CONTENT REPRESENTATION IN LATERAL PARIETAL CORTEX

by

YUFEI ZHAO

A DISSERTATION

Presented to the Department of Psychology
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2023

DISSERTATION APPROVAL PAGE

Student: Yufei Zhao

Title: Content Representation in Lateral Parietal Cortex

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Psychology by:

Brice Kuhl	Chairperson
Dasa Zeithamova	Core Member
Margaret Sereno	Core Member
James Murray	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

DISSERTATION ABSTRACT

Yufei Zhao

Doctor of Philosophy

Department of Psychology

June 2023

Title: Content Representation in Lateral Parietal Cortex

While the lateral parietal cortex (LPC) in the human brain is traditionally investigated for its functions in visual perception, more recent evidence has highlighted its substantial contribution to supporting human episodic memory. Early univariate neuroimaging studies suggest that the strength and direction of LPC activation during memory-related tasks is closely related to memory performance. Moreover, recent multivariate fMRI studies show that the neural activity patterns of LPC actively represent mnemonic contents at various granularities. Despite advances in understanding parietal contributions to episodic memory, the relationship between LPC multivariate content representation and univariate activation changes remains unexplored. Moreover, the mechanisms through which the LPC content representation supports episodic memory success are yet unidentified. In the current dissertation, I aim to investigate these topics by incorporating fMRI techniques with neural networks and multivariate pattern analysis methods in a set of two experiments. In chapter II, I demonstrate that repetition-related neural activity differences in the lateral parietal cortex represent stimulus-specific content information, and a greater amount of decodable content information contributes to memory success. In chapter III, I show that content representations in lateral parietal cortex can be adaptively distorted along a feature dimension in order to resolve memory interference, and

the degree of such adaptive change contributes to memory success. Together, these studies provide new insights into the nature of content representation in the lateral parietal cortex and how it supports memory success.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor **Dr. Brice Kuhl**. It would have been impossible for me to survive doctoral life without your generous advice and support. From academic knowledge to professional skills, I am consistently learning from you, and I particularly appreciate your patience and inspiration whenever I get stuck.

I am grateful to my committee members and the professors I have crossed paths with at the University of Oregon. Thank you, **Dr. Dasa Zeithamova**, for serving on my advising committee and providing valuable feedback for all my research projects since I started my doctoral journey at UO. Thank you, **Dr. Margaret Sereno**, for offering the neural network class in the department. Through the inspiring class, I found where my true passion locates and discovered my career path. Thank you, **Dr. James Murray**, for your flexibility and support, and all your input to my dissertation plans and drafts. My appreciation also goes out to **Dr. Ben Hutchinson** for giving me the opportunity to participate in the NSD project and for encouraging me to explore complex deep-learning methods that have expanded my horizons and skills. I would also like to thank **Dr. Daniel Anderson**, for your intuitive classes, and your practice of inclusiveness that not only guided me in how to learn programming but also provided me with a safe and encouraging space at the time when I was struggling with racism.

I also want to express my gratefulness to mentors whom I encountered outside academia during the past six years. To **Dr. Maria Kissa**, thank you for adding your magic positive energy to my three-month internship, and making it the happiest time of my entire doctoral life. The life experience you shared with me stroke me and taught me

how to confront issues and people that oppose my morality with peace and forgiveness. To **Chen Yuan**, thank you for your selfless and continuous help, and for guiding me toward the right way to seeking the truth and meanings of life. Finally, I owe much gratitude to the love and support of my friends and family. Thank you to (almost Dr.) **Xi Yang**. Our friendship is a clump of honey at the bottom of the cup. Whenever I was hit by the bitterness of life, it overwhelmed me with sweetness. To (almost Dr.) **Steffi Hung**, thank you for your sympathetic ear, for typing in simplified Chinese for me, and for always being willing to try out new things with me. To **Dr. Yingying Wang**, the moments when our thoughts and feelings resonated together will always be remembered. To **my mom** and **my boyfriend**, thank you for always supporting me, and keeping learning and growing together with me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

CHAPTER I

GENERAL INTRODUCTION

The last section of this chapter is from Zhao, Y., & Kuhl, B. A. (2023). Content Reinstatement. In *The Oxford Handbook of Human Memory.* (in press)

Overview

Human cognitive neuroscience aims to understand how complicated cognitive processes are guided, formed, and executed in the human brain. Episodic memory, the conscious memory for life events, is one of those complex cognitive processes that is involved in almost all daily events of our life (Tulving, 1985). From finding the right spot where the car is parked to introducing a travel site during small talk, we rely on our memory functions in a precise way to accomplish our daily tasks.

From the results of early lesion studies, the medial temporal lobe (MTL), where the hippocampus is located, as well as the prefrontal cortex (PFC) have been considered as the brain regions that form the neural architecture of episodic memory (Eichenbaum, 2004; Ranganath & Knight, 2003; Scoville & Milner, 1957; Shimamura, 1995; Squire, 1992; Stuss & Benson, 1984). Although traditional research focuses on the neural mechanisms of how MTL and PFC support the encoding and retrieval of episodic memory, a growing number of neuroimaging studies suggest that the lateral parietal cortex's (LPC) contribution may also be crucial to understanding episodic memory. Specifically, the retrieval success effect (also known as old/new effects) have been obtained by both electroencephalogram (EEG) and fMRI studies. These studies show that LPC subregions demonstrate differential activity during successful memory retrieval (Henson et al., 1999; Konishi et al., 2000; Rugg, 1995). Moreover, univariate activity

changes in LPC have also been found to increase with the recollection of memory details (Henson et al., 1999).

Based on the results of these early studies on LPC's contribution to episodic memory, the traditional theory about LPC suggests this region processes memory in a content-general way (Cabeza et al., 2008; Wagner et al., 2005). Specifically, early univariate studies demonstrated that LPC activation is modulated by memory retrieval success, leading to the theoretical account that the lateral parietal cortex may serve as a mnemonic evidence accumulator (Wagner et al., 2005) or it may direct attention to memory representations in other brain regions (Cabeza et al., 2008). With the development of multivoxel pattern analysis (MVPA), accumulating evidence indicates that distributed patterns of LPC activation actively represent retrieved content (Bird et al., 2015; Buchsbaum et al., 2012; Kuhl & Chun, 2014), challenging the traditional theory of content-general representation in LPC. However, little is known about the nature of content representations in LPC, how they relate to content representations in other brain regions, and how they are involved in supporting successful memory.

This proposed dissertation focuses on understanding the nature and functional significance of mnemonic content representations in LPC. In the remaining sections of this chapter, I will review the early findings of how LPC contributes to human memory from univariate neuroimaging studies and lesion studies. Then I will summarize the popular theoretical hypotheses about LPC's contribution to memory. Last, I will consider recent advances in understanding the content representation in LPC based on MVPA methods.

From Sensory to Memory: the Role of Human Lateral Parietal Cortex

The early view of the human LPC focused on its contribution to visuospatial and sensorimotor functions (Culham & Kanwisher, 2001). One of the most convincing pieces of evidence for this view is that unilateral parietal damage leads to hemispatial neglect, referring to the inability to report visual stimuli in the contralesional hemifield (Bisiach & Luzzatti, 1978; Danckert & Ferber, 2006; Driver & Vuilleumier, 2001; Vallar, 1998). Yet with the advance of human neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), LPC has shown involvement in many memory-based high-ordered cognitive tasks, and its role in human episodic memory has gradually been established in the past decade of human cognitive neuroscience research. This section aims to provide an overview of the early human neuroimaging and anatomical findings that set the foundation for investigating LPC as a memory region.

Evidence of LPC's contributions to human memory

The first group of studies that demonstrated memory-related activity in LPC used block-design positron emission tomography (PET) studies. Increased activation was observed in LPC during conditions where subjects were retrieving memories, compared to the control condition (Buckner et al., 1995; Petrides et al., 1995; Tulving et al., 1994). As the event-related experiment design became popular among EEG and fMRI research, measuring item/trial-level brain activity revealed a more interesting role of LPC in supporting memory. A host of fMRI studies found that during the recognition memory task, in which participants are asked to decide whether each stimulus has been encountered before, LPC showed distinct responses for successfully recognized old

stimuli and correctly identified new stimuli (Hutchinson et al., 2009; H. Kim, 2013; Skinner & Fernandes, 2007; Spaniol et al., 2009). Moreover, a similar distinction was observed in EEG studies, with the parietal cortex showing different event-related potentials (ERP) for hit versus correct rejection trials (Mecklinger, 2000; Rugg & Curran, 2007).

The detected old/new effects triggered more investigations into LPC's contribution to the memory process (Cabeza et al., 2008; Wagner et al., 2005) and the follow-up works provided ample evidence of the specific roles of LPC in different stages of the memory process. For example, a few studies suggested that, instead of signaling the true oldness/newness of the encountered stimuli, LPC reflects the degree of subjective oldness. That is, LPC shows greater activation for stimuli that were subjectively perceived as old (compared to new), regardless of whether these old/new judgments were correct or not (Kahn et al., 2004; Wheeler & Buckner, 2003). This evidence indicates that LPC is involved in the process of retrieval, but does not necessarily signal success in retrieval. Moreover, other works showed that LPC activations during memory retrieval correlate with the amount of memory event details recollected (Dobbins et al., 2003; Eldridge et al., 2000; Henson et al., 1999; Wheeler & Buckner, 2004). This evidence foretold LPC's role in counting and representing recollected mnemonic details (see *the output buffer hypothesis* and *the mnemonic accumulator hypothesis* in the next session for more details)

Together, this early evidence from univariate fMRI studies provided strong evidence that LPC contributes to multiple aspects of human memory, including

recognizing old memories, subjective judgment of memory oldness, and tracking details of recollected memories.

Functional Distinctions between Dorsal and Ventral LPC

As investigations of LPC reveal accumulating evidence of its contribution to memory processes, much work has found that LPC does not function as one functionally united entity. Instead, the LPC consistently demonstrated functional separation among its dorsal and ventral components during both encoding and memory retrieval.

The subsequent memory effect is one of the well-investigated phenomena during memory encoding (Uncapher & Wagner, 2009). With this paradigm, neural activities are categorized based on the fate of each stimulus, being either remembered or forgotten. In this way, memory success can be correlated with neural activities during encoding, and therefore, candidate neural correlates of encoding success and failure can hence be identified. Convergent findings from this paradigm showed that dorsal LPC demonstrated a positive subsequent memory effect, with increased activation during encoding in this region for subsequently remembered items compared to forgotten items (Buckner et al., 2001; Davachi et al., 2001; Ranganath et al., 2004). On the contrary, a negative subsequent memory effect is consistently observed in ventral LPC, with decreased activation during encoding related to remembered items (Daselaar et al., 2004; Otten, 2007; Otten & Rugg, 2001; see the cortical binding of relational activity hypothesis in the next session for more details). Notably, the observed functional divergence between dorsal and ventral PPC is consistent across different retention intervals, study materials, and study tasks (H. Kim, 2011; Uncapher & Wagner, 2009).

Similar functional divergence between the dorsal and ventral LPC was observed during retrieval as well. Research showed that these two regions react differently for the recollection-based and familiarity-based memory retrieval processes. A typical research paradigm under this issue is the "remember/know" recognition task. In this task, subjects need to indicate whether they can recollect any encoded details about the learned stimuli (remember), or they only believe the item has been encountered but can't retrieve any details (know) (Tulving, 1985). During this task, dorsal LPC consistently demonstrates greater activation for the "know" response compared to correction rejections, while ventral LPC is more activated for the "remember" response over the "know" response (Montaldi et al., 2006; Sharot et al., 2004; Wheeler & Buckner, 2004; Woodruff et al., 2005). Similar findings were derived from the source recognition task. In this task, subjects were asked to report the incidentally encoded context of learned stimuli (e.g., locations, colors, categories of task), and the successful context judgment is believed to reflect the memory recollection process (Hutchinson et al., 2009). Consistent with previous findings, the ventral LPC was more activated when the source judgment was correct (More memory details required; Cansino et al., 2002; Kahn et al., 2004), the dorsal LPC activated when the recognition was successful without the recollection of the source memory (Frithsen & Miller, 2014; Kahn et al., 2004; see the attention-to-memory model for details).

Anatomy of LPC that supports human memory

The functional relevance of a brain region can be largely attributed to its anatomical structure and connectivity across the rest of the brain. Thus, the works reviewed above suggest 1) specific anatomical wiring routes that bestow LPC its role in

episodic memory and 2) distinct anatomical wiring routes for dorsal and ventral LPC that give rise to their distinct functional properties. Here I reviewed the basic anatomy and functional connectivity evidence of LPC that support the univariate observations reviewed above.

The lateral posterior parietal (LPC) cortex locates between the somatosensory cortex and the visual cortex (Whitlock, 2017). As expected, LPC can be anatomically divided into the dorsal and ventral components, separated by the intraparietal sulcus (Cabeza et al., 2008). The dorsal LPC consists of the intra-parietal sulcus (IPS) and the superior parietal lobule (SPL), while the supramarginal gyrus (SMG) and angular gyrus (ANG) together are considered as the ventral LPC (Sestieri et al., 2017). Notably, the dorsal and ventral components of LPC, although anatomically adjacent, demonstrated distinct anatomical connections with the rest of the brain. Early anatomical connectivity findings of LPC derived from primate research suggested that the monkey dorsal LPC projects to the frontal cortex, which has been long established as the attention and cognitive control-related hub (Lewis & van Essen, 2000; Makris et al., 2005; Petrides & Pandya, 1984, 1999; Schmahmann et al., 2007). On the other hand, ventral LPC demonstrates strong anatomical connections with regions in the medial temporal lobe (MTL), which has historically been considered the memory processing hub. Specifically, LPC demonstrated bidirectional connectivity with the entorhinal cortex (Insausti & Amaral, 2008; Muñoz & Insausti, 2005; Wellman & Rockland, 1997), the perirhinal cortex and the parahippocampal cortex (Blatt et al., 2003; Lavenex et al., 2002; Seltzer & Pandya, 1984; Suzuki & Amaral, 1994), and the hippocampus subregions (Clower et al., 2001; Rockland & van Hoesen, 1999).

Human functional connectivity studies have further corroborated the anatomical connections identified in the primate studies. In particular, dorsal LPC demonstrated strong functional relevance with the prefrontal regions. Together, they serve as parts of the dorsal attention network (DAN) (Corbetta et al., 2008; Corbetta & Shulman, 2002), which is important for orienting attention aligned with top-down goals (Kincade et al., 2005; Majerus et al., 2018). On the other side, ventral LPC is functionally coupled with the medial temporal lobe (H. Kim, 2012; Rushworth et al., 2006; Sestieri et al., 2011; Takahashi et al., 2008; Vincent et al., 2006). These regions collaborate together for the default mode network (DMN), which has been proven to be critical for behaviors under external goals (Raichle et al., 2001).

This anatomical and functional connectivity evidence clearly suggested that the dorsal LPC collaborates with frontal regions for goal-oriented behaviors, while ventral LPC works closely with memory-related regions. This evidence rationalizes the functional difference between dorsal and ventral LPC and the important role LPC plays in retrieving goal-relevant episodic memory (see *the attention-to-memory model* reviewed below for more details).

Lesion and TMS studies further support how LPC contributes to human memory

It is worth noting that early lesion studies often failed to provide evidence of LPC's contribution to episodic memory. That is, lesions in the parietal lobe rarely lead to obvious memory-related deficits, such as retrograde or anterograde amnesia (i.e., patients with LPC lesions do remember things in the past, and can form new memories; Berryhill, 2012; Critchley, 1953). Nevertheless, with fast-growing evidence from the human neuroimaging literature, more recent lesion and transcranial magnetic stimulation (TMS)

studies on LPC suggested that the malfunction of LPC indeed negatively influences episodic memory, but in a subtle, yet very significant way.

For example, with the recognition memory paradigm, one study found that although the patients with bilateral parietal lesions didn't exhibit any deficit in recognition accuracy, they demonstrated significantly lower confidence in rating their "old" responses compared to healthy controls (Hower et al., 2014). Similarly, another lesion study showed that compared with the healthy control group, patients with LPC damage were more willing to respond with "know" compared to "remember" even when all the source memory judgments were correct (Ciaramelli et al., 2017). These results matched findings from functional studies that LPC is involved in subjective judgment of memory oldness, and is responsible for tracking the amount of detail recollected (see earlier sub-sections).

Moreover, it was shown that lesions or disturbances in dorsal versus ventral LPC tend to affect different aspects of memory processes. Specifically, lesions in the ventral LPC tend to be more directly associated with memory retrieval whereas lesions in the dorsal LPC affect the effectiveness of goal-directed attention (Humphreys et al., 2021). For example, studies have found that ventral LPC damage or TMS disturbance decreases memory vividness, confidence, and recollect details compared to the healthy controls (Berryhill et al., 2007). On the other hand, evidence from TMS showed that disturbance to dorsal LPC leaves memory-related performances relatively intact but significantly affects goal-directed, perceptual attention performance (Capotosto et al., 2017).

Traditional theories on how LPC supports memory

Thus far I have reviewed early neuroimaging evidence of LPC's involvement during various memory processes. Along with these findings, multiple theories regarding *how* LPC supports memory have then been established. These theories proposed the roles of LPC during memory encoding and retrieval and how dorsal and ventral LPC work together to allow successful memory processes. In particular, here I review four important accounts of how LPC supports memory. Note that each of these accounts was developed to explain a specific type of findings, but are not necessarily mutually exclusive and instead often share key concepts and lead to similar conclusions.

The mnemonic accumulator hypothesis was first introduced by Wagner et al. (2005), accounting for LPC's role in recognition memory (see old/new effect reviewed above). This account proposed that LPC works as an information accumulator that temporally integrates memory-based evidence retrieved from MTL until a criterion is reached that leads to the decision (Shadlen & Newsome, 2001). This account resembles the drift-diffusion process of memory retrieval (Ratcliff et al., 1978). In the case of recognition memory, evidence of a stimulus being a previously encountered object is accumulated within LPC; once the accumulated memory strength reaches the decision threshold, the judgment of oldness, or otherwise newness, is subsequently made (Dunn, 2004). This account explains how LPC is responsible for the observed old/new effect during recognition memory and why LPC favor's subjective oldness instead of the ground truth. Notably, although proposed as an accumulator, the current account suggests that the LPC holds the mnemonic evidence in a content-general way, without representing retrieved information per se (Kuhl et al., 2014). However, this account does

not readily explain why some subregions in LPC are more sensitive to source information compared to item information itself (Dobbins et al., 2002).

Consistent with the idea that LPC temporarily stores mnemonic evidence, the *output buffer hypothesis* proposed that LPC is the buffer for retrieved contents before a behavioral decision is rendered, yet the temporary stored evidence is in a content-specific manner (Vilberg & Rugg, 2008; Wagner et al., 2005). In particular, the output buffer hypothesis resembles the multi-modal episodic buffer for working memory, where information is maintained and manipulated and serves as an interface between long-term memory and executive functions (Baddeley, 2000). According to the current account, LPC does not hold long-term memory but instead, it maintains and represents multi-modal information content recollected from other parts of the brain during memory retrieval until a decision threshold is reached. This hypothesis is supported by the evidence that ventral LPC is involved in the recollection-based but not the familiarity-based recognition process. That is, recollection requires solid mnemonic evidence represented in the buffer (i.e., LPC) in order to reach such a decision threshold. Likewise, this hypothesis is also consistent with the finding that LPC activation during memory retrieval is necessary for vivid autobiographical memories (Cabeza, 2008).

The *cortical binding of relational activity (CoBRA) hypothesis* extends the output buffer account by proposing the role of LPC during both memory retrieval and encoding processes and in interacting with other brain regions to bind multi-modal mnemonic evidence (Shimamura, 2011). Multi-modal binding is an established process in the MTL during episodic memory encoding through rapid associative processes via long-term potentiation (Morris, 2006). In this way, the MTL binding process creates the initial link

for multi-model episodic features and consolidates such links through reactivation and memory replay (Frankland & Bontempi, 2005). During episodic memory retrieval, the strongly linked memory features will be reactivated together, leading to the creation of neocortical bindings within the LPC, despite the geographical separation among features (Shimamura, 2011). Note that the anatomical connectivity of LPC facilitates such a multi-modal binding process as it is intricately connected to multiple brain regions such as both the dorsal and ventral visual pathway, MTL, and the prefrontal cortex (Andersen et al., 1990). The current account can be used to understand the negative subsequent memory effect observed in ventral LPC, such that less activity during encoding leads to better retrieval performances (Daselaar et al., 2004; Otten, 2007; Otten & Rugg, 2001). According to the CoBRA hypothesis, it is MTL that performs multi-modal binding during memory encoding whereas the LPC binding is created during retrieval. As a result, minimal LPC activity during encoding indicates a prioritization of MTL binding whereas increased LPC activity can disturb the encoding process by accidentally retrieving goal-irrelevant information (Vannini et al., 2011).

The functional dissociation and integration between dorsal and ventral LPC are proposed by *the attention-to-memory model* (AtoM; Cabeza, 2008; Ciaramelli et al., 2008). The attention component of AtoM states that the dorsal PPC mediates top-down, goal-directed cognitive tasks whereas the ventral PPC responds to bottom-up, salience-driven stimuli (Corbetta & Shulman, 2002). Generalizing the functional dissociation in PPC from attention to memory, the AtoM hypothesis accounts for the positive and negative subsequent memory effects observed in dorsal and ventral LPC, respectively. Specifically, the positive subsequent memory effect observed in the dorsal PPC reflects

executive processing of allocating goal-directed attention for mnemonic encoding (Wager & Smith 2003). On the other hand, the negative subsequent memory effect detected in the ventral PPC likely indicates a form of attentional competition, in which the attentional resources necessary for encoding success were shifted away for some goal-irrelevant yet salient distractors, or due to the nature of the task such as context switching (Uncapher & Wagner, 2009). Moreover, the AtoM hypothesis also explains how the dorsal LPC is more engaged during the "know" responses whereas the ventral LPC is more active during the "remember" responses (Ciaramelli et al., 2017). That is, "know" responses are often associated with relatively poor memory details and thus benefit more exhaustive memory search, requiring dorsal LPC to allocate attention internally to search for memory evidence. On the contrary, "remember" responses indicate ampler details associated with the stimulus, with episodic features being retrieved rather automatically and thus resembling a "bottom-up" attentional process.

