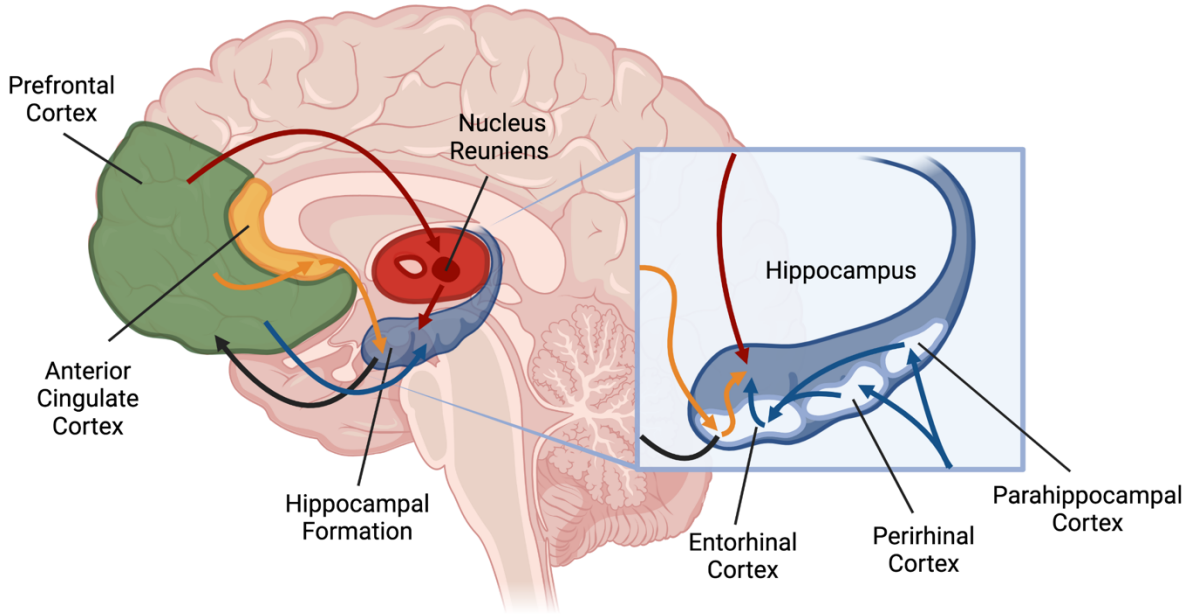

FIGURE 2: The Hippocampal-Prefrontal Circuitry. The blue arrows represent the Input Modulation pathway: Information flows from the PFC to the perirhinal and parahippocampal cortices to EC to the hippocampus. The yellow arrows represent the Output Gating pathway: Information flows from the PFC to ACC to EC to the hippocampus. The red arrows represent the Direct Modulation pathway: Information flows from the PFC to the nucleus reuniens of the thalamus to the hippocampus. Green = PFC; yellow = ACC; red = thalamus; dark red = nucleus reuniens; blue = hippocampal formation.

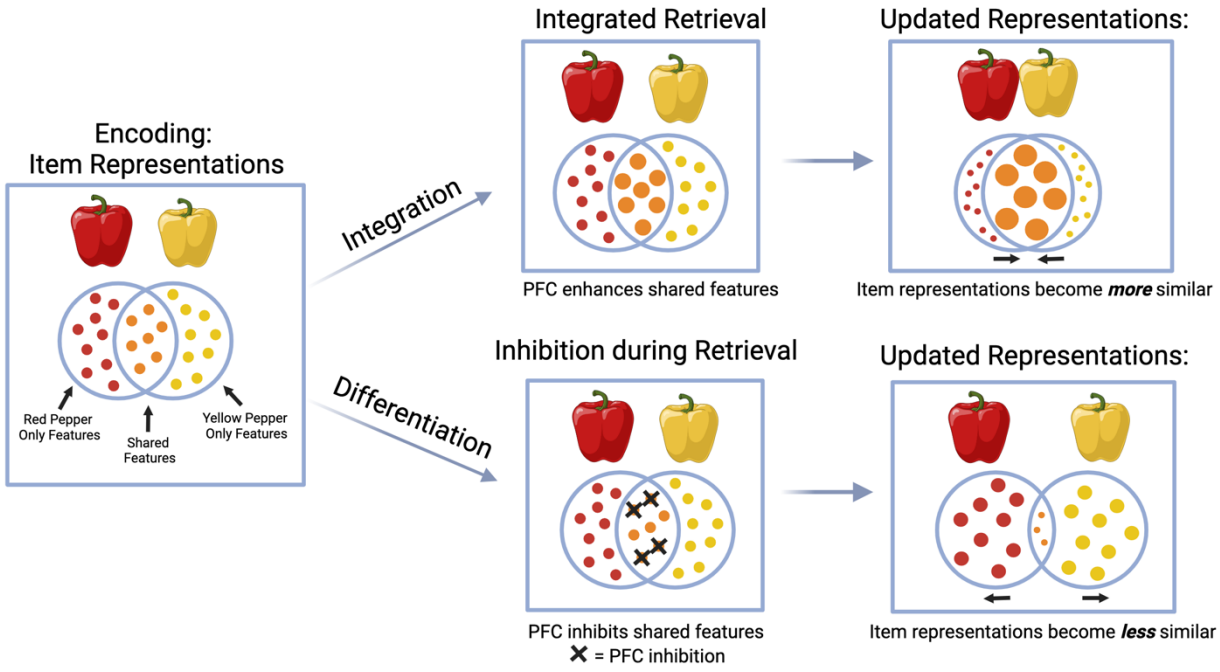

FIGURE 3: Integration vs Differentiation Using PFC. During encoding, the hippocampus encodes different representations for items with shared and distinct features. For integration processes, shared features are enhanced leading to more similar item representations after updating. For differentiation processes, shared features are inhibited to allow for more distinct updated representations.

References

[1] H. Eichenbaum, T. Otto, and N. J. Cohen, “The hippocampus—what does it do?,” *Behav. Neural Biol.*, vol. 57, no. 1, pp. 2–36, Jan. 1992, doi: 10.1016/0163-1047(92)90724-I.

[2] D. Marr, “Simple memory: a theory for archicortex,” *Phil. Trans. R. Soc. Lond. B*, 1971.

[3] E. Rolls, “The mechanisms for pattern completion and pattern separation in the hippocampus,” *Front. Syst. Neurosci.*, vol. 7, 2013, Accessed: Jan. 12, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2013.00074

[4] S. M. Stark, C. B. Kirwan, and C. E. L. Stark, “Mnemonic Similarity Task: A Tool for Assessing Hippocampal Integrity,” *Trends Cogn. Sci.*, vol. 23, no. 11, pp. 938–951, Nov. 2019, doi: 10.1016/j.tics.2019.08.003.

[5] J. L. Klippenstein, S. M. Stark, C. E. L. Stark, and I. J. Bennett, “Neural substrates of mnemonic discrimination: A whole-brain fMRI investigation,” *Brain Behav.*, vol. 10, no. 3, p. e01560, 2020, doi: 10.1002/brb3.1560.

[6] A. Bakker, C. B. Kirwan, M. Miller, and C. E. L. Stark, “Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus,” *Science*, Mar. 2008, doi: 10.1126/science.1152882.

[7] J. W. Lacy, M. A. Yassa, S. M. Stark, L. T. Muftuler, and C. E. L. Stark, “Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity,” *Learn. Mem.*, vol. 18, no. 1, pp. 15–18, Jan. 2011, doi: 10.1101/lm.1971111.

[8] J. K. Leutgeb, S. Leutgeb, M.-B. Moser, and E. I. Moser, “Pattern Separation in the Dentate Gyrus and CA3 of the Hippocampus,” *Science*, Feb. 2007, doi: 10.1126/science.1135801.

[9] J. F. Guzowski, J. J. Knierim, and E. I. Moser, “Ensemble Dynamics of Hippocampal Regions CA3 and CA1,” *Neuron*, vol. 44, no. 4, pp. 581–584, Nov. 2004, doi: 10.1016/j.neuron.2004.11.003.

[10] I. Lee, G. Rao, and J. J. Knierim, “A Double Dissociation between Hippocampal Subfields: Differential Time Course of CA3 and CA1 Place Cells for Processing Changed Environments,” *Neuron*, vol. 42, no. 5, pp. 803–815, Jun. 2004, doi: 10.1016/j.neuron.2004.05.010.

[11] A. Vazdarjanova and J. F. Guzowski, “Differences in Hippocampal Neuronal Population Responses to Modifications of an Environmental Context: Evidence for Distinct, Yet Complementary, Functions of CA3 and CA1 Ensembles,” *J. Neurosci.*, vol. 24, no. 29, pp. 6489–6496, Jul. 2004, doi: 10.1523/JNEUROSCI.0350-04.2004.

[12] C. B. Kirwan and C. E. L. Stark, “Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe,” *Learn. Mem.*, vol. 14, no. 9, pp. 625–633, Sep. 2007, doi: 10.1101/lm.663507.

[13] P. E. Gilbert and R. P. Kesner, “The role of the dorsal CA3 hippocampal subregion in spatial working memory and pattern separation,” *Behav. Brain Res.*, vol. 169, no. 1, pp. 142–149, Apr. 2006, doi: 10.1016/j.bbr.2006.01.002.

[14] R. P. Kesner, “Role of the hippocampus in mediating interference as measured by pattern separation processes,” *Behav. Processes*, vol. 93, pp. 148–154, Feb. 2013, doi: 10.1016/j.beproc.2012.09.018.

[15] A. J. H. Chanales, A. Oza, S. E. Favila, and B. A. Kuhl, “Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations,” *Curr. Biol.*, vol. 27, no. 15, pp. 2307-2317.e5, Aug. 2017, doi: 10.1016/j.cub.2017.06.057.

[16] A. J. H. Chanales, A. G. Tremblay-McGaw, M. L. Drascher, and B. A. Kuhl, “Adaptive repulsion of long-term memory representations is triggered by event similarity,” *Psychol. Sci.*, vol. 32, no. 5, pp. 705–720, 2021, doi: 10.1177/0956797620972490.

[17] S. E. Favila, A. J. H. Chanales, and B. A. Kuhl, “Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning,” *Nat. Commun.*, vol. 7, no. 1, Art. no. 1, Apr. 2016, doi: 10.1038/ncomms11066.

[18] H. R. Dimsdale-Zucker, M. Ritchey, A. D. Ekstrom, A. P. Yonelinas, and C. Ranganath, “CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields,” *Nat. Commun.*, vol. 9, no. 1, p. 294, Dec. 2018, doi: 10.1038/s41467-017-02752-1.

[19] A. C. Schapiro, L. V. Kustner, and N. B. Turk-Browne, “Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities,” *Curr. Biol.*, vol. 22, no. 17, pp. 1622–1627, Sep. 2012, doi: 10.1016/j.cub.2012.06.056.

[20] A. C. Schapiro, E. Gregory, B. Landau, M. McCloskey, and N. B. Turk-Browne, “The Necessity of the Medial Temporal Lobe for Statistical Learning,” *J. Cogn. Neurosci.*, vol. 26, no. 8, pp. 1736–1747, Aug. 2014, doi: 10.1162/jocn_a_00578.

[21] A. Schapiro and N. Turk-Browne, “Statistical Learning,” in *Brain Mapping*, Elsevier, 2015, pp. 501–506. doi: 10.1016/B978-0-12-397025-1.00276-1.

[22] A. R. Preston, Y. Shrager, N. M. Dudukovic, and J. D. E. Gabrieli, “Hippocampal contribution to the novel use of relational information in declarative memory,” *Hippocampus*, vol. 14, no. 2, pp. 148–152, 2004, doi: 10.1002/hipo.20009.

[23] D. Kumaran and J. L. McClelland, “Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system.,” *Psychol. Rev.*, vol. 119, no. 3, p. 573, 20120709, doi: 10.1037/a0028681.

[24] M. L. Schlichting, J. A. Mumford, and A. R. Preston, “Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex,” *Nat. Commun.*, vol. 6, no. 1, p. 8151, Aug. 2015, doi: 10.1038/ncomms9151.

[25] D. Zeithamova, M. Schlichting, and A. Preston, “The hippocampus and inferential reasoning: building memories to navigate future decisions,” *Front. Hum. Neurosci.*, vol. 6, 2012, Accessed: Feb. 06, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2012.00070

[26] Y. Zheng, X. L. Liu, S. Nishiyama, C. Ranganath, and R. C. O’Reilly, “Correcting the hebbian mistake: Toward a fully error-driven hippocampus,” *PLOS Comput. Biol.*, vol. 18, no. 10, p. e1010589, Oct. 2022, doi: 10.1371/journal.pcbi.1010589.

[27] C. Fernandez, J. Jiang, S.-F. Wang, H. L. Choi, and A. D. Wagner, “Representational integration and differentiation in the human hippocampus following goal-directed navigation,” *eLife*, vol. 12, p. e80281, Feb. 2023, doi: 10.7554/eLife.80281.

[28] R. C. O'Reilly and K. A. Norman, "Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework," *Trends Cogn. Sci.*, vol. 6, no. 12, pp. 505–510, Dec. 2002, doi: 10.1016/S1364-6613(02)02005-3.

[29] K. A. Norman and R. C. O'Reilly, "Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach.," *Psychol. Rev.*, vol. 110, no. 4, pp. 611–646, 2003, doi: 10.1037/0033-295X.110.4.611.

[30] J. L. McClelland and R. C. O'Reilly, "Why There Are Complementary Learning Systems in the Hippocampus and Neocortex:InsightsFrom the Successesand Failuresof Connectionist Models of Learning and Memory," p. 39.

[31] J. M. Barnes and B. J. Underwood, "'Fate' of first-list associations in transfer theory," *J. Exp. Psychol.*, vol. 58, no. 2, pp. 97–105, Aug. 1959, doi: 10.1037/h0047507.

[32] Z. Zhou, D. Singh, M. C. Tandoc, and A. C. Schapiro, "Building integrated representations through interleaved learning." bioRxiv, p. 2021.07.29.454337, Feb. 11, 2023. doi: 10.1101/2021.07.29.454337.

[33] A. C. Schapiro, N. B. Turk-Browne, M. M. Botvinick, and K. A. Norman, "Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 372, no. 1711, p. 20160049, Jan. 2017, doi: 10.1098/rstb.2016.0049.

[34] J. Sučević and A. C. Schapiro, “A neural network model of hippocampal contributions to category learning.” bioRxiv, p. 2022.01.12.476051, Jan. 13, 2022. doi: 10.1101/2022.01.12.476051.

[35] V. J. H. Ritvo, N. B. Turk-Browne, and K. A. Norman, “Nonmonotonic Plasticity: How Memory Retrieval Drives Learning,” *Trends Cogn. Sci.*, vol. 23, no. 9, pp. 726–742, Sep. 2019, doi: 10.1016/j.tics.2019.06.007.

[36] J. Wammes, K. A. Norman, and N. Turk-Browne, “Increasing stimulus similarity drives nonmonotonic representational change in hippocampus,” *eLife*, vol. 11, p. e68344, Jan. 2022, doi: 10.7554/eLife.68344.

[37] V. J. H. Ritvo, A. Nguyen, N. B. Turk-Browne, and K. A. Norman, “Differentiation and Integration of Competing Memories: A Neural Network Model.” bioRxiv, p. 2023.04.02.535239, Apr. 13, 2023. doi: 10.1101/2023.04.02.535239.

[38] M. P. Witter, “Organization of the entorhinal—hippocampal system: A review of current anatomical data,” *Hippocampus*, vol. 3, no. S1, pp. 33–44, Oct. 1993, doi: 10.1002/hipo.1993.4500030707.

[39] R. S. Blumenfeld and C. Ranganath, “Prefrontal Cortex and Long-Term Memory Encoding: An Integrative Review of Findings from Neuropsychology and Neuroimaging,” *The Neuroscientist*, vol. 13, no. 3, pp. 280–291, Jun. 2007, doi: 10.1177/1073858407299290.

[40] A. R. Preston and H. Eichenbaum, “Interplay of Hippocampus and Prefrontal Cortex in Memory,” *Curr. Biol.*, vol. 23, no. 17, pp. R764–R773, Sep. 2013, doi: 10.1016/j.cub.2013.05.041.

[41] Q. Lu, U. Hasson, and K. A. Norman, “A neural network model of when to retrieve and encode episodic memories,” *eLife*, vol. 11, p. e74445, Feb. 2022, doi: 10.7554/eLife.74445.

[42] A. P. Shimamura, P. J. Jurica, J. A. Mangels, F. B. Gershberg, and R. T. Knight, “Susceptibility to Memory Interference Effects following Frontal Lobe Damage: Findings from Tests of Paired-Associate Learning,” *J. Cogn. Neurosci.*, vol. 7, no. 2, pp. 144–152, Apr. 1995, doi: 10.1162/jocn.1995.7.2.144.

[43] C. Ranganath, M. K. Johnson, and M. D’Esposito, “Left Anterior Prefrontal Activation Increases with Demands to Recall Specific Perceptual Information,” *J. Neurosci.*, vol. 20, no. 22, p. RC108, Nov. 2000, doi: 10.1523/JNEUROSCI.20-22-j0005.2000.

[44] B. Melo, G. Winocur, and M. Moscovitch, “False Recall and False Recognition: An Examination of the Effects of Selective and Combined Lesions to the Medial Temporal Lobe/Diencephalon and Frontal Lobe Structures,” *Cogn. Neuropsychol.*, vol. 16, no. 3–5, pp. 343–359, May 1999, doi: 10.1080/026432999380825.

[45] P. E. Wais, O. Montgomery, C. E. L. Stark, and A. Gazzaley, “Evidence of a Causal Role for mid-Ventrolateral Prefrontal Cortex Based Functional Networks in Retrieving High-Fidelity Memory,” *Sci. Rep.*, vol. 8, no. 1, Art. no. 1, Oct. 2018, doi: 10.1038/s41598-018-33164-w.

[46] S. A. Johnson, S. Zequeira, S. M. Turner, A. P. Maurer, J. L. Bizon, and S. N. Burke, “Rodent mnemonic similarity task performance requires the prefrontal cortex,” *Hippocampus*, vol. 31, no. 7, pp. 701–716, 2021, doi: 10.1002/hipo.23316.

[47] D. Swick and R. T. Knight, “Contributions of prefrontal cortex to recognition memory: Electrophysiological and behavioral evidence,” *Neuropsychology*, vol. 13, no. 2, pp. 155–170, 1999, doi: 10.1037/0894-4105.13.2.155.

[48] M. Aly, A. P. Yonelinas, M. M. Kishiyama, and R. T. Knight, “Damage to the lateral prefrontal cortex impairs familiarity but not recollection,” *Behav. Brain Res.*, vol. 225, no. 1, pp. 297–304, Nov. 2011, doi: 10.1016/j.bbr.2011.07.043.

