Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	_____
Rafi Haque	**Date**

Neocortical-Medial Temporal Lobe Interactions during Visuospatial Memory Formation and Implications for Screening of Memory Impairment in Alzheimer's Disease

Rafi Haque
Doctor of Philosophy
Graduate Division of Biological and Biomedical Science
Neuroscience

_____

Allan Levey, MD/Ph.D.
Advisor

_____

Annabelle Singer, Ph.D.
Committee Member

_____

Chethan Pandarinath, Ph.D.
Committee Member

_____

Gari Clifford, Ph.D.
Committee Member

_____

James Lah, MD/Ph.D.
Committee Member

_____

Joseph Manns, Ph.D.
Committee Member

_____

Kareem Zaghloul, MD/Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

Neocortical-Medial Temporal Lobe Interactions during Visuospatial Memory Formation and Implications for Screening of Memory Impairment in AD

By

Rafi Haque
B.S.E, University of Pennsylvania, 2013

Advisor: Allan Levey, MD/Ph.D.

An abstract of a dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Neuroscience

2020

Abstract

Neocortical-Medial Temporal Lobe Interactions during Visuospatial Memory Formation and Implications for Screening of Memory Impairment in Alzheimer's Disease

By Rafi Haque

Alzheimer's disease (AD) is a chronic neurodegenerative disorder characterized by pathological changes prior to onset of clinical symptoms. The earliest site of cortical pathology is the medial temporal lobe (MTL), a region critical for the formation of memories. The presence of pathology within the MTL has generated interest in developing memory assessments that involve the MTL and may identify individuals with memory impairment due to AD. To address this need, we developed a visuospatial memory paradigm that requires participants to retrieve their past memories and recognize differences between these memories and the current experience. We presented this memory paradigm while recording electrophysiological activity in visual association areas and the MTL. Successful recognition of visuospatial memories was associated with increases in 80-120 Hz power within visual association areas and the MTL and also accompanied by 8-12 Hz communication between these regions. We then developed a version of this visuospatial memory paradigm that strictly uses eye movements as an index of retrieval. Visuospatial memory performance based on eye movements was substantially reduced in participants with AD and could differentiate the two populations with high sensitivity and specificity. To improve the scalability of this assessment, we developed an iPad-based version of the memory task that utilizes iTracker, a convolutional neural network (CNN) architecture used to track eye movements on Apple devices. We found that the iPad-based implementation could also differentiate cognitively impaired participants from healthy controls with high sensitivity and specificity. This work advances our understanding of the neocortical-MTL interactions underlying visuospatial memory formation and provides a passive, sensitive, and efficient memory assessment that may be used to identify individuals at risk of future memory impairment.

Neocortical-Medial Temporal Lobe Interactions during Visuospatial Memory Formation and Implications for Screening of Memory Impairment in Alzheimer's Disease

By

Rafi Haque
B.S.E, University of Pennsylvania, 2013

Advisor: Allan Levey, MD/Ph.D.

A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Neuroscience

2020

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Allan Levey for his continuous support of my Ph.D study and research. Your patience with me, non-stop encouragement, and clear vision throughout the last few years were deeply inspiring. You were available no matter what time of day and have shaped my future trajectory not just in research but more importantly in life. I could not have imagined having a better mentor for my growth as a scientist and person.

I would like to extend my regards to Kareem Zaghloul. You have been an absolutely incredible mentor for me with your foresight and passion for research. You have taught me to think critically, deeply, and out of the box while pushing me to pursue what I love. For your encouragement and support I cannot possibly thank you enough. I feel fortunate to have worked with you over the last several years and only hope that our work together continues in the future.

I would also like to thank my committee members Gari Clifford, Joseph Manns, Chethan Pandanirath, Annabelle Singer, and James Lah. You are all mentors who I admired prior to starting my dissertation. Now having interacted with you that admiration has only grown. Your suggestions often left me thinking about problems more carefully and have shaped many of the ideas in this dissertation. I appreciate all the support and stimulating discussions and this dissertation would not have been possible without you

I thank my fellow co-workers Ceecee Manzaneres and Alvince Pongos along with rest of the Emory Healthy Aging Study team. Your dedication to this work has allowed me to accomplish more than I could have ever imagined.

Last but not the least, I would like to thank my parents Anwarul and Ashrafi Haque and my brother Tarif Haque. Your love, dedication, and sacrifice gave me the strength to keep moving forward and you are who I aspire to be every day.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disorder characterized histopathologically by the presence of amyloid-beta (Aβ) peptides in extracellular senile plaques and the formation of intracellular, neurofibrillary tangles (NFTs) composed of microtubule-associated protein tau [1,2]. The presence of AD pathology prior to the onset of clinical symptoms has generated considerable interest in identifying individuals during the earliest stages in the AD neuropathological spectrum before severe cognitive impairment. Memory tasks that are mediated by the medial temporal lobe (MTL), the earliest site of pathology in AD, offer promise for early detection of decline in AD. When initially exposed to an experience, memories are generated and stored through feedforward and feedback interactions between the neocortex and MTL. When presented with a similar experience, the previous experience can be retrieved through reactivation of these neocortical-MTL interactions. These interactions therefore provide an internally generated memory to which new experiences may be compared. Memory paradigms have recently been developed that require the detection of violations between past and present experience. These memory paradigms may identify the neocortical-MTL interactions underlying violations between past and present experience and whether the detection of such violations are impaired in AD.

1.1 Molecular Pathogenesis and Pathophysiology of Alzheimer's Disease

Insights into the pathogenesis of AD arose from the pioneering neuropathological observations of senile plaques and NFTs by Alois Alzheimer[9], and the discovery eighty years later that these plaques consist of a 39-43 amino acid peptide now known as amyloid-beta (Aβ) (Fig. 1) [3]. Aβ is derived from the sequential cleavage of the Aβ-precursor protein (APP) by β-secretase and γ-secretase to primarily yield two major isoforms, Aβ40 and Aβ42, along with various C- and N-

terminally truncated and/or modified isoforms. In the disease state, Aβ acquires a β-strand-rich molecular conformation, a state prone to self-assemble into oligomers, diffuse plaques, and dense-core (amyloid) plaques. Many dense-core plaques, which are thought to represent a late stage of development, are surrounded by degenerating neurites as well as activated microglia and astrocytes[2]. Soluble Aβ oligomers, rather than dense core plaques, are now thought to be a key toxic intermediary linked to early synaptic injury in loss and reflect the earliest stages of AD [4]. The role of Aβ in AD pathogenesis is further supported by genetic mutations in rare, autosomal-dominant, familial cases of AD. These mutations in APP and presenilin 1 and 2 components of the γ-secretase complex are associated with the overproduction and/or accumulation of Aβ as well as a more aggressive disease course, including an early age of onset compared to sporadic cases of AD[5–8]. These studies implicate the Aβ peptide as a causative agent in triggering a sequence of events that ultimately leads to NFTs, neuronal dysfunction, synaptic loss, cerebral atrophy, and dementia[4,9] – at least in these rare families with autosomal dominant mutations.

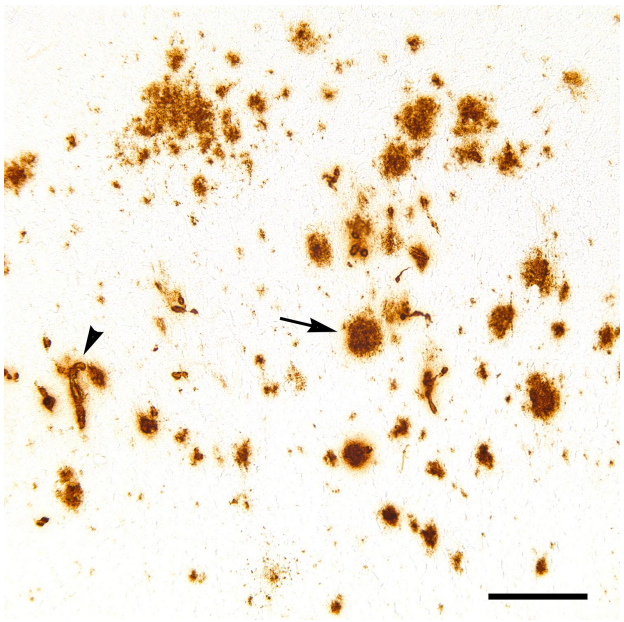

Figure 1. Aβ-immunoreactive deposits (brown) in the neocortex

NFTs are also a core lesion that, along with Aβ plaques, defines all cases of AD [2,10–13]. NFTs are intracellular deposits of hyperphosphorylated, microtubule-associated protein, tau. Tau typically binds tubulin to control the stability of microtubules and regulate axonal transport. In its hyperphosphorylated form, tau self-assembles to form paired helical filaments at the ultrastructural level and NFTs at the light microscopic level [2,10–13]. The emergence of NFTs follows a characteristic spatiotemporal progression across brain regions, known as the Braak stages [12]. Cortical NFTs initially appear in layer II and III of the trans-entorhinal region (stage I) and progress to the hippocampal formation (stage II) [13]. NFTs then deposit in basal areas of the temporal lobe (stage III) and neocortical association areas (stage IV) followed by deposition in motor and sensory areas in stages V and VI. Unlike amyloid pathology, NFTs often correlate with neuronal loss and symptom progression in AD [13], beginning with subtle memory deficits followed by multi-domain cognitive impairment [14–17]. The presence of NFTs and associated MTL neurodegeneration prior to the onset of clinical symptoms suggest individuals at risk of cognitive decline may be identified during the earliest stages of AD.

1.2 Clinical Assessment of Cognitive Impairment in Alzheimer's Disease

Memory loss is the earliest detectable clinical feature of AD and characterized by insidious anterograde long-term episodic amnesia. As the disease progresses, individuals develop other cognitive impairments, including executive dysfunction, attentional impairment, and visuospatial dysfunction. In its later stages, individuals exhibit severe dementia and rely solely on their caregivers to carry out simple tasks. As these symptoms progress, care becomes overwhelmingly difficult and often necessitates placement in long-term care facilities. In the terminal-stages, complications such as dehydration, malnutrition, and infection frequently culminate in death.

A clinical evaluation dementia begins once there are concerns about an individual's memory or cognition. An important part of the diagnostic evaluation of dementia is the clinical history from both the patient and the caregiver. Clinical criteria for dementia includes a slow progressive history of insidious onset, cognitive decline in more than one domain [18]. To obtain an objective measurement of cognition, clinicians administer general cognitive screenings such as the Montreal Cognitive Assessment (MoCA) [19]. The MoCA is a 30-point assessment administered in approximately 15 minutes that measures cognitive domains including memory, visuospatial, executive, attention, language, abstract reasoning, and orientation. MoCA performance for healthy controls, MCI and AD populations are 25.57 ± 2.75, 23.41 ± 3.38, and 15.30 ± 5.51[20]. The MoCA shows robust differences between healthy controls and AD participants however fails to reliably differentiate healthy controls from MCI, especially in the early stages. Even when most predictive sections of the MoCA are used -- memory, executive, and orientation, the MoCA shows a sensitivity and specificity of only 0.73 and 0.62 in differentiating healthy controls from MCI [21]. The MoCA also suffers from severe ceiling effects, depends considerably on socio-educational variables and is influenced by the language of administration. These short-comings have generated considerable interest in the development of new screening assessments that assess memory, the cognitive domain affected earliest in AD.

Detailed neuropsychological testing is also sometimes used for identifying cognitive impairment. Memory deficits are established through performance on neuropsychological tasks measuring verbal recall such as the Rey Auditory Verbal Learning Test (RAVLT) [22] and Free Cued Selective Reminding Test [23,24]. These tests have been successful in detecting memory impairment [25–30].

However, these tests require at least 30-45 minutes to be administered by trained personnel and are often underused for symptomatic individuals because of the resource demands necessary to implement in clinical settings. Participants often do not like the experience of verbal learning tests, due to the perceived poor performance on such tests. The development of new memory assessments that involve the MTL and that can be administered rapidly may identify individuals at risk for future memory decline compared to current cognitive assessments.

1.3 The Medial Temporal Lobe and Memory Formation

The MTL system consists of the hippocampus, entorhinal cortex, perirhinal cortex, and parahippocampal circuits, cortical areas that are organized hierarchically based on their neuroanatomical projection patterns [31,32]. Initial insights into the involvement of the MTL in memory formation arose from studies in patient HM who received bilateral resection of the MTL in order to treat the symptoms of intractable epilepsy [33]. Following removal of the MTL, HM exhibited severe memory impairment for everyday experiences, without any deficits in visual perception, language abilities, or motor learning. Early studies in non-human primates rigorously tested the MTL's involvement in memory formation through the development of the delayed object recognition memory tasks. During these tasks, NHPs were presented with an object and then after a delay, presented with the original object along with a novel object. NHPs with large MTL lesions were impaired in identifying the original object for Delay Match to Sample (DMS) task and the novel object in Delay Non-Match to Sample (DNMS) tasks, especially when significant delay exists between the two presentations were present [34–40].

To clarify the role of individual regions of the MTL in memory formation, selective lesions were performed to the perirhinal cortex, parahippocampal cortex, and the hippocampus proper to a variant of the DNMS task that measures spontaneous exploration of novel objects compared to familiar objects. The results from these studies indicated consistent deficits in the exploration of novel objects following selective lesions of the perirhinal cortex [41–44], compared to the inconsistent deficits from lesions of the hippocampus and parahippocampal cortices [41,42,44–46]. In contrast, relatively small lesions of the hippocampus and parahippocampal resulted in deficits in the exploration of on object that is moved from its original location or environment [42,43,47]. Double dissociation experiments in rats revealed that parahippocampal damage produced a deficit in object-location recognition memory while perirhinal damage resulted in a deficit in object-object recognition memory [43]. These studies have led to the hypothesis that interactions between the neocortical areas and the hippocampus may underlie the formation of visuospatial memories [43,48–50]

1.4 Neocortical and Medial Temporal Lobe Interactions during Visuospatial Memory Formation

The hippocampal-neocortical system consists of a set of cortical areas that are arranged in a hierarchy based on their neuroanatomical projection patterns [31,32,51,52]. In the neocortex, feedforward projections typically originate in supragranular layers and terminate in granular layer IV while feedback (FB) connections begin in infragranular layers and terminate outside of the granular layer IV [51,52]. The entorhinal cortex lacks a granular layer IV and therefore, once reaching the entorhinal cortex, feedforward projections from perirhinal and parahippocampal cortices terminate on layer II neurons [31]. Neocortical input from layer II cells of the entorhinal cortex are transmitted in feedforward direction through the hippocampal trisynaptic pathway consisting of

the dentate gyrus, the recurrent CA3-pyramidal system, and CA1 [53]. The hippocampus returns this feedforward input back to the neocortex via feedback projections onto neurons in layer V of the entorhinal cortex. Layer V entorhinal neurons relays this input back to supragranular layers of perirhinal and parahippocampal cortices via feedback projections [31].