Content representation in LPC

The empirical evidence for LPC's involvement in memory processes reviewed above is mostly based on univariate neuroimaging findings. That is, the primary concern of these studies is the relationship between cognitive variables and the averaged amplitude of BOLD activities of brain regions. Although univariate analysis is productive in identifying important neural substrates underlying specific cognitive functions (e.g., establishing the involvement of LPC in memory processes), it is limited in investigating and understanding the issue of *representations* (Norman et al., 2006). Namely, the exact information carried in different brain structures (e.g., the memory features stored in LPC), as represented by fine-grained patterns of distributed activity, instead of the

indiscriminate average, of cortical neural activities. Indeed, as suggested by the output buffer and cortical binding hypothesis, the LPC temporarily maintains memory features in a content-specific way.

Two multivariate-based analyses techniques—pattern similarity analysis (Kriegeskorte et al., 2008) and multi-voxel classification (Norman et al., 2006)—were developed to investigate the neural representation in a content-specific way. Pattern similarity analysis measures the correlation (or similarity) between neural activity patterns in two different presentations of the stimulus in the same brain region. The two presentations could be both encoding/perception trials (measuring content representation strength during encoding) or memory retrieval trials (measuring content representation strength during retrieval). Content representation (during both encoding and memory retrieval) is quantified as greater neural activity pattern similarity between different presentations of the same stimuli compared to those of different stimuli. The multi-voxel classification relies on neural activity patterns during perception trials to train classifiers for distinguishing various contents of the stimuli; the trained classifiers are then tested to predict the specific content categories or identities based on retrieval trials' neural activity patterns. Content representation, in this case, is quantified as the above-chance prediction accuracy of the classifiers trained on classifying content information. Neuroimaging studies that have leveraged these two methods, which have surged in the past decade, have suggested that distributive patterns of neural activities in LPC encode memory features of remembered stimuli at different levels, including category level (e.g., faces vs. scenes), event-level (e.g., which face), or even feature-level (e.g., color; Bird et al., 2015; Buchsbaum et al., 2012; Jost et al., 2012; Khader et al., 2005; Polyn et al., 2005; St-

Laurent et al., 2015). Yet such content representation does not exclusively exist in the parietal areas; in fact, the sensory cortex also carries the property of content representation throughout many memory processes (Kuhl et al., 2011). This raises the question of how specific properties or mechanisms of content representations in LPC differ from those in sensory regions and allow LPC to contribute to the memory processes. While this topic is still an active area of investigation, here I outline two developing perspectives for which LPC content representations demonstrate unique memory-oriented properties, which show systematically different properties compared to sensory regions.

First, content representation observed in LPC tends to be stronger during retrieval than during encoding/perception, whereas the sensory regions' content representations are stronger during encoding/perception than memory retrieval (Chen et al., 2017; Favila et al., 2018; Xiao et al., 2017). Interestingly, recent evidence indicates that the rodent LPC also more strongly represents past sensory experience than the current sensory experience (Akrami et al., 2018), suggesting that this may be a fundamental property of LPC. The fact that LPC shows stronger content representations during memory retrieval than encoding is surprising in that it does not fit with the intuitive idea that represented information during retrieval is a degraded version of perception—that retrieval can only *approximate* perception. However, these findings align well with evidence that LPC exhibits asymmetries in univariate BOLD responses during memory retrieval vs. memory encoding, echoing the content representation asymmetry of LPC during encoding versus retrieval. As discussed in the previous section, regions within the ventral LPC exhibit increased BOLD responses during successful memory retrieval (strong LPC content

representation during retrieval), but decreased activity during successful memory encoding (relatively weak LPC content representation during perception; Daselaar, 2009; Kim et al., 2010; Lee et al., 2017; Vannini et al., 2011). While there is, to date, no definitive account of why LPC might show stronger representations during memory retrieval than during encoding, it suggests a transformation in how information is represented during encoding vs. retrieval (Favila et al., 2018; Xiao et al., 2017). Note that such transformation happens in the LPC, but not in any sensory regions, which might be important in shaping perceptions to form memories.

The second mechanistic difference in content representations between LPC and sensory regions that may contribute to LPC's role in memory processes is that the content representations in LPC are sensitive to top-down goals (Favila et al., 2018; Kuhl et al., 2013). That is, content representations in LPC are biased in favor of goal-relevant information. In one study (Kuhl et al., 2013), participants encoded face and scene images that appeared either on the left- or right-hand side of a computer screen. During retrieval, subjects were either asked to retrieve (and behaviorally report) the visual category of the retrieved image (face or scene) or the original spatial location of the retrieved image (left or right). In visual cortical areas, representation of visual category information was highly robust but was completely insensitive to subjects' top-down goals: visual category representation was as strong when visual category information was relevant compared to when it was irrelevant (location trials). In contrast, the representation of visual category information in LPC was markedly stronger when that information was relevant. This pattern of results suggests that the representation of visual category information in visual cortical areas was relatively automatic, regardless of top-down goals, but that LPC played

some role in filtering information according to task demands. These findings were conceptually replicated in a study by Favila et al. (2018) in which subjects were cued to retrieve either the category of an object image (e.g., backpack, flower, etc.) or the color of the object image (e.g., red, green, etc.). Strikingly, visual cortical areas represented both feature dimensions (object category and color) equally strongly, regardless of top-down goals. In contrast, dorsal LPC exhibited a stronger (and selective) representation of task-relevant feature information.

Approach and structure of the dissertation

As suggested in the previous section, LPC possess two important properties when representing content information. First, the multivariate neural representation of an episode differs between when it was first encountered (i.e., encoding) and when it was the target of episodic retrieval. Second, LPC content representations are sensitive to and thus can be modulated by top-down goals. Yet it remains elusive how LPC multivariate neural representations change across encoding and retrieval and how LPC sensitivity to top-down goals contribute to human memory success. The primary goal of this dissertation is to fill this gap by extending our understanding of the nature and behavioral relevance of LPC content representations.

These investigations are conducted in a set of two human fMRI studies. Chapter II aims to examine how differences in LPC neural representations between encoding and retrieval are indicative of recognition memory success. Specifically, I show that the *differences* in LPC multivariate neural representation patterns across encoding and retrieval inherently contain decodable, feature-level information about stimuli. Importantly, I provided evidence that the degree of the decodable information embedded

in the encoding-retrieval neural representation *differences* is predictive of behavioral memory success. Chapter III aims to test how adaptive (i.e., goal-relevant) content representation modulations in LPC can solve memory interference and contribute to episodic memory success. Specifically, I show that LPC content representations engaged adaptive distortions that differentiate confusing memories, and the extent of such differentiation process is predictive of behavioral memory success. Lastly, in chapter IV, I summarize and discuss how these findings advance our understanding of the functional role of LPC in supporting memory.

CHAPTER II

REPETITION-RELATED MEMORY SIGNALS IN PARIETAL CORTEX INTEGRATE INFORMATION ABOUT STIMULUS CONTENT

This chapter contains unpublished co-authored material. I am the primary author of this chapter, and I incorporated editing advice from Brice A. Kuhl. and J. Ben Hutchinson. The experiment design and data collection were done by Emily J. Allen, Yihan Wu, Thomas Naselaris, and Kendrick Kay. My contributions to this chapter include data processing, model building, data analyses, and manuscript writing.

Introduction

Research on recognition memory has long shown that repeated presentations of the same stimulus tend to induce systematic changes in brain activity compared to when it was first encountered (Grill-Spector et al., 2006). Specifically, sensory modalities, such as visual and auditory processing, consistently show reduced univariate activation strength in response to repeated presentations of past stimuli or features (i.e., repetition suppression; Grill-Spector & Malach, 2004; Segaert et al., 2013). For example, evidence from both human fMRI and primate single-cell recording studies show that the primate visual memory area (i.e., inferior temporal cortex) and human visual cortex responded to the repeated stimulus with reliably less BOLD signal or spiking rate (Henson et al., 2000; Jiang et al., 2000; Li et al., 1993; Miller et al., 1991; Sobotka & Ringo, 1996). On the other hand, parietal regions demonstrate the opposite univariate pattern of repetition suppression in memory research. That is, the parietal cortex shows greater univariate neural activity in response to old stimuli (i.e., repetition enhancement; Wagner et al., 2005). Importantly, the enhancement of neural activity is only present when memory retrieval of the old stimulus is successful, and, similarly, memory strength for the stimulus is predictive of the degree of univariate activity enhancement (Hutchinson et al.,

2009). Together, although visual and parietal regions respond to repeated stimuli in opposite manners, both sets of results suggest that repetition-induced changes in univariate neural activities (i.e., repetition suppression and enhancement) are memory-based in nature. However, how repetition-related neural activity changes are related to memory components remains elusive.

With the recent development of multivariate pattern analysis, it has been shown that patterns of neural activity in both visual and parietal regions reflect specific contents of episodic memories (Bird et al., 2015; Buchsbaum et al., 2012; Jost et al., 2012; Khader et al., 2005; St-Laurent et al., 2015). For example, previous research has succeeded in decoding memory contents of various granularity (e.g., from event to feature level) from both regions (Norman et al., 2006). Moreover, evidence suggests that the neural activity patterns of the same stimulus across two repetitions are consistently more similar compared to those of different stimuli (Kriegeskorte et al., 2008), suggesting that the neural activity patterns in visual and parietal regions must contain specific content information specific to each stimulus. Notably, the degree to which the memory content is reinstated or decoded can reliably predict subsequent memory performances (Kuhl & Chun, 2014; Lee et al., 2017). These advances in multivariate content representation raise an important question regarding the nature of repetition-induced univariate effects (i.e., suppression and enhancement). Specifically, content information is indexed by either characterizing patterns at encoding or retrieval separately or by testing for the similarity between encoding and retrieval. But, if recognition memory signals are fundamentally expressed as a difference across repetitions (encoding to retrieval), this raises the question of whether that difference metric is also related to the content, and whether the

neural activity patterns of repetition suppression and enhancement may also contain content information.

In the current study, we aim to examine whether the repetition-related neural activity changes contain content information. We analyzed data from an open-source dataset, the Natural Scene Dataset (NSD), where each of the eight subjects viewed 10,000 images three times over a year's time course (Allen et al., 2022). In particular, we focused on the first and second presentations of each stimulus. Regions that demonstrated univariate repetition-related changes (repetition enhancement or repetition suppression) were selected as the ROIs. We quantified the content of the images based on weights from a late layer of a popular convolutional neural network, VGG16 (Simonyan & Zisserman, 2014). We tried to decode the content information from the neural activity pattern difference between the two presentations of each stimulus, and we predicted regions that demonstrate univariate activity changes would demonstrate content coding within the activity difference. We also predicted that behavioral recognition memory success would be critical for decoding content information from repetition-related differences in activation.

Methods

Participants

Eight participants (six female, mean age = 26.5, age range = 19-32 years) with normal or corrected-to-normal vision were recruited for the experiment. None of the participants had cognitive deficits or color blindness. Informed consent was obtained in accordance with the University of Minnesota Institutional Review Board.

Experimental Paradigm

All participants performed a long-term continuous recognition task. Specifically, each participant viewed 10,000 distinct images over 40 sessions with each image repeating 3 times (including the first presentation). For each trial, participants were asked to indicate whether they have seen the image before via a button box (i.e., recognition task). Among the 10,000 images presented to each participant, 9000 images were unique for the given participant while the remaining 1000 images were shared across participants. Note that the time intervals between image repetitions were identical across all participants, thus matching the recognition task difficulties.

During each trial, the image was on for 3 seconds and off for 1 second. Each run contained 75 trials, resulting in a total of 300 seconds. The first 3 trials and last 4 trials of each run were always blank trials, and the remaining 68 trials included another 5 randomly positioned blank trials. Additionally, for even-numbered runs, the 63rd stimulus trial was always a blank trial, resulting in 63 and 62 stimulus trials for odd and even runs, respectively. Note that the blank trials were positioned such that the continuous number of stimulus trials in each run always ranged from 9 to 14 trials. Each scanning session contained 12 runs with 750 stimulus trials in total. This temporal ordering of stimulus and blank trials was consistent across 40 sessions. Due to the time restriction, 4 out of 8 participants finished all 40 sessions. The remaining participants finished at least 30 sessions. Since not all participants finished viewing each image for the third time, for the following analyses, we only used the first and second presentations of the images. To avoid the extremely long time spacing between the first and second representation, we

only used images that have a lag within 20 sessions (~5 months) between the first and second presentation. More comprehensive details can be found in (Allen et al., 2022).

Stimuli

All images were selected from Microsoft's Common Objects in Context (COCO) image database (Lin et al., 2014). Images were cropped into squares along the largest dimension when necessary, resulting in a total of 73,000 colored RGB images with a size of 425 pixels x 425 pixels.

fMRI data acquisition

Imaging data were collected on a 7T Siemens Magetom passive-shielded scanner with 32 channels head coil at the Center for Magnetic Resonance Research at the University of Minnesota. Functional data were acquired using a high-resolution T2 gradient-echo EPI sequence at 1.9-mm isotropic resolution with whole-brain coverage (84 axial slices, slice thickness = 1.8mm, slice gap = 0 mm, field-of-view = 216 mm (FE) x 216 mm (PE), phase-encode direction anterior-to-posterior, matrix size = 120 x 120, TR = 1600 ms, TE = 22.0 ms, flip angle = 62°, echo spacing = 0.66 ms, bandwidth = 1736 Hz/pixel, partial Fourier = 7/8, in-plane acceleration factor (iPAT) = 2, multiband slice acceleration factor = 3). Several dual-echo EPI fieldmaps were acquired periodically over each scan session (2.2mm × 2.2mm × 3.6mm resolution, TR = 510ms, TE1 = 8.16ms, TE2 = 9.18ms, flip angle = 40°, partial Fourier = 6/8). Anatomical images were collected using a 3T Siemens Prisma scanner with a standard 32 channels head coil. Several (6-10) whole brain T1 weighted scans were acquired for each participant across the experiment using MPRAGE sequence (0.8mm isotropic resolution, TR = 2400ms, TE = 2.22ms, TI = 1000ms, flip angle = 8°, inplane acceleration factor (iPAT) = 2).

fMRI data preprocessing

T1 weighted images were corrected for gradient nonlinearities using the Siemens gradient coefficient file from the scanner. Multiple volumes acquired for a given subject were first co-registered and then averaged to create a single T1-weighted anatomical image. The T1 volume was processed by FreeSurfer 6.0.0 with *-hires* option. Manual edits of segmentation were performed to improve the accuracy of surface reconstruction. Several additional cortical surfaces were generated at 25%, 50%, and 75% positions of the distance between the pial surface and the boundary between white and gray matter. These surfaces were used to generate surface representations of the fMRI data.

fMRI data preprocessing was performed with customized scripts in MATLAB in the subject native space. Temporal resampling was performed to correct slice differences and spatial resampling was performed to correct head motion, EPI distortion, and gradient non-linearities. In the current paper, we used the upsampled 1-mm high-resolution version of the data. For each subject, all later fMRI data sessions were co-registered to the mean fMRI volume of the first fMRI data session.

Single-trial estimation

GLM analysis was performed on the pre-processed time series data in a 1mm functional space with a customized package GLMsingle (Allen et al., 2022). The best-fitting HRFs were chosen for each voxel from the GLMsingle library for compensating the differences in hemodynamic time courses across voxels. GLMdenoise was adapted for single-trial estimation for removing noise from beta estimates (Kay et al., 2013). The estimated single-trial betas were further resampled to three different cortical surface

depths and averaged together using cubic interpolation. The result was then transformed to fsaverage space using nearest neighbor interpolation.

Region of interest

fMRI data analyses were conducted with a set region of interest (ROIs) located in the visual, parietal, and motor cortices derived from the HCP MMP1 atlas (Glasser et al., 2016). The ROIs were selected to be sensitive to memory success. Specifically, for each participant and parcel in the atlas, we computed the univariate activation contrast of hit vs. miss trials during the second presentation. IP1, IP2, and PFm from the lateral parietal cortex (LPC) demonstrated a retrieval success effect (significantly greater activation for the second presentation compared to the first), whereas PHA1 and PHA2 from the ventral temporal region showed a significant repetition suppression effect. We also included low-level visual region V1 as the control region.

Univariate analysis of memory success

All parcels from the lateral parietal region and ventral temporal region, as well as V1, were involved in this analysis. For each subject and ROI, we first computed the averaged Beta for the first presentation and the second presentation separately, and then took the difference between averaged activation of the first presentation over the second presentation. One-sample t-tests were conducted for each subject and ROI on whether the activation difference between the presentations is different from 0. Images were excluded from analysis if the response to the first presentation was not a correct rejection (i.e., 'new') or if the lag between the 1st and 2nd presentation of an image was greater than 20 sessions.

Quantifying image content

Each image was passed through a pre-trained VGG16 to derive feature maps at convolutional layer 4 (referred to as the early layer), convolutional layer 7 (referred to as the middle layer), and fully connected layer 3 (referred to as the late layer) (Simonyan & Zisserman, 2014). For the two convolutional layers, we concatenated the feature map matrices at each layer to obtain the 1-D vector format feature map. PCA was conducted on the vectorized feature map of each layer across all images. We kept the top 10 principal components for each layer, which explained 4.2%, 3.2%, and 66.3% of the variance for the early, middle, and late layers respectively. The content of each image was quantified as the loading of these 10 principal components for the given image (see figure 2.1).

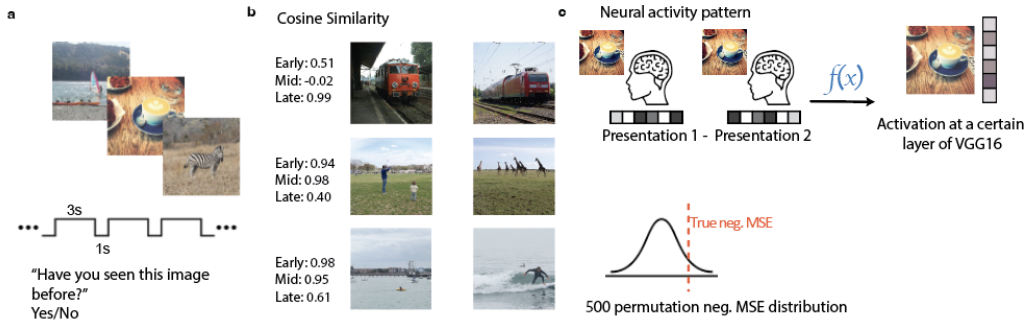

Figure 2.1. Paradigm, stimuli, and illustration of machine learning models.
a, Experiment paradigm. Subjects viewed each image for 3 seconds, while making the old/new judgment. **b**, Pairwise picture stimuli similarities measured by feature maps from different layers of VGG16. **c**, Predict the image content by using the neural activity difference from the two presentations.

Mapping brain signals to image content

To map the brain signal to image content, we trained a ridge regression model for each ROI, using the image-specific neural activity patterns as input features to predict the image content, as represented by the loadings of the 10 PCs of the feature maps derived from VGG16. Model estimation was done within each subject and each ROI. To quantify

the model performance, we used a 10-fold cross-validation framework with the negative mean squared error (neg. MSE) at the trial level as the indicator of the model performance. As described previously, the model outcome, Y, is the 10 principal components of a certain VGG16 layer feature map. Then neg. MSE was computed by (Predicted Y – True Y)/10, since we have 10 components for each Y. Specifically, for each train-test split, we built a true model with neural activity pattern as input, and 10 PCs of the feature maps from VGG16 as output. In addition to the true model, we built 500 null models with permutation data (by randomly shuffling the order of the label) to form a null distribution of the negative MSE of each trial. We computed the z-score of the true negative MSE (z-neg. MSE) of each trial using the mean and variance of the corresponding null distribution.

To confirm the ROI selection, we used the first presentation of each image with the correct response (correct rejection) as the input to predict the top 10 PCs of the early (Conv4), middle (Conv7), and late layer (FC3) of VGG16. For each subject and ROI, we compute the model performance (z-neg. MSE) and tested whether the model performance is significantly above the chance level for each selected ROI with a one-sample t-test. Only images with correct rejection during the first presentation and the two presentations of the image's lag was smaller than 20 sessions were included in the analysis.

Content decoding from repetition-related differences

For memory content decoding based on repetition-related brain activity differences, we subtracted the brain activity during the second presentation from the first presentation of each image. Then we used these activity differences as the model input to predict the top 10 PC of the late layer of VGG16. For each subject and ROI, we compute

the model performance (z-neg. MSE) and tested whether the model performance is significantly above the chance level for each selected ROI with a one-sample t-test. Only images with correct rejection during the first presentation and the two presentations of the image's lag was smaller than 20 sessions were included in the analysis.

Content decoding as a function of behavioral expressions of memory

To investigate whether the repetition-related differences are related to behavior, we separately computed the z-neg. MSE for hit trials and miss trials (whether the response to the second presentation is correct) for each ROI and subject, and then tested the model performance for hit and miss trials separately with a one-sample t-test. We next subtracted the averaged z-neg. MSE of miss trials from hit trials and tested whether the model performance was better for hit trials over miss trials with a one-sample t-test for each ROI. For control analysis, we first tested whether the result was driven by the imbalanced trials. We computed the ratio of hit trials over miss trials for each subject and calculated the Pearson correlation for the hit trials ratio and the hit/miss effect (z-neg. MSE). Next, for each subject and parietal ROI (IP1, IP2, PFm), we built models with hit trials and tested them on the hit trials, and built models with miss trials and tested it on the miss trials. We compared the model performance (z-neg. MSE) built with hit trials and miss trials at three parietal ROIs by two-way ANOVA with ROI and hit/miss as the two main factors. Then, we built and tested three different types of models with data from presentation 1 and presentation 2 neural activity of the presentations separately. For each of these three types of models, we measured the z-neg. MSE separately for hit trials and miss trials at each ROI for each subject. Then we subtracted the averaged z-neg. MSE of

the miss trials from the hit trials, and tested whether these values were different from 0 with one-sample t-tests.