[49] A. Farovik, L. M. Dupont, M. Arce, and H. Eichenbaum, “Medial Prefrontal Cortex Supports Recollection, But Not Familiarity, in the Rat,” *J. Neurosci.*, vol. 28, no. 50, pp. 13428–13434, Dec. 2008, doi: 10.1523/JNEUROSCI.3662-08.2008.

[50] C. Wendelken and S. A. Bunge, “Transitive Inference: Distinct Contributions of Rostrolateral Prefrontal Cortex and the Hippocampus,” *J. Cogn. Neurosci.*, vol. 22, no. 5, pp. 837–847, May 2010, doi: 10.1162/jocn.2009.21226.

[51] L. M. DeVito, C. Lykken, B. R. Kanter, and H. Eichenbaum, “Prefrontal cortex: Role in acquisition of overlapping associations and transitive inference,” *Learn. Mem.*, vol. 17, no. 3, pp. 161–167, Mar. 2010, doi: 10.1101/lm.1685710.

[52] T. R. Koscik and D. Tranel, “The Human Ventromedial Prefrontal Cortex Is Critical for Transitive Inference,” *J. Cogn. Neurosci.*, vol. 24, no. 5, pp. 1191–1204, May 2012, doi: 10.1162/jocn_a_00203.

[53] E. A. Wing, M. C. D’Angelo, A. Gilboa, and J. D. Ryan, “The Role of the Ventromedial Prefrontal Cortex and Basal Forebrain in Relational Memory and Inference,” *J. Cogn. Neurosci.*, vol. 33, no. 9, pp. 1976–1989, Aug. 2021, doi: 10.1162/jocn_a_01722.

[54] K. N. Spalding *et al.*, “Ventromedial Prefrontal Cortex Is Necessary for Normal Associative Inference and Memory Integration,” *J. Neurosci.*, vol. 38, no. 15, pp. 3767–3775, Apr. 2018, doi: 10.1523/JNEUROSCI.2501-17.2018.

[55] S. Becker and J. Lim, “A Computational Model of Prefrontal Control in Free Recall: Strategic Memory Use in the California Verbal Learning Task,” *J. Cogn. Neurosci.*, vol. 15, no. 6, pp. 821–832, Aug. 2003, doi: 10.1162/089892903322370744.

[56] J. C. Hulbert, R. N. Henson, and M. C. Anderson, “Inducing amnesia through systemic suppression,” *Nat. Commun.*, vol. 7, no. 1, Art. no. 1, Mar. 2016, doi: 10.1038/ncomms11003.

[57] W. Xu and T. C. Südhof, “A Neural Circuit for Memory Specificity and Generalization,” *Science*, vol. 339, no. 6125, pp. 1290–1295, Mar. 2013, doi: 10.1126/science.1229534.

[58] D. Zeithamova, A. L. Dominick, and A. R. Preston, “Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference,” *Neuron*, vol. 75, no. 1, pp. 168–179, Jul. 2012, doi: 10.1016/j.neuron.2012.05.010.

[59] M. C. Anderson, J. G. Bunce, and H. Barbas, “Prefrontal–hippocampal pathways underlying inhibitory control over memory,” *Neurobiol. Learn. Mem.*, vol. 134, pp. 145–161, Oct. 2016, doi: 10.1016/j.nlm.2015.11.008.

[60] M. J. D. der Weel *et al.*, “The nucleus reuniens of the thalamus sits at the nexus of a hippocampus and medial prefrontal cortex circuit enabling memory and

behavior," *Learn. Mem.*, vol. 26, no. 7, pp. 191–205, Jul. 2019, doi: 10.1101/lm.048389.118.

[61] R. A. Diana, A. P. Yonelinas, and C. Ranganath, "Imaging recollection and familiarity in the medial temporal lobe: a three-component model," *Trends Cogn. Sci.*, vol. 11, no. 9, pp. 379–386, Sep. 2007, doi: 10.1016/j.tics.2007.08.001.

[62] M. Petrides and D. N. Pandya, "Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey," *Eur. J. Neurosci.*, vol. 16, no. 2, pp. 291–310, 2002, doi: 10.1046/j.1460-9568.2001.02090.x.

[63] H. Kondo, K. S. Saleem, and J. L. Price, "Differential connections of the perirhinal and parahippocampal cortex with the orbital and medial prefrontal networks in macaque monkeys," *J. Comp. Neurol.*, vol. 493, no. 4, pp. 479–509, Dec. 2005, doi: 10.1002/cne.20796.

[64] H. T. Ito, S.-J. Zhang, M. P. Witter, E. I. Moser, and M.-B. Moser, "A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation," *Nature*, vol. 522, no. 7554, pp. 50–55, Jun. 2015, doi: 10.1038/nature14396.

[65] M. P. Witter *et al.*, "Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways," *Hippocampus*, vol. 10, no. 4, pp. 398–410, 2000, doi: 10.1002/1098-1063(2000)10:4<398::AID-HIPO6>3.0.CO;2-K.

[66] R. Koster *et al.*, "Big-Loop Recurrence within the Hippocampal System Supports Integration of Information across Episodes," *Neuron*, vol. 99, no. 6, pp. 1342-1354.e6, Sep. 2018, doi: 10.1016/j.neuron.2018.08.009.

[67] T. J. Teyler and P. DiScenna, “The hippocampal memory indexing theory,” *Behav. Neurosci.*, vol. 100, no. 2, pp. 147–154, Apr. 1986, doi: 10.1037//0735-7044.100.2.147.

[68] H. Eichenbaum and M. Bunsey, “On the Binding of Associations in Memory: Clues From Studies on the Role of the Hippocampal Region in Paired-Associate Learning,” *Curr. Dir. Psychol. Sci.*, vol. 4, no. 1, pp. 19–23, Feb. 1995, doi: 10.1111/1467-8721.ep10770954.

[69] T. J. Teyler and J. W. Rudy, “The hippocampal indexing theory and episodic memory: Updating the index,” *Hippocampus*, vol. 17, no. 12, pp. 1158–1169, 2007, doi: 10.1002/hipo.20350.

[70] S. Mackey and M. Petrides, “Architecture and morphology of the human ventromedial prefrontal cortex,” *Eur. J. Neurosci.*, vol. 40, no. 5, pp. 2777–2796, 2014, doi: 10.1111/ejn.12654.

[71] D. M. Barch, T. S. Braver, F. W. Sabb, and D. C. Noll, “Anterior Cingulate and the Monitoring of Response Conflict: Evidence from an fMRI Study of Overt Verb Generation,” *J. Cogn. Neurosci.*, vol. 12, no. 2, pp. 298–309, Mar. 2000, doi: 10.1162/089892900562110.

[72] M. S. Fleck, S. M. Daselaar, I. G. Dobbins, and R. Cabeza, “Role of Prefrontal and Anterior Cingulate Regions in Decision-Making Processes Shared by Memory and Nonmemory Tasks,” *Cereb. Cortex*, vol. 16, no. 11, pp. 1623–1630, Nov. 2006, doi: 10.1093/cercor/bhj097.

[73] M. Herrmann *et al.*, “Control of semantic interference in episodic memory retrieval is associated with an anterior cingulate-prefrontal activation pattern,” *Hum. Brain Mapp.*, vol. 13, no. 2, pp. 94–103, 2001, doi: 10.1002/hbm.1027.

[74] S.-H. Wang, D. Tse, and R. G. M. Morris, “Anterior cingulate cortex in schema assimilation and expression,” *Learn. Mem.*, vol. 19, no. 8, pp. 315–318, Aug. 2012, doi: 10.1101/lm.026336.112.

[75] F. A. Mansouri, E. Koechlin, M. G. P. Rosa, and M. J. Buckley, “Managing competing goals — a key role for the frontopolar cortex,” *Nat. Rev. Neurosci.*, vol. 18, no. 11, Art. no. 11, Nov. 2017, doi: 10.1038/nrn.2017.111.

[76] J.-C. Dreher, E. Koechlin, M. Tierney, and J. Grafman, “Damage to the Fronto-Polar Cortex Is Associated with Impaired Multitasking,” *PLOS ONE*, vol. 3, no. 9, p. e3227, Sep. 2008, doi: 10.1371/journal.pone.0003227.

[77] K. Christoff and J. D. E. Gabrieli, “The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex,” *Psychobiology*, vol. 28, no. 2, pp. 168–186, Jun. 2000, doi: 10.3758/BF03331976.

[78] M. Ritchey, L. A. Libby, and C. Ranganath, “Cortico-hippocampal systems involved in memory and cognition,” in *Progress in Brain Research*, vol. 219, Elsevier, 2015, pp. 45–64. doi: 10.1016/bs.pbr.2015.04.001.

[79] P. S. Goldman-Rakic, L. D. Selemon, and M. L. Schwartz, “Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey,” *Neuroscience*, vol. 12, no. 3, pp. 719–743, Jul. 1984, doi: 10.1016/0306-4522(84)90166-0.

[80] F. Kouneiher, S. Charron, and E. Koechlin, “Motivation and cognitive control in the human prefrontal cortex,” *Nat. Neurosci.*, vol. 12, no. 7, Art. no. 7, Jul. 2009, doi: 10.1038/nn.2321.

[81] A. A. Taren, V. Venkatraman, and S. A. Huettel, “A Parallel Functional Topography between Medial and Lateral Prefrontal Cortex: Evidence and Implications for Cognitive Control,” *J. Neurosci.*, vol. 31, no. 13, pp. 5026–5031, Mar. 2011, doi: 10.1523/JNEUROSCI.5762-10.2011.

[82] D. Kumaran, “Schema-driven facilitation of new hierarchy learning in the transitive inference paradigm,” *Learn. Mem.*, vol. 20, no. 7, pp. 388–394, Jul. 2013, doi: 10.1101/lm.030296.113.

CHAPTER 2: Prefrontal Control of the Hippocampus: A Computational Model

INTRODUCTION

The ability to integrate [1]–[3] and differentiate [4]–[6] across similar items or events is canonically associated with the hippocampus. Integration – or the increased similarity between item or event representations – is typically investigated using inference [7]–[9] or statistical learning tasks [2]. For example, in a standard associative inference task, participants encode a set of "AB" pairs, where A is an image of an item and B is scene image, and a set of "BC" pairs, where B is the same scene image and C is a new item image. At test, the participants are given item A and asked to pair it with C or a novel item D. If integration is successful, when given A, participants should be able to recognize its relationship with C (i.e., a shared B). In the brain, this integration is shown through an increased similarity in the representational patterns for items A and C from encoding to test [7]–[9]. fMRI studies investigating integration show overall increased activity in the anterior hippocampus (namely, the CA1) [10], or more specifically, an increase in voxel pattern similarity in the CA1 for A and C items from encoding to test [11].

In contrast, differentiation – or the decreased similarity between item or event representations – is investigated through mnemonic similarity tasks. In such a task, participants encode a set of items, then during test, they are presented with the same items (i.e., studied), items similar to studied items (i.e., lures), and novel items (i.e., foils) and have to mark them as "old" (already studied) or "new" (novel). For successful differentiation, the participants must be able to recognize that the lure items are different

from the studied items, thus marking them as “new” [6], [12], [13]. Differentiation is often associated with the dentate gyrus (DG) and CA3 subfields of the hippocampus, such that voxel pattern similarity for studied and lure items become comparable [14]–[17]. These studies typically include a small number of repetitions (once or twice) during encoding and test, however, when there is repeated retrieval, another phenomenon emerges: repulsion. Repulsion is the decrease in representational similarity in the hippocampus for similar items with repeated retrieval such that the voxel pattern similarity for these dissimilar items is greater than similar items [18], [19]. Ultimately, both standard differentiation and repulsion tasks showcase a change in representational similarity between similar items related to better performance as the representations decrease in similarity.

Due to the implication of hippocampal processing for these opposing tasks – integration and differentiation – there are still outstanding questions: How is the hippocampus able to achieve both tasks? And how can it switch between these two task goals? We propose that a control system, namely the prefrontal cortex (PFC), is interacting with the hippocampus to switch between integration and differentiation tasks. This is supported by literature indicating that various subregions of the PFC are necessary in both integration [20]–[24] and differentiation [25]–[27]. During both transitive inference (e.g., if A<B and B<C, then A<C) [20], [22] and associative inference tasks [24], damage or inactivation of the ventromedial PFC (VMPFC) is associated with impaired performance. Similarly, for mnemonic discrimination tasks, damage or inactivation of the medial PFC [25]–[27] causes increased false alarms, indicating a difficulty in determining that a lure item has not been studied.

Retrieval suppression literature, which focuses on diverting attention from distracting information to focus on task goals, suggests that the PFC may be controlling the hippocampus through two pathways: 1) PFC -> ACC -> entorhinal cortex (EC) or 2) PFC -> nucleus reuniens -> hippocampal CA1 [28], [29]. These pathways are in addition to the well-studied pathway from PFC to areas, such as the parahippocampal and perirhinal cortices, which feed into the EC, ultimately landing in the hippocampal circuitry [30], [31]. We suggest that, through these pathways, the PFC can inhibit the information that is "distracting" (i.e., the overlapping feature information between similar items) so that differentiation can occur. Here, we propose a simple, proof-of-concept hippocampal-PFC model that uses inhibitory connections to switch between integration and differentiation task goals.

METHODS

Model Structure

The current model, is an iteration of a previous hippocampal model , based on the Complementary Learning Systems (CLS) model [2-3] as part of the Leabra framework using Emergent software [32], [33]. See https://github.com/emer/leabra for all related documentation and code. Briefly, the previous hippocampal model includes an Input layer, which contains "pools" of patterns (i.e., subset of patterns representing item features). This layer feeds into the EC superficial layer (ECin), which then projects to the DG and CA3. The DG and CA3 have high levels of inhibition, which allow them to showcase their sparse coding and, therefore, any competition that arises between items. The CA3 includes projections to itself (i.e., recurrent connections) and to CA1.

The CA1 then projects to the deep layers of the EC (ECout). Using these pattern representations, we can compare the patterns during learning and retrieval to determine how well the model is able to represent learned items at the end of the hippocampal loop.

Our current model expands upon this hippocampal model by adding a simple PFC, as a proof of concept. Fig. 1 shows the model structure with a 6-unit PFC, where each unit corresponds to an entire pool in the input layer. Importantly, this model uses these pools to represent features of the item. Here, we use these pools to represent overlapping and non-overlapping information between items. Items that share information have very similar patterns (i.e., 80% units the same) in pools that represent shared features. Other pools represent other unshared ("distinct") information, which is related to the type of task (explained further below). When the PFC is "off" there is no modulation of the hippocampus. When the PFC is "on", specific units can be turned on, which then inhibit the patterns in corresponding pools in the ECin and CA1. More specifically, the PFC is turned on during the plus phase of the alpha cycle, which is the period after the expectation is presented (i.e., minus phase) where the outcome is observed [34].

Model Testing

Integration and Differentiation

To model the ways that the PFC may be influencing the hippocampus to integrate vs differentiate similar items, we used a task that involves switching between these two processes [12], [13]. In this task, participants learn common, easily

recognizable items. After each learning block, participants performed a test phase, which included previously learned items (“studied”), items similar to previously learned items (i.e., altered dimensions or reduced/enlarged versions; “lure”), or novel items (“foil”). During this test phase, they were either given a general test or a specific test. During the general test, participants were expected to respond to both studied and lure items as “old” and foils as “new”. This type of test represents an integration task – the goal is to integrate the representations of studied and lure items. In contrast, in one version of the specific test [12], participants were expected to respond to studied items as “old” and lure and foil items as “new”. In another version [13], the specific test required participants to respond to studied items as “old/larger”, lure items as “old/smaller”, and foil items as “new”. In both versions, the specific test represents a discrimination task – the goal is to discriminate between studied items and similar lures.

Figure 2 shows an example of the studied, lure, and foil pattern representations. For studied and lure items, half of the pools (i.e., 3 out of 6) are the same patterns, representing the shared information between studied and lure items, in a paired fashion. In contrast, the foil patterns are completely different from both studied and lure items [may want to get into the percentages here and compare the patterns to get the exact numbers]. The other half of the pools represented information that was distinct to each item. The model received 10 item patterns per run (“studied” items) during the training phase, then during test, the model received 5 studied, 5 lures, and 5 foil item patterns. During the general test, the PFC was off, indicating no PFC modulation of the hippocampus. In contrast, during the specific test, the PFC units corresponding to the “shared” pools of the input layer was turned on to inhibit the possible shared information

between objects. The reason for turning the PFC on during the specific test only is because in an fMRI study using this task, researchers found that the PFC was preferentially active during the specific test compared to the general test [13]. Importantly, this PFC is not discriminatory across items, meaning that the same pools are inhibited for studied, lure, *and* foil items, which is consistent with empirical findings [13]. For each run (of which there are 20, representing separate participants), there were four epochs. Each epoch represents a cycle through the model. For the first two epochs, a general test was performed (i.e., PFC off), and for the final two epochs, a specific test was performed (i.e., PFC on). When the PFC was on, the inhibition occurred specifically during the third quarter of the alpha cycle, which is associated with the "plus" phase. The plus phase is the period in the cycle where there is a comparison between the expectation and the outcome, allowing for error-driven learning to occur [34].

To determine the change in similarity between studied and lure items, we compared the representational patterns for each item type: 1) studied patterns during test to studied item patterns during training, 2) lure item patterns during test to its similar studied item during training, and 3) foil item patterns during test to a studied item during training. To quantify these differences, we analyzed the difference in unit-by-unit similarity in the ECout. We then analyzed the change in ECout similarity between the general test and specific test to show how the item types change in their similarity to studied item patterns. Accuracy for determining "old" responses versus "new" responses were calculated by characterizing "old" as sharing over 50% of the studied item patterns and "new" as sharing less than 50% of those units.

Repulsion

To model repulsion, we used a modified version of a task from Chanales and colleagues [18]. In the original task, participants viewed 4 routes in a city, each with distinct destinations. For each route, participants were shown 98 sequential images along the path of the route. These routes were paired such that two sets of routes shared many similar images (74-77 shared images that were similar but not identical before diverging), but the sets did not overlap with each other (e.g., Route 1 and Route 2 overlapped, Route 3 and Route 4 overlapped, but Routes 1 and 2 did not overlap with Routes 3 and 4). Participants viewed these routes twice per run for a total of 14 runs. Following each run, participants were shown an image along one of the learned routes and were asked if that image was related to the target destination, the lure destination (the route which shared a great deal of similarity), or a destination from a non-overlapping route.