Based on these feedforward and feedback interactions, a model of visuospatial memory formation has been developed that involves a distributed set of interactions involving the neocortex and hippocampus [53–57]. When exposed to an experience, feedforward and feedback interactions between the neocortex and the MTL form a neural representation of that experience [54,56–58]. When presented with a similar visual experience, the previous experience can be retrieved through autoassociative reaction of these neocortical-MTL representations[54,56,59,60]. Any difference between the current experience and our memories is hypothesized to be detected by neocortical-medial temporal lobe circuits. Based on this framework, neocortical-hippocampal circuits are constantly building and updating an internal model of the world in order to detect when current experience is in violation with past experience [53,61–65]. While this framework presents computationally efficient, generalizable approach for memory formation, the neocortical-medial temporal lobe interactions that underlie the detection of violation between current and past experiences remain unknown.

These interactions may be elucidated using memory paradigms that require participants to compare old experiences to the current experience. In one such paradigm, participants are presented with natural scenes and objects that they encode into memory [66–68]. The participants are then presented with repeated versions of the images or the images with an object added or removed. Successfully recognizing this manipulation requires participants to not only retrieve the past visual experience but to also then compare the retrieved memory with the present image. The recognition of this

manipulation can be measured by asking participants whether the image changed during the second viewing and identifying the number of correct responses. MTL damage impairs the explicit identification of these manipulations providing preliminary evidence that neocortical-MTL interactions may be involved in detecting differences between past and present experience [66–68]

1.5 Visuospatial Memory Assessments Screening for Memory Impairment

An important advantage of these visuospatial memory paradigms is that memory retrieval can be assessed simply using eye movements. Successful recognition of a manipulation has been associated with increased viewing of the manipulated region compared to unchanged parts of the image [66–68]. Viewing of the manipulation also increases during successful compared to unsuccessful recognition. Participants with MTL damage also show impairment in viewing these manipulations compared to age-matched controls [66–69]. These studies suggest that visuospatial memory paradigms based on eye movements are sensitive indicators of entorhinal-hippocampal function, the cortical region vulnerable to NFT deposition and neurodegeneration. These paradigms may serve as an early indicator of memory impairment in AD compared to general cognitive screening methods.

Using eye movements as an index of memory offers a number of practical benefits compared to verbal learning tests used to assess memory impairment in AD [70–72]. Memory paradigms that use eye movements as an index of retrieval allow for rapid assessment of memory with little instruction and minimal user input while predicting memory loss with high sensitivity and specificity [73,74]. These paradigms appear to be strikingly less distressing and frustrating to both research participants and clinical patient populations than traditional neuropsychological tasks and also

avoids issues related to task comprehension and explicit memory judgements. Conventional tasks are also subject to differences in effort, literacy, cultural variation, and decision-making capacity which can confound the measures of memory. Eye-movement based memory paradigms may provide a more practical solution to tracking longitudinal cognitive status in early stages of AD, especially when compared with current methods to measure cognitive and memory impairment.

1.6 Thesis Aims

In this thesis, a visuospatial memory paradigm was developed based on prior work showing increased exploration of added or removed objects to a previously viewed image depends on the MTL. This task may identify the neurophysiological interactions between the neocortex and MTL during visuospatial memory formation and provide a passive, sensitive, and efficient assessment of memory impairment in individuals with Alzheimer's disease and other disorders involving MTL circuits. To these hypotheses, the aims of the thesis are provided below:

Aim 1. To identify the neocortical-MTL interactions that underlie visuospatial memory performance in humans.

The neocortical-MTL interactions that underlie the detection of violations between current and past experiences (i.e., memory) remain unknown. We will administer a visuospatial recognition memory paradigm to participants undergoing seizure monitoring for medically refractory epilepsy. We will record neural activity from visual association areas and the MTL using intracranial EEG electrodes and quantify the differences in spectral power and coherence that occur within and between these regions during successful recognition of a manipulation.

Aim 2. To assess visuospatial memory performance based on eye movements during healthy aging, MCI, and AD

MTL damage produces deficits in visuospatial memory performance based on eye movements. Visuospatial memory tasks based on eye movements may offer a passive, sensitive, and efficient assessment of memory performance in AD. Visuospatial Memory EyeTracking Test (VisMET) will be administered in a cohort of healthy, MCI and AD populations. We hypothesize that visuospatial memory performance will be impaired in participants with MCI and AD compared to healthy controls. To test this hypothesis, we will quantify differences in viewing time of objects removed or added to an image between these populations and develop a logistic regression model to estimate cognitive impairment and disease status.

Aim 3. To develop a tablet-based version of visuospatial memory performance based on eye movements in order to identify cognitive impairment in AD

Current administration of eye-movement based memory paradigms require face-to-face administration by a technician within a clinical setting using a standalone eye-tracker. We will develop an iPad based version of VisMET that utilizes iTracker, a deep convolutional neural network used to track eye movements on the iPad. We will administer the mobile version of VisMET in a cohort of control and cognitively-impaired populations. We hypothesize that visuospatial memory performance will be impaired in healthy controls compared to cognitively-impaired participants. To test this hypothesis, we will quantify differences in viewing time of objects removed or added to an image between these populations and develop a logistic regression model to estimate cognitive impairment and disease status.

Chapter 2: Neocortical-Medial Temporal Lobe Interactions during Visuospatial Memory Formation

2.1 Introduction

When initially exposed to any experience, we rely upon our memories to set our expectations. These expectations determine the extent to which any experience involves new information. For example, when we enter our home, we rarely consider the fact that the couch is still next to the wall since that arrangement has been embedded in our memory. Conversely, coming home to find the couch on the other side of the room would be surprising and violate our expectations for how the room should look based on our memory.

From a computational perspective, it would be efficient to dedicate more cognitive resources when an experience is novel or if it violates our expectations [62,75–79]. This is because identical experiences contain no new information whereas novel or unexpected experiences do. This hypothesis, termed predictive coding, has been articulated in computational and theoretical accounts of brain processing and posits that neural activity is optimized to maximize information [62,65,77,80,81]. A key requirement of this hypothesis is that the expectations and predictions to which new experiences are compared are stored in memory. For simple visual and auditory stimuli, sensory expectations are learned through a lifetime of observing and processing the statistical regularities of the natural world and are stored as memory in the brain networks and synaptic weights of primary visual and auditory cortex [82–85]. Efficiently processing new sensory inputs involves comparing those inputs to our expectations for how the visual and auditory world should appear [62,65,77,80,81].

However, it is unclear if and how such expectations are established when considering single episodes or events that we experience. We rely upon episodic memory to encode and remember these events [86]. An important property of episodic memory is that we can form episodic memories even when the experiences we are remembering arise from just a single exposure. From this single exposure, episodic memories may enable us to form expectations of present and future experience. When initially exposed to an experience, interactions between the neocortex and MTL store a distributed representation of experience [54,87–92]. When presented with a similar experience, the previous experience can be retrieved through autoassociative reactivation of these neocortical-MTL representations [53,54,59,93,94]. This framework therefore provides an internally generated memory to which new experiences may be compared. Any difference between past and present experience should violate the expectations set by our episodic memory and therefore signal an error in the predictions we had established for the new experience.

Here we examine whether comparing a new experience to an episodic memory indeed evokes a prediction error signal, and the neural mechanisms that underlie this process in the human brain. We were specifically interested in how such error signals are represented in the neocortex and the MTL, as the interactions between them underlie our ability to encode and retrieve episodic memories [53,58]. We presented participants with images of natural scenes and objects that they encoded into memory. We then tested their memory for these images by presenting them with the same images. We manipulated some of the images to either remove objects from or add objects to the original scene during the testing phase. We explicitly asked participants to determine if the scene had been manipulated. Successfully recognizing this manipulation therefore requires participants to not only retrieve the past visual experience but to also then compare the retrieved

memory with the present image. We examined changes in intracranial EEG (iEEG), captured through subdural electrodes implanted for seizure monitoring, and how these changes were temporally related to eye movements that participants made as they scanned the new scenes during recognition testing. We found that recognizing manipulated images, and therefore successfully identifying a difference between past and present experience, evoked a narrowband high frequency prediction error signal in visual association cortex that then propagated towards the MTL. During successful recognition of these manipulations, this error signal was also accompanied by elevated low frequency coherence between the neocortex and MTL. These results therefore provide a direct account of how violations of the expectations set by the episodic memory of a previous experience are encoded in the human brain.

2.2 Materials and Methods

2.2.1 Participants

14 participants (7 male; 40.9 ± 12 years) participants with drug resistant epilepsy underwent a surgical procedure in which platinum recording contacts were implanted subdurally on the cortical surface as well as within the brain parenchyma. In each case, the clinical team determined the placement of the contacts to localize epileptogenic regions. The Institutional Review Board (IRB) approved the research protocol, and informed consent was obtained from the participants and their guardians. All analyses were performed using custom built Matlab code (Natick, MA).

2.2.2 Visuospatial Recognition Memory Task

Participants performed a visuospatial recognition task (Fig. 2a). During the encoding portion of the task, we sequentially presented participants a set of four color images. Each image was a natural scene containing between one to five items, such as an animal, person, or object, and we instructed the participants to remember each scene. We presented each image for five seconds, and a white fixation cross appeared for one second before each image. Immediately following the list of four images, we then presented the same images during the recognition testing portion of the task and tested the participants on their memory for the images. The images during recognition testing were presented in the same order. Critically, we manipulated some of the images presented during recognition testing by either adding or removing an item from the scene. We therefore presented three different types of images during recognition testing - added, removed, or repeated - depending on whether a manipulation had been performed and the type of manipulation. For most analyses, we considered added and removed trials together as the manipulated condition. We selected images from an open access database of images from Pixabay (Munchen, Germany) and Pexel (Fuldabruck, Germany). We used Adobe Photoshop (San Jose, CA) to remove an item from each of the original images. We used the image with the removed item as the image to be tested during recognition testing for the removed condition, and as the image to be remembered during the encoding period for the added condition.

During recognition testing, participants viewed the images until they made their response. They indicated whether the image was the same or changed using left and right arrow keys. We divided the recognition trials into manipulated correct, manipulated incorrect, and repeated correct trials depending on whether there was a manipulation of the image and on the participant's response. If the participant indicated the image was manipulated, we then presented a mouse cursor at the

center of the screen and instructed the participant to identify the location of the manipulated item using a mouse click. We removed all trials in which the response time for identifying the location was greater than 10 seconds. For each image, we defined a critical region as a rectangular region around each item. Based on whether the mouse click fell within this rectangular region, we therefore determined if the participant was able to correctly identify the location of the manipulated item (Fig. 2d). Participants completed one to two sessions during the monitoring period. Each session was approximately an hour of testing and contained 60 lists of images, where each list contained the sequential presentation of four images during encoding and the same (or manipulated) four images during recognition testing. Participants completed a total of 784 ± 195 trials during the monitoring period.

2.2.3 Intracranial EEG (iEEG) Recordings

Depending on the amplifier and the discretion of the clinical team, intracranial EEG (iEEG) signals were sampled at 1000 or 2000 Hz. For clinical visual inspection of the recording, signals were referenced to a common contact placed subcutaneously, on the scalp, or on the mastoid process. The recorded raw iEEG signals used for analyses were referenced to the system hardware reference, which was set by the recording amplifier (Nihon Kohden, Irvine CA) as the average of two intracranial electrode channels. We re-referenced these raw signals using bipolar referencing (see below) in order to mitigate any effects of volume conduction or any biases introduced by the system hardware reference. All recorded traces were resampled at 1000 Hz, and a fourth order 2 Hz stopband butterworth notch filter was applied at 60 Hz to eliminate electrical line noise.

We collected electrophysiological data from a total of 1716 subdural and depth recording contacts (122 ± 5.9 per participant; PMT Corporation, Chanhassen, MN). Subdural contacts were arranged in both grid and strip configurations with an inter-contact spacing of 10 mm. Contact localization

was accomplished by co-registering the post-op CTs with the post-op MRIs using both FSL Brain Extraction Tool (BET) and FLIRT software packages and mapped to both MNI and Talairach space using an indirect stereotactic technique and OsiriX Imaging Software DICOM viewer package. The resulting contact locations were subsequently projected to the cortical surface of a Montreal Neurological Institute N27 standard brain [95]. Pre-operative MRIs were used when postoperative MR images were not available.

We divided projected electrode contacts into four regions of interest based on their location relative to the Desikan-Killiany atlas[96]: lateral occipital (LOC), parietal (PAR), posterior temporal (PT), and medial temporal lobe (MTL). We assigned all electrodes with locations in the lateral occipital complex as LOC electrodes, and all electrode contacts in the superior and inferior parietal lobe as PAR electrodes. We identified all electrodes with locations in the posterior part of the superior, middle, or inferior lateral temporal cortex or that lay over the posterior aspect of the fusiform gyrus as PT electrodes [97]. We designated all depth electrode contacts within the hippocampus and all subdural contacts that lay along MTL structures, including the parahippocampal gyrus and entorhinal cortex, and that were medial to the collateral sulcus as MTL electrodes.