Statistical modeling and analysis

Statistical analyses were performed using Python 3.7. The pre-trained VGG16 model from the Pytorch module was used for the feature map extraction. Principle component analysis, ridge regression model, and cross-validation were conducted with the Scikit-Learn module. All t-tests were done with the Scipy module two-tailed, with $\alpha = 0.05$. All ANOVAs were computed with the Statsmodels module. All error bars in the figures represent S.E.M.

Results

Univariate effects of recognition memory in parietal and ventral temporal cortices

We first sought to identify candidate ROIs from lateral parietal and ventral temporal cortices that reflected repetition-related univariate changes. To do this, we selected ROIs that exhibited differences in univariate activation for the first presentation versus the second presentation of the same stimulus. For each anatomical parcel in lateral parietal and ventral temporal cortices, the mean activation was computed for the two presentations and these values were compared via two-tailed paired-sample *t*-tests. Three lateral parietal regions exhibited greater activation for the second presentation compared to the first (i.e., retrieval success; IP1: $t_7 = -3.14$, $p = 0.016$; IP2: $t_7 = -4.26$, $p = 0.004$; PFm: $t_7 = -3.46$, $p = 0.011$) whereas two regions in the ventral temporal cortex (PHA1, $t_7 = 2.58$, $p = 0.036$; PHA2, $t_7 = 2.40$, $p = 0.048$) exhibited the opposite pattern (i.e., repetition suppression). These five ROIs were used for all subsequent analyses. For comparative purposes, we also included the early visual cortex as a control region. Note

that the early visual cortex does not demonstrate significant differences in univariate activation between first and second representations (V1: $t_7 = -1.55$, $p = 0.166$).

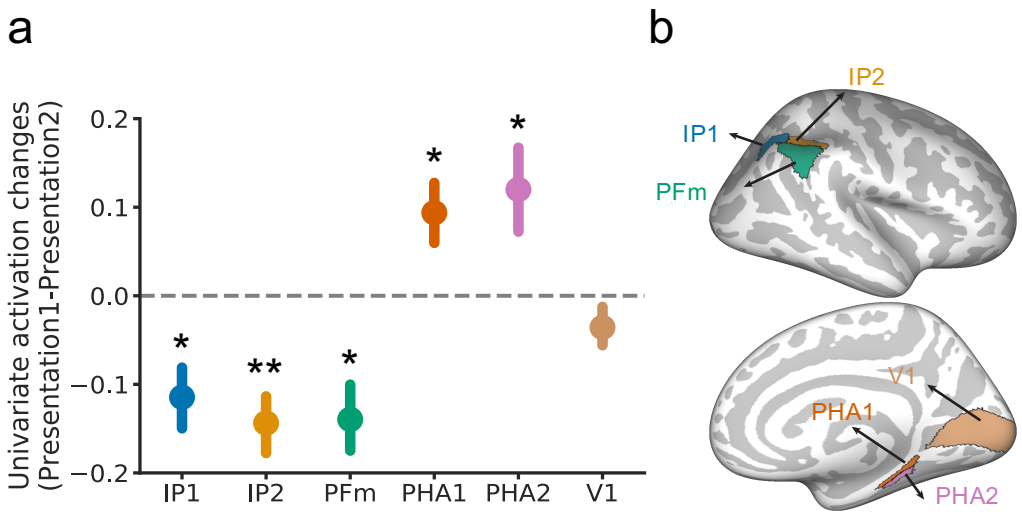

Figure 2.2. Univariate results.
a,The univariate activity changes across the two presentations. **b**, Illustration for ROIs. Error bars reflect +/- S.E.M.; ** $p < 0.01$, * $p < 0.05$

Measuring content information with VGG16

To quantify the content of each scene image, we used the activation map of the selected hidden layer in a pre-trained deep convolutional neural network (VGG16) as an unbiased measure for image content information (Simonyan & Zisserman, 2014). The activation map in the selected layer aims to capture the high-level semantic information encoded within the lateral parietal and ventral temporal regions. To achieve this property, we focused on the final fully connected layer (the deepest layer in the network). The rationale comes from the nature that the deep convolutional neural network resembles the functional hierarchy of the human ventral visual stream, with shallow layers capturing low-level visual features (e.g., color) and deep layers being more sensitive about visual content information of semantic meanings (see Figure 2.1b). To quantitively validate our choice of the deep layer, we first tested the degree to which information from the last

VGG16 layer (FC3), in comparison to that from shallow regions, is reflected in the neural activity patterns expressed in parietal and visual cortical areas during the first presentation of each image. In particular, we expect to see greater relevance between shallow layers and early visual region (i.e., V1), and between FC3 and the remaining ROIs.

To construct a decoding model, we first reduced the dimensionality of the hidden layer activation maps using Principal Components Analysis (PCA) for each scene image. Specifically, we selected the top 10 components, allowing the content of each scene image to be expressed as a vector of 10 component scores for each layer (early, middle, late). For each ROI, we trained ridge regression models on stimulus-induced neural activity patterns to predict the 10-vector content representation of each stimulus, with respect to each layer. The model was tested using a ten-fold cross-validation framework, and the model performance for each fold was measured as the averaged negative MSE (neg. MSE, where a larger number represents better prediction; see Methods) between the predicted and actual component scores across all stimuli. Statistical significance was derived from a permutation test where labels (i.e., content representation vectors) were randomly shuffled (see Methods).

As shown in Figure 2.3, the prediction performance of the model constructed with FC3 content representations was well above chance for each of the lateral parietal (IP1, IP2, PFm) and visual (PHA1, PHA2, V1) ROIs. Notably, model performances for the lateral parietal (IP1, IP2, PFm) and ventral temporal ROIs (PHA1, PHA2) gradually increased as content representations were computed from shallower to deeper layers, as shown by significant main effects of the layer in one-way repeated measures ANOVAs

(IP1: $F_{2,14} = 22.60$, $p < 0.0001$; IP2: $F_{2,14} = 24.00$, $p < 0.0001$; PFm: $F_{2,14} = 28.86$, $p < 0.0001$; PHA1: $F_{2,14} = 67.05$, $p < 0.0001$; PHA2: $F_{2,14} = 49.65$, $p < 0.0001$). In contrast, model performance in V1 was better off from content representations computed from shallower compared to the deeper layers layer ($F_{2,14} = 575.14$, $p < 0.0001$). Separate two-way ANOVAs confirmed that the pattern of data across layers in V1 significantly differed from the pattern of data across layers in each of the lateral parietal and ventral temporal ROIs (V1 vs. IP1: $F_{2,14} = 162.06$, $p < 0.0001$; V1 vs. IP2: $F_{2,14} = 335.75$, $p < 0.0001$; V1 vs. PFm: $F_{2,14}$ 370.49, $p < 0.0001$; V1 vs. PHA1: $F_{2,14} = 202.86$, $p < 0.0001$; V1 vs. PHA2: $F_{2,14} = 170.24$, $p < 0.0001$). These results provide evidence that the content information represented in the lateral parietal and ventral temporal ROIs is best captured by activation maps of a deep (compared to shallow) CNN layer. It is also worth noting that V1 showed relatively weak decoding performance for deep layer content representations, indicating that there was not a global bias toward better performance for FC3 (e.g., as an artifact of the dimensionality reduction step).

Table 2.1. Prediction performance of models constructed with different layers of VGG16.
Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

ROI	Early layer					Middle layer					Late layer				
	Mean	SD	t_7	p	sig.	Mean	SD	t_7	p	sig.	Mean	SD	t_7	p	sig.
IP1	0.37	0.23	4.45	0.003	**	0.59	0.39	4.28	0.004	**	0.84	0.49	4.85	0.002	**
IP2	0.22	0.12	5.08	0.001	**	0.33	0.20	4.74	0.002	**	0.49	0.27	5.01	0.002	**
PFm	0.17	0.11	4.56	0.003	**	0.27	0.17	4.44	0.003	**	0.44	0.24	5.09	0.001	**
PHA1	0.74	0.23	9.26	0.000	***	1.16	0.33	9.89	0.000	***	1.59	0.50	8.97	0.000	***
PHA2	0.78	0.35	6.29	0.000	***	1.23	0.49	7.13	0.000	***	1.76	0.72	6.90	0.000	***
V1	1.49	0.16	26.44	0.000	***	1.77	0.19	27.06	0.000	***	0.73	0.14	15.20	0.000	***

Content decoding from repetition-related differences

The preceding analyses provided evidence that content information encoded in the lateral parietal and ventral temporal ROIs can be best quantified using the activation

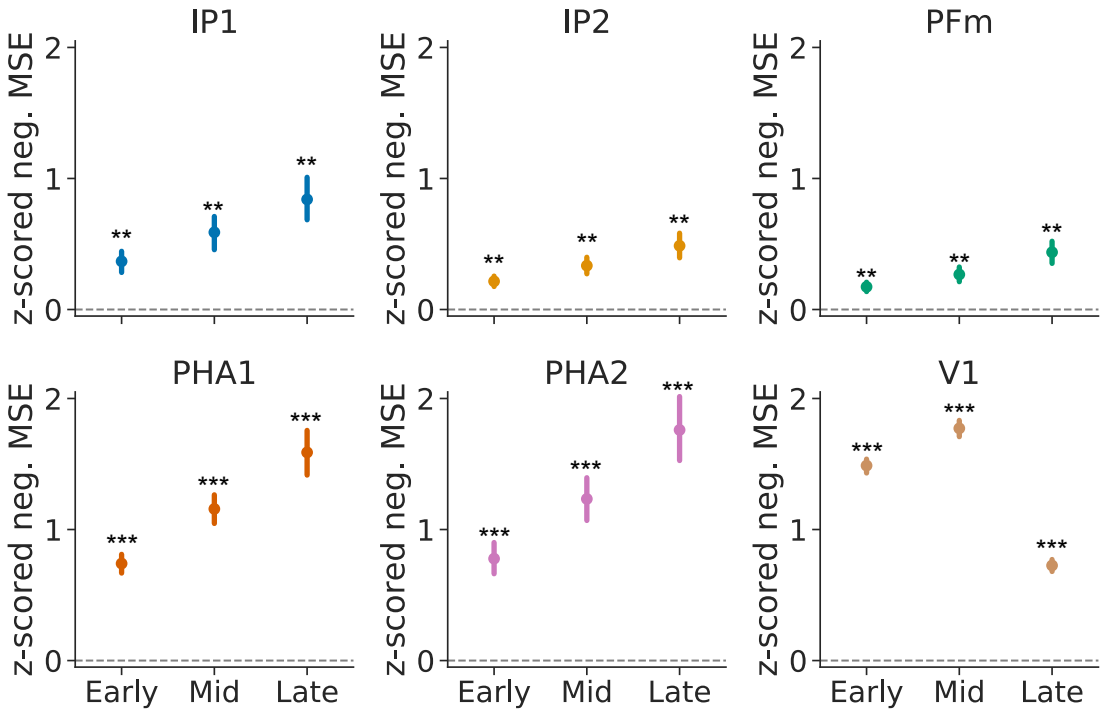

Figure 2.3. Prediction performance of the model constructed with different layers of VGG16. To confirm the ROI selection, we used the first presentation of each image with the correct response (correct rejection) as the input to predict the top 10 PCs of the early (Conv4), middle (Conv7), and late layer (FC3) of VGG16. Error bars reflect +/- S.E.M.; *** $p < 0.001$, ** $p < 0.01$

profile of the selected layer in VGG 16 (FC3). Here, we aim to examine the critical question of whether the *difference* in repetition-induced neural activity —from the first presentation to the second presentation—contained content information about image stimuli, as represented by VGG16. Note that if the stimulus-evoked neural activity patterns are identical every time the stimulus was encountered, content information decoding from repetition-related neural differences would not be possible, even if content information robustly presents in each individual encounter. Moreover, even if there are robust changes in repetition-induced neural activity signals (i.e., memory-related signals), these differences could be orthogonal to content information. Thus, the successful decoding performance of an ROI suggests that the nature of repetition-induced neural

activity in this region is memory-related signals that *incorporate* stimulus content information.

Repetition-induced activation difference was computed for each image stimulus (first presentation – second presentation) and used to predict content information captured by the VGG FC3 layer. The model performance was cross-validated and statistical significance was derived from the permutation test. We found that content information can be successfully decoded from lateral parietal and visual ROIs, as assessed by permutation tests and z-neg. MSE (lateral parietal: $t_7 > 2.38$, $p < 0.049$; V1: $t_7 = 4.37$, $p = 0.003$ one-sample t-tests, see Table 2.2) but not from ventral temporal ROIs ($t_7 < 2.34$, $p > 0.052$).

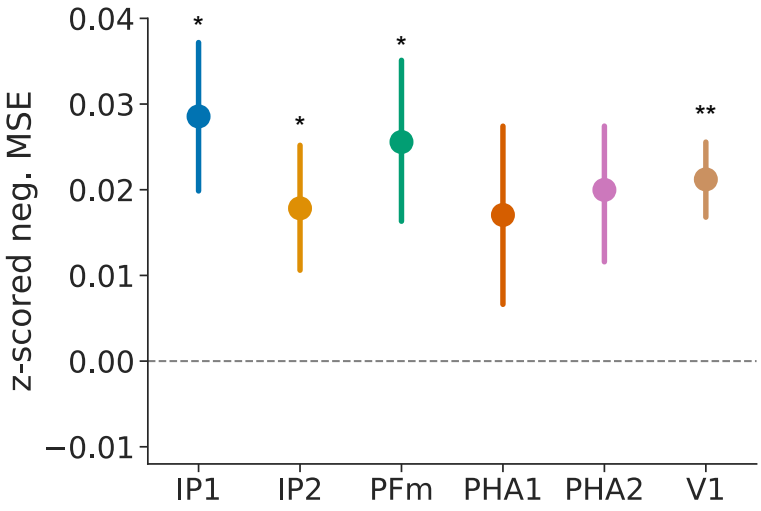

Figure 2.4. Content decoding from repetition-related brain activity changes.
For memory content decoding based on repetition-related brain activity differences, we subtracted the brain activity during the second presentation from the first presentation of each image. Then we used these activity differences as the model input to predict the top 10 PC of the late layer of VGG16. For each subject and ROI, we compute the model performance (z-neg. MSE) and tested whether the model performance is significantly above the chance level for each selected ROI with a one-sample t-test. Error bars reflect +/- S.E.M.; ** $p < 0.01$, * $p < 0.05$

Content decoding as a function of behavioral expressions of memory

The preceding results suggested that the repetition-induced neural activity differences can be memory-based signals that contain content information. It is yet

Table 2.2. Content decoding from repetition-related brain activity changes.
Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

ROI	Mean	SD	t_7	p	sig.
IP1	0.03	0.03	3.13	0.017	*
IP2	0.02	0.02	2.38	0.049	*
PFm	0.03	0.03	2.49	0.041	*
PHA1	0.02	0.03	1.56	0.162	
PHA2	0.02	0.02	2.34	0.052	
V1	0.02	0.01	4.37	0.003	**

unclear to which degree such memory-based signals can contribute to memory success. Thus, we next aim to examine whether the ability to decode content information from repetition-induced neural activity is related to recognition memory success for an image stimulus. Specifically, we used the same ridge regression models trained above but sorted the test trials based on their recognition memory outcomes, either being recognized as a familiar image ('hit') or a new image ('miss'). The rationale is that if the memory-based nature of repetition-induced neural activity differences indeed contributes to recognition memory success, the model performance should be better for hit trials.

We directly compared performance accuracy for hit vs. miss trials by subtracting the mean z-neg. MSE of miss trials from the mean z-neg. MSE of hit trials. Our results revealed that for ROIs that demonstrate significant repetition-induced univariate changes (i.e., either repetition suppression or retrieval success), the model performance was significantly better for hit compared to miss trials (IP1: $t_7 = 4.89$, $p = 0.002$, IP2: $t_7 = 3.33$, $p = 0.013$, PFm: $t_7 = 4.74$, $p = 0.002$; PHA1: $t_7 = 4.74$, $p = 0.002$; PHA2: $t_7 = 2.46$, $p = 0.044$; one-sample t-tests; Table 2.3; Figure 2.5). A follow-up one-way repeated measure ANOVA indicated no performance difference between parietal and ventral temporal ROIs ($F_{1, 7} = 0.33$, $p = 0.582$). More broadly speaking, the repetition-induced

activity changes only contain content information for successfully remembered hit trials, but not for forgotten stimuli. On the contrary, the early visual cortex showed no difference for hit versus miss trials ($t_7 = 0.84$, $p = 0.43$), suggesting that the content information in V1 does not contribute to the behavioral differences in recognition memory.

Next, we found that when separately considering hit trials, content information could be successfully decoded from almost all ROIs, with trending statistical significance in one ventral temporal ROI (Table 2.3; Figure 2.5; IP1: $t_7 = 4.80$, $p = 0.002$; IP2: $t_7 = 3.25$, $p = 0.014$; PFm: $t_7 = 3.43$, $p = 0.011$; PHA1: $t_7 = 2.47$, $p = 0.092$; PHA2: $t_7 = 2.47$, $p = 0.043$; V1: $t_7 = 3.14$, $p = 0.016$; one-sample *t*-test). Notably, there was no difference between averaged model performance between parietal and ventral temporal cortices ($F_{1, 7} = 1.32$, $p = 0.288$, one-way repeated measures ANOVA). On the other hand, for miss trials, only two visual regions, V1 and PHA2 exhibited above-chance performance (V1: $t_7 = 3.84$, $p = 0.006$; PHA1: $t_7 = -2.5$, $p = 0.041$; one-sample *t*-test; Table 2.3; Figure 2.5). There was no observable difference between parietal and temporal cortices ($F_{1, 7} = 0.33$, $p = 0.582$, one-way repeated measures ANOVA). These results indicate that repetition-induced activity differences in ventral temporal and parietal ROIs contain content information, but only for hit trials.

Distinct content decoding profile between first and second presentations

To better understand how repetition-induced neural activity changes contribute to behavioral recognition memory success, we examined the content decoding performances of models relying on the neural activity patterns from only a single presentation (rather than repetition-related difference). We found that model performances during the first

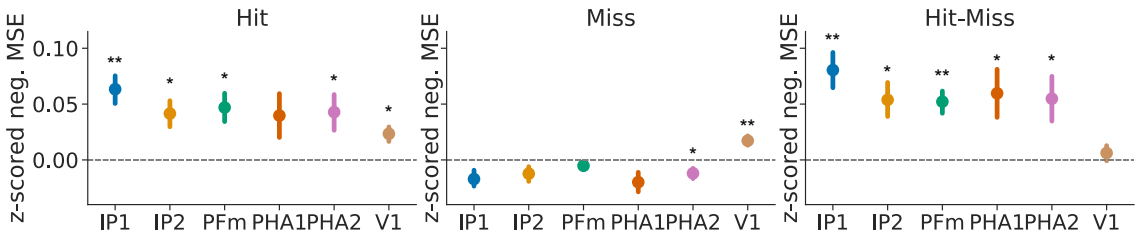

Figure 2.5. Content decoding as a function of behavioral expressions of memory.
We tested whether the model performance was different for hit trials and miss trials. Error bars reflect +/- S.E.M.; ** $p < 0.01$, * $p < 0.05$

Table 2.3. Content decoding as a function of behavioral expressions of memory.
Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

ROI	Hit					Miss					Hit-Miss				
	Mean	SD	t_7	p	sig.	Mean	SD	t_7	p	sig.	Mean	SD	t_7	p	sig.
IP1	0.06	0.04	4.8	0.002	**	-0.02	0.02	-2.33	0.053		0.08	0.05	4.89	0.002	**
IP2	0.04	0.04	3.25	0.014	*	-0.01	0.02	-1.67	0.138		0.05	0.05	3.33	0.013	*
PFm	0.05	0.04	3.43	0.011	*	-0.01	0.01	-1.28	0.240		0.05	0.03	4.74	0.002	**
PHA1	0.04	0.06	1.95	0.092		-0.02	0.03	-2.12	0.072		0.06	0.07	2.46	0.044	*
PHA2	0.04	0.05	2.47	0.043	*	-0.01	0.01	-2.50	0.041	*	0.05	0.06	2.64	0.033	*
V1	0.02	0.02	3.14	0.016	*	0.02	0.01	3.84	0.006	**	0.01	0.02	0.84	0.429	

presentation resemble the subsequent memory effect. That is, the degree to which information about an image stimulus could be decoded from each ROI strongly depended on whether the stimulus was later successfully remembered. Specifically, models trained on all regions demonstrated significantly better performance for subsequently-remembered stimuli compared to subsequently-forgotten stimuli (IP1: $t_7 = 6.39$, $p = 0.0003$, one-sample t-test; IP2: $t_7 = 6.13$, $p = 0.0004$; PFm: $t_7 = 2.89$, $p = 0.023$; PHA1: $t_7 = 4.74$, $p = 0.002$; PHA2: $t_7 = 2.46$, $p = 0.044$; V1: $t_7 = 3.40$, $p = 0.009$). There was no significant difference for averaged parietal regions and ventral temporal regions ($F_{1, 7} = 0.003$, $p = 0.957$, one-way repeated measures ANOVA). Interestingly, the content decoding profile demonstrated the opposite pattern during the second presentation. That

is, content decoding model performances were greater for forgotten stimuli compared to remembered ones for all ROIs, with trending statistical significance in parietal regions (IP1: t_7 = -1.5, p = 0.177, IP2: t_7 = -2.17, p = 0.066; PFm: t_7 = -1.68, p = 0.137; PHA1: t_7 = -2.81, p = 0.026; PHA2: t_7 = -2.92, p = 0.022; V1: t_7 = -2.66, p = 0.032; one-sample t-test; Table 2.4; Figure 2.6). The average model performance of hit-over-miss was significantly different in averaged parietal and ventral temporal regions ($F_{1,7}$ = 11.08, p = 0.013, one-way repeated measures ANOVA). Specifically, the three visual ROIs demonstrated significantly better decoding results for miss trials (perceived as new trials to subjects) compared to hit trials (perceived as old stimuli to subjects), but not for the parietal regions. We argue that this profile indicates an efficient episodic memory encoding strategy. That is, the brain, especially the visual regions, treats the forgotten stimuli as new images and thus automatically encodes the content information of the "miss" trials, thus leading to greater content decoding performance for miss trials during the second presentation. Together, these results suggested that repetition-induced neural differences contribute to recognition memory success by adopting a specific encoding strategy over old vs. new image stimuli.