In our modified version, we had the model learn a full input pattern representing the each of the 4 training routes (Fig. 3), with pools representing overlapping information, information distinct to each route, and the route destinations. The model was trained on these patterns across 28 epochs to represent 28 interactions with the route, in line with the original task. After each learning session, the model was tested on the routes by receiving an input pattern with a missing destination, mimicking a cued recall task rather than a forced-alternative choice task as with the original. For the first half of the epochs (14), the PFC was off, which was followed by 14 epochs with the PFC on. When the PFC was on, the pools corresponding to overlapping information were inhibited in the ECin and CA1. Like the study above, this PFC inhibition occurred during

the plus phase of the alpha cycle [34]. For each layer, representational similarity analyses were performed to determine how similar the representations for overlapping vs non-overlapping routes changed when the PFC was off vs on. As a control, we also performed a similar simulation where the distinct pools were inhibited in ECin and CA1 to show that it is specifically the manipulation of the overlapping pools that showcase repulsion.

To determine the similarity changes between overlapping and non-overlapping routes, we performed representational similarity analyzes in EC, CA1, CA3, and DG. To compare directly to the empirical data [18], we averaged across CA1, CA3, and DG to show the full hippocampal representational similarities across route types in addition to each subfield alone. We performed a repeated measures ANOVA (with/without PFC inhibition x route type) to show the differences between route types between the first half (without PFC) and second half (with PFC) of training and test. To further characterize the changes in similarity between routes as it relates to repulsion, we analyzed the patterns similarity differences in ECout for the overlapping pools only. By doing this analysis, we can investigate how PFC inhibition of these pools changes the similarity for overlapping information.

RESULTS

Integration and Differentiation

To investigate the accuracy of the model (i.e., old vs new responses), we analyzed the similarity of test items to studied items during training (see Fig 4a for single representative run). If the similarity was above 50%, the trial was marked as “old”,

and if it was below, it was marked as “new”. We calculated accuracy for both the general and specific tests. The general test results showed that the model was 100% accurate for studied and foil items while lure items’ accuracy was 86%. The specific test results showed that the model was 100% accurate for lure and foil items while studied items’ accuracy was 92% (Fig. 4b). These results are in line with the empirical results [6-7], such that both the model and participants are able to perform well at both tasks.

Repulsion

To investigate the differences in representational similarity between overlapping and non-overlapping routes as a function of PFC inhibition, we analyzed the similarity in the hippocampus and EC (Fig. 5). Results revealed that in the hippocampus (all subfields averaged; Fig. 5a), there was a main effect of route type ($F(1)=46.85$, $p<0.001$) and PFC inhibition ($F(1)=36.17$, $p<0.001$), such that overlapping routes were more similar than non-overlapping routes and similarity being greater when there was no PFC inhibition. There was also an interaction ($F(1)=53.98$, $p<0.001$), where there was significantly greater similarity for overlapping routes compared to non-overlapping routes when there was no PFC inhibition ($p<0.001$), which was not present when there was PFC inhibition ($p=0.44$). In addition, there was greater similarity for overlapping routes when there was no PFC inhibition compared to when there was PFC inhibition ($p<0.001$). This was not the case for non-overlapping routes ($p=0.059$). This indicates that there a significant decrease in similarity between the without PFC and with PFC conditions for overlapping routes specifically. Similarly, in EC (Fig. 5b), there was a main effect of route type ($F(1)=1323.3$, $p<0.001$) and PFC inhibition ($F(1)=100.2$, $p<0.001$), such that overlapping routes were more similar than non-overlapping routes

and similarity being greater when there was no PFC inhibition. There was also an interaction ($F(1)=2397.8$, $p<0.001$), where there was significantly greater similarity for overlapping routes compared to non-overlapping routes when there was no PFC inhibition ($p<001$), but similarity for non-overlapping routes was greater than overlapping routes when there was PFC inhibition ($p<0.001$). In addition, there was greater similarity for similarity for overlapping routes without PFC inhibition compared to when there was PFC inhibition ($p<0.001$). The opposite was true for non-overlapping routes, such that similarity was greater with PFC inhibition compared to when there was no PFC inhibition ($p<0.001$).

Additionally, we performed ANOVAs for each subfield of the hippocampus. In DG (Fig. 5c), there was a main effect of route type ($F(1)=5.29$, $p=0.028$) and PFC inhibition ($F(1)=12.91$, $p=0.001$), such that overlapping routes were more similar than non-overlapping routes and similarity being greater when there was no PFC inhibition. There was also an interaction ($F(1)=5.28$, $p=0.028$), where there was significantly greater similarity for overlapping routes compared to non-overlapping routes when there was no PFC inhibition ($p=0.011$), which was not present when there was PFC inhibition ($p=0.15$). In addition, there was greater similarity for overlapping routes when there was no PFC inhibition compared to when there was PFC inhibition ($p=0.002$). This was also the case for non-overlapping routes ($p<0.001$). This indicates that there a significant decrease in similarity between the without PFC and with PFC conditions. In CA3 (Fig. 5d), there was a main effect of route type ($F(1)=4.95$, $p=0.033$), such that overlapping route similarity was greater than non-overlapping route similarity. There was no effect of PFC inhibition or interaction. This indicates that in CA3, similarity for overlapping routes

was greater than non-overlapping routes regardless of PFC inhibition. In CA1 (Fig. 5e), there was a main effect of route type ($F(1)=92.43$, $p<0.001$) and PFC inhibition ($F(1)=182.08$, $p<0.001$), such that overlapping routes were more similar than non-overlapping routes and similarity being greater when there was no PFC inhibition. There was also an interaction ($F(1)=195.19$, $p<0.001$), where there was significantly greater similarity for overlapping routes compared to non-overlapping routes when there was no PFC inhibition ($p<0.001$) However, when there was PFC inhibition, non-overlapping routes showed greater similarity than overlapping routes ($p=0.0098$). In addition, there was greater similarity for overlapping routes when there was no PFC inhibition compared to when there was PFC inhibition ($p<0.001$). This was not the case for non-overlapping routes ($p=0.52$). This indicates that there a significant decrease in similarity between the without PFC and with PFC conditions for overlapping routes specifically.

To show that it was specifically the inhibition of the overlapping feature pools, we ran two control simulations: one with no inhibition of any pools (Fig. 6) and one with inhibition of the distinct pools instead of overlapping (Fig. 7). In the no inhibition control (route type x first vs second half of train/test), in the whole hippocampus, there was a main effect of route type, such that overlapping route similarity was greater than non-overlapping routes ($F(1)=87.96$, $p<0.001$; Fig. 6a). There was no effect from the first half of train/test to the second half, and there was no interaction. In EC (Fig. 6b), there was a main effect of route type, such that overlapping route similarity was greater than non-overlapping routes ($F(1)=33994.4$, $p<0.001$), and a main effect of train/test half ($F(1)=226.02$, $p<0.001$), such that similarity overall across routes was greater during the first half. There was also an interaction ($F(1)=65.94$, $p<0.001$). In DG (Fig. 6c), there

was a main effect of first half vs second half of training/test ($F(1)=5.4$, $p=0.03$), such that similarity was higher across route types for the first half compared to the second half. This indicates that over time, the DG is working to pattern separate the similar routes. There was no main effect of route type or an interaction between the two variables. Conversely, in CA3 (Fig. 6d), there was a main effect of first half vs second half of training/test ($F(1)=6.84$, $p=0.013$), such that similarity was higher across route types for the second half compared to the first half. There was no main effect of route type or an interaction between the two variables. In CA1 (Fig. 6e), there was a main effect of route type ($F(1)=377.4$, $p<0.001$), such that similarity was greater for overlapping routes compared to non-overlapping routes, and a main effect of first vs second half ($F(1)=14.7$, $p<0.001$), such that similarity was greater in the first half compared to the second half. There was no interaction. This control shows that repulsion does no occur when there is no PFC inhibition of overlapping pools.

In the distinct feature pool inhibition control (route type x PFC inhibition; Fig. 7), in the whole hippocampus (Fig. 7a), there was a main effect of route type ($F(1)=352.82$, $p<0.001$), such that similarity for overlapping routes was greater than non-overlapping routes. There was no main effect of PFC inhibition, but there was an interaction ($F(1)=14.76$, $p<0.001$), which showed a greater difference in similarity between route types with PFC inhibition compared to without. In EC (Fig. 7b), there was a main effect of route type ($F(1)=28998.6$, $p<0.001$) and PFC inhibition ($F(1)=3829.3$, $p<0.001$), such that overlapping routes were more similar than non-overlapping routes and similarity being greater when there was PFC inhibition. There was also an interaction ($F(1)=649.4$, $p<0.001$), which showed a greater difference in similarity between route

types with PFC inhibition compared to without. In DG (Fig. 7c), there was a main effect of route type ($F(1)=10.87$, $p=0.0024$), such that overlapping route similarity was greater than non-overlapping routes. There was no main effect of PFC inhibition or an interaction. In the CA3 (Fig. 7d), there was no main effect of route type or PFC inhibition. Additionally, there was no interaction, suggesting that in CA3, there was no significant difference between all variables. In CA1 (Fig. 7e), there was a main effect of route type ($F(1)=872.3$, $p<0.001$), such that overlapping route similarity was greater than non-overlapping routes. There was no main effect of PFC inhibition, however there was an interaction ($F(1)=20.67$, $p<0.001$), showcasing a greater difference between overlapping and non-overlapping route similarity when there was PFC inhibition compared to when there was no PFC inhibition. This control showcases that when there is inhibition of the distinct pools, instead of the overlapping pools, there is no repulsion in any region.

Additionally, we investigated the similarity in ECout of the overlapping pools during test to their training patterns over the 28 runs for the main simulation and the controls (Fig. 8). During the first half of training/test, all three model simulations reach 100% similarity. Then during the second half, both the overlapping pool inhibition simulation and the distinct pool inhibition simulation show a decrease in similarity (plateauing at 75% similar) to training patterns for the overlapping pools. In contrast, for the no inhibition simulation, similarity to training patterns for overlapping pools remains at 100%. This indicates that when there is no inhibition, the overlapping feature representation reaches 100% and stays constant. In contrast, when there is inhibition of any feature (overlapping or distinct), the overlapping pool training to test similarity

decreases, which shows an update in the route representations. Despite the similar trend for both simulations, the results above show that repulsion only occurs when there is overlapping pools inhibition.

DISCUSSION

Historically, the hippocampus is viewed as the key to episodic memory processing with the PFC being a region to store representations during consolidation [4], [32]. However, our work shows that the hippocampus and PFC work together, dynamically, to update representations to match task goals. Specifically, our model showcases the possibility that the PFC is working as a controller of the hippocampus to enable switching between integration and differentiation – two opposing concepts. Our results support the idea that the PFC suppresses the overlapping features between highly similar items to divert attention away from the information distracting from differentiation processes. Through the Ranganath et al. studies [12], [13], we show that when the overlapping features between studied and lure items are inhibited, the model is better able to differentiate the differences between the items. This is indicated by the shift for the lure items being treated as “old” for the general test and “new” for the specific test (Fig. 4). In the Chanales et al. repulsion study [18], we show that repulsion occurs in the EC and CA1 when there is inhibition of overlapping features (Fig. 5).

As a simple proof-of-concept, this first iteration of a hippocampal-PFC model has a few limitations and leaves room for future models to expand upon the findings discussed here. One such limitation is the similarity between the original empirical tasks and the tasks in the model, especially in the repulsion task [18]. As described in detail in

the Methods, we modeled the simplest version of the task. In this simple version, the 98 stimuli associated with each of the routes was compressed into a single input pattern into the computational model. In doing so, the routes were easier for the model to learn, leading accuracy to be at ceiling regardless of repulsion. In the literature, repulsion typically leads to better performance in distinguishing between two similar routes [18], [19], which we could not produce due to this compression. To reproduce the performance effects in future iterations of the model, we would create separate inputs for each stimulus in each route with “route” pools to represent which route the image was associated. However, as we were focusing on the replication of the representational effects, the results currently presented are sufficient to make that point. Relatedly, for both tasks, there is an additional limitation in relation to the stimuli. As we did not use the actual stimuli, instead using a pattern to represent the stimuli, we are not able to capture the variability in similarity between studied and lure items. In a more complex version of the model, it may be worth it to more directly translate the stimuli into input patterns. By doing this, an additional variable – the similarity (e.g., visual, semantic, etc.) between items – can be investigated to show how integration/differentiation/repulsion effects change as item similarity changes.

One of the biggest limitations of the model is the lack of empirical research to suggest that the PFC can specifically suppress the neuronal populations associated with overlapping content across similar stimuli. As discussed in the introduction, there is research to support the idea that the PFC works as a controller of the hippocampus in both integration and differentiation processes, as when the connection between the hippocampus and PFC is severed or specific inhibition of related PFC regions [20]–[24],

there is an inability to complete integration and differentiation tasks at the same capacity as an intact PFC. This suggests that the PFC is changing/updating representations to help the hippocampus complete these tasks. Logically, we suggest that in order to differentiate between similar stimuli, there must be either an increased excitation of the neurons corresponding to the distinct features for each item or inhibition of the overlapping features, as supported by the retrieval suppression literature [28]. To fully test this proof-of-concept model, we suggest creation of a study that 1) investigates integration and differentiation such that both tasks are performed and 2) analyzes the functional connectivity between the hippocampus and PFC on the item feature level. With such a study, the assumption we are making here can be investigated further and, ultimately, be used to strengthen (or possibly disprove) the current model.

As described in the Methods, our PFC is a simple 6-unit layer that is used as an inhibitory connection to the EC and CA1, as supported by the literature [28], [29]. This leaves room for future iterations of the model to include a more complex PFC. One way to increase the complexity of the model is to have a larger PFC layer that can learn representations, maintain those representations, and update them according to task goals. In the current model, the PFC is manually turned on and off, but in future iterations, the model would simply need to be given the goal of the task, prompting the PFC to turn on and off as needed per task demands. Another way to increase the complexity of this proposed hippocampal-PFC control system would be to create layers (e.g., VMPFC, ACC, nucleus reuniens) representing the various routes to/from the PFC, as discussed in the Introduction. Additionally, the current model's PFC inhibition is kept

high and steady. With a more complex iteration, the PFC would be able to vary this inhibition accordingly.

Despite these limitations, the current model supports the possibility two concepts: 1) updating representations via PFC control allows for switching between integration and differentiation tasks and 2) repulsion can occur through inhibition of overlapping features. In proving these concepts, this model is the first of its kind to showcase these types of interactions between the hippocampus and PFC. As the first brick in the foundation of hippocampal-PFC models, we have shown the possibilities for more complex versions to come, which can then be used to describe populations with hippocampal and/or PFC dysfunction.

FIGURES

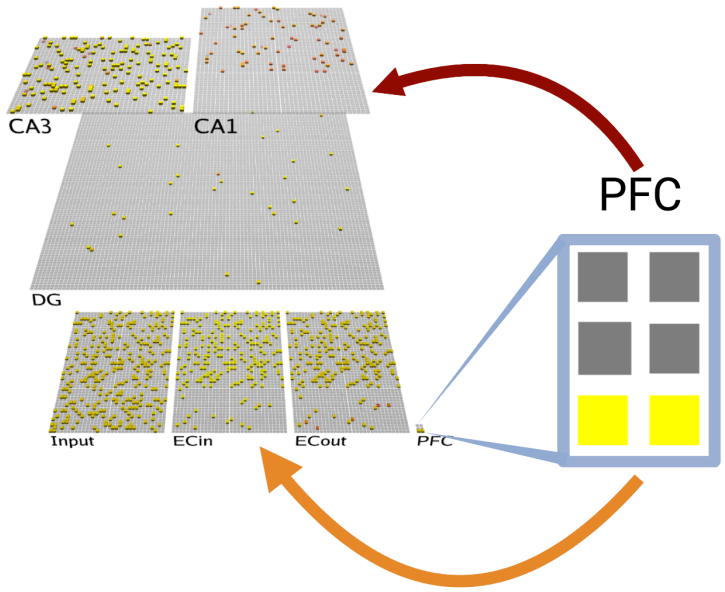

Figure 1. Schematic of the Hippocampal-PFC Model. The hippocampal aspect of the model includes includes the hippocampal loop: Input -> ECin -> DG -> CA3 -> CA1 -> ECout -> ECin. The PFC is a simple 6-unit layer. When the units of this PFC are “on”, there is inhibition of the pools representing the overlapping content between item representations in both the EC and CA1.

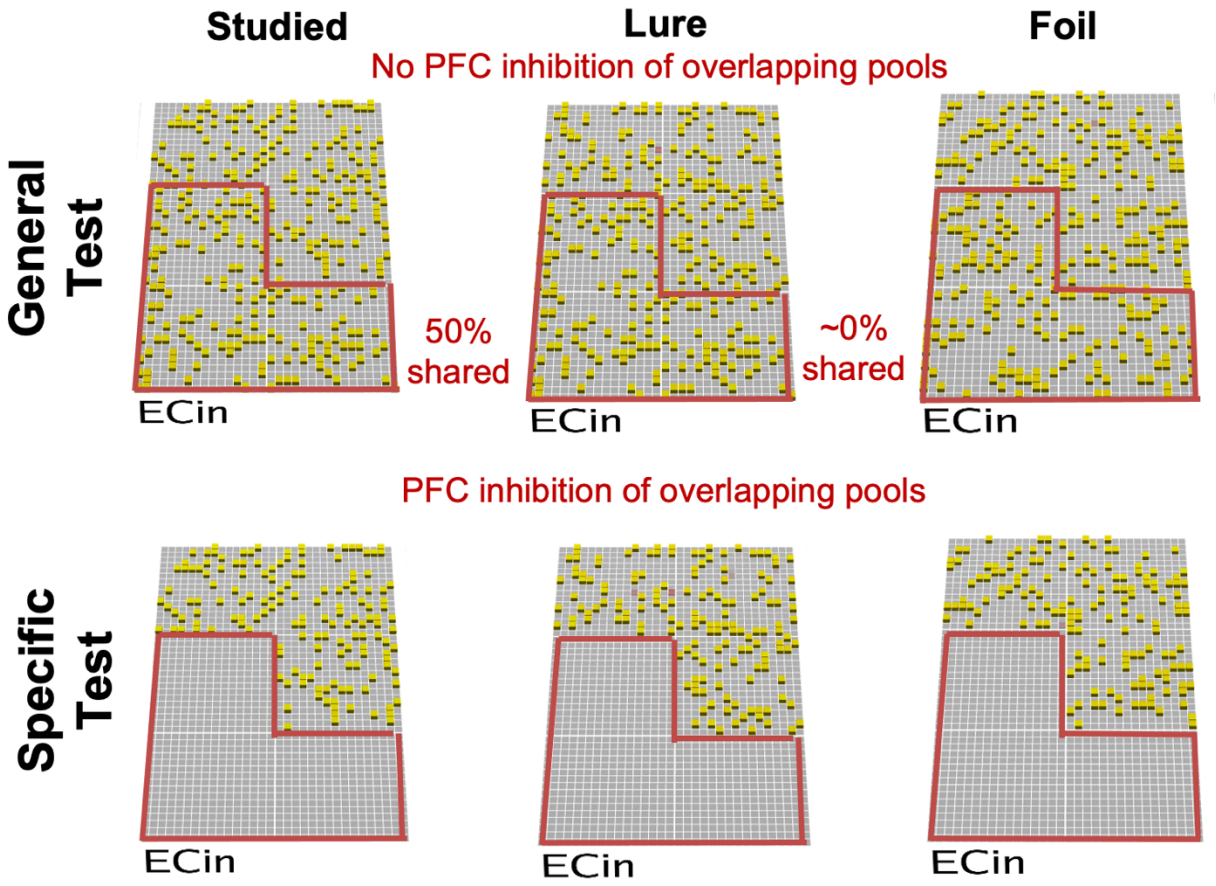

Figure 2. Item Representation Patterns in the Entorhinal Cortex Input Layer.