We analyzed iEEG data using bipolar referencing to reduce volume conduction and spurious signals introduced by the system reference. The choice of referencing largely depends on the assumptions regarding the spatial distribution of the signal of interest [98]. Bipolar referencing offers a practical approach for examining local events with spatial distributions that are smaller than the inter-electrode distance of our recordings since it will filter out activity at the larger spatial scale that is common to both electrodes that may be introduced by the system reference or that may

result from volume conduction. In addition, because each neighboring bipolar channel records activity from similar brain regions through similar electrode contacts, bipolar referencing ensures that any referencing that is applied to each recorded iEEG trace is performed using a reference electrode that shares similar impedance and noise profiles. Finally, bipolar referencing has also been noted to be superior to the average reference montage in reducing muscular artifacts in iEEG [99]. We defined the bipolar montage in our data set based on the geometry of iEEG electrode arrangements. For every grid and strip, we isolated all pairs of contacts that were positioned immediately adjacent to one another. Bipolar signals were then calculated by finding the difference in the signal between each pair of immediately adjacent contacts. The resulting bipolar signals were treated as new virtual electrodes (henceforth referred to as electrodes throughout the text), originating from the midpoint between each contact pair. All subsequent analyses were performed using these derived bipolar signals.

High frequency activity can be associated with epileptiform activity in addition to cognitive processes. Therefore we implemented several measures to provide the most conservative sampling of non-pathological signals possible. We implemented a previously reported automated trial and electrode rejection procedure based on excessive kurtosis or variance of iEEG signals [97]. We calculated and sorted the mean iEEG voltage across all trials, and divided the distribution into quartiles. We identified trial outliers by setting a threshold, Q3+w*(Q3-Q1), where Q1 and Q3 are the mean voltage boundaries of the first and third quartiles, respectively. We empirically determined the weight w to be 2.3. We excluded all trials with mean voltage that exceeded this threshold. The average percent removed across all sessions in each participant due to either system-

level noise or transient epileptiform activity was 1.7 ± 0.2% of all electrodes and 2.8 ± 0.1% of all trials.

In addition to system level line noise, eye-blink artifacts, sharp transients, and inter-ictal discharges (IEDs) can confound the interpretation of our results. We therefore implemented a previously reported automated event-level artifact rejection [100]. We calculated a z-score for every iEEG time point based on the gradient (first derivative) and amplitude after applying a 250 Hz high pass filter (for identification of epileptogenic spikes). Any time point that exceeded a z-score of 5 with either gradient or high frequency amplitude was marked as artifactual, and 100 ms before and after each identified time point was also classified as an artifact. We visually inspected the resulting iEEG traces and found that the automated procedure reliably removed IEDs and other artifacts. In total, following bipolar referencing and exclusion of electrodes because of artifact, our pre-processed data set consisted of 1716 bipolar electrodes (123 ± 5.9 per participant).

2.2.4 Eye Movement and Fixation Detection

In a subset of participants (n=8), we tracked the locations of their gaze on the screen using a Tobii X3-120 EyeTracker (Stockholm, Sweden) that sampled eye movements at 120 Hz. At the start of each session, participants performed a calibration procedure in order to convert eye rotations into a set of gaze positions relative to the screen. For each participant, we extracted raw eye movement data during the experimental session and converted the movements into a set of fixations using a dispersion-based algorithm [101]. We defined each fixation point as a point on the screen upon which gaze continually remained within 2 degrees of visual angle for a period of 100 ms or more. We excluded the remaining six participants from the analysis due to the inability to calibrate the participants or collect eye movement data due to clinical constraints.

2.2.5 Spectral Power

We quantified spectral power and phase by convolving iEEG signals with complex valued Morlet wavelets (wavelet number 6) [102]. We extracted data from all encoding and recognition trials, beginning with the presentation of the image on the screen until the image was removed during encoding or until the response during recognition testing and localization, for our analyses. In all trials, we included a 1000 ms buffer on both sides of the clipped data. To generate corresponding power spectrograms, we calculated spectral power using 32 logarithmically spaced wavelets between 2 and 431 Hz. We then squared and log-transformed the continuous-time wavelet transform to generate a continuous measure of instantaneous power. To account for changes in power across experimental sessions, we z-scored power values separately for each frequency and for each session using the mean and SD of all respective values for that session. We binned the continuous time z-scored power for each frequency into 200 ms epochs spaced every 100 ms (50% overlap) and averaged the instantaneous power over each epoch, and performed subsequent analyses on these binned values.

2.2.6 Spectral Coherence

We computed the magnitude squared spectral coherence between every electrode pair using one second temporal epochs during the recognition period (MATLAB function 'mscohere') [103,104]. We computed the coherence between individual electrode contacts, rather than between bipolar virtual contacts, since bipolar referencing has been shown to remove low frequency coherence between iEEG electrodes [105]. In this case, before computing coherence between any electrode pair, we re-referenced the signal from each electrode to a global common average in order to eliminate common mode signals that would arise from the system level reference or from artifacts. We

calculated the coherence between two time series, x(t) and y(t), in two electrode contacts as a function of frequency:

$$C(f) = \frac{|Pxy(f)|^2}{Pxx(f)Pyy(f)}$$

where P_{XX} and P_{YY} are the power spectral densities and P_{XY} is the cross-spectral density. We generated a coherence spectrum for each temporal epoch, frequency, electrode pair and trial. We then z-scored coherence values separately for each electrode pair using the mean and SD of all coherence values for the session.

2.2.7 Generation and Characterization of Cross-Correlograms

We computed a cross-correlation of the spectral power time series between pairs of electrodes spanning PT and MTL and spanning PAR and MTL in order to examine the temporal relation of high frequency activity between visual association cortex and MTL. We used the first second of data following image presentation during recognition testing for this analysis. In each electrode, we extracted the continuous time series of spectral power and divided this trace into non-overlapping 10 ms bins by averaging the time series over each bin. We used these 10 ms bins to compute the cross-correlations for computational efficiency and to generate more temporally smoothed representations of the cross-correlations between electrode pairs. For every electrode pair, in each trial, we computed the time-lagged cross-correlation between the time series of binned values. We then averaged these cross-correlations across trials, thus generating a true cross-correlogram for each pair of electrodes in each participant that we can compare to a chance distribution and that we can use to identify the time lag of high frequency activity between the electrode pair.

We generated a chance cross-correlogram for each electrode pair characterizing the baseline cross-correlation that would be expected by chance given the presentation of a stimulus and to which the true correlogram could be compared [94,106,107]. For every pair of electrodes, we generated this chance distribution by computing the cross-correlation of the power time series of one electrode during a randomly chosen individual trial with the time series of the other electrode from another randomly chosen trial. We repeated this procedure 100 times, and averaged across all permutations to generate an average chance cross-correlogram for that electrode pair. The difference between the true cross-correlogram and the chance cross-correlogram reflects the extent to which two signals are cross-correlated greater than chance given the presentation of a stimulus.

To assess significant coupling for a single electrode pair, we compared the true distribution of cross-correlation values between -50 ms to 50 ms to the chance distribution in this same window using a paired t-test. To assess significant coupling between two regions across participants, we first averaged the true and chance cross-correlation values over this window for each electrode pair, and then computed the average difference between the true and chance cross correlograms across all electrode pairs between two regions for a single participant. We then compared the distribution of these average differences across participants to 0 to assess significance ($p < .05$, paired t-test). To determine the relative timing of high frequency power between two regions, we identified the peak time of each correlogram for every pair of electrodes between two regions. We computed the average peak time across electrode pairs between the two regions for each participant, and assessed whether the distribution of average peak times across participants was significantly different than zero ($p < .05$, paired t-test).

2.2.8 Metrics of Reinstatement

To quantify reinstatement of representations during the recognition period, we conducted a representational similarity analysis using methods described previously [93,108]. Briefly, we binned the continuous time z-scored power for each frequency into 200 ms epochs spaced every 100 ms (50% overlap) and averaged the instantaneous power over each epoch. For each temporal epoch, we subsequently averaged the z-scored power across bins within two frequency bands contained within the 80-120 Hz band. For every temporal epoch in each trial, we constructed a feature vector composed of the average z-scored power for every electrode within a given region of interest and for every frequency band. For each encoding temporal epoch, i, and for each retrieval temporal epoch, j, we define feature vectors as follows:

$$E_i = [z_{1,1}(i) \dots z_{1,F}(i) \dots z_{L,F}(i)]$$

$$R_j = [z_{1,1}(j) \dots z_{1,F}(j) \dots z_{L,F}(j)]$$

where $z_{l,f}(i)$ is the z-scored power of electrode $l = 1 \dots L$ at frequency band $f = 1 \dots F$ in temporal epoch i. For L electrodes and F frequency bands, we thus create a feature vector at each temporal epoch that contains $K = L*F$ features, which represents the distributed spectral power across all electrodes and across the two frequency bands at a single moment in time.

To quantify reinstatement during trial n, we calculated the cosine similarity between encoding and recognition feature vectors E_i and R_j for all pairs of encoding and recognition temporal epochs during that trial. Cosine similarity gives a measure of how close the angles of two vectors are in a multidimensional space. We chose cosine similarity over Pearson's correlation to measure

reinstatement because if all of the elements of two feature vectors show increases in power from baseline, with small additional random noise, then these two vectors should have high measured reinstatement. Pearson's correlation, a centered version of cosine similarity, would give a low correlation in this case because of the noise fluctuations, whereas the cosine similarity would be high, consistent with our interpretation of reinstatement. Thus, for each trial, *n*, we generate a temporal map of reinstatement values:

$$C_n(i,j) = \frac{E_i * R_j}{||E_i||\,||R_j||}$$

where $C_n(i,j)$ corresponds to the reinstatement of neural activity across all electrodes and all frequencies between encoding epoch i and retrieval epoch j during trial n. We computed the reinstatement maps separately for all trials for each participant and averaged the reinstatement maps across participants.

2.2.9 Temporal Dynamics of Spectral Power

We compared the peak times of the spectral power time series for LOC, PT, PAR, and MTL in order to examine the temporal relation of high frequency activity through the visual hierarchy. We identified the peak time for a particular region by calculating the time during which the average spectral power time series across participants reached its maximum value. We then computed the difference in the peak times between two regions to identify the temporal relation of spectral power between them. To assess if this difference was significant, we generated a chance distribution to which the true difference in peak times could be compared. We computed this chance distribution

by randomly switching the spectral power time series for one region with the spectral power time series of the other region in each participant. Hence, in each permutation, some participants would retain their original power time series traces in their original regions, and some participants would have the labels for the regions randomly switched. We then averaged these shuffled spectral power time across participants for each region and then computed the difference in peak times between the two regions in each permutation. We repeated this procedure 1000 times to generate a shuffled distribution of differences in peak times. We assigned p-values that characterize the difference in peak times between any two brain regions by comparing the true difference in peak times to the shuffled distribution of differences.

To estimate how quickly high frequency activity increased in each brain region and how this compared across LOC, PT, PAR and MTL, we computed the instantaneous slope of the increases we observed in the time series of high frequency spectral power. We computed the difference in spectral power between adjacent time bins (200 ms overlapping bins incremented by 100 ms) and then averaged these estimates of instantaneous slopes across all time points within the first 500 ms after image presentation. Within each brain region in each participant, we computed the average instantaneous slope across all visually responsive electrodes. We compared the distribution of average values across participants between two brain regions using an unpaired t-test ($p < .05$) in order to assess whether the rise in high frequency activity was different between the regions across participants.

To determine whether the differences in high frequency 80-120 Hz spectral power that we observed between conditions arose at different times in different brain regions, we performed two

analyses. In both cases, we explicitly generated a time course of the average difference between conditions that showed any significant difference between conditions in each brain region in each participant. In the first analysis, we identified the time points that exhibited the first significant difference in spectral power between conditions in each electrode within a region. We then averaged these first time points across all significant electrodes within each region in each participant. We compared the distribution of these time points of first differences across participants between brain regions (unpaired t-test, $p < .05$). In the second analysis, we used the rise in the average time series across all significant electrodes of the differences in spectral power to estimate the first time point when this difference deviates from zero and to estimate the time point when the increase in high frequency power reached 50% of its peak. We used this approach to generate a more temporally precise estimate of when this signal first increased above baseline since in our main analysis we generated the time series using overlapping 200 ms bins incremented every 100 ms. To estimate this initial time of deviation, we identified the time point of the peak difference between conditions and the time point of the local minimum that immediately preceded the peak difference. We then fit a line to these two time points and identified the time point when that line intersected with zero. We designated this as the time point at which the difference between conditions first deviates from baseline. We compared the distribution of these first significant time points across participants between brain regions using an unpaired t-test. We similarly identified when the rise of spectral power reached 50% of the peak and compared the distribution of 50% time points across participants between each brain region. Finally, we compared the estimated time points at which we first observed a rise in the difference in high frequency power between conditions to the time points at which we observed overall increases in high frequency power in the MTL across conditions. In a similar manner, we used the average time series of spectral power

across significant MTL electrodes to estimate the first time point when overall 80-120 Hz power deviated from baseline in the MTL.

2.2.10 Statistical Analyses

We employed a non-parametric clustering based procedure to identify significant time, frequency, or time frequency epochs for differences in power, coherence, and reinstatement between conditions (Fig. 2d-f, 3a-d, 5b-c, 7b, Supplementary Fig. 10-13)[109] . The procedures for all analyses were identical with the exception that clusters identified for coherence and reinstatement analysis were generated across the two dimensions of encoding and retrieval time. The clustering procedure identifies contiguous temporal or time-frequency clusters exhibiting significant differences between two conditions (e.g. manipulated correct and repeated correct), with the null hypothesis that across participants, each epoch showed no difference between the conditions. For each time or time-frequency window, we computed the true t-statistic and p-value across participants between the two conditions by comparing the distribution of average values across all visually responsive electrode within each brain region across participants. The p-value for each individual time point or time-frequency window in the true case, however, does not take into account the multiple comparisons that are made in across time points.