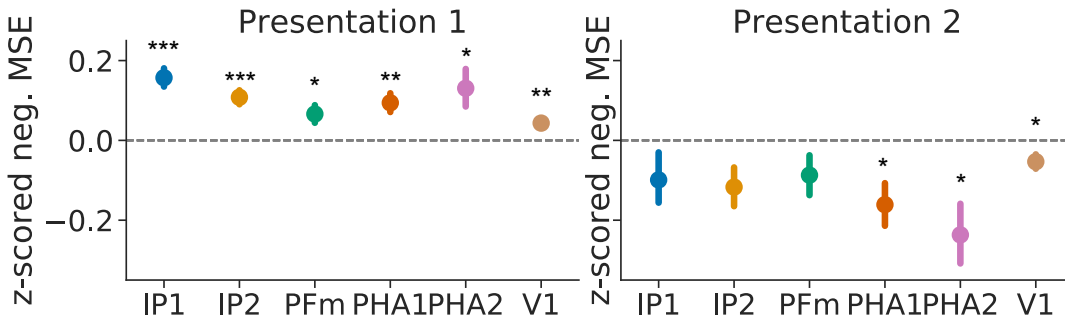

Figure 2.6. Content decoding by the two presentations separately.
We built models by presentation 1 / presentation 2 separately to test whether the results from the difference model were driven by a certain presentation. Error bars reflect +/- S.E.M.; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 2.4. Content decoding by the two presentations separately.
Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

ROI	Presentation 1					Presentation 2				
	Mean	SD	t_7	p	sig.	Mean	SD	t_7	p	sig.
IP1	0.16	0.07	6.39	0.0003	***	-0.1	0.19	-1.5	0.177	
IP2	0.11	0.05	6.13	0.0004	***	-0.12	0.15	-2.17	0.066	
PFm	0.07	0.06	2.89	0.023	*	-0.09	0.15	-1.68	0.137	
PHA1	0.09	0.07	3.59	0.009	**	-0.16	0.16	-2.81	0.026	*
PHA2	0.13	0.14	2.67	0.032	*	-0.24	0.23	-2.92	0.022	*
V1	0.04	0.03	3.60	0.009	**	-0.05	0.06	-2.66	0.032	*

Discussion

Previous studies have provided consistent evidence that the human brain, when encountering an old stimulus, responds differently compared to when the stimulus was first perceived. Specifically, visual regions exhibit less univariate activation when processing an old stimulus (repetition suppression) whereas parietal regions show the opposite pattern (repetition enhancement). The current study investigated the nature of these repetition-related univariate effects from a multivariate perspective. Specifically, regions of interest (ROIs) were identified in visual and parietal regions that demonstrate robust repetition-induced univariate effects (i.e., repetition enhancement and suppression). We then examined whether the pattern of repetition-related neural differences represents content information about the stimuli, measured using a pre-trained convolutional neural network (VGG16), and how the degree of the content representation relates to recognition memory success. First, our results suggest that both parietal and visual regions that show significant univariate-activation changes also robustly reflect memory contents in the multi-voxel activity patterns. Importantly, the degree of content

information in regions that demonstrated either repetition enhancement (parietal regions) or repetition suppression (ventral visual regions) is predictive of recognition memory success, suggesting these regions' role in memory-related processes. Interestingly, we found that content representation during the first presentation (i.e., encoding) alone can predict subsequent recognition memory success in both parietal and ventral temporal regions, while content representation during the second presentation (i.e., retrieval) favors forgotten stimuli, especially in ventral temporal regions.

Repetition-related neural differences reflect content information

Recognition memory refers to the ability to identify stimuli that have been previously encountered. Previous studies have suggested that recognition memory success is coupled with repetition-related univariate changes in the visual and lateral parietal regions. Here, we examined the nature of these univariate activation changes and how they may contribute to recognition memory success. Our results show that, when participants successfully recognized a stimulus as an old image, predictive models could rely on the pattern of repetition-related neural differences to reconstruct content information pertaining to the very stimulus. In other words, the multivoxel patterns of repetition enhancement and repetition suppression for remembered trials carry stimulus-specific information. Note that the below-threshold content representation in the two visual areas observed in Figure 2.4 was driven by the minimum content representations for the miss trials. These results suggest the univariate activation changes in both visual and parietal regions do not demonstrate an invariant form of neuronal excitement for fatigue that uniformly acts on all LPC or visual voxels. Instead, voxels are enhanced or suppressed at different degrees during the re-exposure of an old stimulus, and such a repetition-related

multi-voxel modulation is largely attributed to the content information of the given stimulus. Note that this modulation can be mostly driven by how the stimulus is encoded during its first presentation, which will be discussed in details in the following subsections. Interestingly, our results also show that content representation does not limit to regions that demonstrate robust repetition-related univariate changes. In particular, the early visual regions (V1), though not showing significant univariate changes in response to repeatedly presented stimulus, also show content representation in its repetition-related neural differences.

The degree of content representation in the pattern of repetition enhancement and repetition suppression predicts recognition memory success

Previous work suggests that the degree of repetition-related univariate activation changes in the parietal regions, but not the visual areas, are indicative of recognized memory success. Specifically, the effect of repetition enhancement is stronger for hit compared to miss trials in the parietal regions, but no difference in the magnitude of repetition suppression was observed in the visual regions (Ward et al., 2013). Here, we examined whether the degree of content representations in the visual and parietal regions is related to recognition memory success and whether this brain-behavioral relationship varies between parietal and visual regions. Contrary to what was observed in previous univariate analyses, our results suggested that the degree to which content information in both visual and parietal regions is indicative of recognition memory success. Specifically, the predictive models demonstrated significantly better performances for remembered (i.e., hit) compared to forgotten (i.e., miss) trials. That is, the degree to which the multivoxel pattern of repetition-related changes, either repetition enhancement or

suppression, resembles true stimulus-specific content information determines whether a given stimulus can be successfully recognized by participants. These results suggest that although the univariate activation of the visual areas does not show greater suppression in response to "remembered" old stimuli, the information carried in the pattern of repetition suppression may be sharpened, which then contributes to recognition memory success. It is worth noting that although the repetition-related neural differences in the early visual cortex (V1) do show content representation, the degree of V1 content representation does not differ between remembered and forgotten trials. It is thus suggested that in LPC and VTC, only the neural differences pattern in regions that show repetition-related univariate effects can be predictive of recognition memory success.

Repetition-related neural differences as a result of encoding strategy

We also looked at the first and second presentations to investigate the source of the content representation embedded in the repetition-related neural differences. We speculate that the differences in content representation profiles between the two presentations are due to certain properties of memory encoding. It has been suggested that the amount and accuracy of stimulus-specific information encoded in the brain when stimuli are first encountered will determine whether the stimulus can be subsequently remembered (Lee et al., 2017). Consistent with the previous theory, our results show that when participants encountered the stimuli for the first time, greater content representations were observed for subsequently remembered items. Conversely, when participants re-encountered the stimuli, greater content representations were found for the forgotten items (participants forgot they have seen the image before). This bias toward encoding subjectively new stimuli may be due to the fact that the brain favors new

information over the old (Nyberg, 2005). That is, the remembered items were sharply represented in the parietal and visual areas only when it was first encountered but were not the main target of encoding when it was re-encountered. In this way, the repetition-related differences contained content representations carried over from the first presentation. On the other hand, by staying subjectively new, the forgotten items were the target of memory encoding of both presentations, although the sharpness of content presentations may vary. As a result, content representations from the two presentations can be canceled out when computing the repetition-related neural differences, leading to weaker content representation strength compared to the remembered items.

Using Convolutional neural network to quantity memory content

In the current study, memory contents were quantified using feature maps from a deep convolutional neural network (CNN), VGG16 (Simonyan & Zisserman, 2014). VGG16 is a typical deep CNN that consists of 16 convolutional layers that resembles the human ventral visual stream. Neurons in each convolutional layer extract information (by performing convolution computation) from the previous layer's output (feature maps) with a certain receptive field (filters) (Güçlü & van Gerven, 2015). In this way, early convolutional layers in a VGG16 tend to capture relatively low-level visual features, such as angle and color whereas deeper layers were able to capture increasingly complex visual features that convey semantic meanings, such as "ocean" or "baseball games" (Zeiler & Fergus, 2014). Thus, feature maps of a CNN reveal image content at different granularity, and this property makes it possible to extract different levels of features in an image stimulus.

Although some research has found that late layers from CNNs show similar categorization performance as the ventral visual system in primates (Cadieu et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2014), little is known about the mapping relationships between human visual system to layers from CNN. Here we chose convolutional layer 4 as the early layer for capturing low-level features like color and angles, convolutional layer 7 as the middle layer for capturing combined shapes and colors, and fully connected layer 3 as the late layer for capturing complex semantic-level features. This decision was made based on the visualization of image projection on each feature map (Figure 2.1b). Future works could provide more insights into explaining the connections between the human visual system and the CNN structures.

Results from this current study showed that using a CNN is a valid method to quantify stimulus content perceived by participants. In recent years, a growing body of cognitive neuroscience research is trying to utilize advances in deep learning to understand human brain mechanisms. While recurrent neural networks and transformers might be more similar to the human memory system due to their storage or attention properties, CNNs resemble the human visual system better given their feedforward structure.

CHAPTER III

ADAPTIVE MEMORY DISTORTIONS ARE PREDICTED BY FEATURE REPRESENTATIONS IN PARIETAL CORTEX

From Zhao, Y., Chanales, A. J., & Kuhl, B. A. (2021). Adaptive memory distortions are predicted by feature representations in parietal cortex. *Journal of Neuroscience*, *41*(13), 3014-3024.

Introduction

Given the vast number of memories that humans store, the overlap between memories is inevitable. For example, one may have taken multiple vacations to the same town or parked in the same garage on many occasions. There is a long history of behavioral studies in psychology documenting the many contexts in which this type of overlap leads to memory interference and forgetting (Anderson & Spellman, 1995; Barnes & Underwood, 1959; Mensink & Raaijmakers, 1988; Osgood, 1949; Wixted, 2004). As a result, a primary focus of theoretical models of memory has been to specify the computational mechanisms by which interference is resolved (Colgin et al., 2008; O'Reilly & McClelland, 1994; Treves & Rolls, 1994). These models have largely focused on how memories are encoded so that the content of memories is protected against interference. An alternative perspective, however, is that instead of protecting memories from interference, there is adaptive value in allowing the content of memories to be *shaped* by interference (Hulbert & Norman, 2015; G. Kim et al., 2017). Specifically, to the extent that overlap across memories is the root cause of interference, then distorting memories to reduce this overlap is a potentially effective remedy.

Evidence from recent neuroimaging studies hints at the idea that memory representations are distorted as an adaptive response to interference. Namely, several

studies have found that when similar events are encoded into memory, this triggers a targeted exaggeration of differences in patterns of activity in the hippocampus (Ballard et al., 2019; Chanales et al., 2017; Dimsdale-Zucker et al., 2018; Favila et al., 2016; Hulbert & Norman, 2015; G. Kim et al., 2017; Schapiro et al., 2012; Schlichting et al., 2015). The key observation in these studies is that similar memories 'move apart' from each other in representational space, suggesting a form of memory repulsion. Yet, a critical limitation of these studies is that the feature dimensions along which memories move are underspecified. That is, do changes in neural representations correspond to changes in the information content of memories? On the one hand, neural activity pattern may become separated without any changes to underlying memories. Alternatively, changes in neural activity patterns may reflect adaptive changes in memory content. For example, if two vacations to the same city were associated with different weather conditions, then weather-related information may be a salient component of corresponding memories and weather-related *differences* between those vacations may be exaggerated to improve memory discriminability (e.g., "That was the year it was *really cold*," vs. "That was the year it was *really hot*").

While it has proven difficult to translate hippocampal activity patterns to explicit feature dimensions (LaRocque et al., 2013; Liang et al., 2013), feature dimensions are far more accessible in (or decodable from) neocortical regions involved in memory retrieval. In particular, there is rapidly growing evidence that lateral parietal cortex carries detailed information about the content of retrieved memories (Chen et al., 2017; Long et al., 2016; Xiao et al., 2017) and amplifies behaviorally relevant information (Favila et al., 2018; Kuhl et al., 2013). Moreover, recent studies have shown that memory representations in

parietal cortex can be decomposed into separable feature dimensions (Bone et al., 2020; Favila et al., 2018; Lee et al., 2019). Thus, lateral parietal cortex may provide a unique window into how memory representations are shaped by interference.

Here, we tested whether interference between highly similar memories triggers adaptive distortions in parietal memory representations and corresponding behavioral expressions of memories. Our motivating theoretical perspective was that subtle differences between similar memories are prioritized and exaggerated to reduce the potential for interference. To test these ideas, we modified a recent behavioral paradigm that demonstrated adaptive biases in long-term memory for objects (Chanales et al., 2021). We predicted that competition between memories for similar objects would trigger a memory-based exaggeration of subtle differences between those objects, and that greater exaggeration would be associated with lower memory interference. Using pattern-based fMRI analyses, we tested whether memory representations in lateral parietal cortex (a) preferentially express features that are critical for discriminating similar objects and (b) predict feature-specific distortions in behavioral expressions of memory.

Materials and Methods

Participants

Thirty-two (21 female; mean age = 23.5 years) right-handed, native English speakers from the University of Oregon community participated in the experiment. Three participants were excluded from analysis (two due to falling asleep inside the scanner, one due to technical error), resulting in a final set of 29 participants (19 female; mean age = 23.7 years) included in data analysis. Participants were screened for motion during the scanned recall tasks, but no participants exceeded the exclusion criteria (mean framewise

displacement > 0.25) for any of the runs. The sample size was comparable to similar fMRI studies in the field. All participants had normal or corrected-to-normal vision. Informed consent was obtained in accordance with the University of Oregon Institutional Review Board.

Overview of Experimental Paradigm

We modified a paradigm from a recent behavioral study that was used to demonstrate adaptive biases in long-term memory for object colors (Chanales et al., 2021). In the prior (and current) study, participants learned associations between faces and object images. Critically, the objects contained 'pairmates' for which the object images were identical except for their color (e.g., a blue backpack and a purple backpack), and successful learning required discriminating between these pairmates. In the current study, we used a two-day procedure in which participants received extensive behavioral training on face-object associations on Day 1 and then returned on Day 2 for additional behavioral training, followed by an fMRI session, and finally a behavioral color memory test (Figure 3.1). A critical feature of our design is that we held color similarity between pairmates constant (24 degrees apart), but we included a competitive and non-competitive condition (Figure 3.1b). In the competitive condition, pairmate images corresponded to the same object category (e.g., two beanbags of slightly different colors). In the non-competitive condition, pairmates corresponded to distinct object categories (e.g., a pillow and a ball of slightly different colors). Thus, in both conditions the pairmates were 24 degrees apart in color space; but, for the competitive condition,

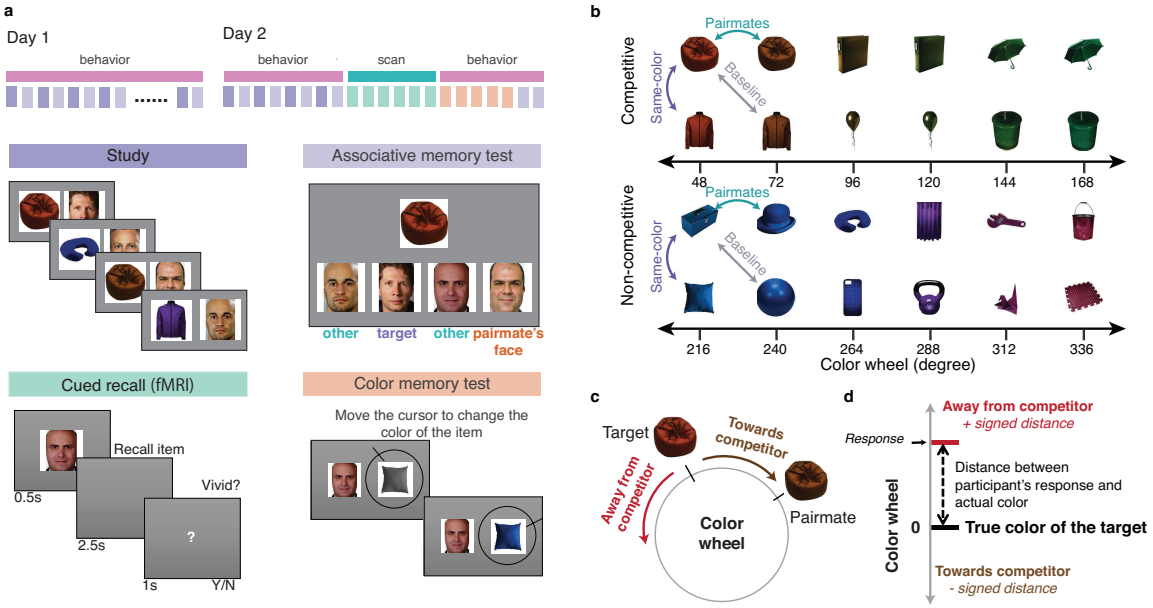

Figure 3.1. Experimental design and procedure.
a, Overview of paradigm. On Day 1, participants completed 14 Study and Associative Memory Test rounds. During Study, participants were shown object-face pairs and during Associative Memory Test, participants were shown an object and selected the corresponding face from a set of four choices. The set of four choices included the target face along with the face associated with the object's pairmate. On Day 2, participants completed four additional Study and Associative Memory Test rounds before entering the fMRI scanner. During scanning, participants completed a Cued Recall task during which face images were shown and participants recalled the corresponding image and indicated, by button press, the vividness of their recall. After exiting the scanner, participants completed a Color Memory Test during which a face image was shown alongside a greyscale version of the corresponding object. Participants used a continuous color wheel to indicate their memory for the object's color. Finally participants completed 2 more Associative Memory Test rounds. **b**, Sample structure of object stimuli. For both the competitive and non-competitive conditions, pairmate stimuli were 24 degrees apart in color space. For the competitive condition, pairmates were from the same object category; for the non-competitive condition, pairmates were from distinct categories. For both conditions, some objects had identical colors (Same-color). fMRI pattern similarity for Pairmate and Same-color comparisons were compared against a Baseline comparison of stimuli that were from different object categories and 24 degrees apart in color space. **c**,**d**, Responses on the color memory test were used to categorize memory for each object's color as being biased toward or away from the color of the competing object (**c**) and to measure the signed distance, in degrees, between participants' responses and the true color of the target (**d**).

color was the only feature dimension on which the pairmates differed. In contrast, for the non-competitive condition, object category also differed between pairmates. Thus, although color distance between pairmates was matched across conditions, color information was more important in the competitive condition. For the fMRI session, participants were shown faces, one at a time, with the only instruction being to retrieve corresponding objects as vividly as possible. An important feature of our procedure is

that participants were not explicitly instructed to retrieve color information during the fMRI scans, nor had color memory been tested at any point prior to scanning. Rather, we only tested color memory after participants exited the scanner.

Stimuli

Participants learned associations between 24 object images and 24 images of white male faces. The 24 object images corresponded to 18 distinct object categories (e.g., beanbag, hat, umbrella, balloon) and 12 distinct color values. Thus, some of the 24 object images were from the same object category (e.g., two beanbags) or had the same color value. The object images were generated from an image set that allowed for each image's color to be rotated along a 360° color wheel (Brady et al., 2013). To assign colors to each object, the 360° color wheel was divided into 15 evenly spaced color values (0°, 24°, 48°, etc.). These 15 values were arbitrarily chosen but were fixed across participants. For each participant, 6 consecutive color values were selected (randomly positioned among the set of 15 color values) for the competitive condition. For example, color values of 48°, 72°, 96°, 120°, 144°, and 168° might be selected for the competitive condition (Figure 3.1b). Likewise, 6 consecutive color values were selected for the non-competitive condition. The 6 values for the non-competitive condition always 'started' 48° after the competitive color values 'ended.' For example, if the color values for the competitive condition spanned 48° to 168°, then the color values for the non-competitive condition would be 216°, 240°, 264°, 288°, 312°, 336° (Figure 3.1b).

For both conditions, the 6 color values were clustered into 3 sets of consecutive color values: e.g., 48° and 72°, 96° and 120°, 144° and 168°. Each of these sets included a total of 4 object images (resulting in 12 object images for each condition). For the

competitive condition, the four images in each set represented two color values (e.g., 48° and 72°) and two object categories (e.g., beanbag and jacket). For example, the set might include a 48° beanbag, a 72° beanbag, a 48° jacket and a 72° jacket (Figure 3.1b). Object images within each set that were from the same object category (e.g., the 48° beanbag and the 72° beanbag) are referred to as 'pairmates.' For the non-competitive condition, the four images in each set represented two color values (e.g., 216° and 240°) and four distinct object categories (Figure 3.1b). Although none of the object images in the non-competitive condition were from the same object category, the four images in each set were also divided into pairmates, with pairmates being images from distinct object categories and, as in the competitive condition, with color values 24° apart. For example, if a set in the non-competitive condition included a 216° lunchbox, a 216° pillow, a 240° hat, and a 240° ball, the 216° lunchbox and the 240° hat might be arbitrarily designated as one set of pairmates and the 216° pillow and the 240° ball as the other set of pairmates. These non-competitive pairmates functioned as a critical control condition for behavioral and fMRI analyses (see fMRI Pattern Similarity Analyses, below).