During the general test, 50% of the item representation is shared between the studied and lure items. Foil representations do not include any shared content between items. During the specific test, this shared information (or where the shared information would be stored for foil items) is inhibited.

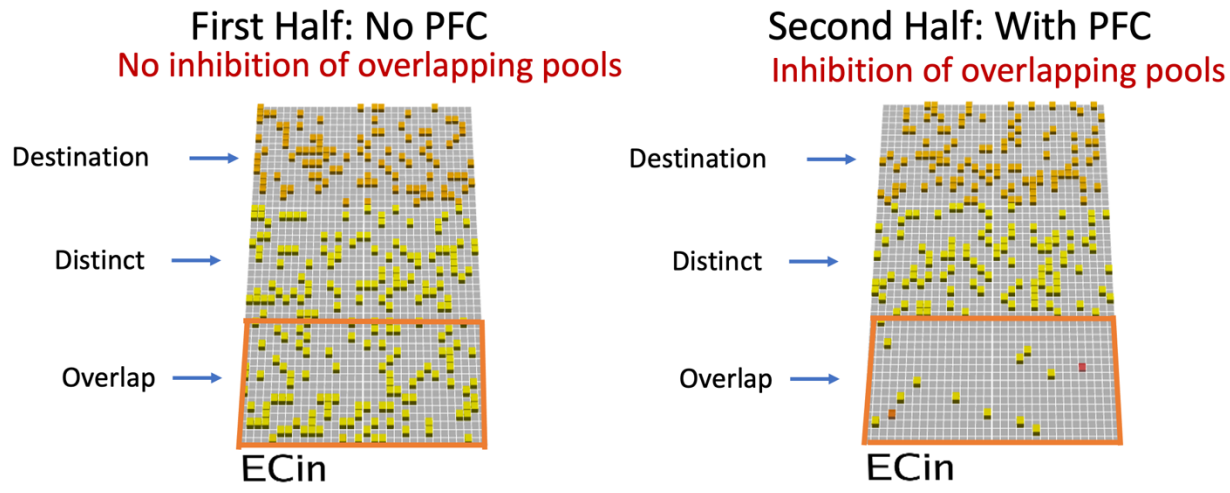

Figure 3. Repulsion Task Patterns. Each route is represented as patterns with pools representing the overlapping features, distinct features, and the route destination. During the first half of training and test, there is no PFC inhibition, and during the second half, there is inhibition of the overlapping pools.

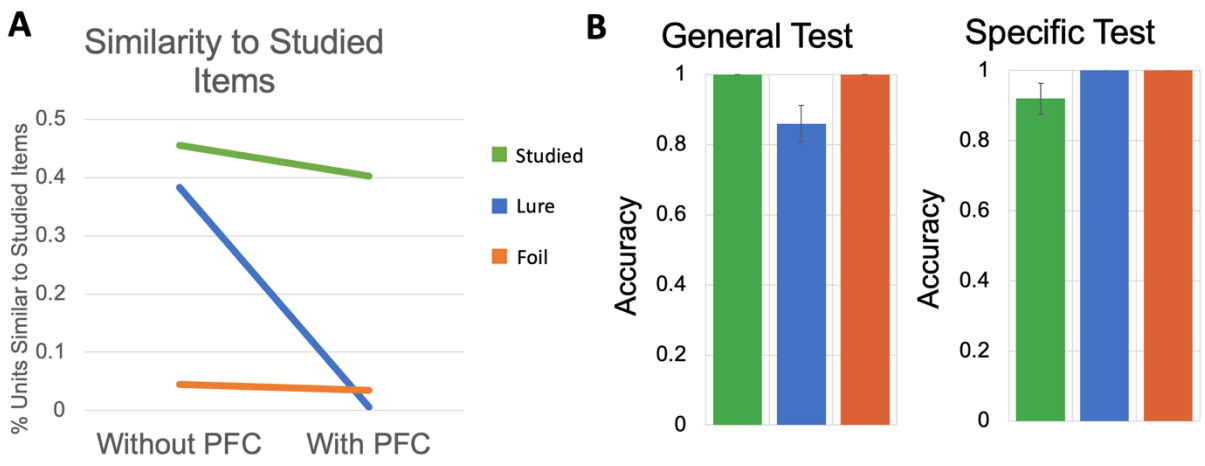

Figure 4. Integration and Differentiation Results. A) The % units similar to studied items changes depending on PFC inhibition. When the PFC is off (i.e., the general test), the lure similarity to comparable to the studied items. When the PFC is on (i.e., the specific test), the lure similarity to studied items decreases to be in line with foil items. **B)** Accuracy (calculated by the similarity to studied items) shows that the model can perform well for both the general test (where lures are considered "old") and specific test (where lures are "new").

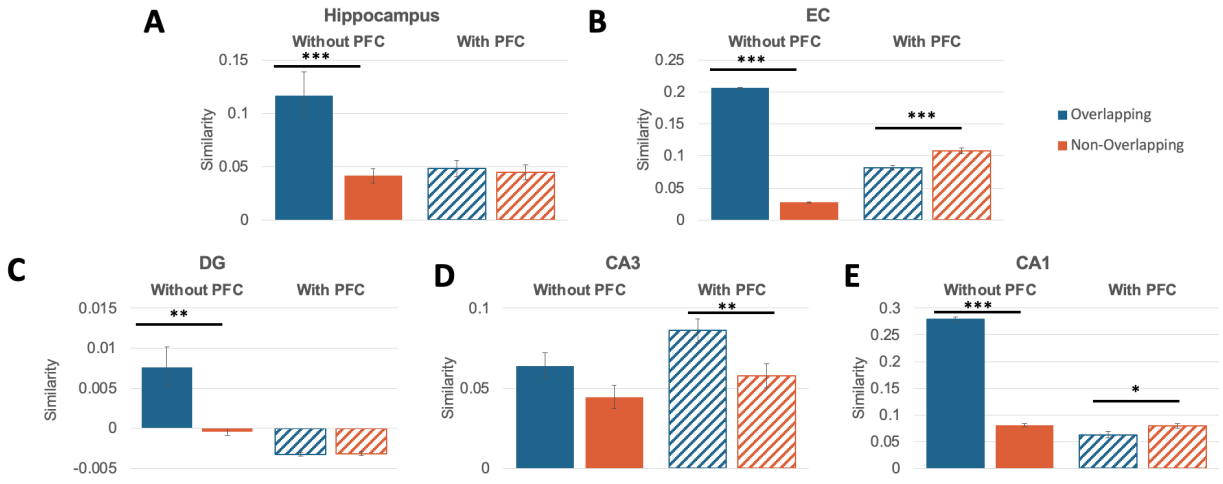

Figure 5. Pattern Similarity by ROI: Inhibition of overlapping pools. A) The whole hippocampus analysis reveals greater pattern similarity for overlapping routes compared to non-overlapping routes when there is no PFC inhibition ($p<0.001$) that is not present when there is PFC inhibition. **B)** In EC, when there is no PFC inhibition, pattern similarity is greater for overlapping routes compared to non-overlapping routes ($p<0.001$), and this relationship inverses when there is PFC inhibition ($p<0.001$), indicating repulsion. **C)** In DG, when there is no PFC inhibition, pattern similarity is greater for overlapping routes compared to nonoverlapping routes ($p=0.011$), which is not present when there is PFC inhibition. **D)** In CA3, regardless of PFC inhibition, pattern similarity is greater for overlapping routes compared to non-overlapping routes ($p=0.033$). **E)** In CA1, when there is no PFC inhibition, pattern similarity is greater for overlapping routes compared to non-overlapping routes ($p<0.001$), and this relationship inverses when there is PFC inhibition ($p=0.0098$).

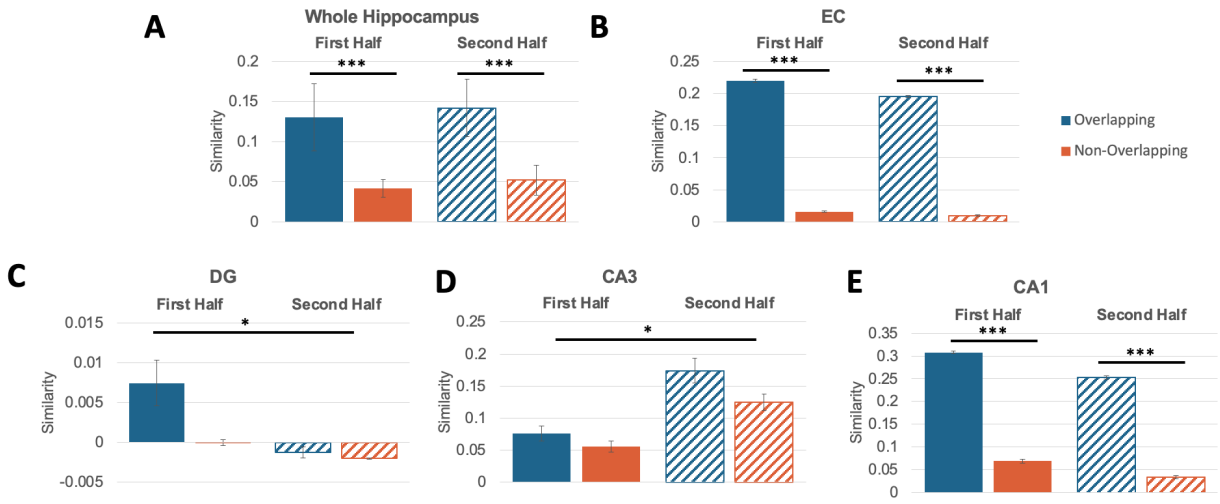

Figure 6. Pattern Similarity by ROI: No inhibition. A) In the whole hippocampus, regardless of training/test half, pattern similarity is greater for overlapping routes compared to non-overlapping routes ($p<0.001$). **B)** In EC, regardless of training/test half, pattern similarity is greater for overlapping routes compared to non-overlapping routes ($p<0.001$). **C)** In DG, overall pattern similarity is higher during the first half compared to the second half. This indicates that the DG is working to do pattern completion. **D)** In CA3, pattern similarity is overall greater in the second half compared to the first half. This indicates that the CA3 is working to pattern complete across routes. **E)** In CA1, regardless of training/test half, pattern similarity is greater for overlapping routes compared to non-overlapping routes ($p<0.001$).

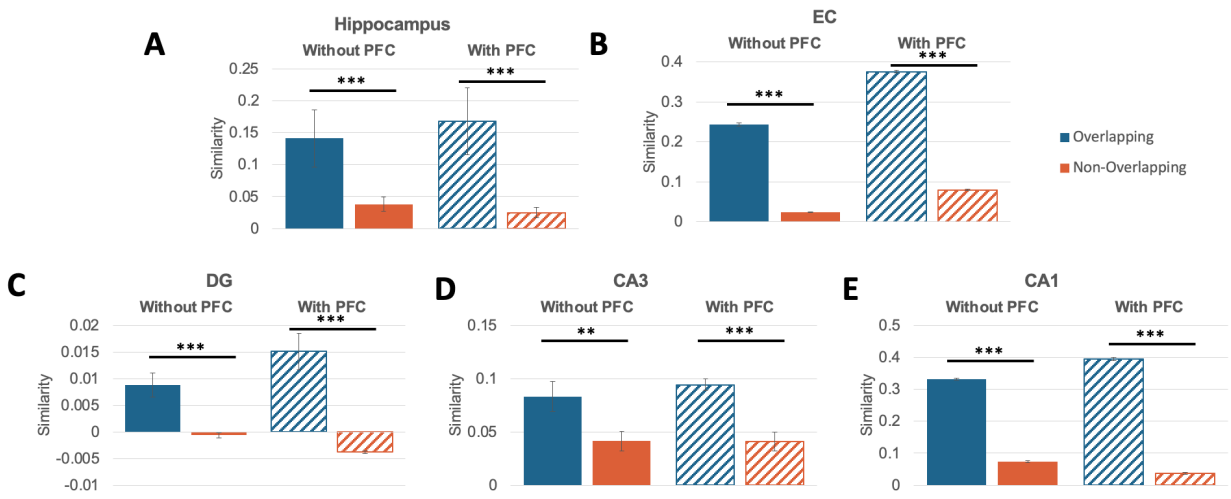

Figure 7. Pattern Similarity by ROI: Inhibition of distinct pools. A-E) Regardless of inhibition of the distinct pools, there is greater pattern similarity for overlapping routes compared to non-overlapping routes in every ROI.

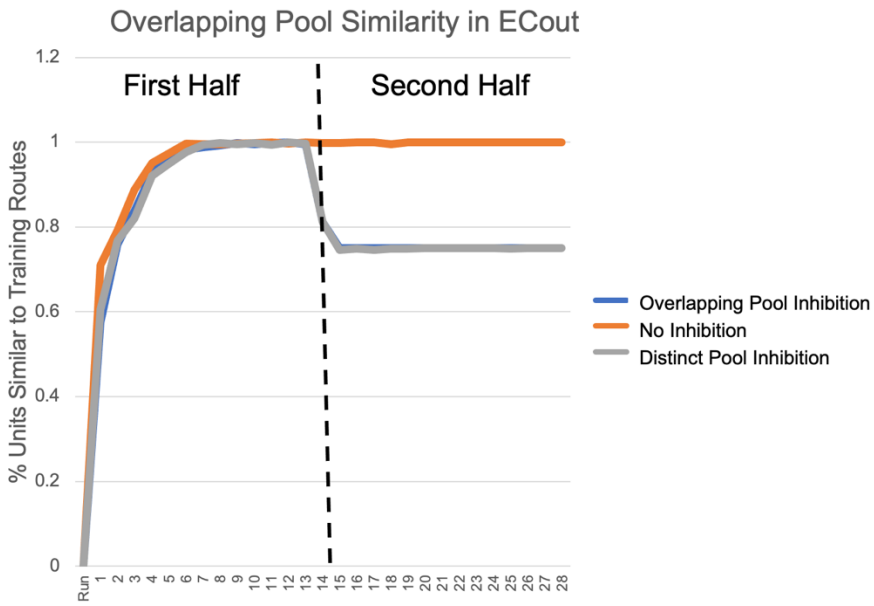

Figure 8. Change in overlapping pool similarity over repeated retrieval. When the overlapping pools or distinct pools are inhibited, similarity to training patterns decreases with inhibition. When there is no inhibition, similarity to training patterns for all routes reaches 100%, indicating perfect accuracy for overlapping features.

References

[1] A. C. Schapiro, L. V. Kustner, and N. B. Turk-Browne, “Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities,” *Curr. Biol.*, vol. 22, no. 17, pp. 1622–1627, Sep. 2012, doi: 10.1016/j.cub.2012.06.056.

[2] A. C. Schapiro, E. Gregory, B. Landau, M. McCloskey, and N. B. Turk-Browne, “The Necessity of the Medial Temporal Lobe for Statistical Learning,” *J. Cogn. Neurosci.*, vol. 26, no. 8, pp. 1736–1747, Aug. 2014, doi: 10.1162/jocn_a_00578.

[3] A. R. Preston, Y. Shrager, N. M. Dudukovic, and J. D. E. Gabrieli, “Hippocampal contribution to the novel use of relational information in declarative memory,” *Hippocampus*, vol. 14, no. 2, pp. 148–152, 2004, doi: 10.1002/hipo.20009.

[4] D. Marr, “Simple memory: a theory for archicortex,” *Phil. Trans. R. Soc. Lond. B*, 1971.

[5] E. Rolls, “The mechanisms for pattern completion and pattern separation in the hippocampus,” *Front. Syst. Neurosci.*, vol. 7, 2013, Accessed: Jan. 12, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2013.00074

[6] S. M. Stark, C. B. Kirwan, and C. E. L. Stark, “Mnemonic Similarity Task: A Tool for Assessing Hippocampal Integrity,” *Trends Cogn. Sci.*, vol. 23, no. 11, pp. 938–951, Nov. 2019, doi: 10.1016/j.tics.2019.08.003.

[7] D. Kumaran and J. L. McClelland, “Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system.,” *Psychol. Rev.*, vol. 119, no. 3, p. 573, 20120709, doi: 10.1037/a0028681.

[8] M. L. Schlichting, J. A. Mumford, and A. R. Preston, “Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex,” *Nat. Commun.*, vol. 6, no. 1, p. 8151, Aug. 2015, doi: 10.1038/ncomms9151.

[9] D. Zeithamova, M. Schlichting, and A. Preston, “The hippocampus and inferential reasoning: building memories to navigate future decisions,” *Front. Hum. Neurosci.*, vol. 6, 2012, Accessed: Feb. 06, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2012.00070

[10] A. Bakker, C. B. Kirwan, M. Miller, and C. E. L. Stark, “Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus,” *Science*, Mar. 2008, doi: 10.1126/science.1152882.

[11] H. R. Dimsdale-Zucker, M. Ritchey, A. D. Ekstrom, A. P. Yonelinas, and C. Ranganath, “CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields,” *Nat. Commun.*, vol. 9, no. 1, p. 294, Dec. 2018, doi: 10.1038/s41467-017-02752-1.

[12] C. Ranganath and K. A. Paller, “Frontal Brain Potentials during Recognition Are Modulated by Requirements to Retrieve Perceptual Detail,” *Neuron*, vol. 22, no. 3, pp. 605–613, Mar. 1999, doi: 10.1016/S0896-6273(00)80714-X.

[13] C. Ranganath, M. K. Johnson, and M. D’Esposito, “Left Anterior Prefrontal Activation Increases with Demands to Recall Specific Perceptual Information,” *J. Neurosci.*, vol. 20, no. 22, p. RC108, Nov. 2000, doi: 10.1523/JNEUROSCI.20-22-j0005.2000.