To correct for multiple comparisons across time points, we randomly permuted the participant-specific averages between the two conditions. In practice, this translates to randomly reversing the sign of the difference within each participant and recomputing the mean difference across participants. For n participants, this results in an empiric distribution of 2^n possible mean differences that are all equally probable under the null hypothesis. We generated the empiric distribution from 1000 permutations for every time point and calculated t-statistics for each time

point in each permutation. We identified clusters containing time points or time-frequency windows that were adjacent in time (or in time-frequency space) that exhibited a significant difference between trial types (where in each time point, $p < .05$ unless specified otherwise) in both the true case and in each permutation. For each cluster of significant time points identified in the true and permuted cases, we defined a cluster statistic as the sum of the t-statistics within that temporal cluster. We retained the maximum cluster statistic during each of the 1000 permutations to create a distribution of maximum cluster statistics. We assigned p-values to each identified cluster of the true data by comparing its cluster statistic to the distribution of maximum cluster statistics from the permuted cases. Clusters were determined to be significant if their p-value calculated in this manner was less than .05.

We used a similar procedure to identify electrodes showing a significant difference between conditions (Fig. 4b,c Supplementary Fig. 9). The clustering procedure identifies contiguous temporal clusters exhibiting significant differences between the two conditions with the null hypothesis that across trials, each epoch showed no differences between the participants. In this case, we created a permuted distribution by randomly switching the condition of one trial with the condition of another and then computing the mean difference between the conditions. We repeated this procedure 1000 times to create an empirical distribution and compared the true difference to the empiric distribution of mean difference. We then identified the maximum cluster statistic of the true data and compared to the cluster statistic of the permuted distribution to correct for multiple comparisons. Clusters were determined to be significant if their p-value calculated in this manner was less than .05.

2.3 Results

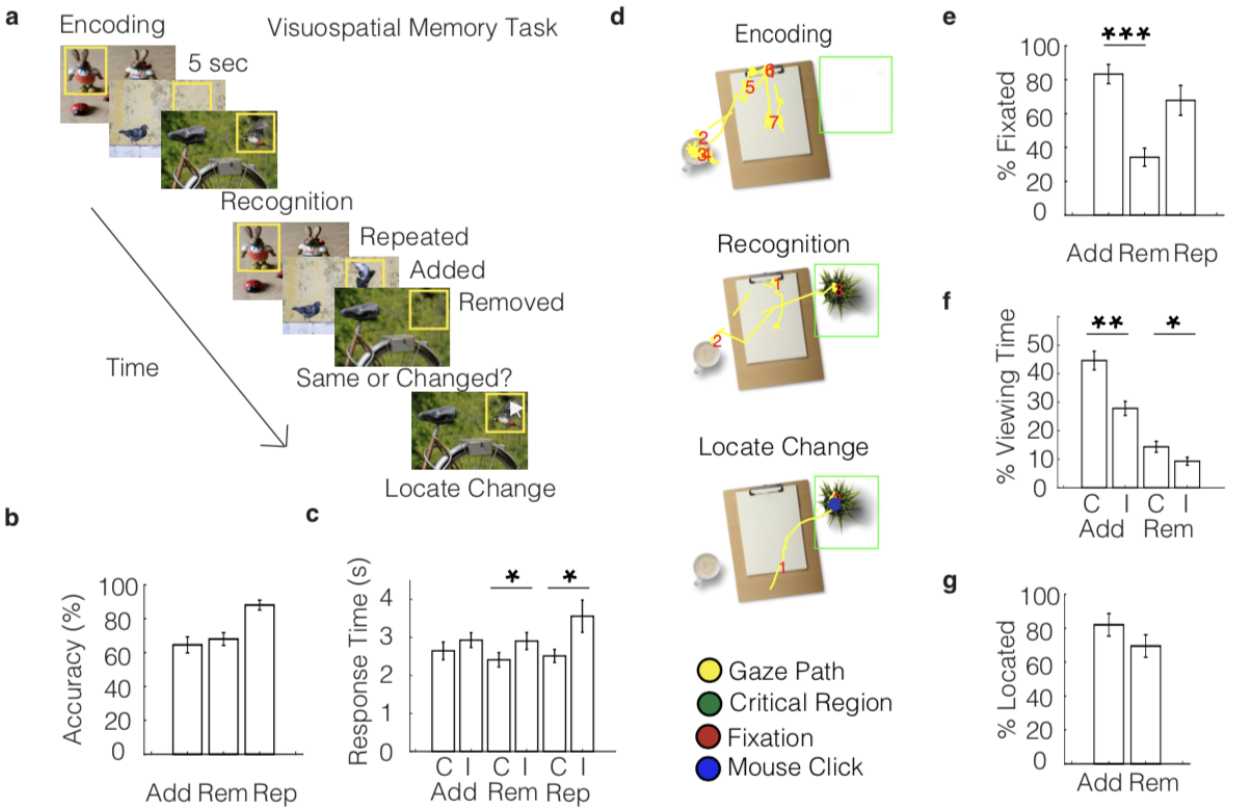

Figure 2. Visuospatial Memory Task Behavior Performance.

a) Participants viewed a set of images followed by the same set of images containing either a removed, added, or repeated object. Participants indicated whether the image was the same or changed. If changed was selected, participants identified the location of the change using a mouse click**. b)** Probability of recognition for the added, removed, and repeated conditions across participants. **c)** Mean response times were compared for correct and incorrect responses for added, removed, and repeated conditions across participants. **d)** Representative eye-movement recordings during the encoding, recognition, and identification of the manipulation. The gaze path (yellow) was converted into fixations (red) and the percentage of fixations within the critical region (green) was calculated for each condition. The location of the mouse click (blue) was used to determine the number of manipulated or repeated objects that were successfully located. **e)** Probability of fixation within critical region for correctly recognized added, removed, and repeated objects across

participants. The critical region for the condition was selected to be a random object within the image to assess the probability of baseline fixations on an object **f)** Probability of viewing time within critical region for correctly and incorrectly identified images. **g)** Probability of mouse click within critical region for correctly recognized added and removed conditions across participants. All error bars indicate standard error of mean. Asterisks (*,**,***) indicate significance at $p < 0.05$, $p < 0.01$, and $p < 0.001$. using two-sided paired t-test.

14 participants (7 male, 40.9 ± 12 years) with intracranial electrodes placed for seizure monitoring performed a visuospatial recognition memory task (Fig. 2a; see Methods). During the encoding portion of the task, we presented images of natural scenes containing different items to participants and instructed them to remember the images. We then subsequently presented the same images during the recognition phase of the task and tested their memory for the images. We either added or removed an item from some of the images that were presented during the recognition phase and instructed the participants to indicate whether each image was identical to the one they had encoded or if it had been manipulated. We therefore designated the three different types of images presented during recognition testing as repeated, added, or removed versions of the original images based on the manipulation we performed. If the participant indicated that the image had been manipulated, we then instructed them to identify the location of the manipulation using a mouse click on the screen (Fig. 2a).

Participants successfully recognized 65 ± 5, 68 ± 4, and 88 ± 3% of the added, removed, and repeated images, respectively, during testing with a mean response time of 2.65±0.23, 2.41±0.19, and 2.51±0.17s (Fig. 2b,c). Response times were significantly faster when participants correctly recognized removed and repeated images compared to when they were incorrect (removed, $t(13) = -2.31$, $p = .038$, paired t-test; repeated $t(12) = -2.85$, $p = .014$) but not when comparing correct and incorrect added images ($t(13) = -1.36$, $p = .195$).

In a subset of participants (n = 8), we recorded the location of each participant's gaze as they scanned the images during recognition testing (Fig. 2d; see Methods). Participants spent at least one fixation on 83 ± 4% of successfully recognized added items, but on only 34±4% of the regions in which items were removed when successfully recognizing a removed image ($t(7) = 11.2$, $p = .00001$; Fig. 2e). However, participants spent a greater percentage of time viewing both the added

item and the region of the removed item during correct compared to incorrect trials (added, $t(7) = 4.93$, $p = .002$; removed, $t(7) = 2.52$, $p = .040$, paired t-test; Fig. 2f). In addition, participants were able to correctly identify the location of the manipulation when successfully recognizing that items had been added or removed in 82 ± 6 and 70 ± 6% of the trials, respectively (Fig. 2g). These data together suggest that participants were able to successfully recognize when and where an image was manipulated even though they were more likely to explicitly fixate on the manipulation only when an item was added.

2.3.1 80-120 Hz power progresses down the visual hierarchy and reflects specific visual experience

We examined intracranial EEG (iEEG) recordings captured from intracranial electrode contacts in all participants as they performed the visuospatial recognition task (Fig. 3a; see Methods). In a representative example electrode in the posterior temporal cortex, we found a narrowband increase in 80-120 Hz power that was reliably detected across all trials during the recognition period (Fig. 3b). We divided electrode contacts in each participant into four regions of interest - lateral occipital cortex (LOC), parietal cortex (PAR), posterior temporal cortex (PT), and the medial temporal lobe (MTL) - based on the known feedforward organization within the visual hierarchy [51,52] (Supplementary Fig. 8). In each of these regions across participants, we observed a consistent increase in high frequency power centered around 80-120 Hz when averaged across all trials during the recognition period (Fig. 3c). We focused subsequent analyses on changes in spectral power within this narrowband of high frequency activity.

When examining all recognition trials, we found that 80-120 Hz power exhibited a significant increase above baseline within 200 ms after image presentation in all regions ($p < .001$,

permutation procedure; see Methods; Fig. 3d). This rise in 80-120 Hz peaked within 600 ms in all regions, but peaked at a significantly earlier time following image presentation in LOC and PT compared to MTL ($p < .05$, permutation procedure). The rise in 80-120 Hz activity was significantly faster in LOC, PAR, and PT compared to MTL (LOC v MTL, $t(14) = 2.26$, $p = .04$; PAR v MTL, $t(16) = 3.14$, $p = .006$; PT v MTL $t(17) = 2.84$, $p = .011$; see Methods), suggesting that image presentation during the recognition period evokes a rise in 80-120 Hz power that progresses down the visual hierarchy.

Given that the images presented during recognition testing were similar, but not identical, to the images presented during encoding, we investigated whether the patterns of 80-120 Hz power that arose along the visual hierarchy following image presentation during recognition testing were also similar to the patterns of 80-120 Hz power present during encoding. For each image presented during encoding and during retrieval, we constructed a distributed representation of 80-120 Hz power across all electrode contacts at each time point following image presentation. We computed how similar the distributed representation of power at each time point during recognition testing to the representation present at every time point when encoding the same image (see Methods). We found that viewing the same image, regardless of whether the image was manipulated, reinstated the distributed pattern of 80-120 Hz power that was present during encoding across participants (Fig. 3e).

We compared the true reinstatement of these distributed patterns of 80-120 Hz power to the reinstatement observed after shuffling the trial labels in order to assess whether this reinstatement was specific to each individual image (see Methods). Viewing the images during

recognition testing significantly reinstated the specific distributed patterns of activity for each image as compared to the shuffled trials beginning 100 ms after the image presentation ($p < .001$, permutation test; Fig. 3e). We then computed a time series of the mean level of reinstatement during recognition testing within each region by using all encoding epochs that demonstrated item-specific reinstatement when considering all electrodes. Across participants, we found that viewing the image during recognition testing resulted in significant item-specific reinstatement in the LOC, PT, PAR, and MTL ($p < .05$, permutation test; Fig. 3e), suggesting that the progression of high frequency power down the visual hierarchy contains information regarding the specific image being viewed.

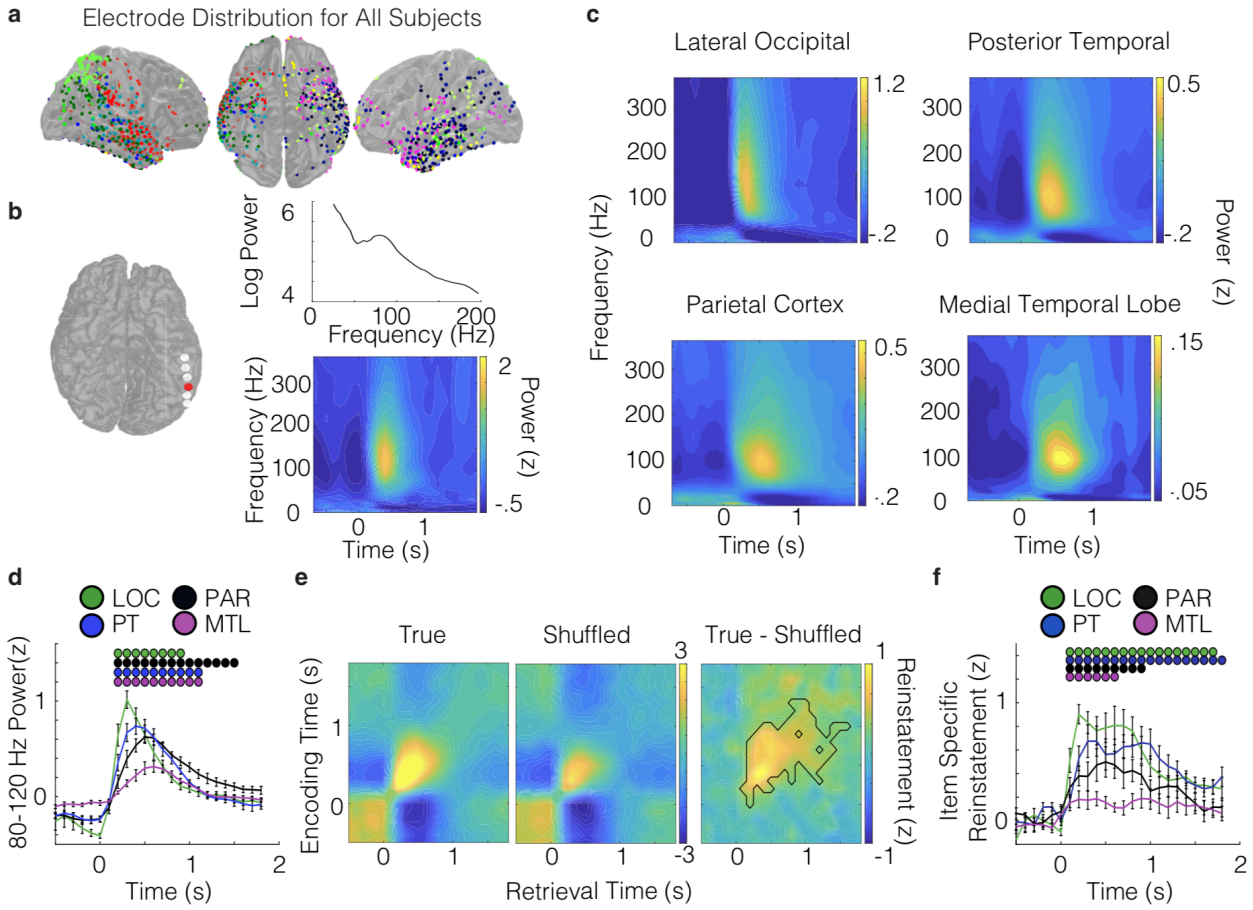

Figure 3. Feedforward transmission of 80-120 Hz power between neocortical-MTL circuits reflect specific visual experiences

a) Intracranial electrode locations for all 14 participants. Each color corresponds to an individual participant. **b)** In a representative electrode in PT, narrowband increases in power in the 80-120 Hz band were observed at the single trial level. Across all trials in this electrode, images during the recognition period elicited increases in power at this frequency as indicated by the warmer colors. **c)** Image presentation during the recognition period elicited increases in 80-120 Hz power across participants in LOC, PT, PAR, and MTL. **d)** Average 80-120 Hz power during the recognition period in lateral occipital complex (LOC), posterior temporal (PT), parietal cortex (PAR), and medial temporal lobe (MTL) across participants. Time courses were generated with a 200 ms sliding window in 100 ms steps (50% overlap) locked the second viewing of the image.