The mapping between the 24 object images and the 24 face images was randomly determined for each participant. All face and object images were 250 * 250 pixels.

Pre-scan face-object training

Participants completed the experiment on two consecutive days (Figure 3.1a). On Day 1, participants learned 24 face-object associations across 14 training rounds. Each training round consisted of a study phase and an associative memory test phase. During study phases, participants were presented with the 24 face-object associations, one association at a time, in random order. Each trial started with a fixation cross presented in

the center of the screen (1.5 s), followed by the face-object association (3.5 s). Faces were presented to the left of the objects. During the associative memory test phases, object images were presented at the top of the screen with four face choices below. The four face choices always included the target face (i.e., the face associated with the presented object image), the pairmate's face (i.e., the face that was associated with the presented object's pairmate), and two foil faces (associated with non-pairmate objects). Participants were asked to select the face that was associated with the presented object. After responding, participants received feedbacks indicating whether or not they were correct and showing the correct face-object association for 1.5 s. Each trial in the associative memory test was self-paced up to a maximum of 8 s. On Day 2, participants completed 4 additional training rounds immediately prior to entering the fMRI scanner. The procedure was the same as on Day 1.

Scanned perception and cued recall tasks

During fMRI scanning, participants completed 6 consecutive rounds of a perception task and 6 consecutive rounds of a cued recall task (each round corresponded to a separate fMRI scan). The order of the perception and cued recall tasks was counterbalanced across participants. In the perception task, each trial presented one of the 24 object images in the center of the screen for 0.5 s followed by a fixation cross for 3.5 s. A black cross was embedded within the object images at a random location on 25% of trials and participants were instructed to make a button press whenever they detected a black cross. In each perception round, each object image was presented twice, in block randomized order. Participants were instructed to remain centrally-fixated, on a white fixation cross, throughout each perception run. Each perception round contained a 10 s

null trial (fixation cross only) at the beginning and end of each scan and 12 null trials (4 s each) randomly distributed throughout the run. Here, we do not consider data from the perception task because (a) our primary hypotheses related to participants' memories for the object images and (b) subtle color differences between were more to detect in the scanner environment.

In the cued recall task, each trial started with one of the 24 face images presented at the center of the screen for 0.5 s, followed by a blank screen for 2.5 s, and then a question mark for 1 s. Participants were instructed to recall the object image that was associated with the presented face as vividly as possible and to hold the image in mind throughout the trial. Participants were instructed to rate the vividness of their memories ('vivid' or 'not vivid') via a button box response when the question mark appeared. The question mark was followed by a fixation cross for 2 s before next trial began. Responses were recorded during the trial and during the 2 s fixation cross between trials. Together, the intertrial interval was 6 s. All face-object associations were tested twice in each retrieval round, in block randomized order. Each retrieval round contained a 10 s null trial (fixation cross only) at the beginning and end of each scan and 12 null trials (4 s each) randomly distributed throughout the run.

Post-scan behavioral tests

After participants completed the perception and cued recall tasks, they exited the scanner and completed five rounds of the color memory test. During the color memory test, each trial began with one of the 24 face images presented on the left side of the screen and the corresponding object image presented on the right of the screen. Importantly, the object image was initially in grey scale. Participants were instructed to

move a cursor along a color wheel (Figure 3.1a, c) to adjust the color of the object to the remembered color value. Participants clicked the mouse to record their response and then moved on to the next trial. Each face-object association was tested once per round and the task was self-paced. After completing the five color memory test rounds, participants completed two final rounds of the associative memory test—the same task they completed during the training rounds on Day 1 and just prior to fMRI scanning. The sole purpose of the post-scan associative memory test was to motivate participants to maintain their effort and memory accuracy throughout the fMRI session as the post-scan associative memory test was used to determine a monetary bonus for participants (a fact which participants were made aware of prior to the fMRI scan).

Measuring color memory bias

The post scan color memory test was used to measure participants' color memory for each object image. However, rather than focusing on the accuracy of recall, we were critically interested in recall bias. Bias was measured in two ways. The first measure—mean signed distance—was computed by first averaging the responses across the 5 color memory test trials for each object image. The difference between the mean response and the actual color value for a given object image reflects the color memory distance for that object image. Critically, if the mean response was biased away from the color of the pairmate object (Figure 3.1c), the distance measure was positively signed; if the mean response was biased toward the color of the pairmate object (Figure 3.1c), the distance measure was negatively signed. By averaging the signed distance measure across the 12 object images within each condition, the mean signed distance was computed for each condition (competitive, non-competitive) and for each participant. The second measure—

percentage of away responses—was computed by ignoring the distance between participants' responses and the actual color values and instead simply computing the percentage of responses that were biased away from the color of the pairmate object. It is important to note that this measure was computed at the trial level. Thus, for a given object image, if a participant recalled the object's color 'away from' the pairmate on 4 out of the 5 test trials for that object image, the percentage of away responses for that object image would be 80%. Although we did not expect (or observe) notable differences between the two measures (mean signed distance and percentage of away responses), the percentage of away responses addressed the concern that any observed effects for the mean signed distance measure were driven by a few extreme responses.

fMRI data acquisition

Imaging data were collected on a Siemens 3 T Skyra scanner at the Robert and Beverly Lewis Center for NeuroImaging at the University of Oregon. Functional data were acquired using a T2*-weighted multiband EPI sequence with whole-brain coverage (repetition time = 2 s, echo time = 36 ms, flip angle = 90°, multiband acceleration factor = 3, inplane acceleration factor = 2, 72 slices, 1.7 × 1.7 × 1.7 mm voxels) and a 32-channel head coil. Note that due to an a priori decision to focus on visual and parietal cortical areas, we used a high-resolution protocol that fully covered visual/parietal regions but only partially covered frontal cortex. Each perception scan (6 total) consisted of 130 total volumes. Each retrieval scan (6 total) consisted of 190 total volumes. Oblique axial slices were aligned parallel to the plane defined by the anterior and posterior commissures. A whole-brain T1-weighted MPRAGE 3D anatomical volume (1 × 1 × 1 mm voxels) was also collected.

fMRI data preprocessing

fMRI data preprocessing was performed using fMRIPrep 1.3.1 (Esteban et al., 2019). The T1-weighted (T1w) image was corrected for intensity non-uniformity with N4BiasFieldCorrection (Tustison et al., 2010) and skull-stripped using antsBrainExtraction.sh (ANTs 2.2.0) with OASIS30ANTs as the target template. Brain surfaces were reconstructed using recon-all from FreeSurfer 6.0.1 (Dale et al., 1999). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0). For the functional data, susceptibility distortion corrections were estimated using 3dQwarp (Cox & Hyde, 1997). The BOLD reference was then co-registered to the T1w reference by bbregister (FreeSurfer) using boundary-based registration with nine degrees of freedom (Greve & Fischl, 2009). Head-motion parameters were estimated by mcflirt from FSL 5.0.9 (Jenkinson et al., 2002). Slice-time correction was done by 3dTshift from AFNI 20160207 (Cox & Hyde, 1997). Functional data were smoothed with a 1.7 mm FWHM Gaussian kernel and high pass filtered at 0.01Hz. Smoothing and filtering were done with the Nipype pipeline tool (Gorgolewski et al., 2011).

Response estimates were obtained for each trial (one regressor per trial, 4 s duration) in each cued recall run using the "least-squares separate" method (Mumford et al., 2012). With this method, each item was estimated in a separate GLM as a separate regressor while all remaining items were modeled together with another regressor. The six movement parameters and framewise displacement were included in each GLM as confound regressors. This resulted in t maps that were used for the pattern similarity analysis. Given that all analyses averaged data across multiple trials—mitigating the

influence of any one trial—we did not perform any data exclusion for outliers at the trial level.

Regions of interest

fMRI analyses were conducted using a set of visual and parietal regions of interest (ROIs) that were identical to those used by Favila, Samide, Sweigart, & Kuhl (2018) to measure object and color representations during memory recall. While our primary focus was on the parietal ROIs, we anticipated that visual regions might also reflect feature-specific information during memory retrieval. For low level visual regions, we combined bilateral V1v and V1d as V1 and combined bilateral LO1 and LO2 as LO based on Wang, Mruczek, Arcaro, & Kastner (2014). For high level visual regions, we generated a VTC ROI by combining bilateral fusiform gyrus, collateral sulcus, and lateral occipitotemporal sulcus derived from the output of Freesurfer segmentation routines. For lateral parietal cortex, we referenced Yeo et al. (2011)'s 17-network resting state atlas. The parietal nodes from Network 12 and 13 (subcomponents of the frontoparietal control network) are referred to as dorsal lateral intraparietal sulcus (dLatIPS) and ventral lateral intraparietal sulcus (vLatIPS), respectively. For the parietal node of Network 5 (dorsal attention network), we separated it along the intraparietal sulcus to create a dorsal region we refer to as posterior intraparietal sulcus (pIPS) and a ventral region we refer to as ventral IPS (vIPS) (Sestieri et al., 2017). The vertices in lateral occipital cortex were eliminated in these two regions. The parietal nodes of Networks 15–17 (subcomponents of the default mode network) were combined into a region we refer to as angular gyrus (AnG).

For post hoc analyses, we generated medial temporal and hippocampus subfield ROIs using ASHS (Yushkevich et al., 2015). We selected bilateral CA1, subiculum, entorhinal cortex, and parahippocampal cortex. We combined CA2, CA3 and dentate gyrus into a single ROI (CA23DG) and combined BA35 and BA36 into a perirhinal cortex ROI.

fMRI Pattern similarity analyses

Pattern similarity analyses were used to measure the similarity of fMRI activity patterns for various pairs of object images during the cued recall task. To calculate pattern similarity, we first computed the mean activity pattern for each of the 24 recalled objects by averaging t maps for odd runs and even runs separately. Pearson correlations were then computed between the mean t map of odd runs and even runs. All the correlations were z-transformed (Fisher's z) before subsequent analyses. All analyses were performed in the participant's native T1w space and were done separately for each ROI. Pattern similarity analyses focused on three specific correlations within each 'set' of 4 object images (see Figure 3.1b and Stimuli for explanation of 'sets'): (1) 'Pairmate correlations' (see Stimuli for definition of pairmates), (2) 'Same-color correlations,' which refer to correlations between object images from different object categories but with identical color values (Figure 3.1b), and (3) 'Baseline correlations,' which refer to object images from different object categories and different color values (24 degrees apart; Figure 3.1b). Again, it is important to emphasize that all pattern similarity analyses were performed within the sets of 4 object images and, critically, the same correlations were applied for the competitive and non-competitive conditions.

Neural representation of color information

To test whether representation of color information was stronger in the competitive condition than the non-competitive condition, we first obtained (for each condition, ROI, and participant) the mean 'Same-color correlation' and the mean 'Baseline correlation.' Both of these correlations reflect correlations between object images from different object categories (Figure 3.1b), but the same-color correlation reflects images with identical color values whereas the baseline correlation reflects images with a 24° difference in color. Thus, the difference between these measures (same-color – baseline) isolates color-related similarity. Of critical interest was whether this color-related similarity was stronger in the competitive condition than the non-competitive condition. Critically, color similarity was objectively identical across conditions, but we predicted stronger color representation in the competitive condition owing to its greater diagnostic value in the competitive condition. It is important to note that the inclusion of a separate baseline correlation for each condition (competitive, non-competitive) controlled for potential global similarity differences between conditions (i.e., that correlations among all pairs of object images might be higher in one condition vs. the other).

Neural similarity between pairmates

To test whether similarity between pairmates was stronger in the competitive condition than the non-competitive condition, we first obtained (for each condition, ROI, and participant) the mean 'Pairmate correlation' and the mean 'Baseline correlation.' For the competitive condition, pairmate correlations reflect object images from the same object category but with a 24° difference in color (Figure 3.1b). For the non-competitive

condition, pairmate correlations reflect object images from different object categories, again with a 24° difference in color (Figure 3.1b). Thus, pairmate similarity was objectively greater in the competitive condition than the non-competitive condition. For both conditions, the baseline correlations reflect object images from different object categories and with a 24° difference in color. Thus, the difference between these measures (pairmate – baseline) was intended to isolate object-related similarity (specifically for the competitive condition). As with the color information analysis, the condition-specific baseline correlations controlled for potential global similarity differences between conditions.

Neural measures of pairmate similarity predict color memory repulsion

To test whether similarity between vIPS representations of pairmates during competitive recall predicted the degree to which there was repulsion of color memories (as measured in the post-scan color memory test), we first computed the mean signed color memory distance for the two objects in each set of pairmates. This yielded a single value representing the distance between a given set of pairmates, with greater distance reflecting greater repulsion. Next, for vIPS we computed dissimilarity between each set of pairmates, as defined by: 1 – the Pairmate correlation. (Note: for this analysis we used dissimilarity, as opposed to similarity, simply for ease of interpretation). Thus, for each participant and for each condition (competitive, non-competitive), this resulted in 6 values representing color memory distance between each set of pairmates and 6 values representing vIPS dissimilarity between each set of pairmates. We then performed a Spearman correlation between these two measures. For each condition, one-sample t-tests were performed on the participants' z-transformed Spearman's rs values to test whether

the mean correlation between color memory distance and vIPS dissimilarity differed from 0. For comparison, similar analyses were also performed for other ROIs (Table 1).

Table 3.1. Summary of key statistical analyses.
Color representation analyses refer to paired-samples *t*-tests comparing color similarity effects (see Methods) for the competitive vs. non-competitive conditions. Pairmate similarity analyses refer to paired-samples *t*-tests comparing pairmate similarity effects (see Methods) for the competitive vs. non-competitive conditions. The relation to mean signed distance refers to one-sample *t*-tests comparing *z*-transformed correlations between fMRI pairmate dissimilarity and mean signed color memory distance to a test statistic of 0 (no relationship). Results from individual visual and parietal ROIs are presented in separate rows. Note: * $p < .05$, uncorrected; ** $p < .05$, Bonferroni corrected; *** $p < .01$, Bonferroni corrected.

ROI	Color representation		Pairmate similarity		Relation to mean signed distance			
					Competitive		Non-competitive	
	t_{28}	p	t_{28}	p	t_{28}	p	t_{28}	p
V1	1.22	0.232	0.89	0.382	0.82	0.417	-0.34	0.734
LO	**2.27**	**0.031***	1.71	0.098	1.34	0.190	-0.75	0.458
VTC	1.16	0.257	0.45	0.653	**2.13**	**0.042***	0.59	0.558
pIPS	1.85	0.075	0.84	0.409	**3.08**	**0.005****	1.08	0.289
dLatIPS	1.68	0.104	0.73	0.472	1.50	0.145	0.65	0.520
vLatIPS	1.69	0.101	0.52	0.609	**2.92**	**0.007****	-1.89	0.069
AnG	0.57	0.573	0.36	0.720	0.75	0.462	-0.72	0.475
vIPS	**2.67**	**0.012***	**3.12**	**0.004****	**3.75**	**0.0008*****	0.78	0.443

To better visualize the relationship between color memory distance and vIPS dissimilarity, for each participant the 6 pairmates in the competitive condition were divided into three bins (2 pairmates per bin) based on vIPS pairmate dissimilarity (low, medium, high). We then computed the mean signed color memory distance (from the post-scan color memory test) and the mean associative memory accuracy (from the pre-scan associative memory test) for each of these bins. One-way ANOVA was used to test whether mean signed distance and/or mean associative memory accuracy varied as a function of vIPS dissimilarity bin. Finally, we performed a multilevel mediation analysis

to test whether color memory mediated the relationship between vIPS pairmate dissimilarity and associative memory accuracy. This analysis was performed by obtaining, for each participant, the mean color memory distance, vIPS dissimilarity, and associative memory performance for each of the 6 pairmates in each condition. Mediation analyses included a random intercept for each participant, but random slopes were not included due to the small number of data points per condition/participant.

Statistical analysis

Statistical analyses were performed using R version 3.6.3. All t-tests were two-tailed, with $\alpha = 0.05$. All repeated measures ANOVAs were computed with the afex package using Type III sums of squares. Effect sizes for t-tests were estimated using the effsize package. Multilevel mediation analyses were computed using the mediation package. Multilevel models were built using the lme4 package. All error bars in the figures represent S.E.M.

Results

Associative Memory Performance

Participants completed three separate sessions that tested memory for object-face associations (14 rounds on Day 1; 4 rounds before scanning on Day 2; 2 rounds after scanning on Day 2; Figure 3.1a). Participants showed improved accuracy across test rounds in the Day 1 session, from a mean of 56.9% (*SD* = 12.8%) on round 1 to a mean of 95.5% (*SD* = 4.8%) on round 14 (main effect of test round: $F_{5.56,\ 155.73} = 91.29$, $p < 0.0001$, $\eta^2 = 0.55$). Accuracy did not vary by test round for either of the Day 2 sessions (Day 2 pre-scan: $F_{2.77,\ 77.63} = 1.63$, $p = 0.194$, $\eta^2 = 0.01$; Day 2 post-scan: $F_{1,\ 28} = 0.14$, $p = 0.713$, $\eta^2 = 0.0009$). Critically, accuracy was lower in the competitive condition than the

non-competitive condition for each of the sessions (Day 1: $F_{1, 28} = 115.89$, $p < 0.0001$, $\eta^2 = 0.29$; Day 2 pre-scan: $F_{1, 28} = 21.8{,}1$ $p < 0.0001$, $\eta^2 = 0.15$; Day 2 post-scan: $F_{1, 28} = 22.25$, $p < 0.0001$, $\eta^2 = 0.20$; Figure 3.2a). For subsequent analyses, we focused on associative memory performance from the Day 2 pre-scan session (an *a priori* decision; see Methods). Notably, for the Day 2 pre-scan session, lower accuracy in the competitive condition ($M = 93.2\%$, $SD = 6.9\%$) than the non-competitive condition ($M = 98.9\%$, $SD = 2.1\%$) was driven by an increased rate of selecting faces that were associated with the pairmate image (competitive condition: $M = 6.0\%$, $SD = 6.6\%$; non-competitive condition: $M = 0.2\%$, $SD = 0.6\%$; $t_{28} = 4.74$, $p < 0.0001$, 95% CI = [0.03 0.08], Cohen's $d = 1.16$, paired *t*-test; Figure 3.2a). The rate of other errors did not differ in the competitive vs. non-competitive conditions (competitive: $M = 0.8\%$, $SD = 1.4\%$; non-competitive: $M = 0.98\%$, $SD = 1.6\%$; $t_{28} = -0.18$, $p = 0.861$, 95% CI = [-0.01 0.01], Cohen's $d = -0.04$, paired *t*-test). Thus, as intended, the competitive condition specifically increased interference between pairmate images.

Color Memory Bias

Immediately after the fMRI session, participants completed a color memory test. Color memory was indexed in two ways: (1) using a continuous, signed measure of distance, in degrees, between the reported and actual color; positive values indicate a bias *away from* the competing memory and negative values indicate a bias *toward* the competing memory, and (2) using a categorical measure of the percentage of responses that were biased away from the competing memory (see Methods for details of each measure). We refer to these two measures as the *signed distance* and *percentage of away responses*, respectively.

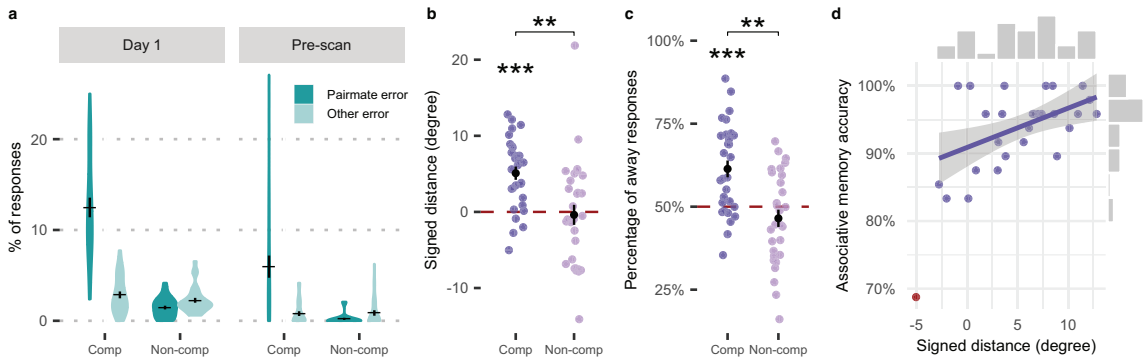

Figure 3.2. Behavioral results.
a, Associative memory performance across the experiment. The overall error rate (pairmate error + other error) was higher in the competitive condition than the non-competitive condition for each of the associative memory test sessions (Day 1, Day 2 pre-scan, Day 2 post-scan (not shown); all p's < 0.0001). Subsequent analyses focused on associative memory performance from the Day 2 pre-scan session. For the Day 2 pre-scan session, participants were significantly more likely to select faces that were associated with the pairmate image (pairmate error) in the competitive condition (M = 6.0%, SD = 6.6%) compared to the non-competitive condition (M = 0.2%, SD = 0.6%; p < 0.0001), confirming that similarity between pairmates was a source of interference. **b**, Signed distance of responses in the color memory test. For the competitive condition, mean signed distance was significantly greater than 0 (p = 0.000003), reflecting a bias away from the color of the pairmate object (repulsion). Signed distance did not differ from 0 in the non-competitive condition (p = 0.771). The difference between the competitive and non-competitive conditions was also significant (p = 0.007). **c**, Percentage of away responses in the color memory test. The percentage of color memory responses 'away from' the color of the pairmate object was significantly greater than 50% for the competitive condition (p = 0.0001), but not for the non-competitive condition (p = 0.189). The difference between the competitive and non-competitive conditions was also significant (p = 0.001). **d**, Relationship between associative memory accuracy and mean signed color memory distance. For the competitive condition, participants with greater mean signed color memory distance (greater repulsion) exhibited better associative memory accuracy [r = 0.50, p = 0.007, one outlier (red dot) excluded for associative memory performance < 3 SD below mean]. Notes: colored dots reflect data from individual participants. Error bars reflect +/- S.E.M.; *** p < 0.001, ** p < 0.01

For the competitive condition, mean signed distance was significantly greater than 0 (5.09 ± 4.69, mean ± SD; t_{28} = 5.84, p = 0.000003, 95% CI = [3.30 6.87], Cohen's d = 1.08, one-sample t-test; Figure 3.2b), indicating that participants' color memory was systematically biased away from the color of the pairmate. In contrast, for the non-competitive condition—where the only difference was that pairmates were not from the same object category—signed distance did not differ from 0 (-0.39 ± 7.08; t_{28} = -0.29, p = 0.771, 95% CI = [-3.08 2.31], Cohen's d = -0.05, one-sample t-test). Signed distance was significantly greater (i.e., a stronger bias away from the pairmate) in the competitive condition compared to the non-competitive condition (t_{28} = 2.90, p = 0.007, 95% CI =

[1.61 9.34], Cohen's $d = 0.92$, paired t-test). These data clearly demonstrate that similarity between images triggered the color memory bias.