[14] J. W. Lacy, M. A. Yassa, S. M. Stark, L. T. Muftuler, and C. E. L. Stark, “Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity,” *Learn. Mem.*, vol. 18, no. 1, pp. 15–18, Jan. 2011, doi: 10.1101/lm.1971111.

[15] J. K. Leutgeb, S. Leutgeb, M.-B. Moser, and E. I. Moser, “Pattern Separation in the Dentate Gyrus and CA3 of the Hippocampus,” *Science*, Feb. 2007, doi: 10.1126/science.1135801.

[16] J. F. Guzowski, J. J. Knierim, and E. I. Moser, “Ensemble Dynamics of Hippocampal Regions CA3 and CA1,” *Neuron*, vol. 44, no. 4, pp. 581–584, Nov. 2004, doi: 10.1016/j.neuron.2004.11.003.

[17] C. B. Kirwan and C. E. L. Stark, “Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe,” *Learn. Mem.*, vol. 14, no. 9, pp. 625–633, Sep. 2007, doi: 10.1101/lm.663507.

[18] A. J. H. Chanales, A. Oza, S. E. Favila, and B. A. Kuhl, “Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations,” *Curr. Biol.*, vol. 27, no. 15, pp. 2307-2317.e5, Aug. 2017, doi: 10.1016/j.cub.2017.06.057.

[19] S. E. Favila, A. J. H. Chanales, and B. A. Kuhl, “Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning,” *Nat. Commun.*, vol. 7, no. 1, Art. no. 1, Apr. 2016, doi: 10.1038/ncomms11066.

[20] C. Wendelken and S. A. Bunge, “Transitive Inference: Distinct Contributions of Rostrolateral Prefrontal Cortex and the Hippocampus,” *J. Cogn. Neurosci.*, vol. 22, no. 5, pp. 837–847, May 2010, doi: 10.1162/jocn.2009.21226.

[21] L. M. DeVito, C. Lykken, B. R. Kanter, and H. Eichenbaum, “Prefrontal cortex: Role in acquisition of overlapping associations and transitive inference,” *Learn. Mem.*, vol. 17, no. 3, pp. 161–167, Mar. 2010, doi: 10.1101/lm.1685710.

[22] T. R. Koscik and D. Tranel, “The Human Ventromedial Prefrontal Cortex Is Critical for Transitive Inference,” *J. Cogn. Neurosci.*, vol. 24, no. 5, pp. 1191–1204, May 2012, doi: 10.1162/jocn_a_00203.

[23] E. A. Wing, M. C. D’Angelo, A. Gilboa, and J. D. Ryan, “The Role of the Ventromedial Prefrontal Cortex and Basal Forebrain in Relational Memory and Inference,” *J. Cogn. Neurosci.*, vol. 33, no. 9, pp. 1976–1989, Aug. 2021, doi: 10.1162/jocn_a_01722.

[24] K. N. Spalding *et al.*, “Ventromedial Prefrontal Cortex Is Necessary for Normal Associative Inference and Memory Integration,” *J. Neurosci.*, vol. 38, no. 15, pp. 3767–3775, Apr. 2018, doi: 10.1523/JNEUROSCI.2501-17.2018.

[25] B. Melo, G. Winocur, and M. Moscovitch, “False Recall and False Recognition: An Examination of the Effects of Selective and Combined Lesions to the Medial Temporal Lobe/Diencephalon and Frontal Lobe Structures,” *Cogn. Neuropsychol.*, vol. 16, no. 3–5, pp. 343–359, May 1999, doi: 10.1080/026432999380825.

[26] P. E. Wais, O. Montgomery, C. E. L. Stark, and A. Gazzaley, “Evidence of a Causal Role for mid-Ventrolateral Prefrontal Cortex Based Functional Networks in Retrieving High-Fidelity Memory,” *Sci. Rep.*, vol. 8, no. 1, Art. no. 1, Oct. 2018, doi: 10.1038/s41598-018-33164-w.

[27] S. A. Johnson, S. Zequeira, S. M. Turner, A. P. Maurer, J. L. Bizon, and S. N. Burke, “Rodent mnemonic similarity task performance requires the prefrontal cortex,” *Hippocampus*, vol. 31, no. 7, pp. 701–716, 2021, doi: 10.1002/hipo.23316.

[28] M. C. Anderson, J. G. Bunce, and H. Barbas, “Prefrontal-Hippocampal Pathways Underlying Inhibitory Control Over Memory,” *Neurobiol. Learn. Mem.*, vol. 134, no. Pt A, pp. 145–161, Oct. 2016, doi: 10.1016/j.nlm.2015.11.008.

[29] M. J. D. der Weel *et al.*, “The nucleus reuniens of the thalamus sits at the nexus of a hippocampus and medial prefrontal cortex circuit enabling memory and behavior,” *Learn. Mem.*, vol. 26, no. 7, pp. 191–205, Jul. 2019, doi: 10.1101/lm.048389.118.

[30] M. Petrides and D. N. Pandya, “Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey,” *Eur. J. Neurosci.*, vol. 16, no. 2, pp. 291–310, 2002, doi: 10.1046/j.1460-9568.2001.02090.x.

[31] H. Kondo, K. S. Saleem, and J. L. Price, “Differential connections of the perirhinal and parahippocampal cortex with the orbital and medial prefrontal networks in macaque monkeys,” *J. Comp. Neurol.*, vol. 493, no. 4, pp. 479–509, Dec. 2005, doi: 10.1002/cne.20796.

[32] R. C. O’Reilly and K. A. Norman, “Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework,” *Trends Cogn. Sci.*, vol. 6, no. 12, pp. 505–510, Dec. 2002, doi: 10.1016/S1364-6613(02)02005-3.

[33] R. C. O'Reilly and J. W. Rudy, "Conjunctive representations in learning and memory: Principles of cortical and hippocampal function," *Psychol. Rev.*, vol. 108, no. 2, pp. 311–345, 2001, doi: 10.1037/0033-295X.108.2.311.

[34] R. C. O'Reilly, Y. Munakata, M. J. Frank, and T. E. Hazy, "Computational Cognitive Neuroscience".

CHAPTER 3: Testing the Boundaries: The Testing Effect and Event Segmentation

INTRODUCTION

It's the night before a big exam, and you're scrambling to figure out the best way to prepare. Your options are to either reread your notes over and over or use a stack of flashcards to self-test. Which is the most effective? According to research, the best way to retain information long-term is to self-test. The testing effect – the phenomenon that testing (or retrieval practice), such as cued recall, results in greater performance in the long term compared to restudy – has been examined in many forms. Within the past two decades, there has been a concerted effort to research different aspects of the testing effect to determine 1) how robust the effect is, and 2) what type of test results in the greatest memory retention. The testing effect has been shown to have different results depending on the method of test. Multiple choice questions over short-answer responses at initial testing leads to greater memory performance at final test [1], [2]. However, there is evidence to show that this is highly dependent on feedback. Feedback, in this case, means that the participant is shown the correct answer immediately after answering during initial testing. With feedback at initial testing, investigators showed that ~~the~~ short-answer questions yielded better final performance. This indicates that the type of test and amount of feedback greatly determine how robust the memory retention will be [3]. Not only does type of test matter but the amount of time spent restudying/doing retrieval practice impacts memory retention. Karpicke and Roediger showed that the highest recall at final test was associated when participants had been tested multiple times [1]. These results stress the idea that more opportunities to be tested is more beneficial than a single testing session.

This leads to the question: if we know that items that undergo retrieval practice are subsequently remembered better in the long-term, what about untested items related to these retrieval-practiced items? Either untested items are remembered better, which is referred to retrieval-induced facilitation (RIFA), or if interference may cause these untested items to be remembered less, which is referred to as retrieval-induced forgetting (RIF) [4]. Research shows that RIF and RIFA are dependent on more than one factor. In one passage-reading study, it was found that retrieval-practiced passages presented in a random order of sentences showed RIF for untested items at 20 minutes. However, at a 24-hour delay, there was RIFA for untested items in retrieval-practiced passages that were read in the correct order [4]. This indicates that RIFA is not only related to the delay between initial retrieval practice and final test, as seen with the testing effect, but also how integrated/related the untested information is to the retrieval-practiced information. Liu and Ranganath [5] investigated what "related" means by separating relatedness into temporal or semantic relatedness. In this behavioral study, participants learned scene-word associations where scenes were associated with two words, either semantically related or unrelated, and presented at varying temporal distances to compare temporally distant/semantically related pairs, temporally distant/semantically unrelated pairs, temporally close/semantically related pairs, and temporally close/semantically unrelated pairs. Results showed that regardless of semantic relatedness, there was RIF for temporally far pairs and RIFA for temporally close pairs at a 10-minute delay. At a 24-hour delay, there was RIF for temporally far/semantically unrelated items and RIFA for all other conditions (temporally far/semantically related, temporally close/semantically unrelated, and temporally

close/semantically related). These results suggest that RIF occurs when items are unrelated and far, but RIFA occurs when items are related and/or temporally close [5].

Most studies on the effects of retrieval practice have used relatively simple stimuli, such as word pairs or static images, or passages. But how do these phenomena occur in real life? Unlike snapshots, we perceive continuous stimulus input that we then must separate into discrete portions – or events. Research into event segmentation suggests that segmentation occurs automatically and that the way we segment events ultimately affects what information is retained and able to be recalled [6], [7]. One way to mimic this segmentation is to use a continuous stimulus set (e.g., videos) that has discrete boundaries. Radvansky and colleagues have repeatedly shown that the event of walking through a doorway is sufficient as a boundary between events and that location-updating (i.e., moving from one room to another via a doorway) causes forgetting of information from a previous event. For these tasks, participants must interact with items and move those items to another room. Results showed that once they entered the next room, interference occurred such that it was harder to remember information about the prior room [8]–[11]. However, using doorways as an event boundary can also lead to memory facilitation as well [12], [13]. In one such study, participants either learned two word lists in a one room or one word list in one room followed by learning of another word list after moving to walking to a new room. The results revealed that there was less retroactive interference of the first list for the group that moved to a separate room [12], leading to overall greater memory retention in that group.

In this study, we hypothesize that unpracticed items within an event with an item that has been practiced will get a boost in recall memory due to being in close proximity to a practiced item (i.e., RIFA), however unpracticed items across a boundary will not receive a boost. Alternatively, the unpracticed items within an event may be forgotten (i.e., RIF) due to competition between practiced and unpracticed items within an event or because of the event segmentation (i.e., doorways). To test this hypothesis, we created virtual reality-based videos that included semantically unrelated items that were either retrieval practiced (RP+) or restudied (RS+), unpracticed but in a room with a practiced/restudied item (RP-/RS-), or unpracticed across a boundary from a practiced/restudied item (CTRL), where events were separated by doorways. Using these conditions, we examined the testing effect (RP+ vs RS+) and the effect of boundaries (RP-/RS- vs CTRL). The testing effect would be demonstrated by better cued recall of items that had undergone retrieval practice (RP+) than cued recall of items that had been re-studied (RS+). A boundary effect would be demonstrated if there was a difference in RP-/RS- items compared to CTRL items. If cued recall of items accompanying re-studied (RS-) or retrieval practiced items (RP-) is better than items in the next room (CTRL), then that would suggest RIFA. If the inverse relationship is found, that would suggest RIF. No difference in memory from control items would suggest no boundary effect.

METHODS

Participants

Healthy undergraduates (N=84, mean age = 19.6, 39 female, 38 male, 7 other) were recruited from the University of California, Davis through the SONA Psychology Research Recruitment System and were randomly sorted into two groups: Retrieval Practice (RP; N = 42) and Restudy (RS; N = 42). Inclusion criteria required that participant must be over 18 years of age and native English speakers (characterized by learning English prior to 5 years of age). Compensation for participation was two class credits. Three participants (3 RS) were excluded because behavioral performance during encoding (explained further below) was below two standard deviations of the mean. Following this exclusion, data are presented for a final sample of 81 participants (42 RP, 39 RS; 39 female, 36 male, 6 other).

Procedure and Design

Stimuli. We created 10 videos using Cinema4D. In each video, there were 4 rooms, where rooms 1 and 2 shared a wall and floor color/texture and rooms 3 and 4 shared a different wall and floor color/texture. In each room, there were two real-world items, collected from sketchfab.com and cgtrader.com. In each video, 2 rooms were designated “Retrieval Practice/Restudy” (RP, RS), where participants either performed RP or RS on one item (RP+/RS+; other item: RP-/RS-) in the room, and the other 2 rooms contained control items (CTRL), where there was no RP or RS. These rooms were designed to be in interleaved order. Event boundaries were simulated by hallways and numbered doors, and importantly, the temporal distance between items both within and between rooms was the same.

Encoding Phase. Participants were randomly assigned to either the RP or RS group and were instructed to pay attention to both the items and the surroundings. They then viewed 5 videos in a virtual reality environment. In each room, participants viewed the first item for 8 seconds, then would be shown the second item for the same amount of time. The name of the item was presented underneath each item. Following the second item, the video would then show the door, which led to a hallway, before entering the door for the next room (Figure 1a). Importantly, the time between viewing the items within a room was the same as the time between viewing the items between rooms to preserve temporal distance. After each video, participants were given a chance to RP or RS the RP+/RS+ items (2 items per video). For RP, participants viewed a still image of the background of the room and numbered door corresponding to the RP+ item with the first letter of the RP+ item name followed by dashes representing the number of letters in the RP+ item name. Participants were then given an unlimited amount of time to type in the correct item name. Following this, the correct name of the item briefly (8 seconds) appeared on the screen as feedback. For RS, participants were given the name of the RS+ item and had to retype the name of the item (Figure 1b). In both groups, participants performed this twice per item before moving on to the next video. Participants viewed a total of 40 items.

Retrieval Phase. The retrieval phase occurred immediately after encoding. During this phase, participants were tested on every item viewed during the encoding phase, in pseudorandom order, in the same form as the previously described RP+ items. To eliminate interference between RP+ and RP- items, the RP+ items were presented at the end of the retrieval phase.

Analysis. To determine if performance was better on RP+ items compared to all other items (RP- and CTRL), the RP- and CTRL item performance was averaged, and a two-way ANOVA (group x item type) performed. To assess the effect of event boundaries, item trials were grouped into threes in one of two ways depending on how they were viewed during encoding: 1) CTRL, boundary, RP+, RP- or 2) RP-, RP+, boundary, CTRL (Figure 1c). A two-way ANOVA (group x item type) was then performed between the RP and RS group for the RP- and CTRL items. This way, we could determine if/how retrieval practice or restudy in one room affected unpracticed/unstudied items across a boundary compared to within an event. Post-hoc correlation analyses were performed to further characterize the relationship between RP+ and RP-/CTRL performance, using a difference score (RP- - CTRL).

RESULTS

Overall performance on the final memory test for both groups were similar with the RP group correctly identifying 30.9% of items and the RS group correctly identifying 29.1%. To determine if there was a testing effect (RP+ vs RS+) [1], [2] and retrieval induced facilitation/forgetting (RP-/RS- vs CTRL) [4], we performed a two-way ANOVA with factors of Group (retrieval practice vs. restudy) and Item type (Practiced or Unpracticed). Results showed that performance on the final test was higher for practiced than for unpracticed items ($F(1,79)=28.86$, $p<0.001$). These results show that doing retrieval practice or restudy is better than no practice or restudy. Results also show an interaction ($F(1,79)=4.28$, $p=0.042$; Figure 2) where the difference between the RP+ and other item performance was significantly greater for the RP group compared to

the RS group. This demonstrates the ‘testing effect’ [1], [2] in demonstrating that memory retention was greater for retrieval practiced versus restudied items.

The main analysis of this experiment is to determine if there is an effect of event boundaries. To investigate whether there is RIFO or RIFA within and across event boundaries, we used a two-way ANOVA (group x item type) focused on unstudied items (CTRL and RP-) within and across events. Results revealed a main effect of item type ($F(1,79)=5.05$, $p=0.027$; fig), where participants performed better on CTRL items compared to RP- items. These results show that participants were better at retaining the memory of items across a boundary compared to items in an event with an RP+/RS+ item, suggesting that retrieval practice and restudy cause interference (i.e., forgetting) for items within an event. Correlations between correct RP+ performance and the difference between RP- and CTRL showed that both RP group and RS (Figure 3) had a small negative correlation ($r=-0.20$ and $r=-0.23$, respectively), indicating that when performance on RP+ items was high, participants performed better on CTRL items compared to RP- items (Figure 4). Together, these results show that although RP+ performance is significantly higher for the RP group (i.e., the testing effect), both RP and RS cause interference within an event that does not cross event boundaries, and this interference is correlated with RP+ performance.

DISCUSSION

Our results support both the testing effect and RIF within boundaries. The testing effect is showcased by better performance on RP+ items in the RP group compared to RS+ items in the RS group. RIF is shown by better performance on CTRL items

compared to RP-/RS- item performance. This is further supported by the correlation results, which suggest that performance on RP+/RS+ items is related to the amount of RIF. Together, these results overall support the idea that doorways “cause” forgetting [8]. Because we controlled for temporal distance, we can confidently assert that this forgetting is due to a change in spatial location and, therefore, unrelated to how close in time the items are viewed within and across boundaries.

The use of videos of a virtual environment has its advantages in that we can better represent real-world instances, while still controlling for variables of interest (e.g., temporal distance between items). However, there are few limitations for real-world applications. Although we often interact with items in a room, we typically do not interact with one item at a time, perfectly spaced. Additionally, due to the timing constraints needed to showcase boundary effects, true VR where participants are able to explore the environment on their own was unavailable. In a future study, using an actual VR apparatus, perhaps these concerns can be controlled for by employing eye-tracking techniques to determine both the time the participant viewed each item and the time between viewing items.

As a future step, we suggest three changes to the current task to ask further questions. One change is to use a different recall task. In the current study, we utilized a stem-completion type cued recall, which is harder than other recall/recognition tasks such as a forced-alternative choice, but easier than free recall. To boost overall performance, it would be worth it to explore either a recognition test (e.g., “yes”/”no” to “Have you seen this item?”) or a forced-alternative choice (e.g., “Did you see Item A, B, or C?”). Another change, which could delve deeper into the boundary effects, would be

to include more than two items per room. By doing this, we can investigate the difference between items close to a boundary vs items in the middle of an event. A final suggested change is to change the order of the paradigm so that there is not an additional chance to encode the items after participants have already done retrieval practice or restudy. By doing it this way, the competition between items in a room could be preserved in a way that doesn't allow participants a way to resolve it (i.e., by being able to re-view the items).

Overall, the current results suggest that doing either retrieval practice or restudy lead to RIF, such that there is competition between items within an event where the practiced/restudied item wins out. Further research is necessary to better characterize this relationship and apply it to disordered populations.