Dots indicate significant increases in power compared to average session power that survived corrections for multiple comparisons **e)** Average observed and shuffled reinstatement across all participants. The difference in average reinstatement between the observed and chance reinstatement reflects item-specific reinstatement across all participants. The black outline constitutes all epochs that exhibited significant differences between the observed and chance representations that survived corrections for multiple comparisons **f)** Average item-specific reinstatement in LOC, PAR, PT, and MTL across participants. Dots indicate significant increases in item-specific reinstatement that survived corrections for multiple comparisons. All error bars indicate standard error of mean. All multiple comparisons corrections were performed using a non-parametric permutation procedure.

2.3.2 80-120 Hz power increases within visual association areas and the MTL when present visual experience differs from the remembered experience

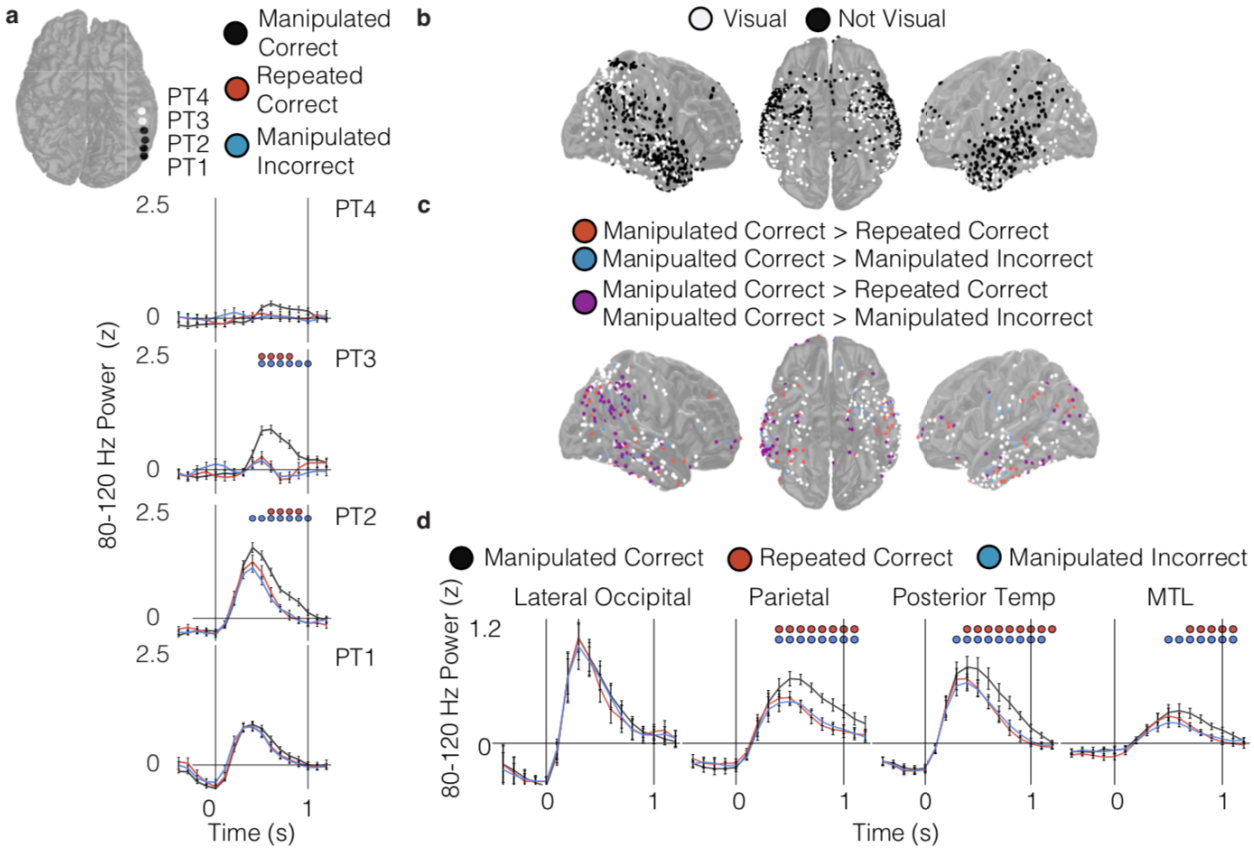

Figure 4. Visual association areas and MTL circuits increase in 80-120 Hz power when present and past visual experience are in violation

a) Average 80-120 Hz power for a set of posterior temporal electrodes across manipulated correct (black), manipulated incorrect (blue), and repeated correct (red) trials. Time courses were generated with a 200 ms sliding window in 100 ms steps (50% overlap) locked the second viewing of the image. Blue and red dots indicate significant increases in power for the manipulated correct compared to the manipulated incorrect and repeated correct condition, respectively, that survived corrections for multiple comparisons. **b)** Visualization of electrodes showing significant increases

in 80-120 Hz power from baseline (white) that survived corrections for multiple comparison. **c)**Visually responsive electrodes (white) that showed significant increases in 80-120 Hz power for the manipulated correct compared to the manipulated incorrect (blue), manipulated correct compared to the manipulated repeated (red), or both (purple) that survived corrections for multiple comparisons. **d)** Time courses of 80-120 Hz power across participants for LOC, PAR, PT, and MTL for the manipulated correct, manipulated incorrect, and repeated correct conditions locked to the second viewing of the image. All error bars indicate standard error of mean. All multiple comparisons corrections were performed using a non-parametric permutation procedure.

Participants in our task were able to correctly identify when a visual image was manipulated compared to the image they had remembered. We were interested in examining the neural mechanisms underlying this ability to recognize the difference between past and present visual experience. We therefore compared the item-specific increases in 80-120 Hz power that progress down the visual hierarchy between conditions. In an example participant, we examined a set of electrodes arranged in linear strip from posterior to anterior regions of the PT. Viewing the manipulated images resulted in a significantly greater and more prolonged increase in 80-120 Hz power in individual electrode contacts when compared to viewing a repeated image or viewing an image that had been manipulated but that was incorrectly identified as being repeated ($p < .001$, permutation test; see Methods; Fig. 4a). This difference was specific to two electrode contacts and was not present on the most posterior contact and was more attenuated in the anterior contact, suggesting that correctly identifying a difference between a presented and remembered image results in difference in 80-120 Hz power only within specific regions of the visual association cortex.

We examined the changes in 80-120 Hz between conditions in all electrode contacts in all participants. We first identified any electrode contact that demonstrated a significant difference in 80-120 Hz power at any point during recognition testing compared to baseline when averaged across all trials (Fig. 4b). We found that 43% of electrode contacts across participants exhibited a significant increase in 80-120 Hz power from baseline ($p < .05$, permutation test; see Methods; 17% of electrode contacts showed a significant decrease). We designated these electrode contacts as visually responsive. Visually responsive electrode contacts were primarily located in LOC, PAR, PT, and MTL and were relatively absent from the anterior lateral temporal cortex. We then

investigated how activity within these visually responsive electrodes changed between conditions. We found that 35% and 31% of visually responsive electrode contacts demonstrated significantly greater 80-120 Hz power at some point when viewing manipulated images that were correctly identified as compared to repeated images and as compared to images that were not correctly identified as manipulated, respectively ($p < .05$, permutation test; Fig. 4c, red and blue electrodes respectively). 21% of the electrodes exhibited a significant increase in both comparisons ($p < .05$, permutation test; Fig. 4c, purple electrodes). In contrast, less than 5% of electrode contacts demonstrated a significant decrease in 80-120 Hz power in any of these comparisons ($p < .05$, permutation test; Supplementary Fig. 9).

Although these data suggest that a large subset of electrode exhibit significantly greater 80-120 Hz power at some time point when viewing manipulated images, we were specifically interested in understanding the time course of these changes. We therefore examined the average time series of 80-120 Hz power in each condition across all visually responsive electrode contacts in each region. Across participants, as in the example set of electrodes, we found that electrodes within PT, PAR, and MTL exhibited significantly higher and more prolonged 80-120 Hz power when viewing and correctly identifying manipulated images than when viewing repeated or incorrectly identified manipulated images ($p < .01$, permutation test; Fig. 4d). We did not observe a significant difference between conditions in the LOC ($p > .05$, permutation test). The differences observed between the conditions were specific to the 80-120 Hz frequency band (Supplementary Fig. 10) and reflected temporally discrete increases in 80-120 Hz power within each trial (Supplementary Fig. 11). Moreover, these differences were not present during the encoding period (Supplementary Fig. 12),

as overall 80-120 Hz power was enhanced when viewing the images during recognition testing compared to encoding in PT, PAR, and MTL in all three conditions (Supplementary Fig. 13).

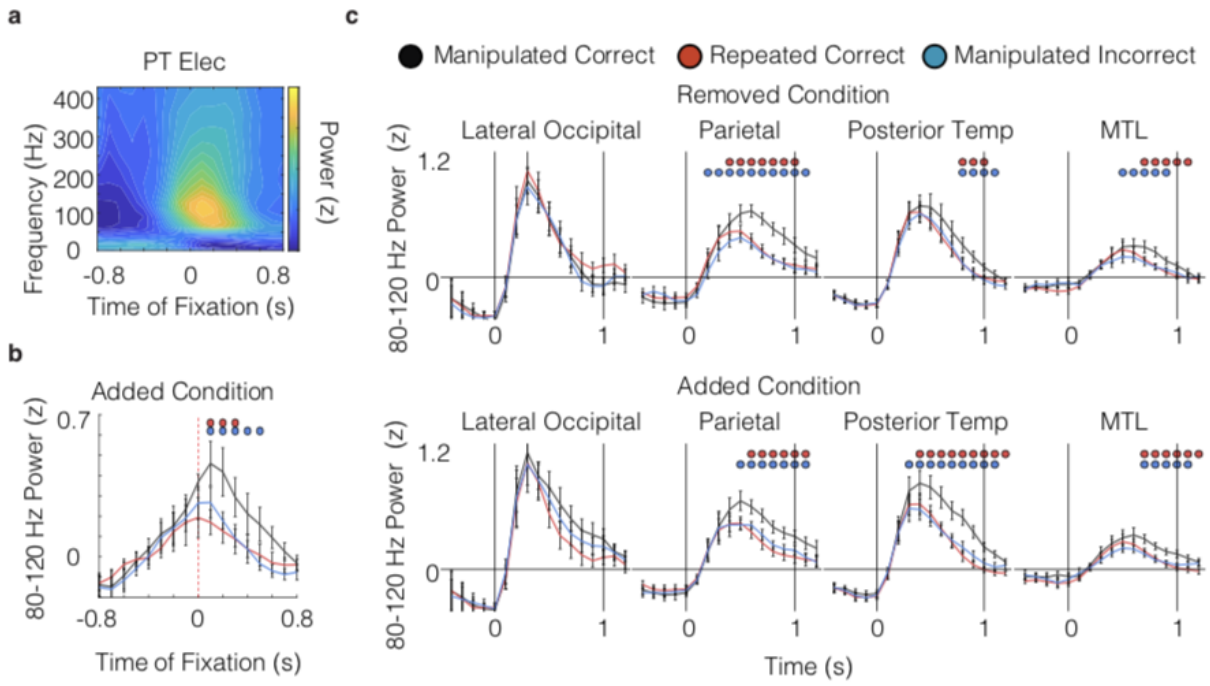

Figure 5. Visual association areas and MTL circuits increase in 80-120 Hz power when present and past visual experience are in violation for both added and removed conditions.

a) Average spectrogram relative to the first fixation on successfully recognized added objects for a representative electrode in PT. **b)** Time courses of 80-120 Hz for visually responsive electrodes across participants for the added correct, added incorrect, and repeated correct conditions locked to the first fixation on the added object. **c)** Time courses of 80-120 Hz power across participants for LOC, PAR, PT, and MTL for the removed correct, removed incorrect, and repeated correct conditions locked to the second viewing of the image. Time courses of 80-120 Hz power across participants for LOC, PAR, PT, and MTL for the added correct, added incorrect, and repeated correct conditions locked to the second viewing of the image. Blue and red dots indicate significant increases in power for the added/removed correct compared to the added/removed incorrect and repeated correct condition, respectively, that survived corrections for multiple comparisons via permutation procedure.

In a subset of participants, we recorded eye movements during recognition testing in order to determine whether the observed changes in 80-120 Hz power were temporally related to viewing the manipulated item. We focused on the added condition because participants were significantly more likely to make a fixation to the added item than to the location of the removed item in manipulated images (Fig. 3e). In a representative example, we found that the increase in 80-120 Hz power locked to the time of the fixation on the manipulated item across all trials (Fig. 5a). Across participants, we found that fixating on the added item resulted in a significantly higher level of 80-120 Hz power that was locked to the fixation during trials that were correctly identified as manipulated compared to incorrect trials and to trials with no manipulation ($p < .001$, permutation test; Fig. 5b). Hence, the observed differences in 80-120 Hz power appear to be triggered by viewing an item that was not present in a remembered visual image.