The pattern of data was identical when considering the percentage of away responses. Namely, the percentage of away responses was significantly greater than 50% for the competitive condition (61.4 ± 3.6%; $t_{28} = 4.49$, $p = 0.0001$, 95% CI = [56.2% 66.6%], Cohen's $d = 0.83$, one-sample t-test; Figure 3.2c), but not for the non-competitive condition (46.5 ± 14%; $t_{28} = -1.35$, $p = 0.189$, 95% CI = [41.2% 51.8%], Cohen's $d = -0.25$, one-sample t-test). The difference between the two conditions was also significant ($t_{28} = 3.58$, $p = 0.001$, 95% CI = [0.06 0.23], Cohen's $d = 1.08$, paired t-test). While the percentage of away responses does not contain information about the magnitude of the bias in color memory, it rules out the possibility that the effects observed with the signed distance measure were driven by a minority of trials with very high bias.

Relationship between associative memory and color memory bias

A key component of our theoretical framework is that exaggerating the color distance (in memory) between similar objects plays an adaptive role in reducing memory interference. To test this idea, we correlated each participant's associative memory performance (from the Day 2 pre-scan session) with their color memory performance. For the competitive condition, mean associative memory performance was positively correlated with mean signed distance ($r = 0.50$, $t_{26} = 2.91$, $p = 0.007$, 95% CI = [0.15 0.73], one outlier excluded for associative memory performance < 3 SD below mean; Figure 3.2d), consistent with the idea that stronger color memory repulsion (i.e., a bias in color memory away from the pairmate) supports lower associative memory interference.

For the non-competitive condition, this correlation was not significant (r = -0.31 t_{26} = -1.63, p = 0.114, 95% CI = [-0.61 0.08], one outlier excluded for signed distance > 3 SD above the mean). Thus, a bias in color memory away from the pairmate was not beneficial if the pairmate was not similar to (competitive with) the target. An identical pattern of data was observed when considering the percentage of away responses as an index of color memory. Namely, for the competitive condition there was a positive correlation between associative memory performance and the mean percentage of away responses (r = 0.42, t_{26} = 2.39, p = 0.025, 95% CI = [0.06 0.69], one outlier excluded for associative memory performance < 3 SD below mean) and no significant correlation for the non-competitive condition (r = -0.37, t_{27} = -2.05, p = 0.050, 95% CI = [-0.65 -0.002]).

Neural representation of color information during recall

The key design feature of the competitive condition was that color information was critical for discriminating between pairmates. Specifically, in the competitive condition the only difference between pairmates was a 24-degree color difference. This contrasts with the non-competitive condition where pairmates differed in color (again 24 degrees) *and* object category. Because color information was therefore more important in the competitive condition, we predicted that representation of color information during the scanned recall trials would be relatively stronger in the competitive condition than the non-competitive condition. Notably, participants' only instruction on the recall trials was to bring each stimulus to mind as vividly as possible (mean percentage of vivid responses = 95.42%, *SD* = 5.43%). Participants were not explicitly oriented to color information nor had participants' memory for color been tested in any way to that point in the experiment.

To test for representation of color information, we computed the mean correlation of activity patterns evoked during recall of non-pairmate stimuli that shared an identical color value (e.g., red bean bag and red jacket; 'same-color' comparison, see Figure 3.1b) and subtracted from this value the mean correlation between non-pairmate stimuli that were 24 degrees apart in color space (e.g., red bean bag and brown jacket; 'baseline' comparison, see Figure 3.1b). Thus, the difference between these two measures (same-color – baseline) provided an index of color information. We then compared this index across the competitive and non-competitive trials. Critically, in terms of physical properties of the stimuli, the comparison between the competitive and non-competitive trials was perfectly matched: there was no objectively greater similarity between the *stimuli included in this analysis* in the competitive condition compared to the non-competitive condition—there was only a difference in the importance of the information.

For this and subsequent fMRI analyses we used a set of visual and parietal regions of interest (ROIs) previously described in Favila et al. (2018) (see Methods; Figure 3.3a). Critically, these ROIs were previously shown to contain color and object feature representations during a memory recall task very similar to the current study. The set of ROIs included three visual ROIs (V1, LO, VTC) and five lateral parietal ROIs (pIPS, dLatIPS, vLatIPS, AnG, vIPS).

An ANOVA with factors of condition (competitive, non-competitive) and ROI (all eight ROIs) revealed a significant main effect of condition, with relatively stronger color information in the competitive condition than the non-competitive condition ($F_{1, 28} = 5.03$, $p = 0.033$, $\eta^2 = 0.04$). Neither the main effect of ROI nor the condition x ROI interaction were significant (ROI: $F_{4.55, 127.36} = 0.12$, $p = 0.984$, $\eta^2 < 0.001$; condition x

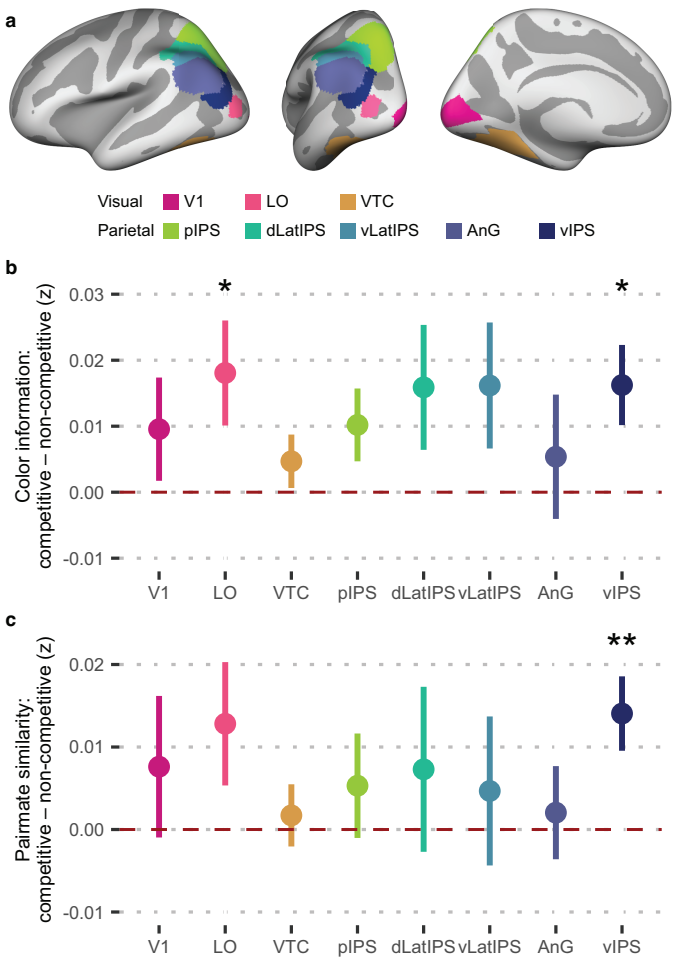

Figure 3.3. Neural feature representations as a function of memory competition.
a, Anatomical ROIs visualized on the Freesurfer average cortical surface. **b**, Color information as a function of memory competition. Color information was defined as the fMRI pattern similarity between pairs of same-color objects relative to pattern similarity between baseline pairs of objects (see **Figure 3.1b**). Color information was significantly stronger in the competitive than non-competitive condition (i.e., values greater than 0) across the set of ROIs as a whole and in LO and vIPS individually (p's < .05). **c**, Pairmate similarity as a function of memory competition. Pairmate similarity was defined as the fMRI pattern similarity between pairmate objects relative to pattern similarity between baseline pairs of objects. Only vIPS showed significantly greater pairmate similarity in the competitive than non-competitive conditions ($p = 0.004$). Error bars reflect +/- S.E.M.; ** $p < 0.01$, * $p < 0.05$

ROI: $F_{4.10,\ 114.92} = 0.78$, $p = 0.542$, $\eta^2 = 0.008$). Considering individual ROIs, only LO and vIPS exhibited significantly stronger color representation in the competitive than non-competitive condition (LO: $t_{28} = 2.27$, $p = 0.031$, 95% CI = [0.002 0.03], Cohen's $d = 0.69$; vIPS: $t_{28} = 2.67$, $p = 0.012$, 95% CI = [0.004 0.03], Cohen's $d = 0.63$; paired t-tests,

uncorrected; Figure 3.3b). Thus, as predicted, the greater relevance of color information in the competitive condition resulted in stronger representation of color information during recall, despite the fact that participants had not been explicitly oriented to color information in any way by this point of the experiment (the critical behavioral test of color memory occurred after fMRI scanning).

Post-hoc analyses of medial temporal and hippocampal ROIs (see Methods) did not reveal stronger color representation in the competitive than non-competitive condition for any of the ROIs ($|t|$'s < 1.66, p's > 0.109).

Neural similarity between pairmates during recall

We next tested whether neural similarity between pairmate stimuli was greater in the competitive than non-competitive condition. In terms of physical stimulus properties, pairmates were, of course, more similar in the competitive condition (e.g., two bean bags 24 degrees apart in color space) than in the non-competitive condition (e.g., a pillow and a ball 24 degrees apart in color space). Thus, based on stimulus properties alone, fMRI pattern similarity between pairmates should be greater in the competitive condition than the non-competitive condition. To measure pairmate similarity we computed the mean correlation between pairmate stimuli ('pairmate' comparison, see Figure 3.1b) and subtracted from this value the mean correlation between non-pairmate stimuli that were also 24 degrees apart in color space ('baseline' comparison, see Figure 3.1b). The difference between these two values (pairmate – baseline) yielded an index of pairmate similarity which was then compared across the competitive and non-competitive conditions.

Although pairmate similarity was numerically greater in the competitive than non-competitive condition across each of the eight ROIs, an ANOVA with factors of ROI and condition did not reveal a significant main effect of condition ($F_{1, 28} = 2.30$, $p = 0.140$, $\eta^2 = 0.016$). The main effect of ROI and the condition x ROI interaction were also not significant (ROI: $F_{4.57, 127.90} = 0.68$, $p = 0.626$, $\eta^2 = 0.006$; condition x ROI: $F_{3.82, 106.85} = 0.58$, $p = 0.670$, $\eta^2 = 0.006$). However, there was a significant effect of condition, corrected for multiple comparisons (Bonferroni corrected), in vIPS, with greater pattern similarity in the competitive than non-competitive conditions ($t_{28} = 3.12$, $p = 0.004$, 95% CI = [0.005 0.02], Cohen's $d = 0.70$, paired *t*-test; Figure 3.3c). Notably, as described above (Figure 3.3b), vIPS also exhibited significantly stronger color representation in the competitive than non-competitive condition. Moreover, vIPS also exhibited significant object and color representations during a recall task in a prior study (Favila et al., 2018). Thus, across two independent studies, we have consistently observed feature representations in this ROI during memory recall.

Post-hoc analyses of medial temporal and hippocampal ROIs (see Methods) did not reveal greater pairmate similarity in the competitive than non-competitive condition for any of the ROIs ($|t|$'s < 1.42, p's > 0.168).

Neural measures of pairmate similarity predict color memory bias

Results from the preceding analysis revealed greater similarity in vIPS representations of pairmates in the competitive condition than the non-competitive condition. While this measure of neural similarity reflects the greater physical similarity between pairmates in the competitive condition than the non-competitive condition, the key finding from our behavioral results is that there is an adaptive benefit to *reducing*

similarity (in memory) between pairmates in the competitive condition. This raises the question of whether similarity between vIPS representations of pairmates during competitive recall predicted the degree to which there was repulsion of color memories (as measured in the post-scan color memory test). To test this, for each condition (competitive, non-competitive) we correlated fMRI measures of pairmate *dissimilarity* (1 – pattern similarity) with behavioral measures of mean signed color memory distance. This analysis was performed within participant (i.e., at the level of individual pairmates). Given that each condition only corresponded to 6 pairmates per participant, Spearman rank correlation was used in order to reduce the influence of any one data point. Correlation coefficients were then *z*-transformed, yielding a single *z*-transformed value for each condition and participant.

For the competitive condition, the mean correlation between pairmate dissimilarity in vIPS during recall and mean signed color memory distance was significantly positive (vIPS: $t_{28} = 3.75$, $p = 0.0008$, 95% CI = [0.34 1.14], Cohen's $d = 0.70$, one-sample *t*-test; Figure 3.4a). In other words, the more dissimilar vIPS activity patterns were when recalling pairmates, the greater the color memory repulsion effect for those pairmates. There was no correlation between pairmate dissimilarity in vIPS and

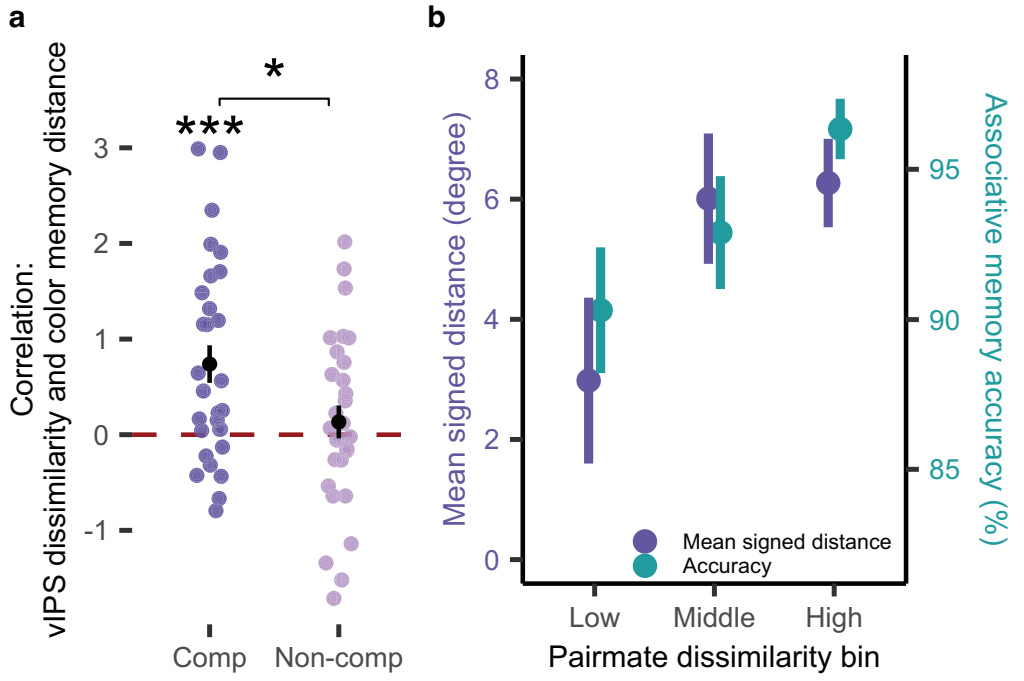

Figure 3.4. Neural measures of pairmate (dis)similarity predict color memory bias in vIPS. **a**, Mean correlation between vIPS pairmate dissimilarity during recall and mean signed color memory distance. Correlations were performed within participant and correlation coefficients were *z*-transformed. For the competitive condition, the mean correlation was significantly positive ($p = 0.004$), indicating that greater pairmate dissimilarity in vIPS was associated with a stronger bias to remember pairmates' colors as *away from* each other. There was no correlation between vIPS pairmate dissimilarity and signed color memory distance for the non-competitive condition ($p = 0.566$). **b**, Relationship between vIPS pairmate dissimilarity (binned into low, medium, high groups) and mean signed color memory distance (purple) and associative memory accuracy (teal). Mean signed color memory distance and associative memory accuracy each significantly varied as a function of vIPS dissimilarity (p's < .05), with greater vIPS dissimilarity associated with greater mean signed color memory distance and higher associative memory accuracy. *** $p < 0.001$, * $p < 0.05$

signed color memory distance for the non-competitive condition ($t_{28} = 0.78$, $p = 0.443$, 95% CI = [-0.22 0.49], Cohen's $d = 0.14$; Figure 3.4a) and the difference between the competitive and non-competitive conditions was significant ($t_{28} = 2.39$, $p = 0.024$, 95% CI = [0.09 1.12], Cohen's $d = 0.61$, paired *t*-test). Significant positive relationships were also observed when pairmate dissimilarity was measured from pIPS, VTC, and vLatIPS—again, only for the competitive condition (see table 1).

As a complementary analysis—and to better visualize the results in vIPS—we binned pairmates, for each participant, based on vIPS dissimilarity (competitive condition

only). We generated three bins per participant: low, medium, and high pairmate dissimilarity. We then computed the mean signed color memory distance for each of these bins. A one-way ANOVA revealed a significant main effect of pairmate dissimilarity in vIPS on mean signed color memory distance (Figure 3.4b; $F_{1.75, 48.90}$ = 4.95, p = 0.014, η^2 = 0.062), with greater dissimilarity between vIPS representations associated with greater distance in remembered color values (i.e., greater repulsion). We also computed mean accuracy on the associative memory test for these same vIPS dissimilarity bins in order to more directly test whether vIPS dissimilarity was associated with lower interference. Indeed, we again found a significant main effect of bin ($F_{1.78, 49.87}$ = 4.52, p = 0.019, η^2 = 0.068), with behavioral accuracy increasing as a function of pairmate dissimilarity in vIPS. Finally, a mediation analysis performed at the level of individual pairmates (see Methods) revealed that the relationship between vIPS dissimilarity and associative memory accuracy was significantly mediated by signed color memory distance (β = 0.12, CI = [0.02 0.23], p = 0.016, 1000 bootstrapped samples), consistent with the interpretation that vIPS dissimilarity reflected the degree of color memory repulsion, which in turn was associated with better associative memory accuracy (lower interference).

Discussion

Here, we show that competition between similar memories triggers biases in their neural representations and corresponding behavioral expressions. Specifically, we demonstrate that subtle, diagnostic differences between events were exaggerated in long-term memory and that this exaggeration reduced interference. Critically, these behavioral

expressions of memory distortion were predicted by adaptive, feature-specific changes to memory representations in parietal cortex.

Our behavioral paradigm was designed to isolate the effect that competition had on color memory. Specifically, the competitive and non-competitive conditions had perfectly matched structures, with equivalent color distances between pairmates in both conditions (Figure 3.1b). The only difference was that pairmates in the competitive condition were from the same object category. As intended, this increased the number of interference-related errors, particularly during early stages of learning (Figure 3.2a). The increase in interference-related errors is consistent with a long history of behavioral studies of memory interference (Anderson & Spellman, 1995; Mensink & Raaijmakers, 1988; Wixted, 2004). Our critical question, however, was whether competition distorted memory for object *features* that were otherwise successfully remembered. Results from the color memory post-test revealed a robust bias in color memory in the competitive condition—that is, participants exaggerated the distance between pairmates—but no systematic bias in the non-competitive condition. We refer to the bias in the competitive condition as a repulsion effect in order to emphasize that the bias was triggered by the representational proximity of competing memories (Bae & Luck, 2017; Chanales et al., n.d., 2017; Golomb, 2015), just as spatial proximity of like-poled magnets triggers magnetic repulsion.

It is important to emphasize that the repulsion effect is distinct from—in fact, opposite to—an interference effect. That is, interference-related errors should lead participants to occasionally recall the color of the competing object—an error that would produce a bias in color memory *toward* the pairmate (Figure 3.1c, d). Here, we did not

test color memory until the very end of the experiment, so as to avoid explicitly orienting participants to color information prior to (or during) the fMRI session, but our speculation is that the repulsion effect only emerged after extensive practice and as interference errors subsided (Chanales et al., 2021). In this sense, the repulsion effect can be thought of as an aftereffect of initial memory interference. Although repulsion reflects a form of memory error, our findings indicate that it is an adaptive error: participants who exhibited a stronger repulsion effect also exhibited fewer interference-related errors (Figure 3.2d). To the extent that objective similarity between stimuli is a root cause of memory interference (Osgood, 1949), then exaggerating the difference between stimuli in memory is a potentially powerful means for reducing interference (Chanales et al., 2021; Favila et al., 2016; Hulbert & Norman, 2015).

Our fMRI analyses, which measured neural activity patterns as participants recalled object images, provided a unique means for covertly probing the qualities of participants' memories. These analyses revealed two forms of adaptive memory representations in parietal cortex. First, despite the fact that participants were not instructed to think about or report objects' colors during these recall trials, we observed stronger color information—across the full set of visual and parietal ROIs, and in vIPS specifically—during competitive than non-competitive recall trials. The stronger representation of color information during competitive trials can be viewed as an adaptive response to competition in that color information was the only (or diagnostic) feature dimension for discriminating pairmates in the competitive condition.