FIGURES

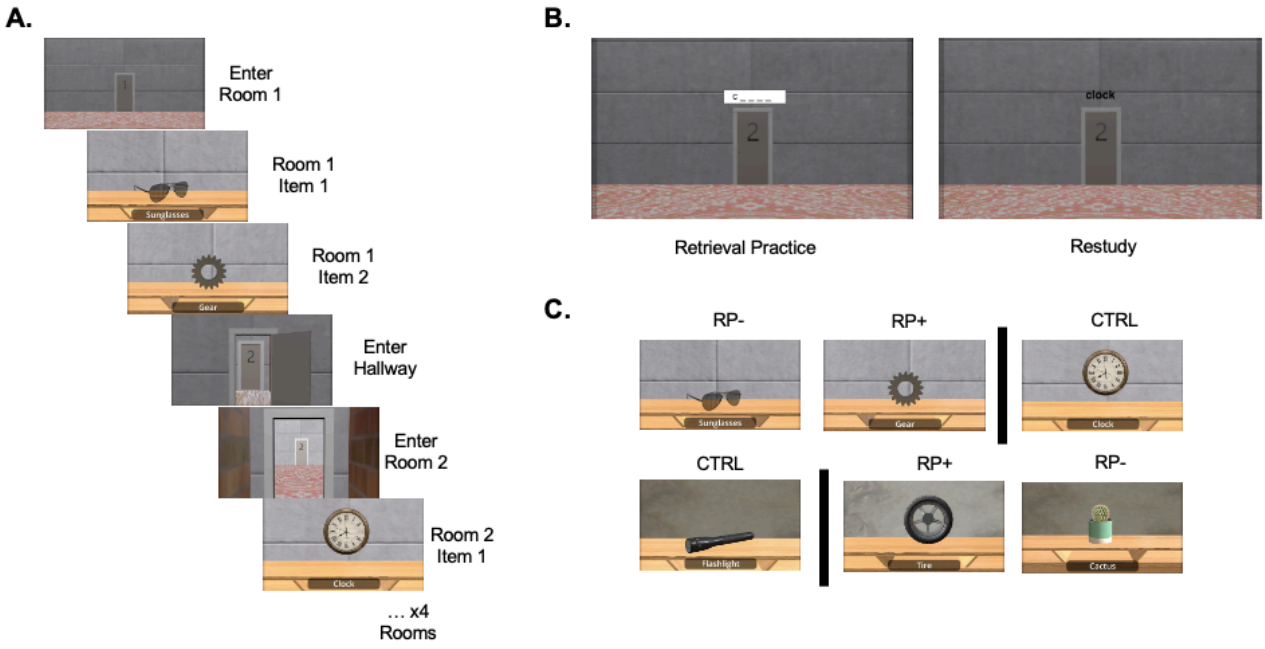

Figure 1. Task Schematic. A) Participants enter Room 1 to view Item 1 followed by Item 2. After viewing Item 2, the participants move through a hallway and enter Room 2, where two new items await. This is repeated for four rooms per building with a total of 5 buildings. Under each item, participants can read the name of the item. **B)** After viewing each building, participants either perform retrieval practice (RP), where they are given the first letter of two previously encoded items (interleaved, such that the RP items are separate by control items) and must complete the name of the item, or restudy (RS), where they see the name of two items they encoded (again, interleaved) which they must retype. This is followed by a feedback screen showing the correct answer. **C)** Event boundaries are shown in two ways. In one, an unpracticed item (RP-) is followed by a practiced item (RP+) in the same room, then a boundary and another unpracticed

item (CTRL) in the adjacent room. In the other, an unpracticed item (CTRL) is followed by boundary, then a practiced item (RP+) is followed by another unpracticed item (RP-) in the same room.

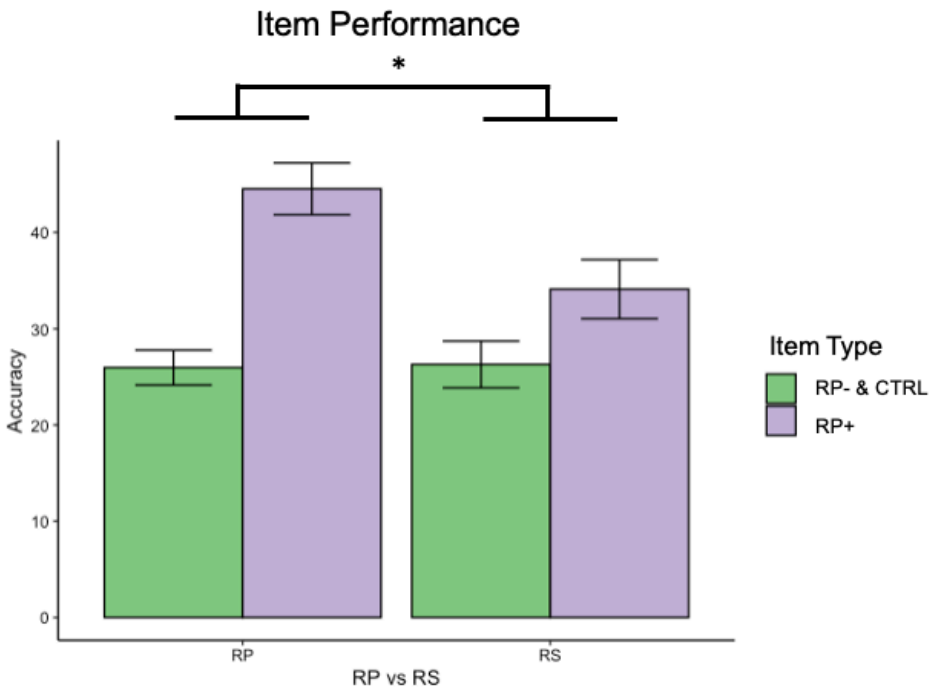

Figure 2. Item Performance Results. Both groups show better performance on RP+ items compared to all other items, however, the difference between the item types is greater for the RP group (p=0.042), suggesting a testing effect.

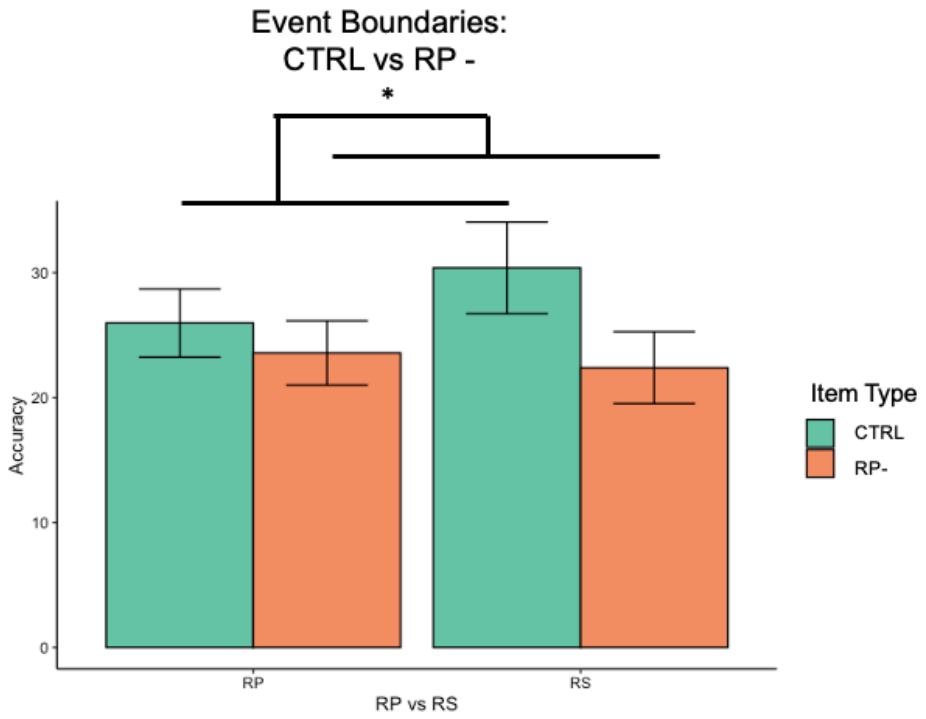

Figure 3. Event Boundary Results. ANOVA results show that, regardless of group, performance on CTRL items was better than RP- items (p=0.027), suggesting RIF.

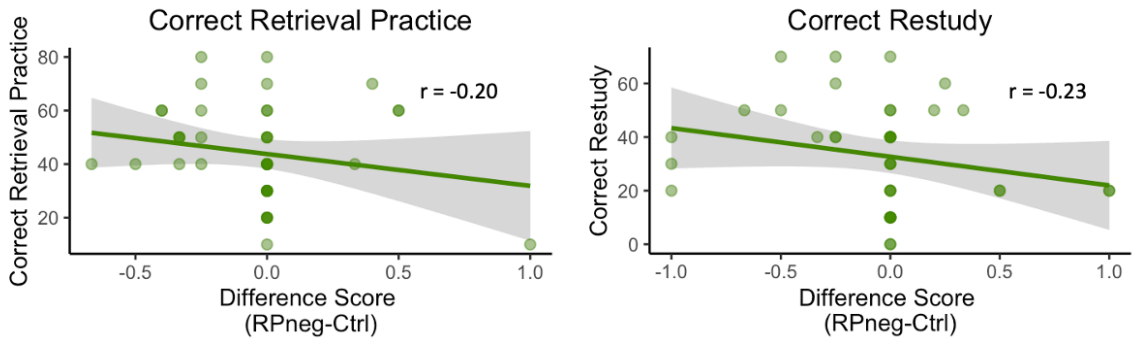

Figure 4. Correlation Between Correct Retrieval Practice/Restudy and the Difference in Performance for RP- and CTRL. Correlation results reveal that there is a small negative correlation between correct retrieval practice/restudy and the difference in RP- and CTRL, such that participants who performed better on RP+/RS+ items also performed better on CTRL items compared to RP-/RS- items. This suggests RIF correlates with correct retrieval practice/restudy.

References

[1] H. L. Roediger and J. D. Karpicke, “Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention,” *Psychol. Sci.*, vol. 17, no. 3, pp. 249–255, Mar. 2006, doi: 10.1111/j.1467-9280.2006.01693.x.

[2] J. D. Karpicke and H. L. Roediger, “The Critical Importance of Retrieval for Learning,” *Science*, vol. 319, no. 5865, pp. 966–968, Feb. 2008, doi: 10.1126/science.1152408.

[3] S. H. K. Kang, K. B. McDermott, and H. L. R. III, “Test format and corrective feedback modify the effect of testing on long-term retention,” *Eur. J. Cogn. Psychol.*, vol. 19, no. 4–5, pp. 528–558, Jul. 2007, doi: 10.1080/09541440601056620.

[4] J. C. K. Chan, “When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing,” *J. Mem. Lang.*, vol. 61, no. 2, pp. 153–170, Aug. 2009, doi: 10.1016/j.jml.2009.04.004.

[5] X. L. Liu and C. Ranganath, “Resurrected memories: Sleep-dependent memory consolidation saves memories from competition induced by retrieval practice,” *Psychon. Bull. Rev.*, Jun. 2021, doi: 10.3758/s13423-021-01953-6.

[6] J. M. Zacks and K. M. Swallow, “Event Segmentation,” *Curr. Dir. Psychol. Sci.*, vol. 16, no. 2, pp. 80–84, Apr. 2007, doi: 10.1111/j.1467-8721.2007.00480.x.

[7] G. A. Radvansky and J. M. Zacks, “Event boundaries in memory and cognition,” *Curr. Opin. Behav. Sci.*, vol. 17, pp. 133–140, Oct. 2017, doi: 10.1016/j.cobeha.2017.08.006.

[8] G. A. Radvansky and D. E. Copeland, “Walking through doorways causes forgetting: Situation models and experienced space,” *Mem. Cognit.*, vol. 34, no. 5, pp. 1150–1156, Jul. 2006, doi: 10.3758/BF03193261.

[9] G. A. Radvansky, K. A. Pettijohn, and J. Kim, “Walking through doorways causes forgetting: Younger and older adults.,” *Psychol. Aging*, vol. 30, no. 2, pp. 259–265, 2015, doi: 10.1037/a0039259.

[10] G. A. Radvansky, S. A. Krawietz, and A. K. Tamplin, “Walking through Doorways Causes Forgetting: Further Explorations,” *Q. J. Exp. Psychol.*, vol. 64, no. 8, pp. 1632–1645, Aug. 2011, doi: 10.1080/17470218.2011.571267.

[11] K. A. Pettijohn and G. A. Radvansky, “Walking through doorways causes forgetting: Event structure or updating disruption?,” *Q. J. Exp. Psychol.*, vol. 69, no. 11, pp. 2119–2129, Nov. 2016, doi: 10.1080/17470218.2015.1101478.

[12] B. Z. Strand, “Change of context and retroactive inhibition,” *J. Verbal Learn. Verbal Behav.*, vol. 9, no. 2, pp. 202–206, Apr. 1970, doi: 10.1016/S0022-5371(70)80051-2.

[13] K. A. Pettijohn, A. N. Thompson, A. K. Tamplin, S. A. Krawietz, and G. A. Radvansky, “Event boundaries and memory improvement,” *Cognition*, vol. 148, pp. 136–144, Mar. 2016, doi: 10.1016/j.cognition.2015.12.013.

CHAPTER 4: Memory Based Prediction Deficits and Dorsolateral Prefrontal Dysfunction in Schizophrenia

INTRODUCTION

As we recollect past events ("episodic memory"), we can unfold an entire sequence of experiences happening at a particular place and time. Neuroscience highlighted the fact that this ability to encode and recall sequences of events serves several purposes beyond retrieval of past events. Past information is often useful for generating predictions helping us navigate an uncertain world. For instance, imagine watching a basketball game the first time. When a player shoots a basket, the score for that player's team goes up. Therefore, the next time a player shoots the ball into the basket, prior knowledge allows us to predict that the score will increase. Humans can be remarkably effective learning about arbitrary sequences of stimuli (1), and information about learned sequences can help us anticipate and efficiently process information even when explicit episodic memory retrieval is not required (2,3).

Episodic memory is impaired in people with schizophrenia (SZ). Although SZ affects a range of cognitive abilities, episodic memory is disproportionately impaired, and severity of memory deficits predicts patients' ability to work and live independently (4). A key finding is that people with SZ can sometimes perform well at recognizing familiar objects or events but are especially impaired at remembering relationships between objects and the context in which they were encountered (5,6,7,8,9). Results from functional magnetic resonance imaging (fMRI) studies suggest two possible explanations for these deficits. In one view, dysfunction might reflect impaired functioning of the hippocampus, which normally supports the ability to bind item and

context information in a manner that can support episodic memory (2,10). Another explanation, that is not mutually exclusive (11), is that prefrontal dysfunction affects control processes that enable one to use learned information to make complex attributions about the context in which events take place (12,13).

Whereas studies of memory emphasize memory for past events, other work has focused on the idea that people with SZ might be impaired generating precise predictions about the future (14,15). Bayesian models propose that, in the healthy brain, higher-order brain areas generate predictions about upcoming sensory information and experience prediction errors encouraging belief updating and better future predictions (14,16,17,18). In this framework, people with SZ generate aberrant prediction errors impairing their learning about the statistical structure of the world. Prediction error research in SZ has informed a broad range of paradigms, including mismatch negativity (19), examination of hallucinations (20), and studies of reinforcement learning - demonstrating impaired learning (21) and dysfunction in dorsal and ventromedial prefrontal cortex and striatum (22,23).

At present, it is not clear whether people with SZ show more global deficits in the ability to predict future events based on learned memory representations. Although studies have shown deficits in explicit memory for temporal or sequential relationships, these deficits might reflect an inability to make complex memory attributions, rather than a prediction deficit per se. Accordingly, in the present study, we investigated the extent to which people with schizophrenia are able to utilize memory of learned sequences to successfully predict future events.

We adapted a paradigm from a recent study of healthy undergraduates which scanned participants while making semantic decisions about sequences of objects (24). In some sequences, object order was *fixed*, such that seeing the first object could enable generation of precise predictions about the remaining objects in the sequence. In other sequences, object order was changed on every repetition *(i.e., random*), allowing participants to become highly familiar with the objects, but unable to make accurate predictions. With this paradigm, healthy individuals had faster reaction times (RTs) when making semantic decisions about objects in fixed versus random sequences. Thus, after a sequence was learned, people used that sequence memory to facilitate response preparation by predicting upcoming objects during fixed sequences, resulting in faster semantic decisions.

Previously (3), we used electroencephalography (EEG) to examine sequence learning and found that both HC and SZ reached criterion for sequence learning and utilized sequence memory to predict future objects and make faster semantic decisions for objects in fixed versus random sequences. This RT facilitation is referred to as the "sequence prediction effect". Although people with SZ also reached criterion during sequence learning, their learning was less efficient and accompanied by decreased alpha and beta1 power prior to stimulus onset for fixed versus random sequences (3). Interestingly, these frequency bands have also been found to mediate prediction feedback (14).

Here, we report the second part of this study, in which fMRI was used to identify brain regions that might underlie hypothesized deficits in sequence-based prediction in SZ (see Fig. 1). Based on previous work, we focused on two regions of interest (ROI):

dorsolateral prefrontal cortex (DLPFC) - identified as dysfunctional during relational memory in fMRI studies of SZ (25), and right posterior hippocampus - identified as a key region mediating sequence representation in our previous sequence memory study (24). Using representational similarity analysis (RSA), we examined the extent to which activity patterns in these regions carry position information for objects in learned sequences, and whether the fidelity of these representations could identify individual differences in sequence-based item prediction for fixed versus random sequences. We hypothesized that neural sequence representations in DLPFC and posterior hippocampus are disrupted in people with SZ, resulting in attenuated sequence prediction effects.

METHODS AND MATERIALS

Participants

As previously (3), forty-four individuals (7 unmedicated) with SZ were recruited from the UC Davis Early Psychosis Programs (EDAPT and SacEDAPT). Sixty-six HC responded to paid advertisements through the UC Davis Imaging Research Center (IRC). Clinical assessments were conducted to confirm SZ diagnosis and symptom severity. Clinicians with master's or doctoral level training confirmed diagnosis using the Structured Clinical Interview for DSM-V (SCID-V). Symptoms were assessed using the Brief Psychiatric Rating Scale (BPRS), Scale for the Assessment of Negative Symptoms (SANS), and Scale for the Assessment of Positive Symptoms (SAPS). Exclusion criteria included: substance abuse in past year, implants that may interfere with MRI scanning (e.g. ferromagnetic implants), neurological defects, loss of consciousness after head trauma, low IQ (i.e., <70), or less than 20/30 vision when

corrected. Four participants were excluded for miscellaneous reasons including; a change in diagnosis (1 HC, 1 SZ), an incidental finding (1 SZ), and refusal to enter the scanner due to anxiety (1 SZ). Eleven participants (7 HC, 4 SZ) were excluded due to operator error which caused a mis-alignment of task onset and scanner onset. An additional 6 participants (2 HC, 4 SZ) were excluded due to poor quality behavioral data (greater than 30% non-responses), and 4 were excluded (2 HC, 2 SZ) for low quality structural/functional scans (signal dropout or excess motion greater than 1 voxel). Following exclusions, data are presented on a final sample of 85 participants (54 HC and 31 SZ). As seen in Table 1, groups were matched on age, gender and parental education, but the SZ sample had lower participant education as illness often interrupts educational attainment. Included in the table are CPZ equivalents for medicated SZ participants. The current study was approved through the University of California, Davis, Institutional Review board and participants provided informed consent prior to study.