One concern regarding the differences we observed in 80-120 Hz power between manipulated and repeated images is that these differences could have been driven by the stimulus properties of the image presented during recognition testing. For example, some of the manipulated images contained items that were added to the original image, and the increases in 80-120 Hz power may simply be due to the additional visual input from the added item. To examine this possibility, we separately analyzed the trials in which manipulated images contained an added item or had an item that was removed and compared them to the repeated images and to the corresponding incorrectly identified manipulated images. Both the added and removed conditions demonstrated similar significant increases in 80-120 Hz power compared to the repeated correct and manipulated incorrect conditions ($p < .05$, permutation test; Fig. 5c). These data therefore confirm that the observed increases in 80-120 Hz power arise due to manipulation of the image and are not related to specifically how the image had changed.

2.3.3 Differences in 80-120 Hz power during manipulated images progress down the visual hierarchy

We were interested in examining the time course of this 80-120 difference signal that arises when viewing a manipulated image across brain regions. The goal was to distinguish whether this signal progresses from posterior to anterior brain regions similar to routine visual processing or whether this difference signal is first detected in higher order brain regions such as the MTL that are explicitly involved in encoding and retrieving the associations present in each image. We first examined the direction of propagation of overall 80-120 Hz power between neocortical association areas and the MTL. In an example pair of electrodes located in the PT and MTL, we computed a cross correlation of 80-120 Hz power using the first second of data following image presentation during recognition testing (Fig. 6a). Across all trials in this example pair, the cross-correlation demonstrated a clear peak that was significantly greater than chance ($p < .01$, permutation procedure; see Methods) and that 80-120 Hz power in PT preceded MTL with a consistent delay. We repeated these cross-correlations across all participants using all visually responsive electrode pairs between brain regions. Across participants, we found a significant peak in the cross-correlation of 80-120 Hz power between PT and MTL ($t(7) = 5.23$, $p = .0012$, paired t–test; see Methods; Fig. 6b) and between PAR and MTL ($t(4) = 5.0$, $p = .008$). We used the time of the peak of each cross-correlation to quantify the delay in 80-120 Hz activity between brain regions and averaged the peak delay across participants (Fig. 6c; see Methods). The increases in 80-120 Hz power in PT significantly preceded the increases in MTL by an average delay of 38 ± 9 ms across participants ($t(7) = 2.06$, $p = .005$, paired t-test). No differences in delay were observed between the PAR and MTL ($t(4) = .30$, $p > .05$, paired t-test).

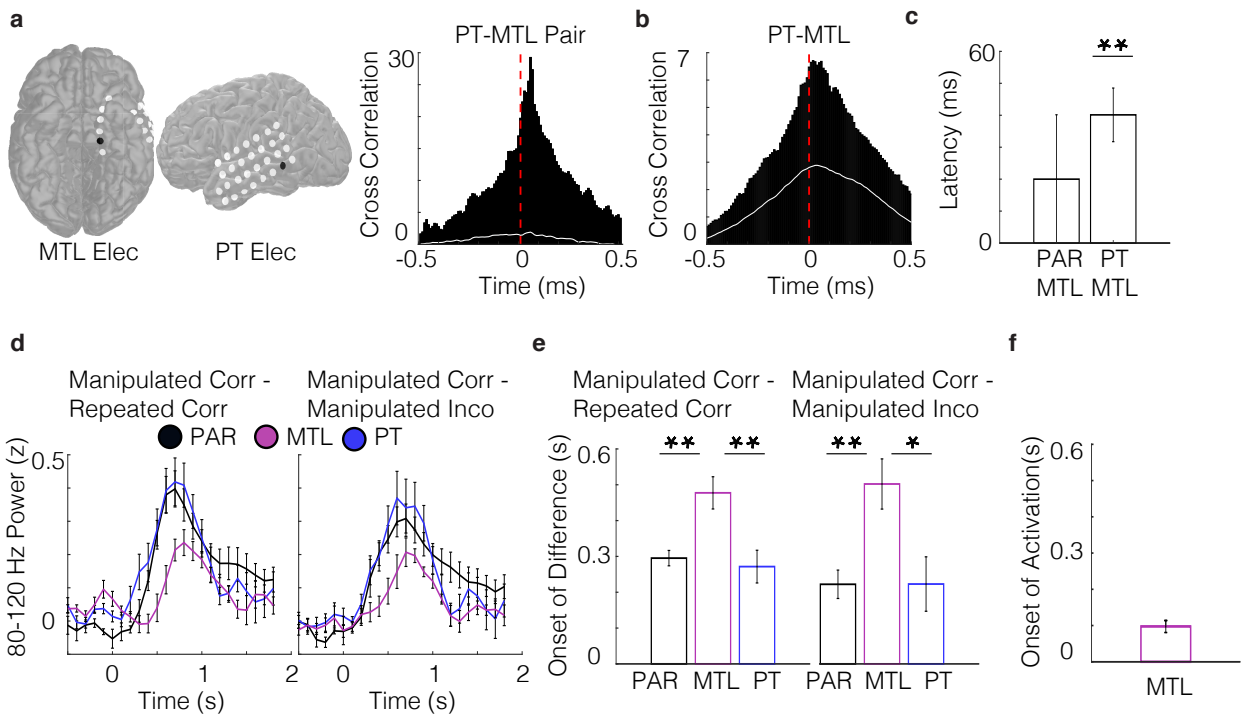

Figure 6. Differences in 80-120 Hz power when present visual experience is different from the past visual experience emerge earlier in the visual association cortex than in the MTL.

a) Representative cross-correlation of 80-120 Hz power between a PT and MTL electrode in an example participant. The chance cross-correlation is indicated by the white line. b) Average cross-correlation of 80-120 Hz power between the PT and MTL electrodes across participants. c) Average peak times (latency) of 80-120 Hz cross-correlograms for PT-MTL and PAR-MTL electrode pairs across participants. Average latency between PT and MTL electrodes was significantly greater than zero ($p < .05$, paired t-test). d) Average time series of differences between the manipulated and repeated correct conditions (left) and manipulated correct and incorrect conditions (right) in PT, PAR, and MTL across participants (200 ms sliding windows, 50% overlap; image appears at t = 0). e) Average estimated onset of differences in 80-120 Hz power between conditions in PT, PAR, and MTL across participants. Asterisks (*,**) indicate significance at $p < .05$ and $p < .01$, two-sided unpaired t-test. f) Average estimated onset of 80-120 Hz MTL power (activation) across participants.

We next examined whether the difference signals that arose in 80-120 Hz power when viewing an image that had been manipulated also progressed in a feedforward direction along the visual hierarchy. We visualized the time course of these differences in each brain region across participants and identified the first time point exhibiting a difference in 80-120 Hz power when comparing manipulated to repeated conditions and when comparing manipulated correct to incorrect trials (Fig. 6d; see Methods). Across participants, the differences in 80-120 Hz power between manipulated and repeated conditions arose significantly earlier in PT and PAR compared to MTL (PT v MTL, $t(16) = -3.21$, $p = .005$, unpaired t-test; PAR v MTL, $t(14) = -3.32$, $p = .005$; Fig. 6e). These differences began at 271 ± 46 and 295 ± 21 ms after the image presentation in PT and PAR, respectively, but only started at 477 ± 45 ms in the MTL. Across participants, the differences in 80-120 Hz power between manipulated and repeated conditions also reached 50% of the peak significantly earlier in PT and PAR compared to MTL (PT v MTL, $t(16) = -2.63$, $p = .018$, unpaired t-test; PAR v MTL, $t(14) = -2.81$, $p = .013$). PT also exhibited significant differences earlier than MTL ($t(16) = -2.44\$$, $p = .026$). We found similar temporal patterns of activation when comparing manipulated correct and incorrect conditions (Fig. 6e), demonstrating that the differences that are detected between past and present visual experience are captured by increases in 80-120 Hz power that also propagate from posterior to anterior brain regions along the visual hierarchy.

Although specific differences in activity between manipulated and repeated conditions arose in the MTL only after those differences were detected in the visual association cortex, we hypothesized that the ability to detect any difference between past and present experience required initially retrieving the past experience, and therefore activation of the MTL. We were therefore interested in when the differences in 80-120 Hz power that we observed in PT and PAR occurred with respect

to overall activity in the MTL. Based on the increase in 80-120 Hz power observed across all trials following the presentation of the image during recognition testing (Fig. 3d), we estimated the first time point at which MTL activity exceeded baseline as above (97.2 ± 16 ms; see Methods)(Fig. 2.5f). We then compared this time point to the first time points exhibiting a difference in activity between conditions in PT and PAR. In both cases, we found that overall activation of the MTL preceded the detection of a difference between past and present visual experience in the visual association cortex (PT v MTL, $t(16) = -3.58$, $p = .003$; PAR v MTL, $t(14) = -7.43$, $p = 3.18 \times 10^{-6}$, unpaired t-test). Substantial previous evidence have linked memory retrieval with activation of the slow gamma band within the MTL [110–115]. We compared activity between the manipulated correct and repeated correct/manipulated incorrect condition across all frequencies. In contrast to the PAR and PIT, MTL also showed a second peak in the 20-50 Hz band when comparing the manipulated correct to the manipulated incorrect and repeated correct conditions (Supplementary Fig. 10b). We found these differences in 20-50 Hz power were significant from 600 to 900 ms within the MTL ($p < .01$, permutation test; Supplementary Fig. 10c) during the successful recognition, but not within the PAR or PT.

Successfully identifying whether an image has been manipulated during recognition testing requires comparing that image to a retrieved memory. Because our data suggest that retrieval may involve activation of the MTL, we were interested in whether neural communication between the MTL and the visual association cortex where differences were first detected might underlie this process. Low frequency oscillatory coherence has previously been linked with neural communication between brain regions [116], and so we examined oscillatory coherence between electrode pairs in the data (see Methods). In a representative single trial, we observed clear evidence of phase-lagged alpha oscillations, and consequently alpha coherence, between a pair of

electrodes in the MTL and PT, suggesting that these two regions may become synchronized during the recognition period (Fig. 7a). We computed the coherence at all frequencies between all visually responsive PT and MTL electrodes during the recognition period across participants and found alpha coherence between the two regions significantly increased 300 to 1600 ms after the presentation of the image when examining all trials ($p < .05$, permutation procedure; Fig. 7b), suggesting that comparing past to present visual experience involves some communication between the MTL and visual association cortex. We did not find evidence of significant coherence between the MTL and PAR ($p > .05$, permutation procedure). We then examined coherence between PT and MTL electrodes separately for the manipulated correct, repeated, and manipulated incorrect conditions (Supplementary Fig. 14). Alpha coherence between these brain regions appeared more robust and more prolonged during the manipulated correct trials compared to the other conditions. We computed the mean level of alpha coherence over the recognition period and found significantly greater alpha coherence for the manipulated correct condition compared to the manipulated incorrect condition ($t(7) = 2.61$, $p = .03$, unpaired t-test; Fig. 7c), but only marginally greater coherence when compared to the repeated condition ($t(7) = 1.91$, $p = .09$).

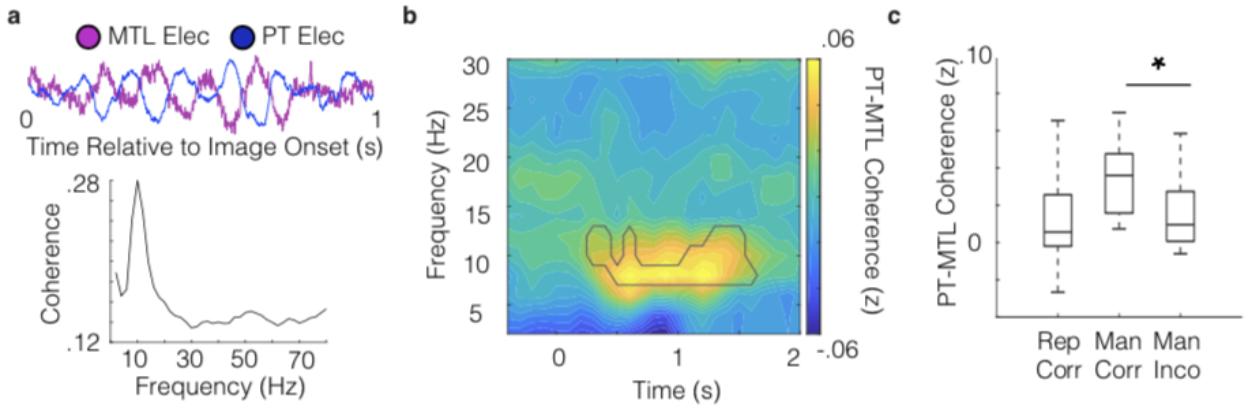

Figure 7. Alpha coherence increases between MTL and PT increases when past and present visual experience are in violation.

a) iEEG traces from representative PT and MTL electrodes indicating low frequency coherence at the single trial level. Across all trials in this electrode pair, coherence spectrum showed a peak in the low frequency band. b) Average coherence spectrum for all PT-MTL across participants. The black outline constitutes all epochs that exhibited significant differences compared to baseline that survived corrections for multiple comparisons ($p < .05$, permutation procedure). c) Comparison of average low frequency PT-MTL coherence across participants between manipulated correct and repeated correct/manipulated incorrect conditions. Asterisks (*) indicate significance at $p<.05$ using two-sided unpaired t-test.

2.4. Discussion

In this work, we observed an increase in 80-120 Hz oscillations that progressed from visual association areas to the MTL that contained information regarding the specific image being viewed. When the present image was in violation with the previously viewed image, we found increases in 80-120 Hz oscillations that specifically occurred after fixating on the region of the image that was in violation. Critically, the violation between past and present experience was detected in visual association areas earlier than the MTL. The detection of this violation was also accompanied by elevated low frequency coherence between the visual association areas and the MTL. Together these observations provide a direct account of how violations of the expectations set by the episodic memory of a previous experience are encoded within neocortical-MTL circuits.