Second, although pairmate similarity in vIPS was stronger during competitive than non-competitive recall trials (indicating that vIPS was sensitive to object similarity;

Figure 3.3c), we found that greater *dissimilarity* between vIPS pairmate representations during competitive recall trials was associated with greater color memory repulsion and less memory interference. In other words, minimizing the overlap of neural representations of pairmates was an adaptive response to competition. This relationship was observed within participants, at the level of individual pairmates, but it is important to emphasize that these measures were temporally offset: vIPS pattern similarity was measured during recall trials in the scanner (with the only instruction being to recall objects as vividly as possible) whereas behavioral expressions of color memory were only tested after scanning was completed. This again makes the point that color information—in this case the subtle *difference* in pairmate colors—was a salient component of activity patterns in vIPS during competitive recall.

Importantly, when our two main fMRI findings are taken together, they indicate that an adaptive response to competition involved an *increase* in similarity between stimuli that shared a diagnostic feature value (i.e., objects of the same color) but a *decrease* in similarity between stimuli that had subtly different values for a diagnostic feature (i.e., pairmates, which had slightly different colors). This indicates that avoiding memory interference does not necessarily require a global reduction in similarity to all other memories (LaRocque et al., 2013), but instead may be accomplished by more *targeted changes* in representational structure that emphasize relevant similarities as well as important differences between events that are stored in memory. Critically, this idea is distinct from—if not fundamentally incompatible with—the traditional, and dominant view that interference is avoided through the orthogonalization of memory representations (Colgin et al., 2008; Yassa & Stark, 2011). Specifically, whereas

orthogonalization emphasizes an initial encoding of new memories as independent from existing memories, our findings instead emphasize that the representation of a given memory is *highly dependent* on representations of other memories (Hulbert & Norman, 2015).

Our fMRI findings also add to a growing body of evidence that implicates parietal cortex in actively representing content during memory retrieval (Kuhl & Chun, 2014; Lee et al., 2019; Lee & Kuhl, 2016; Rugg & King, 2018; Sestieri et al., 2017). Of most direct relevance, in a recent study we found that vIPS (a ventral subregion of parietal cortex) actively represents color and object category information during memory recall (Favila et al., 2018). However, this prior study focused on decoding the *objective properties* of recalled stimuli and did not test whether competition influenced or distorted these representations, nor did it establish a link between vIPS representations and behavioral expressions of memory. The current findings provide unique evidence that representations within this same vIPS subregion reflect subtle distortions in how events are remembered that are *dissociable from the objective properties of the event*. More generally, our findings highlight the behavioral relevance and detailed nature of memory representations in parietal cortex.

While our findings provide strong evidence that representations in parietal cortex reflect the influence that competition had on memory representations, it is not necessarily the case that parietal cortex was the source of this influence. Rather, competition between memories is thought to induce targeted plasticity in the hippocampus (Norman et al., 2007; Ritvo et al., 2019). In fact, hippocampal representations have been shown to specifically exaggerate differences between highly similar stimuli (Ballard et al., 2019;

Chanales et al., 2017; Dimsdale-Zucker et al., 2018; Favila et al., 2016; Hulbert & Norman, 2015; Schapiro et al., 2012; Schlichting et al., 2015). However, these exaggerations in hippocampal activity patterns have generally been observed during memory encoding or perception (Ballard et al., 2019; Chanales et al., 2017; Dimsdale-Zucker et al., 2018; Favila et al., 2016; Hulbert & Norman, 2015; Schapiro et al., 2012; Schlichting et al., 2015), as opposed to memory recall, and they have not been translated to explicit feature spaces. Indeed, attempts to translate hippocampal activity patterns to explicit feature dimensions or categories have tended to be unsuccessful (LaRocque et al., 2013; Liang et al., 2013). In post hoc analyses, we did not find any evidence that competition influenced feature representations in the hippocampus or medial temporal lobe ROIs. That said, one notable aspect of our study is that each object was retrieved from memory many times before fMRI scanning began. Given that repeated retrieval has specifically been shown to hasten the transfer of representations to parietal cortex (Brodt et al., 2016, 2018), this raises the question of whether the observed findings in parietal cortex were dependent on repeated retrieval. For example, it is possible that competition induces exaggerated representations that are initially expressed in the hippocampus but ultimately transformed, via retrieval, into stable representations in parietal cortex (Favila et al., 2020). While the current study cannot address this question, it represents an interesting avenue for future research.

In summary, our findings provide unique evidence that memory-based representations in parietal cortex exhibit adaptive, feature-specific changes in response to competition and that these changes in parietal representations predict distortions in behavioral expressions of memory. More generally, our findings provide unique evidence

in support of the perspective that memory distortions are an adaptive component of the memory system (Schacter et al., 2011).

CHAPTER IV

GENERAL DISCUSSION

The lateral parietal cortex (LPC) was initially studied in the context of attentional and spatial cognition (Culham & Kanwisher, 2001). Nevertheless, more recent research has stressed LPC's contribution to episodic memory (Buckner & Wheeler, 2001; Cabeza et al., 2008; Wagner et al., 2005). Specifically, LPC univariate activation profiles have been observed to relate to episodic memory encoding and retrieval success (Hutchinson et al., 2009; Wagner et al., 2005) and multivoxel activity patterns in LPC have been suggested to represent memory contents (Bird et al., 2015; Buchsbaum et al., 2012; Kuhl & Chun, 2014; Polyn et al., 2005; St-Laurent et al., 2015). In this dissertation, I applied a variety of fMRI data analyses to better understand the nature of content representations in LPC. Collectively, across a set of two fMRI experiments, I compared content representations and their functional relevance in different subregions of LPC; I explored different ways LPC content representations can contribute to behavioral success; and I compared functional significances between LPC and other brain regions in content representation. In the following sections, I discuss these findings in turn.

Functional Heterogeneity of LPC during Content Representation

Research from the past suggests that the dorsal and ventral LPC contribute to memory processes differentially. Specifically, as part of the dorsal attention network (DAN; Corbetta et al., 2008; Corbetta & Shulman, 2002), the dorsal LPC univariate activities track top-down attentional efforts during a memory process, such as allocating attention during memory encoding (i.e., positive subsequent memory effect; Uncapher & Wagner, 2009). On the other hand, as part of the default mode network (DMN; Raichle et

al., 2001), the ventral LPC univariate activities track bottom-up salience-driven memory processes, such as retrieving information about "remembered" items (Montaldi et al., 2006; Sharot et al., 2004; Wheeler & Buckner, 2004; Woodruff et al., 2005). Although this attention-to-memory account (Cabeza, 2008; Ciaramelli et al., 2017) of LPC's contribution to memory processes has been well-explored in terms of univariate activation, it remains unknown whether the functional heterogeneity between dorsal vs. ventral LPC is retained during multivariate, content representation within the LPC. In other words, it is unclear whether memory content representation in the dorsal and ventral LPC contributes to the memory processes differently.

In a set of two experiments, I examined LPC content representation that contributes to recognition memory and interference resolution. Notably, with our experimental design, successful recognition memory relies on accurate neural reinstatement of content information, whereas resolving memory interference relies on goal-related adaptive content representation shifts. Consistent with the attention-to-memory account, I found a double dissociation effect between dorsal and ventral LPC in contributing to attention- and memory-related processes. Specifically, results from chapter II suggest that the accuracy of content representations in ventral, but not dorsal LPC, determines the subsequent success in recognizing an encountered item. On the other hand, results from chapter III suggest that the content representations in dorsal (vIPS), but not ventral LPC, can be adaptively distorted to achieve specific attentional goals.

These results suggest that the functional heterogeneity between dorsal and ventral LPC does not only exist in their univariate activity profiles, but also demonstrate how they represent memory contents. Consistent with what was observed with the univariate

findings, content representation in the dorsal LPC integrate top-down goals. That is, if the goal requires differentiating two episodes on a specific dimension, dorsal LPC content representation can be adaptively distorted along that dimension to serve the goal. Nevertheless, content representation in the ventral LPC focuses specifically on content features. Thus, the fidelity of ventral LPC content representation is the deterministic factor of retrieving past episodes.

Relationship between LPC Content Representations and Behavior

LPC neural activities have been shown to relate to human episodic memory success. For example, early neuroimaging findings have suggested that LPC univariate activation levels during both memory encoding and retrieval can be indicative of final memory success (Cabeza et al., 2008; Wagner et al., 2005). With the development of multivariate investigation approaches content (Bird et al., 2015; Buchsbaum et al., 2012; Kuhl & Chun, 2014), it has been shown that LPC neural activity patterns are meaningful and convey content information about memory episodes. Importantly, previous research showed that the degree to which such content representations can be precisely reconstructed during rehearsal and memory retrieval is deterministic of final memory success (Kuhl & Chun, 2014; Lee et al., 2017).

Chapter II corroborates the relationship between the sharpness of content representations and behavioral memory success. Our findings and approaches were also able to extend our understanding of LPC content representations in two ways. First, previous work has typically quantified the precision of content representations by computing LPC activity pattern similarity. For example, Bird et al. (2015) quantified more precise content representation of complex naturalistic events as greater LPC neural

activity pattern similarity during encoding and rehearsal for any given participant. In other words, previous work did not have a "ground truth" measure of content information. In Chapter II, I attempted to define the common "ground truth" content information for all participants using the feature map from a pre-trained deep convolutional neural network (CNN; VGG16; Simonyan & Zisserman, 2014). Importantly, the feature map was extracted from relatively deep layers to make the "ground truth" of content information rely more on semantic rather than lower-level visual details. More importantly, instead of relying on correlational measures, I adopted a predictive approach to quantify the relationship between LPC content representation and the ground truth of content information. That is, the degree to which the ground truth can be predicted by LPC neural activity patterns was quantified as the preciseness of LPC content representations. With these changes, our results were consistent with previous findings, such that the more comprehensive and precise reinstatement of the target-specific content representation can lead to success in memory retrieval.

In addition to the differences in quantifying neural differences, Chapter II also differs from traditional content representation measures in how LPC neural activity patterns were selected. Specifically, previous studies focused on how much content is represented in LPC neural activity during retrieval or rehearsal, whereas chapter II examined how much content is represented in repetition-related neural differences (i.e., encoding-retrieval differences). Early univariate studies have shown that repetition-related suppression and retrieval success effects differ between later remembered vs. forgotten items (Hutchinson et al., 2009; Uncapher & Wagner, 2009; Wagner et al., 2005). Thus, I argue that repetition-related neural differences in LPC indicate how

memory episodes were processed or transformed after they were initially encoded. Our results from chapter II suggest that content information was indeed represented during the LPC repetition-related neural differences, and the degree of content representation can predict recognition memory success, indicating the potentially important role of such a post-encoding neural transformation process.

Strikingly, however, our findings from chapter III suggested that precise content representation does not always lead to better behavioral memory performances. Specifically, when facing memory interferences (i.e., two episodes being particularly similar), the exact content representations of the two episodes do not necessarily lead to memory success. Instead, the results from chapter III suggest that, such memory interference can be resolved when the LPC content representations of the two episodes are adaptively pushed apart from each other. Moreover, the more they are "pushed apart" in LPC neural representations, the more they are remembered differently. Notably, a similar neural mechanism for resolving memory interference has also been observed in the hippocampus (Chanales et al., 2017; Favila et al., 2016). Specifically, it was found that, when facing memory interference, the hippocampus represents two similar episodes significantly more differently compared to two unrelated episodes (i.e., hippocampal differentiation; Hulbert & Norman, 2015). That is, the neural representations of competing memories are pushed apart in order to be successfully differentiated. Yet the relationship between hippocampal neural differentiation and LPC adaptive content representations remains elusive, and can benefit from future works that aim to understand the dependencies between the two neural processes. Together, works from the current dissertation extend our understanding of LPC content representations by demonstrating

different approaches to quantify reinstated content information, but also provide evidence on how LPC content representations can be adaptively distorted in the service of resolving memory interferences.

Differences and Similarities between Content representations in LPC and VTC

It is worth noting that content representation is not exclusive to the LPC. In fact, early research shows that successfully recalling past events involve reactivating event-specific contents in sensory regions (Grill-Spector & Malach, 2004). For example, the ventral temporal cortex (VTC) reactivates visual category information when participants tried to recall a face or scene image (Kuhl et al., 2011; Kuhl & Chun, 2014). Previous studies have characterized two major differences in content representations between LPC and VTC: 1) content representation is stronger in VTC during perception but stronger in LPC during retrieval (Chen et al., 2017; Xiao et al., 2017), and 2) content representations in LPC during recall, but not VTC, predict success in memory recall (Kuhl & Chun, 2014).

Different from what was observed in previous studies, our results from chapter II suggest that a greater amount of content representation in both LPC and VTC are associated with remembered compared to forgotten stimuli. I speculate that this discrepancy could be caused by two possible factors. The first factor is that chapter II examines content representations within repetition-related neural patterns rather than within neural activity patterns during perception or recall. Furthermore, chapter II uses a recognition memory paradigm instead of a recall paradigm, which means that participants are exposed to the perceptual contents of the visual stimuli during both repetitions. Consequently, in the current study, the content representation may reflect how stimuli

were *encoded*, and the effect on memory, whereas previous studies examined how stimuli were *retrieved*, and the effect on memory (Chen et al., 2017; Xiao et al., 2017). In particular, since the brain prefers to encode subjectively new information (Nyberg, 2005), item-specific contents of the remembered stimuli are encoded only once during the first presentation, but those of the forgotten stimuli are repeatedly encoded across both presentations, which cancels out when neural differences are computed. Thus, the repetition-related neural differences would consist mainly of content representations of the remembered stimuli. Together, our results from chapter II suggest that item-specific content representations in LPC and VTC may contribute to recognition memory in a similar way during encoding, although they may contribute to memory differently during recall according to previous research (i.e., content representation in LPC is more relevant to memory success than VTC during recall; Kuhl & Chun, 2014).

Moreover, our results from chapter III reveal another systematic difference in content representation between LPC and VTC. That is, while both regions show content representations during recall, only those in LPC demonstrate adaptive modulations biased by the goal. Specifically, our results suggest that while resolving memory interference, LPC adaptively distorted content representations of similar memories in order to better separate them. On the contrary, VTC sticks with the unbiased content representation without showing adaptive changes. This finding is consistent with the theory of spatial transformation of content representation (Favila et al., 2020). Specifically, it is suggested that content representation is transformed from VTC during perception to LPC during retrieval. Critically, this spatial transformation also involves systematical changes of the content representation, which tend to shift from perceptual to conceptual representations.

That is, content representations in VTC are primarily concerned with absolute perceptual details, which remain consistent regardless of the goals. On the other hand, content representations in LPC reflect high-level, conceptual properties, which can be flexibly modulated by the goals. Together, works from the current dissertation reveal both similarities and differences in content representations between LPC and VTC. Future works can potentially focus on understanding the hierarchical relationships between LPC and VTC during perception and retrieval and how information is shared between the two regions.

Conclusion

In this thesis, I have investigated the nature and behavioral relevance of content representation in LPC. These experiments explored novel approaches to quantify content representations using pre-trained convolutional neural networks in addition to the conventional pattern similarity measures. Our results demonstrate the functional heterogeneity in content representation between dorsal and ventral LPC. Moreover, I show that the behavioral relevance of LPC content representation can resemble that of the visual region during perception. Nevertheless, LPC content representation also demonstrates its uniqueness in adaptive modulation biased by the high-level goal during memory retrieval. Understanding how perceptual and mnemonic details are similarly or differently represented and transformed between the parietal and visual regions is critical for deepening our understanding of human memory.

REFERENCES CITED

Akrami, A., Kopec, C. D., Diamond, M. E., & Brody, C. D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature 2018 554:7692*, *554*(7692), 368–372. https://doi.org/10.1038/nature25510

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126. https://doi.org/10.1038/s41593-021-00962-x

Andersen, R. A., Asanuma, C., Essick, G., & Siegel, R. M. (1990). Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *The Journal of Comparative Neurology*, *296*(1), 65–113. https://doi.org/10.1002/CNE.902960106

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. In *Psychological Review* (Vol. 102, Issue 1, pp. 68–100). American Psychological Association. https://doi.org/10.1037/0033-295X.102.1.68

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, and Psychophysics*, *79*(8), 2376–2395. https://doi.org/10.3758/s13414-017-1404-8

Ballard, I. C., Wagner, A. D., & McClure, S. M. (2019). Hippocampal pattern separation supports reinforcement learning. *Nature Communications*, *10*(1), 1073. https://doi.org/10.1038/s41467-019-08998-1

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*(2), 97–105. https://doi.org/10.1037/h0047507

Berryhill, M. E. (2012). Insights from neuropsychology: Pinpointing the role of the posterior parietal cortex in episodic and working memory. *Frontiers in Integrative Neuroscience*, *0*(JUNE 2012), 31. https://doi.org/10.3389/FNINT.2012.00031/BIBTEX

Berryhill, M. E., Phuong, L., Picasso, L., Cabeza, R., & Olson, I. R. (2007). Parietal Lobe and Episodic Memory: Bilateral Damage Causes Impaired Free Recall of Autobiographical Memory. *Journal of Neuroscience*, *27*(52), 14415–14423. https://doi.org/10.1523/JNEUROSCI.4163-07.2007

Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J., & Burgess, N. (2015). Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex. *Journal of Neuroscience*, *35*(43), 14426–14434. https://doi.org/10.1523/JNEUROSCI.1774-15.2015

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *14*(1), 129–133. https://doi.org/10.1016/S0010-9452(78)80016-1

Blatt, G. J., Pandya, D. N., & Rosene, D. L. (2003). Parcellation of cortical afferents to three distinct sectors in the parahippocampal gyrus of the rhesus monkey: an anatomical and neurophysiological study. *The Journal of Comparative Neurology*, *466*(2), 161–179. https://doi.org/10.1002/CNE.10866

Bone, M. B., Ahmad, F., & Buchsbaum, B. R. (2020). Feature-specific neural reactivation during episodic memory. *Nature Communications*, *11*(1), 1945. https://doi.org/10.1038/s41467-020-15763-2

Brodt, S., Gais, S., Beck, J., Erb, M., Scheffler, K., & Schönauer, M. (2018). Fast track to the neocortex: A memory engram in the posterior parietal cortex. *Science*, *362*(6418), 1045 LP – 1048. https://doi.org/10.1126/science.aau2528

Brodt, S., Pöhlchen, D., Flanagin, V. L., Glasauer, S., Gais, S., & Schönauer, M. (2016). Rapid and independent memory formation in the parietal cortex. *Proceedings of the National Academy of Sciences*, *113*(46), 13251 LP – 13256. https://doi.org/10.1073/pnas.1605719113

Buchsbaum, B. R., Lemire-Rodger, S., Fang, C., & Abdi, H. (2012). The Neural Basis of Vivid Memory Is Patterned on Perception. *Journal of Cognitive Neuroscience*, *24*(9), 1867–1883. https://doi.org/10.1162/jocn_a_00253

Buckner, R. L., Petersen, S. E., Ojemann, J. G., Miezin, F. M., Squire, L. R., & Raichle, M. E. (1995). Functional anatomical studies of explicit and implicit memory retrieval tasks. *Journal of Neuroscience*, *15*(1 I), 12–29. https://doi.org/10.1523/JNEUROSCI.15-01-00012.1995

Buckner, R. L., & Wheeler, M. E. (2001). The cognitive neuroscience og remembering. *Nature Reviews Neuroscience*, *2*(9), 624–634. https://doi.org/10.1038/35090048

Buckner, R. L., Wheeler, M. E., & Margaret, A. (2001). Encoding Processes during Retrieval Tasks. *Journal of Cognitive Neuroscience*, *13*(3), 406–415. https://doi.org/10.1162/08989290151137430

Cabeza, R. (2008). Role of parietal regions in episodic memory retrieval: The dual attentional processes hypothesis. *Neuropsychologia*, *46*(7), 1813–1827. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2008.03.019

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*(8), 613–625. https://doi.org/10.1038/nrn2459

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLOS Computational Biology*, *10*(12), e1003963. https://doi.org/10.1371/JOURNAL.PCBI.1003963

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain Activity Underlying Encoding and Retrieval of Source Memory. *Cerebral Cortex*, *12*(10), 1048–1056. https://doi.org/10.1093/CERCOR/12.10.1048

Capotosto, P., Baldassarre, A., Sestieri, C., Spadone, S., Romani, G. L., & Corbetta, M. (2017). Task and Regions Specific Top-Down Modulation of Alpha Rhythms in Parietal Cortex. *Cerebral Cortex*, *27*(10), 4815–4822. https://doi.org/10.1093/CERCOR/BHW278

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology*, *27*(15), 2307-2317.e5. https://doi.org/10.1016/j.cub.2017.06.057

Chanales, A. J. H., Tremblay-McGaw, A. G., Drascher, M. L., & Kuhl, B. A. (2021). Adaptive Repulsion of Long-Term Memory Representations Is Triggered by Event Similarity. *Psychological Science*, *32*(5), 705–720. https://doi.org/10.1177/0956797620972490/ASSET/IMAGES/LARGE/10.1177_0956797620972490-FIG2.JPEG

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Ciaramelli, E., Faggi, G., Scarpazza, C., Mattioli, F., Spaniol, J., Ghetti, S., & Moscovitch, M. (2017). Subjective recollection independent from multifeatural context retrieval following damage to the posterior parietal cortex. *Cortex*, *91*, 114–125. https://doi.org/10.1016/J.CORTEX.2017.03.015

Clower, D. M., West, R. A., Lynch, J. C., & Strick, P. L. (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *21*(16), 6283–6291. https://doi.org/10.1523/JNEUROSCI.21-16-06283.2001

Colgin, L. L., Moser, E. I., & Moser, M.-B. (2008). Understanding memory through hippocampal remapping. *Trends in Neurosciences*, *31*(9), 469–477. https://doi.org/https://doi.org/10.1016/j.tins.2008.06.008

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306. https://doi.org/10.1016/J.NEURON.2008.04.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Critchley, M. (1953). The parietal lobes. In *The parietal lobes.* Williams and Wilkins.

Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, *11*(2), 157–163. https://doi.org/10.1016/S0959-4388(00)00191-4

Danckert, J., & Ferber, S. (2006). Revisiting unilateral neglect. *Neuropsychologia*, *44*(6), 987–1006. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2005.09.004

Daselaar, S. M. (2009). Posterior midline and ventral parietal activity is associated with retrieval success and encoding failure. *Frontiers in Human Neuroscience*, *3*. https://doi.org/10.3389/neuro.09.013.2009

Daselaar, S. M., Prince, S. E., & Cabeza, R. (2004). When less means more: deactivations during encoding that predict subsequent memory. *Neuroimage*, *23*(3), 921–927.

Davachi, L., Maril, A., & Wagner, A. D. (2001). When Keeping in Mind Supports Later Bringing to Mind: Neural Markers of Phonological Rehearsal Predict Subsequent Remembering. *Journal of Cognitive Neuroscience*, *13*(8), 1059–1070. https://doi.org/10.1162/089892901753294356

Davidson, P. S. R., Anaki, D., Ciaramelli, E., Cohn, M., Kim, A. S. N., Murphy, K. J., Troyer, A. K., Moscovitch, M., & Levine, B. (2008). Does Lateral Parietal Cortex Support Episodic Memory? Evidence from Focal Lesion Patients. *Neuropsychologia*, *46*(7), 1743. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2008.01.011

Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-017-02752-1

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron*, *35*(5), 989–996. https://doi.org/10.1016/S0896-6273(02)00858-9

Dobbins, I. G., Rice, H. J., Wagner, A. D., & Schacter, D. L. (2003). Memory orientation and success: Separable neurocognitive components underlying episodic recognition. *Neuropsychologia*, *41*(3), 318–333. https://doi.org/10.1016/S0028-3932(02)00164-1

Driver, J., & Vuilleumier, P. (2001). Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition*, *79*(1–2), 39–88. https://doi.org/10.1016/S0010-0277(00)00124-4

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120. https://doi.org/10.1016/J.NEURON.2004.08.028

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, *3*(11), 1149–1152. https://doi.org/10.1038/80671

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*(1), 11066. https://doi.org/10.1038/ncomms11066

Favila, S. E., Lee, H., & Kuhl, B. A. (2020). Transforming the Concept of Memory Reactivation. *Trends in Neurosciences*. https://doi.org/https://doi.org/10.1016/j.tins.2020.09.006

Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal representations of stimulus features are amplified during memory retrieval and flexibly aligned with top-down goals. *Journal of Neuroscience*, *38*(36), 0564–18. https://doi.org/10.1523/JNEUROSCI.0564-18.2018

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience 2005 6:2*, *6*(2), 119–130. https://doi.org/10.1038/nrn1607

Frithsen, A., & Miller, M. B. (2014). The posterior parietal cortex: Comparing remember/know and source memory tests of recollection and familiarity. *Neuropsychologia*, *61*(1), 31–44. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2014.06.011

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Golomb, J. D. (2015). Divided spatial attention and feature-mixing errors. *Attention, Perception, & Psychophysics*, *77*(8), 2562–2569. https://doi.org/10.3758/s13414-015-0951-0

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/https://doi.org/10.1016/j.tics.2005.11.006

Grill-Spector, K., & Malach, R. (2004). THE HUMAN VISUAL CORTEX. *Https://Doi.Org/10.1146/Annurev.Neuro.27.070203.144220*, *27*, 649–677. https://doi.org/10.1146/ANNUREV.NEURO.27.070203.144220

Güçlü, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Henson, R., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *19*(10), 3962–3972. https://doi.org/10.1523/JNEUROSCI.19-10-03962.1999

Hower, K. H., Wixted, J., Berryhill, M. E., & Olson, I. R. (2014). Impaired perception of mnemonic oldness, but not mnemonic newness, after parietal lobe damage. *Neuropsychologia*, *56*(1), 409–417. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2014.02.014

Hulbert, J. C., & Norman, K. A. (2015). Neural Differentiation Tracks Improved Recall of Competing Memories Following Interleaved Study and Retrieval Practice. *Cerebral Cortex*, *25*(10), 3994–4008. https://doi.org/10.1093/cercor/bhu284

Humphreys, G. F., Lambon Ralph, M. A., & Simons, J. S. (2021). A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends in Neurosciences*, *44*(6), 452–463. https://doi.org/10.1016/J.TINS.2021.01.006

Hutchinson, J. B., Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic retrieval: Convergent and divergent effects of attention and memory. *Learning & Memory*, *16*(6), 343–356. https://doi.org/10.1101/LM.919109

Insausti, R., & Amaral, D. G. (2008). Entorhinal cortex of the monkey: IV. Topographical and laminar organization of cortical afferents. *Journal of Comparative Neurology*, *509*(6), 608–641. https://doi.org/10.1002/CNE.21753

Jost, K., Khader, P. H., Düsel, P., Richter, F. R., Rohde, K. B., Bien, S., & Rösler, F. (2012). Controlling Conflict from Interfering Long-term Memory Representations. *Journal of Cognitive Neuroscience*, *24*(5), 1173–1190. https://doi.org/10.1162/jocn_a_00199

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-Neuroanatomic Correlates of Recollection: Implications for Models of Recognition Memory. *Journal of Neuroscience*, *24*(17), 4172–4180. https://doi.org/10.1523/JNEUROSCI.0624-04.2004

Kay, K., Rokem, A., Winawer, J., Dougherty, R., & Wandell, B. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data . In *Frontiers in Neuroscience* (Vol. 7). https://www.frontiersin.org/article/10.3389/fnins.2013.00247

Khader, P., Burke, M., Bien, S., Ranganath, C., & Rösler, F. (2005). Content-specific activation during associative long-term memory retrieval. *NeuroImage*, *27*(4), 805–816. https://doi.org/https://doi.org/10.1016/j.neuroimage.2005.05.006

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLOS Computational Biology*, *10*(11), e1003915. https://doi.org/10.1371/JOURNAL.PCBI.1003915

Kim, G., Norman, K. A., & Turk-Browne, N. B. (2017). Neural differentiation of incorrectly predicted memories. *Journal of Neuroscience*, *37*(8), 2022–2031.

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage*, *54*(3), 2446–2461. https://doi.org/10.1016/J.NEUROIMAGE.2010.09.045

Kim, H. (2012). A dual-subsystem model of the brain's default network: Self-referential processing, memory retrieval processes, and autobiographical memory retrieval. *NeuroImage*, *61*(4), 966–977. https://doi.org/10.1016/J.NEUROIMAGE.2012.03.025

Kim, H. (2013). Differential neural activity in the recognition of old versus new events: An Activation Likelihood Estimation Meta-Analysis. *Human Brain Mapping*, *34*(4), 814–836. https://doi.org/10.1002/HBM.21474

Kim, H., Daselaar, S. M., & Cabeza, R. (2010). Overlapping brain activity between episodic memory encoding and retrieval: Roles of the task-positive and task-negative networks. *NeuroImage*, *49*(1), 1045–1054. https://doi.org/10.1016/J.NEUROIMAGE.2009.07.058

Kincade, J. M., Abrams, R. A., Astafiev, S. v, Shulman, G. L., & Corbetta, M. (2005). An Event-Related Functional Magnetic Resonance Imaging Study of Voluntary and Stimulus-Driven Orienting of Attention. *The Journal of Neuroscience*, *25*(18), 4593. https://doi.org/10.1523/JNEUROSCI.0236-05.2005

Konishi, S., Wheeler, M. E., Donaldson, D. I., & Buckner, R. L. (2000). Neural Correlates of Episodic Retrieval Success. *NeuroImage*, *12*(3), 276–286. https://doi.org/10.1006/NIMG.2000.0614

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 4. https://doi.org/10.3389/neuro.06.004.2008

Kuhl, B. A., & Chun, M. M. (2014). Successful Remembering Elicits Event-Specific Activity Patterns in Lateral Parietal Cortex. *Journal of Neuroscience*, *34*(23), 8051–8060. https://doi.org/10.1523/jneurosci.4328-13.2014

Kuhl, B. A., Johnson, M. K., & Chun, M. M. (2013). Dissociable Neural Mechanisms for Goal-Directed Versus Incidental Memory Reactivation. *Journal of Neuroscience*, *33*(41), 16099–16109. https://doi.org/10.1523/jneurosci.0207-13.2013

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences*, *108*(14), 5903–5908. https://doi.org/10.1073/pnas.1016939108

LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Memory. *Journal of Neuroscience*, *33*(13), 5466–5474. https://doi.org/10.1523/JNEUROSCI.4293-12.2013

Lavenex, P., Suzuki, W. A., & Amaral, D. G. (2002). Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *The Journal of Comparative Neurology*, *447*(4), 394–420. https://doi.org/10.1002/CNE.10243

Lee, H., Chun, M. M., & Kuhl, B. A. (2017). Lower Parietal Encoding Activation Is Associated with Sharper Information and Better Memory. *Cerebral Cortex*, *27*(4), 2486–2499. https://doi.org/10.1093/cercor/bhw097

Lee, H., & Kuhl, B. A. (2016). Reconstructing Perceived and Retrieved Faces from Activity Patterns in Lateral Parietal Cortex. *Journal of Neuroscience*, *36*(22), 6069–6082. https://doi.org/10.1523/JNEUROSCI.4286-15.2016

Lee, H., Samide, R., Richter, F. R., & Kuhl, B. A. (2019). Decomposing Parietal Memory Reactivation to Predict Consequences of Remembering. *Cerebral Cortex*, *29*(8), 3305–3318. https://doi.org/10.1093/cercor/bhy200

Lewis, J. W., & van Essen, D. C. (2000). Corticocortical Connections of Visual, Sensorimotor, and Multimodal Processing Areas in the Parietal Lobe of the Macaque Monkey. *J. Comp. Neurol*, *428*, 112–137. https://doi.org/10.1002/1096-9861(20001204)428:1

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*(6), 1918–1929. https://doi.org/10.1152/jn.1993.69.6.1918

Liang, J. C., Wagner, A. D., & Preston, A. R. (2013). Content Representation in the Human Medial Temporal Lobe. *Cerebral Cortex*, *23*(1), 80–96. https://doi.org/10.1093/cercor/bhr379

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). *Microsoft COCO: Common Objects in Context BT - Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars, Eds.; pp. 740–755). Springer International Publishing.

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal Mismatch Signals Are Modulated by the Strength of Neural Predictions and Their Similarity to Outcomes. *Journal of Neuroscience*, *36*(50), 12677–12687.

Majerus, S., Péters, F., Bouffier, M., Cowan, N., & Phillips, C. (2018). The Dorsal Attention Network Reflects Both Encoding Load and Top–down Control during Working Memory. *Journal of Cognitive Neuroscience*, *30*(2), 144–159. https://doi.org/10.1162/jocn_a_01195

Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., & Pandya, D. N. (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cerebral Cortex*, *15*(6), 854–869. https://doi.org/10.1093/CERCOR/BHH186

Mecklinger, A. (2000). Interfacing mind and brain: A neurocognitive model of recognition memory. *Psychophysiology*, *37*(5), 565–582. https://doi.org/10.1111/1469-8986.3750565

Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434–455. https://doi.org/10.1037/0033-295X.95.4.434

Miller, E. K., Li, L., & Desimone, R. (1991). A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex. *Science*, *254*(5036), 1377–1379. https://doi.org/10.1126/science.1962197

Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, *16*(5), 504–520. https://doi.org/https://doi.org/10.1002/hipo.20178

Muñoz, M., & Insausti, R. (2005). Cortical efferents of the entorhinal cortex and the adjacent parahippocampal region in the monkey (Macaca fascicularis). *The European Journal of Neuroscience*, *22*(6), 1368–1388. https://doi.org/10.1111/J.1460-9568.2005.04299.X

Norman, K. A., Newman, E. L., & Detre, G. (2007). A Neural Network Model of Retrieval-Induced Forgetting. *Psychological Review*, *114*(4), 887–953. https://doi.org/10.1037/0033-295X.114.4.887

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. v. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Nyberg, L. (2005). Any novelty in hippocampal formation and memory? *Current Opinion in Neurology*, *18*(4), 424–428. https://doi.org/10.1097/01.WCO.0000168080.99730.1C

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, *4*(6), 661–682. https://doi.org/10.1002/hipo.450040605

Osgood, C. E. (1949). The similarity paradox in human learning: a resolution. In *Psychological Review* (Vol. 56, Issue 3, pp. 132–143). American Psychological Association. https://doi.org/10.1037/h0057488

Otten, L. J. (2007). Fragments of a Larger Whole: Retrieval Cues Constrain Observed Neural Correlates of Memory Encoding. *Cerebral Cortex*, *17*(9), 2030–2038. https://doi.org/10.1093/CERCOR/BHL111

Otten, L. J., & Rugg, M. D. (2001). When more means less: neural activity related to unsuccessful memory encoding. *Current Biology*, *11*(19), 1528–1530.

Petrides, M., Alivisatos, B., & Evans, A. C. (1995). Functional activation of the human ventrolateral frontal cortex during mnemonic retrieval of verbal information. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(13), 5803–5807. https://doi.org/10.1073/PNAS.92.13.5803

Petrides, M., & Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *The Journal of Comparative Neurology*, *228*(1), 105–116. https://doi.org/10.1002/CNE.902280110

Petrides, M., & Pandya, D. N. (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *The European Journal of Neuroscience*, *11*(3), 1011–1036. https://doi.org/10.1046/J.1460-9568.1999.00518.X

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Neuroscience: Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*(5756), 1963–1966. https://doi.org/10.1126/science.1117645

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Ranganath, C., & Knight, R. T. (2003). Prefrontal cortex and episodic memory: Integrating findings from neuropsychology and functional brain imaging. In *The Cognitive Neuroscience of Memory: Encoding and Retrieval* (pp. 83–99). Psychology Press.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*(1), 2–13. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2003.07.006

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends in Cognitive Sciences*, *23*(9), 726–742. https://doi.org/10.1016/j.tics.2019.06.007

Rockland, K. S., & van Hoesen, G. W. (1999). Some temporal and parietal cortical connections converge in CA1 of the primate hippocampus. *Cerebral Cortex (New York, N.Y. : 1991)*, *9*(3), 232–237. https://doi.org/10.1093/CERCOR/9.3.232

Rugg, M. D. (1995). Event-related potential studies of human memory. In *The cognitive neurosciences.* (pp. 789–801). The MIT Press.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. https://doi.org/10.1016/J.TICS.2007.04.004

Rugg, M. D., & King, D. R. (2018). Ventral lateral parietal cortex and episodic memory retrieval. *Cortex*, *107*, 238–250. https://doi.org/https://doi.org/10.1016/j.cortex.2017.07.012

Rushworth, M. F. S., Behrens, T. E. J., & Johansen-Berg, H. (2006). Connection patterns distinguish 3 regions of human parietal cortex. *Cerebral Cortex*, *16*(10), 1418–1430. https://doi.org/10.1093/CERCOR/BHJ079

Schacter, D. L., Guerin, S. A., & st. Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*(10), 467–474. https://doi.org/10.1016/j.tics.2011.08.004

Schapiro, A. C., Kustner, L. v., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*(1), 8151. https://doi.org/10.1038/ncomms9151

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., & Wedeen, V. J. (2007). Association fibre pathways of the brain: Parallel observations from diffusion spectrum imaging and autoradiography. *Brain*, *130*(3), 630–653. https://doi.org/10.1093/BRAIN/AWL359

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11–21. https://doi.org/10.1136/JNNP.20.1.11

Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2013). The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia*, *51*(1), 59–66. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2012.11.006

Seltzer, B., & Pandya, D. N. (1984). Further observations on parieto-temporal connections in the rhesus monkey. *Experimental Brain Research*, *55*(2), 301–312. https://doi.org/10.1007/BF00237280

Sestieri, C., Corbetta, M., Romani, G. L., & Shulman, G. L. (2011). Episodic Memory Retrieval, Parietal Cortex, and the Default Mode Network: Functional and Topographic Analyses. *Journal of Neuroscience*, *31*(12), 4407–4420. https://doi.org/10.1523/JNEUROSCI.3335-10.2011

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience 2017 18:3*, *18*(3), 183–192. https://doi.org/10.1038/nrn.2017.6

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936. https://doi.org/10.1152/JN.2001.86.4.1916

Sharot, T., Delgado, M. R., & Phelps, E. A. (2004). How emotion enhances the feeling of remembering. *Nature Neuroscience*, *7*(12), 1376–1380. https://doi.org/10.1038/nn1353

Shimamura, A. P. (1995). Memory and the Prefrontal Cortex. *Annals of the New York Academy of Sciences*, *769*(1), 151–160. https://doi.org/10.1111/J.1749-6632.1995.TB38136.X

Shimamura, A. P. (2011). Episodic retrieval and the cortical binding of relational activity. *Cognitive, Affective and Behavioral Neuroscience*, *11*(3), 277–291. https://doi.org/10.3758/S13415-011-0031-4/FIGURES/2

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia*, *45*(10), 2163–2179. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2007.03.007

Sobotka, S., & Ringo, J. L. (1996). Mnemonic Responses of Single Units Recorded from Monkey Inferotemporal Cortex, Accessed via Transcommissural Versus Direct Pathways: A Dissociation between Unit Activity and Behavior. *The Journal of Neuroscience*, *16*(13), 4222 LP – 4230. https://doi.org/10.1523/JNEUROSCI.16-13-04222.1996

Spaniol, J., Davidson, P. S. R., Kim, A. S. N., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. *Neuropsychologia*, *47*(8–9), 1765–1779. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2009.02.028

Squire, L. R. (1992). Memory and the Hippocampuss: A Synthesis From Findings With Rats, Monkeys, and Humans. *Psychological Review*, *99*(2), 195–231. https://doi.org/10.1037/0033-295X.99.2.195

St-Laurent, M., Abdi, H., & Buchsbaum, B. R. (2015). Distributed Patterns of Reactivation Predict Vividness of Recollection. *Journal of Cognitive Neuroscience*, *27*(10), 2000–2018. https://doi.org/10.1162/jocn_a_00839

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, *95*(1), 3–28. https://doi.org/10.1037/0033-2909.95.1.3

Suzuki, W. L., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *The Journal of Comparative Neurology*, *350*(4), 497–533. https://doi.org/10.1002/CNE.903500402

Takahashi, E., Ohki, K., & Kim, D. S. (2008). Dissociated pathways for successful memory retrieval from the human parietal cortex: Anatomical and functional connectivity analyses. *Cerebral Cortex*, *18*(8), 1771–1778. https://doi.org/10.1093/CERCOR/BHM204

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391. https://doi.org/10.1002/hipo.450040319

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology / Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/H0080017

Tulving, E., Kapur, S., Markowitsch, H. J., Craik, F. I. M., Habib, R., & Houle, S. (1994). Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(6), 2012–2015. https://doi.org/10.1073/PNAS.91.6.2012

Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: Insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology of Learning and Memory*, *91*(2), 139–154. https://doi.org/10.1016/J.NLM.2008.10.011

Vallar, G. (1998). Spatial hemineglect in humans. *Trends in Cognitive Sciences*, *2*(3), 87–97. https://doi.org/10.1016/S1364-6613(98)01145-0

Vannini, P., O'Brien, J., O'Keefe, K., Pihlajamaki, M., LaViolette, P., & Sperling, R. A. (2011). What Goes Down Must Come Up: Role of the Posteromedial Cortices in Encoding and Retrieval. *Cerebral Cortex*, *21*(1), 22–34. https://doi.org/10.1093/cercor/bhq051

Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia*, *46*(7), 1787–1799. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2008.01.004

Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., & Buckner, R. L. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of Neurophysiology*, *96*(6), 3517–3531. https://doi.org/10.1152/JN.00048.2006/ASSET/IMAGES/LARGE/Z9K0110677360008.JPEG

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453. https://doi.org/https://doi.org/10.1016/j.tics.2005.07.001

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2014). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Ward, E. J., Chun, M. M., & Kuhl, B. A. (2013). Repetition Suppression and Multi-Voxel Pattern Similarity Differentially Track Implicit and Explicit Visual Memory. *Journal of Neuroscience*, *33*(37), 14749–14757. https://doi.org/10.1523/JNEUROSCI.4889-12.2013

Wellman, B. J., & Rockland, K. S. (1997). Divergent cortical connections to entorhinal cortex from area TF in the Macaque. *Journal of Comparative Neurology*, *389*(3), 361–376. https://doi.org/10.1002/(SICI)1096-9861(19971222)389:3<361::AID-CNE1>3.0.CO;2-Z

Wheeler, M. E., & Buckner, R. L. (2003). Functional Dissociation among Components of Remembering: Control, Perceived Oldness, and Content. *Journal of Neuroscience*, *23*(9), 3869–3880. https://doi.org/10.1523/JNEUROSCI.23-09-03869.2003

Wheeler, M. E., & Buckner, R. L. (2004). Functional-anatomic correlates of remembering and knowing. *NeuroImage*, *21*(4), 1337–1349. https://doi.org/10.1016/J.NEUROIMAGE.2003.11.001

Whitlock, J. R. (2017). Posterior parietal cortex. *Current Biology*, *27*(14), R691–R695. https://doi.org/10.1016/J.CUB.2017.06.007

Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, *55*(1), 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555

Woodruff, C. C., Johnson, J. D., Uncapher, M. R., & Rugg, M. D. (2005). Content-specificity of the neural correlates of recollection. *Neuropsychologia*, *43*(7), 1022–1032. https://doi.org/10.1016/j.neuropsychologia.2004.10.013

Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., & Xue, G. (2017). Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. *The Journal of Neuroscience*, *37*(11), 2986–2998. https://doi.org/10.1523/JNEUROSCI.2324-16.2017

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8619–8624. https://doi.org/10.1073/PNAS.1403112111/SUPPL_FILE/PNAS.201403112SI.PDF

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/https://doi.org/10.1016/j.tins.2011.06.006

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., Mancuso, L., Kliot, D., Das, S. R., & Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, *36*(1), 258–287. https://doi.org/10.1002/hbm.22627

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 818–833). Springer International Publishing.