Procedure and Design

Encoding Phase

Participants learned five sequences (see 24), during EEG (3) prior to entering the MRI scanner. Encoding conditions are illustrated in Figure 1, and details are provided in Zheng et al. (3).

Sequence Retrieval Task

Immediately following encoding, participants were positioned in the MRI scanner. During sequence retrieval, participants viewed previously encoded sequences (two

fixed, two random, and one novel) for three repetitions per run across four runs. Sequence order was randomized and objects appeared in a continuous stream, with no delays between sequences. Each object in each sequence was presented for 1000ms. Before each run, a semantic question was provided, which participants answered for each object. Questions were: 1) Can this object fit in a shoebox? 2) Can you easily lift this object with one hand? 3) Is the presented object living? or 4) Does this object contain visible metal? Semantic questions were asked to maintain attention and gauge RT differences both within and between sequences to index sequence prediction success.

MRI Acquisition

Imaging was conducted at the University of California, Davis, Imaging Research Center (IRC) on a 3T Siemens Trio Total imaging matrix (Tim) MRI system with a 32-channel head-coil. Structural images were acquired using T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) pulse sequence (208 slices, sagittal; voxel size = $1.0\mathrm{mm}^3$; TR = 2000ms; TE = 2.98ms; flip angle = 8°; FoV = 256mm). Functional images were acquired with an echo planar imaging (EPI) sequence (235 time points; voxel size = $3.4\mathrm{mm}^3$; TR = 2000ms; TE = 25ms; FoV = 218mm; 34 slices, interleaved).

Data Processing and Analysis

Behavioral Data

Sequence prediction effects were calculated by averaging RT across objects 2-5 for fixed and random sequences and calculating differences in RT between sequence

types (fixed – random). Repeated-measures analysis of variance (ANOVA) identified main effects of group (HC vs. SZ), sequence type (fixed vs. random) or higher order interactions for averaged RT values. Pearson Product moment correlations examined associations between symptom severity (total scores on the BRPS, SANS, and SAPS) and sequence prediction effects. Significance levels were set at $p<0.05$ for all analyses.

fMRI Data Preprocessing

Preprocessing of fMRI data was modeled after Hsieh et al (2). Preprocessing was accomplished using fMRI Expert Analysis Tool (FSL version 5.0.9). To strip the skull and remove any non-brain tissue, the Brain Extraction Tool (BET) extracted brain volumes. All functional images were slice-time corrected and high-pass filtered with a 0.01 Hz cut-off. MCFLIRT was used for motion correction and functional images were co-registered with each individual's MPRAGE using FLIRT. Resulting transformation matrices were used to transform ROIs into native-space for each participant.

Regions of Interest

A priori regions of interest (ROIs) were bilateral DLPFC and bilateral hippocampus. Hippocampal ROIs utilized probabilistic maps based upon an average of 55 hand-traced T1 images using methods validated by Ritchey and colleagues (10). Hippocampal ROIs included the full body, and sub-regions for head, body and tail based on anatomical landmarks (10). The DLPFC ROI (36 voxels) utilized a probabilistic mask including Brodmann areas 9 and 46 based upon Talairach coordinates functionally defined by MacDonald and colleagues (26). After placing ROIs in standard space, they were transformed into native space prior to statistical analysis as described above.

We also performed an exploratory searchlight whole brain RSA analysis using 400 parcellations acquired from Schaefer and colleagues (27). As previously, parcellations were transformed into native space prior to statistical analyses (FWE corrected at $p<0.05$).

Representational Similarity Analysis

RSA is a multivariate approach correlating patterns of voxel activation across objects that share a similar feature of interest to determine if brain regions are sensitive to that feature (28,29). For example, imagine an area of the brain encoding representations of dogs. Patterns of activation in that area for one dog will be similar to patterns of activation for another dog. In contrast, patterns of activation for a cat in that area might show shared activation across some voxels because of some shared features (e.g. four legs, domestic animal, etc.), but the overall pattern of activation for the cat will be more dissimilar than either of the dog representations, confirming that the area is most sensitive to processing dog-related representations. For the current study, we utilized RSA to understand how DLPFC and posterior hippocampus represented fixed versus random sequences.

To do this, we first assessed patterns of activity across voxels within DLPFC and posterior hippocampus during single trials using parameter estimates (i.e. beta weights) for each object, estimated through the Least-Square2 (LS2) method (30). A general linear model (GLM) computed beta weight estimates for each object. For each functional run, there were 75 GLMs (5 objects/sequence x 5 sequences x 3 repetitions) for a total of 300 beta maps (4 runs x 75 beta maps/run). Outlier beta maps, determined by a signal intensity lying in the 1% of all beta maps, were excluded from analyses.

Next, we examined voxel pattern similarity for fixed and random sequences by calculating Pearson's correlation coefficient between beta weight vectors for pairs of trials, which were Fisher transformed and averaged in three ways (fig. 1B). To examine fixed sequence representations, an average was taken for all fixed sequences across repetitions and runs, where each object was in the same position. For random sequences, an average was taken across all random sequences, where the position of objects varied. To include representations where object information was shared between objects in different sequences regardless of position, voxel patterns were rearranged within random sequences and averaged. These two averages were combined to create a full picture of random sequence representation. For bilateral DLPFC, we performed a three-way ANOVA to identify effects of group (HC vs. SZ), hemisphere (left vs. right), and sequence representation (fixed vs. random). Based on prior research using a similar paradigm (24), we limited analyses of the posterior hippocampus to the right hemisphere. We performed a two-way ANOVA in right posterior hippocampus examining effects of group (HC vs. SZ) and sequence representation (fixed vs. random). Correlation analyses were performed to determine associations between RT sequence prediction effects and similarity values (fixed – random) gleaned from RSA results.

RESULTS

Behavioral Results

Based upon previous results (3), we hypothesized that people with SZ would show reduced sequence prediction effects. This was supported by the ANOVA, which revealed main effects of group ($F(1,83)=8.27$, $p=0.005$) and sequence type

(F(1,83)=21.90, p<0.001), as well as a significant group by sequence interaction (F(1,83)=8.93, p=0.004). As shown in Figure 2, t-tests investigating sequence prediction effects (fixed minus random RT for objects 2-5) revealed that this interaction was due to HC showing a greater speeding of RT for fixed versus random sequences than people with SZ (t=-2.99, p=0.004). Thus, although people with SZ were able to learn and retrieve well-learned sequences and speed their RTs, these memory prediction effects were reduced relative to HC. (Supplemental figure 1 illustrates the difference between objects within each sequence.)

fMRI Results

Based on previous work (24), we hypothesized that DLPFC and posterior hippocampus are involved in supporting memory-based prediction for objects within a temporal context and that these representations would be reduced in people with SZ. We examined this separately for DLPFC and hippocampus using RSA to compare voxel pattern similarities for fixed and random sequences.

Supporting our hypothesis for the DLPFC, ANOVA revealed a main effect of sequence with fixed sequences showed higher similarity than random sequences (F(1,83)=14.84, p<0.001; fig 3). There was no main effect of hemisphere (F(1,83)=3.37, p=0.07). Although there was no main effect of group, there was an interaction between sequence and group (F(1,83)=5.67, p=0.02). Post-hoc tests revealed that in HC only, DLPFC voxel pattern similarity was higher for fixed sequences than for random sequences (t=5.16, p<.001). Conversely, in the SZ sample there were no differences in how the DLPFC represented objects in fixed versus random sequences (t=0.92,

p=0.36). These data suggest that DLPFC dysfunction may be associated with memory-based prediction deficits in individuals with SZ.

In posterior hippocampus, ANOVA revealed no main effects of group or sequence. There was, however, a trend toward an interaction between group and sequence ($F(1,87)=3.20$, $p=0.08$; fig 3). Exploratory post-hoc analyses revealed that this trend-level interaction was due to increased pattern similarity for fixed versus random sequences in the right posterior hippocampus of HC ($t=2.26$, $p=0.03$) but not in people with SZ ($t=-0.55$, $p=0.58$). Results did not strongly support our hypothesis that hippocampal dysfunction contributes to temporal sequence memory deficits in SZ.

Searchlight analyses did not reveal other areas showing increased pattern similarity for fixed versus random sequences. Thus, there were no effects in other parts of the brain.

Association with Performance

To determine the relationship between sequence prediction effects and representational similarity, we performed correlation analyses (Fig 4). In the left DLPFC, fixed sequence representations were significantly correlated with sequence prediction effects in HC ($r=-0.330$, $p=0.011$) but not in SZ ($r=-0.115$, $p=0.392$). In the right DLPFC, there was no correlation between sequence prediction and representational similarity. These data suggest that, in HC only, as sequence prediction increases (indicated by an increasingly negative value), representational similarity increases in the left DLPFC.

Similar analyses were performed in the bilateral posterior hippocampus, however none of the results from these analyses were significant, indicating that pattern similarity in the posterior hippocampus was not associated with sequence prediction effects.

Association with Clinical Symptoms

Correlational analyses did not reveal any significant relationships between DLPFC pattern similarity effects and severity of clinical symptoms (total SANS, SAPS and BPRS) in the SZ sample (all r-values > 0.08).

DISCUSSION

We rely on memory to make accurate predictions about our changing environment. In the present study, we investigated whether people with SZ show deficits in memory-based prediction using a temporal sequence paradigm (2,3,31). Although both groups predicted and responded more quickly to objects within a previously learned sequence (i.e. faster RTs for fixed versus random sequences), this effect was reduced in people with SZ relative to HC. Multivariate analyses of fMRI data revealed that participants with SZ showed disrupted neural representations of learned sequences in the DLPFC. These findings support the conclusion that people with SZ can learn temporal sequences but their ability to utilize this sequence memory to predict future events is dysfunctional.

Sequence prediction effects were measured indirectly as individuals responded to objects contained in sequences that could be learned (i.e., fixed) or could not be learned (i.e., random), with speeding of semantic decisions for learned versus unlearned sequences providing evidence of successful prediction. Behavioral results

indicate that individuals with SZ were capable of forming and using memories to predict the next object in the sequence and, thereby, guide their behavior. As previously (3) both groups showed evidence of sequence learning, but memory prediction effects were reduced in people with SZ relative to HC. Therefore, memory impairments do not appear due to a lack of attention or a generalized memory deficit. However, individuals with SZ did not improve predictions to the same degree as healthy individuals, suggesting that people with SZ were less successful in using memory representations to guide predictions (14).

Although semantic priming deficits have been reported in SZ (32,33), it is notable that, in the present study, objects in learned and random sequences were equally familiar. Thus, differential effects of sequence learning on semantic decisions between patients and controls cannot be explained by differences in semantic priming. Instead, our results are more consistent with the idea that people with SZ have a reduced ability to use sequential regularities to predict upcoming events.

To better understand these memory-related prediction impairments, we used RSA to characterize representations of information from temporal sequences in DLPFC and posterior hippocampus. Multivariate analyses revealed that, in HC, representational similarity was higher across repetitions of objects in learned sequences relative to repetitions of objects in random sequences in the DLPFC. These effects were not significant for people with SZ, and DLPFC voxel pattern similarity differences between objects in learned and random sequences were significantly higher in HC than in SZ. Moreover, the fidelity of DLPFC representations of objects in learned sequences was predictive of sequence prediction success in HC only, although there were no significant

group differences in the size of the association. Results are consistent with a large body of behavioral, eye-tracking, EEG and fMRI research linking both memory (25,34) and prediction (23) impairments to DLPFC dysfunction in people with SZ. Impaired DLPFC control of memory encoding and retrieval has been repeatedly demonstrated in people with SZ on both an individual study (31,34) and meta-analytic level (25).

In addition to the DLPFC, numerous studies supported the idea that hippocampal abnormalities might contribute to relational memory deficits in SZ (35). Several studies documented reductions in hippocampal volume in SZ (36) and others (37,38,39), including results from our group (31,40), demonstrated reduced hippocampal activation during relational memory retrieval in people with SZ relative to HC. In the present study, however, we did not observe any evidence for hippocampal dysfunction in people with SZ. Exploratory analyses of data restricted to the HC group revealed that, consistent with our previous study (24), pattern similarity in right posterior hippocampus was higher for objects in learned relative to random sequences. Although these effects were not significant in people with SZ, we did not observe any significant between group differences in hippocampal sequence representations, nor did we find significant relationships between hippocampal results and sequence prediction effects in either group. Thus, results did not support the hypothesis that sequence-based prediction deficits in people with SZ were related to impaired hippocampal function.

As described in Zheng et al. (3), people with SZ learned sequences more slowly than HC, consistent with a deficit in relational memory. As previously (2,41), participants were highly trained on fixed and random sequences prior to scanning, thus enabling people with SZ to compensate for any learning deficits. During scanning, participants

were not asked to explicitly recall sequences, so we would only expect them to show faster decisions for objects in fixed sequences if they proactively used memory for the learned sequences to accurately predict upcoming objects. Thus, results are consistent with concluding that, even when learning is sufficient to overcome relational memory deficits, people with SZ are impaired using what was learned to predict upcoming events, with this deficit strongly associated with DLPFC dysfunction.

A challenge of studying people with SZ is the heterogeneity of the disorder which can increase variability in addition to potential medication effects and differences in clinical presentations. One might expect that we would have found correlations with positive symptoms given literature linking predictive coding deficits with severity of positive symptoms (18). One notable difference between our study and previous work is that participants in this study were early psychosis patients who were clinically stable with mild to moderate symptoms. Therefore, a restricted range may have contributed to lack of clinical correlations. Most participants in the SZ group were receiving second generation antipsychotics and when analyses were repeated after excluding un-medicated participants there was no difference in the pattern of behavioral or fMRI results. We also did not find any significant correlations with standardized medication dose (i.e., CPZ equivalents). Thus, results do not appear to be influenced by medication or symptom severity effects. We did experience significant data loss due to excess motion and operator error. This was likely related to both operator and participant fatigue as fMRI recordings were obtained immediately following an EEG study (3). Finally, our primary fMRI analysis examined *a priori* regions in the hippocampus and prefrontal cortex based upon previous fMRI studies (2,31), raising the possibility that

there were task effects or group differences within other regions of the brain. To address this, we conducted an exploratory whole-brain searchlight analysis which did not reveal any task effects or group differences in other brain regions.

In conclusion, results indicate a key finding: individuals with SZ are able to learn sequences, but there is dysfunction in using prior knowledge about sequences to aid in prediction of upcoming objects. HC were more successful than people with SZ engaging their DLPFC to form object/sequence representations to facilitate prediction of upcoming objects in learned sequences. These findings support prior theories proposing that there are aberrant prediction processes in people with SZ (14,17). Ongoing efforts to remediate memory-based prediction deficits in SZ using neurostimulation, pharmacology or behavioral interventions may be most successful if they target DLPFC-related control processes.

Table 1. Sample Demographics

	Healthy Controls (n=54)	**Schizophrenia (n=31)**	***p-values***
	Mean ± SD	*Mean ± SD*	
Age (years)	24.10 ± 4.38	23.05 ± 4.23	0.28
Sex (% male/female)	72%/28%	81%/19%	0.39
Education (years)	15.07 ± 1.98	13.50 ± 1.80	0.00047***
Parental Education (years)	13.70 ± 3.03	14.69 ± 2.57	0.13
BPRS (total)		37.77 ± 10.09	
SANS (total)		18.10 ± 10.71	
SAPS (total)		7.45 ± 12.00	
CPZ Equivalents		244.22 ± 161.22	

SZ, schizophrenia; BPRS, Brief Psychiatric Rating Scale; SANS, Scale for the Assessment of Negative Symptoms; SAPS, Scale for the Assessment of Positive Symptoms; CPZ, chlorpromazine

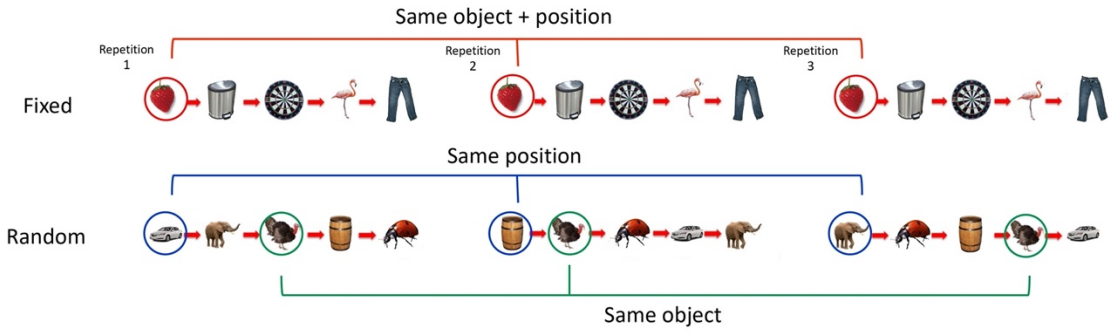

Figure 1. Schematic of the paradigm. Fixed sequences show the same objects in the same position for each repetition. Random sequences show the same objects but in a different position each repetition.