Episodic memory formation relies on distributed set of interactions between the neocortex and MTL [53,58]. During an experience, the MTL receives input from sensory association cortices and transmits this input back to association areas [54,87–92] These forward and feedback interactions between the neocortex and MTL enable the formation of distributed neural trace through associative synaptic modification. When presented with a similar visual experience, neocortical-MTL interactions reactivate the distributed representation present during the original experience[54,89,93,94,117].

Expanding on this framework, these data suggest neocortical-MTL interactions may enable the comparison of the present experience with the remembered experience during episodic memory retrieval. Our data is largely consistent with predictive coding theory, which suggest neocortical-

MTL circuits generate an internal model of the world in order to predict external input [61–65]. Under this framework, higher order neocortical circuits send a feedback signal to lower order neocortical circuits in order to predict external visual input. The comparison of the predicted and actual input are transmitted from low order areas to high order areas via feedforward interactions in order to improve this internal model [81,118].

While this framework presents a computationally efficient, generalizable approach for cortical processing, the principles of predictive coding are largely supported by phenomenon within lower order circuits. For example, surround suppression within the retina and the receptive field properties of end-stopping cells in primary visual areas are canonical examples of how neural activity reflects a comparison between an external input and expectations of that input based on an internal model [62,77,119,120]. The expectations of sensory features are largely learned based on the natural statistics of the world and formed in order to reduce redundancy in representation[62,75–79]. If predictive coding is a computationally, generalizable approach for cortical processing, such principles should extend beyond simple sensory features to single events or episodes. In contrast to simple visual features, episodes are often only experienced just once. The generation of expectations for events may therefore rely on the MTL, a cortical area necessary for rapid one trial learning [121,122]. Our data demonstrate that such expectations for single events are formed from episodic memory and violation of such expectations involve a distributed interactions between neocortical circuits and the MTL.

The oscillatory mechanisms underlying visuospatial memory formation has largely focused on gamma oscillations within the rat hippocampal system [111–113,123]. Hippocampal CA1 shows the

presence of two types of gamma oscillations, slow (20-50 Hz) and fast gamma (60-140 Hz) oscillations [110,124]. Fast gamma oscillations within CA1 are thought to originate from the medial entorhinal cortex and contain representations regarding the current location of the animal within a spatial environment [110,113]. In contrast, slow gamma oscillations within CA1 arise from input from CA3 and codes the past and future trajectories of the animal within a spatial environment [110,113,125]. It has been proposed that fast gamma supports the coding of representations that are currently being experienced while slow gamma promotes the retrieval of memories associated with the current experience[112,113,126,127]. Consistent with this distinction, our data show increases in fast gamma oscillations within the MTL arise from neocortical circuits and contain information regarding the current image being viewed while slow gamma oscillations exhibit increases in activity specifically within the MTL during successful recognition of a manipulation.

However, we propose that fast gamma oscillations within neocortical-MTL circuits reflect a difference between the current experience and expectations of the experience, rather than the current experience alone. The primate visual system consists of a set of cortical areas that exhibit neurophysiological asymmetries [51,52]. Whereas feedforward interactions are characterized by interareal synchronization of higher frequencies, feedback interactions are mediated by interareal synchronization of the lower frequencies [128–131]. Based on these physiological asymmetries, prediction errors are hypothesized to be transmitted in a feedforward direction by high frequencies while predictions are thought to be transmitted in a feedback direction by low frequencies [81,118]. Our data showed differences in 80-120 Hz power that were detected in visual association areas prior to the MTL during the recognition of manipulation and suggests the comparison of the predicted and actual input may be transmitted in a feedforward direction by high frequencies. We

also found increases in alpha coherence between neocortical-MTL circuits during the successful compared to unsuccessful recognition of manipulation consistent with the idea that predictions being transmitted between neocortical-MTL circuits by lower frequencies.

As both past and present experience are represented in the MTL initially, one might expect the detection of differences between past and present experience to occur in MTL prior to visual association areas. Remarkably, the detection of differences between past and present experience occurred in the opposite direction. A possible explanation is that sensory association areas may serve as a comparator circuit that detects and transmits unexpected input to the MTL. The MTL then selectively encodes and retrieves the representations associated with unexpected input to generate future expectations. Based on this framework, feedforward induction of 80-120 Hz activity between visual association areas and the MTL may promote the encoding of future expectations. Feedback induction of alpha oscillations may enable the transmission of these expectations from the MTL to visual association areas. Whether or not the causal induction of these feedforward and feedback interactions between visual association areas and the MTL improves successful detection of a manipulation remains a future direction of this work.

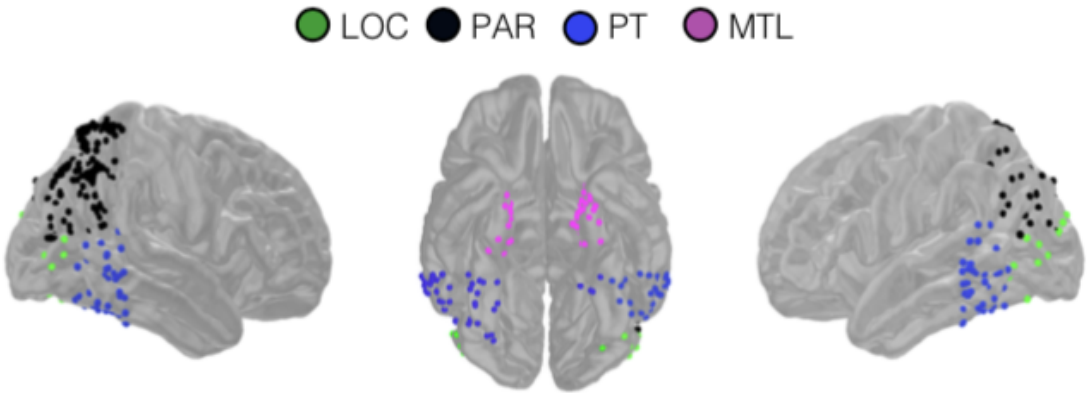

Figure 8. Distribution of Electrodes in LOC, PAR, PT, and MTL

All electrodes within lateral occipital complex (LOC), posterior temporal (PT), parietal (PAR), and medial temporal lobe (MTL) ROIs

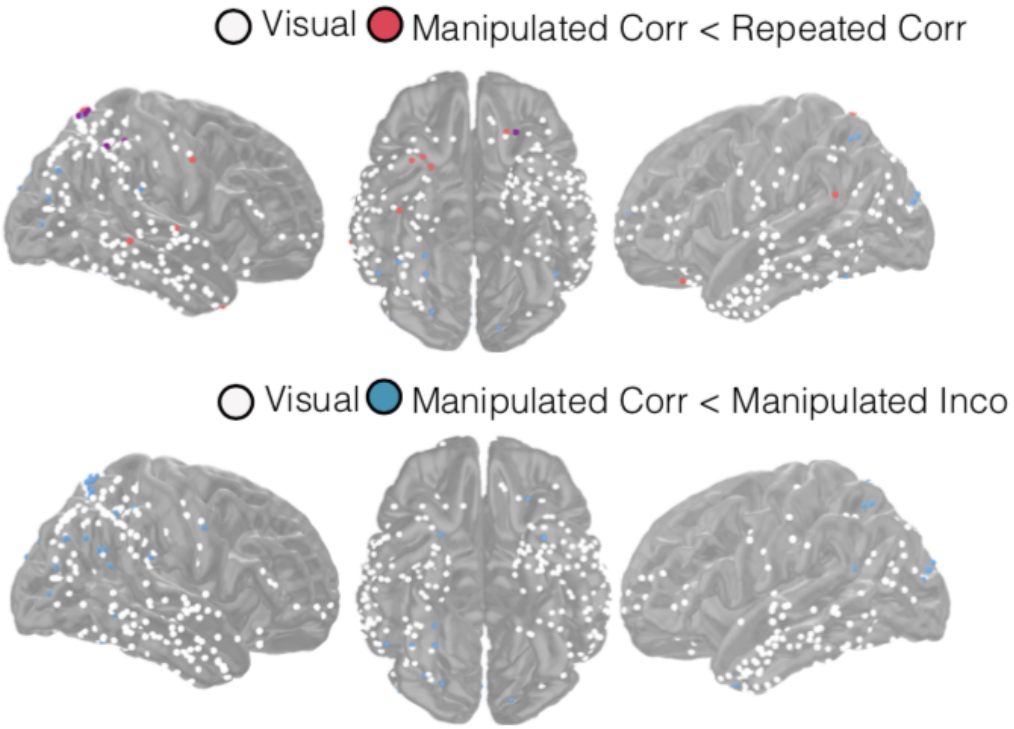

Figure 9. Electrodes showing decreases in 80-120 Hz power between the conditions

Visually responsive electrodes (white) that showed significant decreases in 80-120 Hz power for the manipulated correct compared to the manipulated incorrect (blue), manipulated correct compared to the manipulated repeated (red), or both (purple).

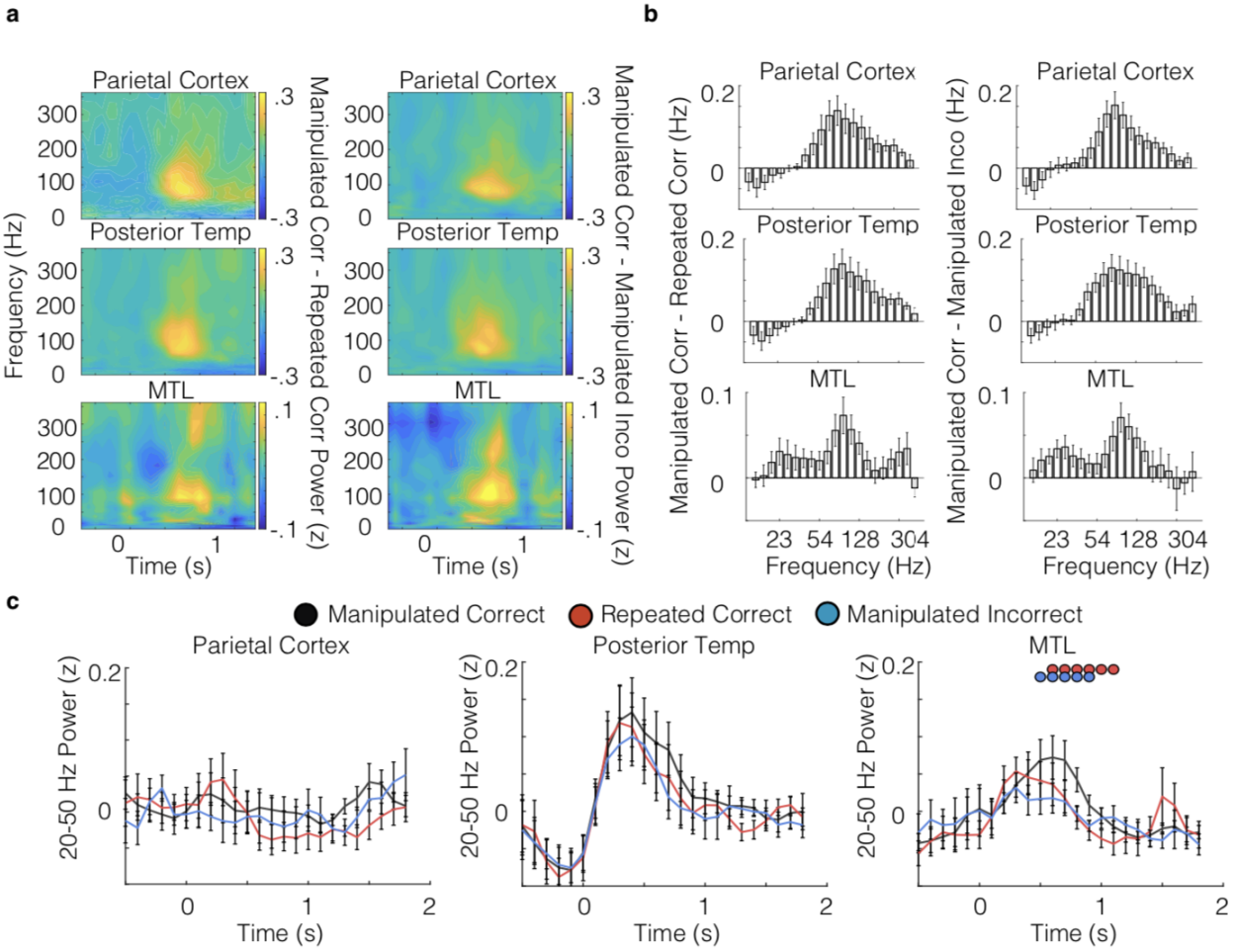

Figure 10. Spectral power differences when predicted and present visual input are in violation **a)** Spectrograms of power differences between the manipulated correct and manipulated incorrect and repeated correct conditions across participants in PAR, PT, and MTL. **b)** Average difference between the manipulated and manipulated incorrect/repeated correct trials across participants for each frequency **c)** Average 20-50 Hz power in PAR, PT, and MTL for manipulated correct (black), repeated correct(red), and manipulated incorrect (blue). Blue and red dots indicate significant increases in power for the manipulated correct compared to the manipulated incorrect and repeated correct condition, respectively, that survived corrections for multiple comparisons ($p < .05$, permutation procedure)