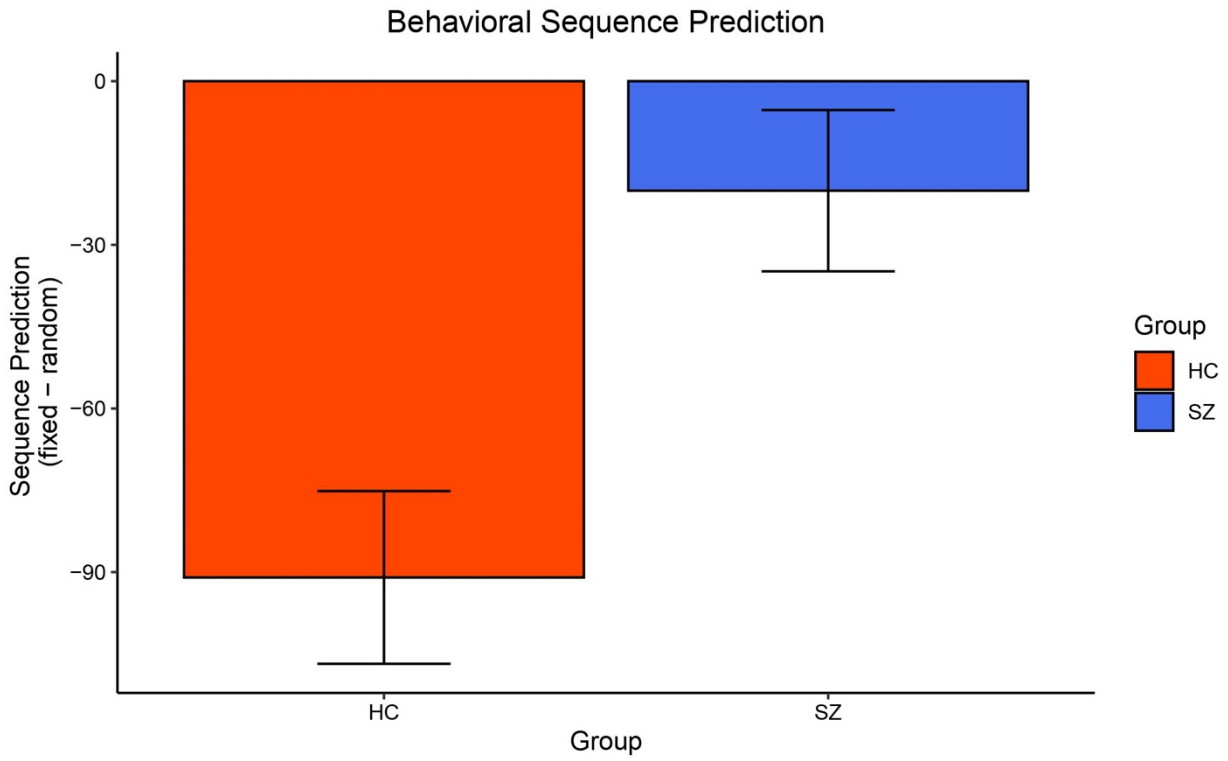

Figure 2. Sequence prediction indicated by the difference in RT for objects 2-5 between fixed and random sequences. RT for fixed sequences was faster than random sequences for both groups with HC showing a significantly greater difference.

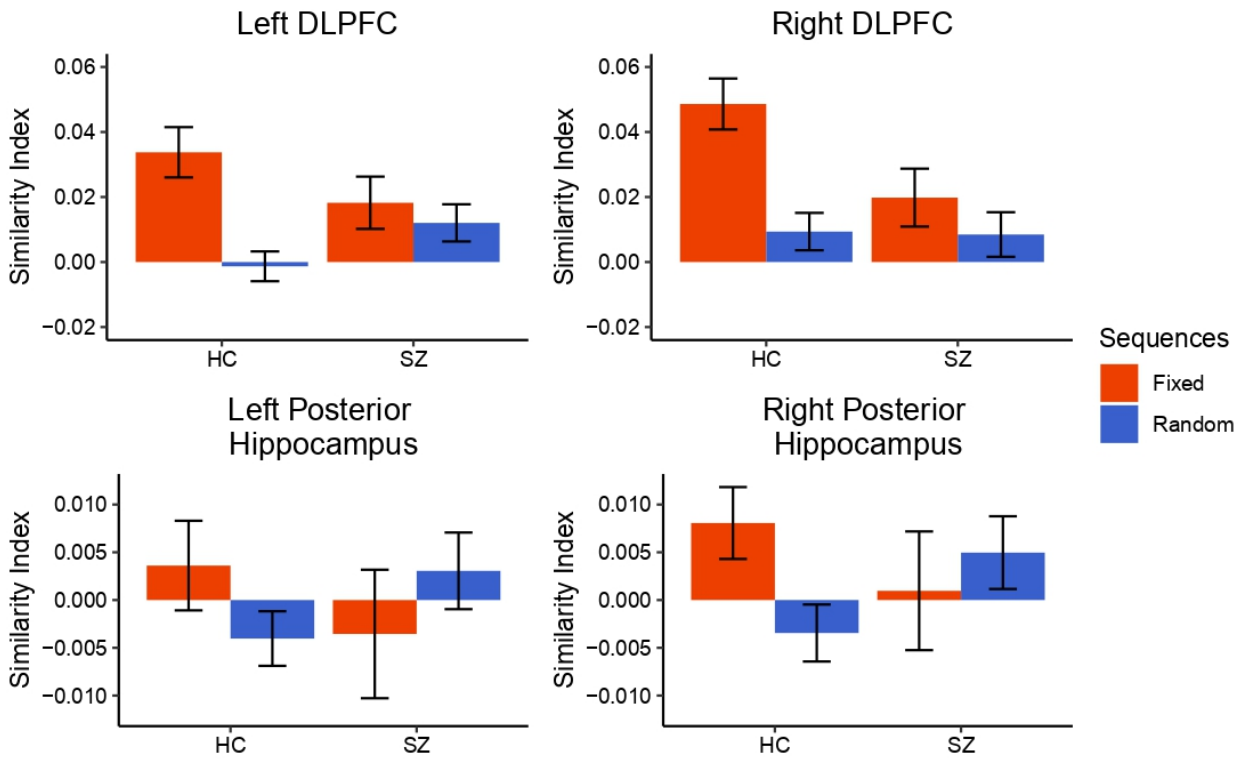

Figure 3. In the bilateral DLPFC (**Fig. 3A and 3B**), HC showed significantly greater similarity for fixed sequences compared to random while SZ did not. The left posterior hippocampus showed no group or sequence differences (**Fig. 3C and 3D**). In the right posterior hippocampus, in HC only, fixed sequences showed significantly greater similarity for fixed versus random sequences.

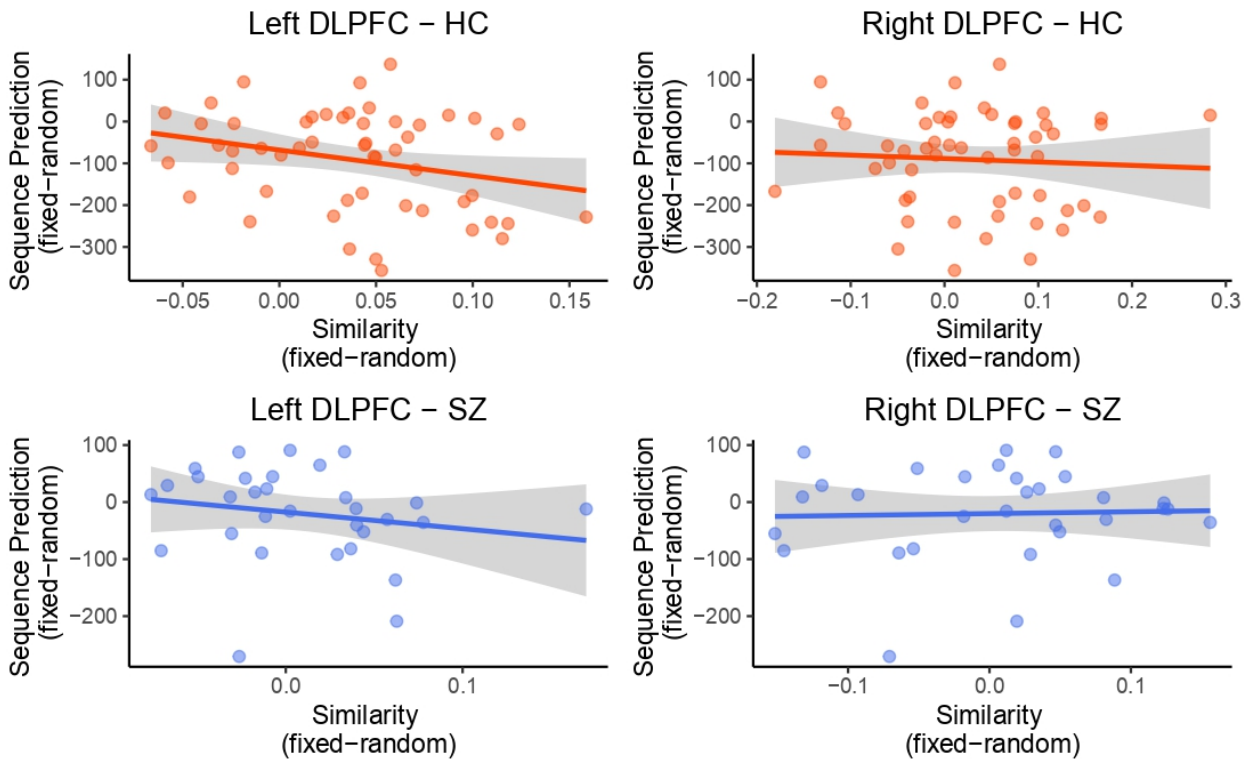

Figure 4. Scatter plot of correlations between pattern similarity in the left DLPFC and sequence prediction success. For HC (**Fig. 4A and 4B**), greater pattern similarity for fixed versus random sequences in the left DLPFC was correlated with better sequence prediction. These correlations were not significant in people with SZ (**Fig. 4C and 4D**).

REFERENCES

1. Schapiro, A., & Turk-Browne, N. (2015): Statistical Learning. In: *Brain Mapping*. Elsevier: pp 501-506.
2. Hsieh, L.-T., & Ranganath, C. (2015): Cortical and subcortical contributions to sequence retrieval: Schematic coding of temporal context in the neocortical recollection network. *NeuroImage* 121: 78–90.
3. Zheng, Y., Liu, X. L., Hsieh, L.-T., Hurtado, M., Wang, Y., Niendam, T. A., *et al.* (2021): Disrupted Modulation of Alpha and Low Beta Oscillations Mediates Temporal Sequence Memory Deficits in People with Schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 6(12): 1157-1164.
4. Aleman, A., Hijman, R., de Haan, E. H., & Kahn, R. S. (1999): Memory impairment in schizophrenia: A meta-analysis. *The American Journal of Psychiatry* 156(9): 1358–1366.
5. Waters, F. A. V., Maybery, M. T., Badcock, J. C., & Michie, P. T. (2004): Context memory and binding in schizophrenia. *Schizophrenia Research* 68(2): 119–125.
6. Titone, D., Ditman, T., Holzman, P. S., Eichenbaum, H., & Levy, D. L. (2004): Transitive inference in schizophrenia: Impairments in relational memory organization. *Schizophrenia Research* 68(2): 235–247.
7. Öngür, D., Cullen, T. J., Wolf, D. H., Rohan, M., Barreira, P., Zalesak, M., & Heckers, S. (2006): The Neural Basis of Relational Memory Deficits in Schizophrenia. *Archives of General Psychiatry* 63(4): 356–365.

8. Avery, S. N., Rogers, B. P., & Heckers, S. (2018): Hippocampal Network Modularity Is Associated With Relational Memory Dysfunction in Schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3(5): 423–432.
9. Avery, S. N., Armstrong, K., Blackford, J. U., Woodward, N. D., Cohen, N., & Heckers, S. (2019): Impaired relational memory in the early stage of psychosis. *Schizophrenia Research* 212: 113–120.
10. Ritchey, M., Libby, L. A., & Ranganath, C. (2015): Cortico-hippocampal systems involved in memory and cognition. In: *Progress in Brain Research*, *Vol. 219*. Elsevier, pp.45-64.
11. Place, R., Farovik, A., Brockmann, M., & Eichenbaum, H. (2016): Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nature Neuroscience* 19(8): 992–994.
12. Fuster, J. M., & Bressler, S. L. (2015): Past Makes Future: Role of pFC in Prediction. *Journal of Cognitive Neuroscience* 27(4): 639–654.
13. Guo, J., Ragland, J., & Carter, C. (2019): Memory and Cognition in Schizophrenia. *Molecular Psychiatry* 24(5): 633–642.
14. Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., *et al.* (2018): The Predictive Coding Account of Psychosis. *Biological Psychiatry* 84(9): 634–643.
15. Heinz, A., Murray, G. K., Schlagenhauf, F., Sterzer, P., Grace, A. A., & Waltz, J. A. (2019): Towards a Unifying Cognitive, Neurophysiological, and Computational Neuroscience Account of Schizophrenia. *Schizophrenia Bulletin* 45(5): 1092–1100.

16. Fogelson, N., Litvak, V., Peled, A., Fernandez-del-Olmo, M., & Friston, K. (2014): The functional anatomy of schizophrenia: A dynamic causal modeling study of predictive coding. *Schizophrenia Research* 158(1): 204–212.
17. Sterzer, P., Voss, M., Schlagenhauf, F., & Heinz, A. (2019): Decision-making in schizophrenia: A predictive-coding perspective. *NeuroImage* 190: 133–143.
18. Fletcher, P. C., & Frith, C. D. (2009): Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience* 10(1): 48–58.
19. Kirihara, K., Tada, M., Koshiyama, D., Fujioka, M., Usui, K., Araki, T., & Kasai, K. (2020): A Predictive Coding Perspective on Mismatch Negativity Impairment in Schizophrenia. *Frontiers in Psychiatry*, *11*: 660.
20. Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. (2019): Hallucinations and Strong Priors. *Trends in Cognitive Sciences*, *23*(2): 114–127.
21. Dowd, E. C., Frank, M. J., Collins, A., Gold, J. M., & Barch, D. M. (2016): Probabilistic Reinforcement Learning in Patients With Schizophrenia: Relationships to Anhedonia and Avolition. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 1(5): 460–473.
22. Koch, K., Schachtzabel, C., Wagner, G., Schikora, J., Schultz, C., Reichenbach, J. R., *et al.* (2010): Altered activation in association with reward-related trial-and-error learning in patients with schizophrenia. *NeuroImage* 50(1): 223–232.
23. Haarsma, J., Fletcher, P. C., Griffin, J. D., Taverne, H. J., Ziauddeen, H., Spencer, T. J., *et al.* (2020): Precision weighting of cortical unsigned prediction

error signals benefits learning, is mediated by dopamine, and is impaired in psychosis. *Molecular Psychiatry*: 1–14.

24. Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014): Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron* 81(5): 1165–1178.
25. Ragland, J. D., Laird, A. R., Ranganath, C., Blumenfeld, R. S., Gonzales, S. M., & Glahn, D. C. (2009): Prefrontal Activation Deficits During Episodic Memory in Schizophrenia. *American Journal of Psychiatry* 166(8): 863–874.
26. MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000): Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science* 288(5472): 1835–1838.
27. Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018): Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex* 28(9): 3095–3114.
28. Kriegeskorte, N. (2008): Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*.
29. Dimsdale-Zucker, H. R., & Ranganath, C. (2018): Representational Similarity Analyses. In *Handbook of Behavioral Neuroscience*, *Vol. 28*. Elsevier: pp. 509-525.
30. Turner, B. O., Mumford, J. A., Poldrack, R. A., & Ashby, F. G. (2012): Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage* 62(3): 1429–1438.

31. Ragland, J. D., Ranganath, C., Harms, M. P., Barch, D. M., Gold, J. M., Layher, E., *et al.* (2015): Functional and Neuroanatomic Specificity of Episodic Memory Dysfunction in Schizophrenia: A Functional Magnetic Resonance Imaging Study of the Relational and Item-Specific Encoding Task. *JAMA Psychiatry* 72(9): 909.
32. Almeida, V. N., & Radanovic, M. (2021): Semantic priming and neurobiology in schizophrenia: A theoretical review. *Neuropsychologia* 163.
33. Sharpe, V., Weber, K., & Kuperberg, G. R. (2020): Impairments in Probabilistic Prediction and Bayesian Learning Can Explain Reduced Neural Semantic Priming in Schizophrenia. *Schizophrenia Bulletin* 46(6): 1558–1566.
34. Lepage, M., Montoya, A., Pelletier, M., Achim, A. M., Menear, M., & Lal, S. (2006): Associative Memory Encoding and Recognition in Schizophrenia: An Event-Related fMRI Study. *Biological Psychiatry* 60(11): 1215–1223.
35. Ranganath, C., Minzenberg, M. J., & Ragland, J. D. (2008): The Cognitive Neuroscience of Memory Function and Dysfunction in Schizophrenia. *Biological Psychiatry* 64(1): 18–25.
36. Heckers, S. (2001): Neuroimaging studies of the hippocampus in schizophrenia. *Hippocampus* 11(5): 520–528.
37. Heckers, S., Rauch, S., Goff, D., Savage, C., Schacter, D., Fischman, A., & Alpert, N. (1998): Impaired recruitment of the hippocampus during conscious recollection in schizophrenia. *Nature Neuroscience* 1(4): 318–323.
38. Zorrilla, L. T. E., Jeste, D. V., & Brown, G. G. (2002): Functional MRI and Novel Picture-Learning Among Older Patients With Chronic Schizophrenia: Abnormal

Correlations Between Recognition Memory and Medial Temporal Brain Response. *The American Journal of Geriatric Psychiatry* 10(1): 52–61.

39. Weiss, A. P., Zalesak, M., DeWitt, I., Goff, D., Kunkel, L., & Heckers, S. (2004): Impaired hippocampal function during the detection of novel words in schizophrenia. *Biological Psychiatry* 55(7): 668–675.
40. Ragland, J. D., Layher, E., Hannula, D. E., Niendam, T. A., Lesh, T. A., Solomon, M., Carter, C. S., *et al.* (2016): Impact of schizophrenia on anterior and posterior hippocampus during memory for complex scenes. *NeuroImage: Clinical*, 13: 82–88.
41. Crivelli-Decker, J., Hsieh, L.-T., Clarke, A., & Ranganath, C. (2018): Theta oscillations promote temporal sequence learning. *Neurobiology of Learning and Memory*, *153:* 92–103.

Concluding Remarks

Together, these chapters highlight the idea that the relationship between the hippocampus and prefrontal cortex (PFC) is dynamic and both are necessary for successful encoding, retrieval, and memory updating. In Chapter 1, I proposed that the PFC flexibly controls hippocampal representations to switch between integration and differentiation when items are similar. I continued this line of thinking in Chapter 2 by proposing a possible mechanism by which these regions may be interacting with each other, computationally. Using studies that investigate integration and differentiation, I showed that the proposed computational model can achieve similar results as the empirical data. In Chapter 3, I focused on retrieval processes and how retrieval practice or restudy of information can affect memory for other information within the same event. This work showcases that retrieval competition can occur within events, leading to forgetting of information related to retrieval practiced/restudied information. And finally, in Chapter 4, I showcased how the hippocampus and PFC are affected in temporal sequence processing in individuals with schizophrenia, which may be related to the behavioral effects observed. Ultimately, the underlining thread connecting these lines of research is the dynamic relationship between the hippocampus and PFC, where the PFC may be controlling hippocampal representations during both encoding and retrieval leading to memory updating that is necessary to achieve task goals and, importantly, flexibly change these representations when task goals change.