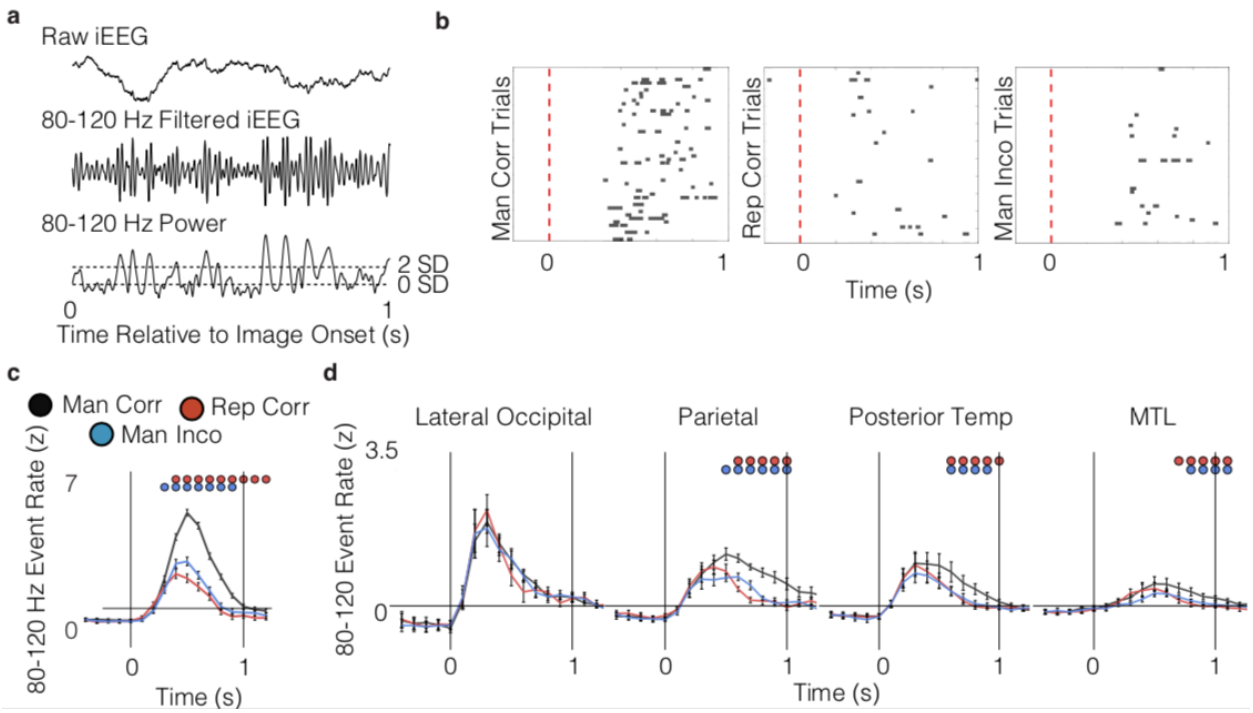

Figure 11. 80-120 Hz events increase when predicted and present visual input are in violation **a)** iEEG signal filtered in the 80-120 Hz band and extraction of instantaneous power using Hilbert transform. Events were defined as time points where the Hilbert envelope exceeded 2 standard deviations above the mean amplitude of the filtered traces for 20 ms. **b)** Rasters plots of 80-120 Hz events in a representative electrode in PT for the manipulated correct, manipulated incorrect, and repeated correct condition. **c)** Average event rate for a PT electrodes across manipulated correct (black), manipulated incorrect (blue), and repeated correct (red) trials. Time courses were generated with a 200 ms sliding window in 100 ms steps (50% overlap) locked the second viewing of the image. Blue and red dots indicate significant increases in rate for the manipulated correct compared to the manipulated incorrect and repeated correct condition, respectively, that survived corrections for multiple comparisons ($p < .05$, permutation procedure) **d)** Time courses of 80-120

Hz event rate across participants for LOC, PAR, PT, and MTL for the manipulated correct, manipulated incorrect, and repeated correct conditions locked to the second viewing of the image.

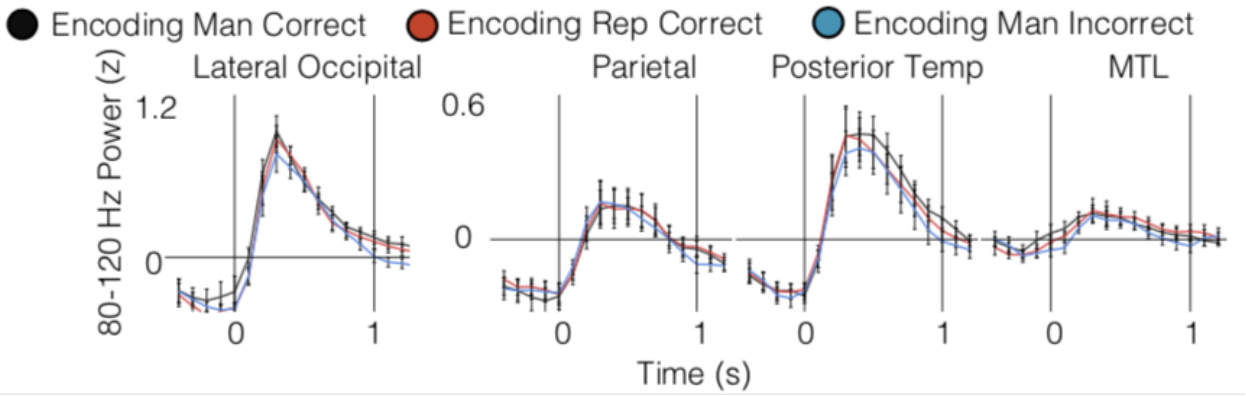

Figure 12. Time courses of 80-120 Hz power during encoding

Time courses of 80-120 Hz power across participants for LOC, PAR, PT, and MTL for the manipulated correct, manipulated incorrect, and repeated correct conditions locked to the first viewing of the image. No significant differences were observed for the manipulated correct compared to manipulated incorrect and repeated correct images ($p < .05$, permutation procedure).

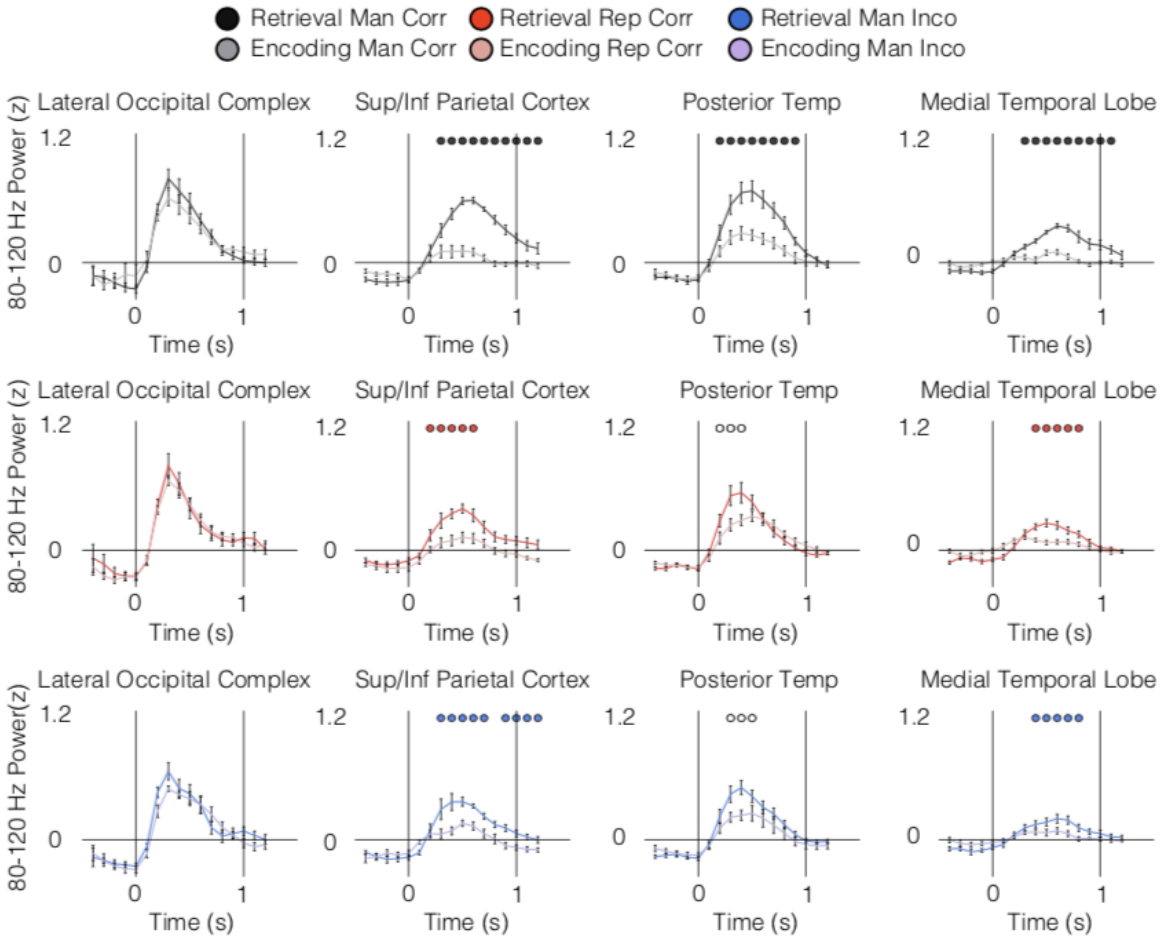

Figure 13. Time courses of 80-120 Hz power during encoding and recognition

Time courses of 80-120 Hz power across participants for LOC, PAR, PT, and MTL for the removed correct, removed incorrect, and repeated correct conditions during the first and second viewing of the image. Blue, red, and black dots indicate significant increases in power for the second viewing compared to the first viewing ($p < 0.05$, permutation procedure).

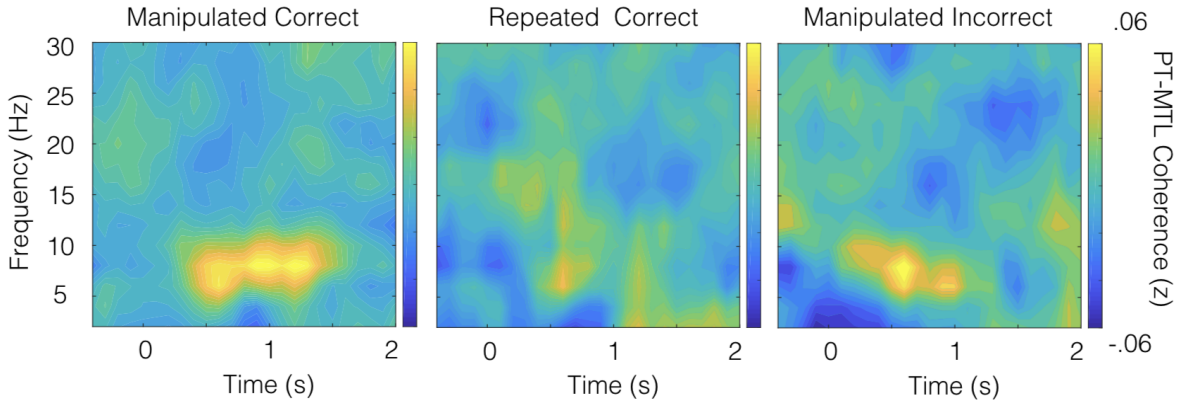

Figure 14. Spectrograms of average PT-MTL coherence

Average coherence spectrum for PT-MTL electrode pairs for the manipulated correct, manipulated incorrect, and repeated correct conditions across participant

Chapter 3: Visuospatial Memory Performance during Healthy Aging, Mild Cognitive Impairment, and Alzheimer's Disease

3.1 Introduction

In the previous chapter, we showed that visuospatial memory formation involves a set of interactions between the visual association areas and MTL. As pathological changes in AD develop within the MTL years before the onset of clinical symptoms, we hypothesized visuospatial memory formation may be impaired in participants with AD. Memory impairment in AD and other conditions has typically been established through performance on neuropsychological tasks measuring verbal recall. These conventional memory tests typically require trained personnel, a considerable amount of time to administer, and are often underused for symptomatic individuals because of the resource demands necessary to implement in clinical settings. Visuospatial memory paradigms based on eye movements may offer a passive, sensitive, and efficient assessment of memory performance in Alzheimer's disease.

The aim of the current study was to develop a passive, efficient, and sensitive paradigm that assesses visuospatial memory and evaluate its performance in healthy controls and memory-impaired subjects. Building on previous work with eye-tracking for assessing memory [73,74] and adapting a previously well studied task [66–68], we developed VisMET, which uses eye movements rather than explicit memory judgements in order to assess memory. Participants were asked to view a set of naturalistic images followed by the same set of images with either an object removed

or a new object added in order to alter the visuospatial relationships between the objects and locations. The amount of time viewing these manipulations compared to unchanged parts of the images was used to measure memory of either a previously viewed object and location (removed condition), or a new object and location (added condition). This four-minute paradigm was administered to 296 control or memory-impaired participants (mild cognitive impairment, MCI or AD) recruited from the Emory Healthy Brain Study and Alzheimer's Disease Research Center. Visuospatial memory performance was compared in healthy aging and at different stages of AD.

3.2 Methods

3.2.1 Participants

296 participants were recruited from research and clinical populations, including the Emory Healthy Brain Study (EHBS) and the Alzheimer's Disease Research Center (ADRC) (Table 1.1). Research participants received detailed evaluations that included neuropsychological testing and for ADRC subjects a diagnosis (healthy controls, MCI, or AD) reached after review by an interdisciplinary team of research neurologists, geriatricians, and neuropsychologists. A group of symptomatic subjects was evaluated in the Emory Memory Disorders Clinic where they received a comprehensive clinical evaluation consisting of standardized neuropsychological testing, neurological exam, brain imaging, and bloodwork, with a diagnosis by a board-certified behavioral neurologist's best clinical judgment. Controls were defined by normal memory and general cognition with preserved functional abilities. MCI was defined by impaired memory (single or multiple domain, based on cognitive testing) with preserved functional abilities, while AD was defined as impairment in two or more cognitive domains and functional abilities. The Montreal Cognitive Assessment (MOCA) was used as a common test across the EHBS, ADRC, and clinic

populations. Both age and education were significantly different between healthy controls and MCI (unpaired t-test, $p<10^{-5}$; $p<10^{-3}$) and AD populations (unpaired t-test, $p<10^{-10}$; $p<10^{-3}$) and therefore used as co-variates in later analysis.

Table 1 Demographics of Controls, MCI, and AD Participants

	Controls	MCI	AD
Number	182	74	40
Age	63.8 ± 7.4	70.8 ± 7.9	74.3 ± 7.5
Gender (M/F)	50/132	38/36	20/20
Race (Cau/AA/NA/Asi/Oth)	154/21/1/4/2	66/6/2/0/0	38/2/0/0/0
Education	16.6 ± 2.2	15.6 ± 2.5	15.3 ± 2.8
MoCA	26.7 ± 2.1	21.8 ± 3.9	14.6 ± 4.8

3.2.2 Visuospatial Memory Eye-tracking Test (VisMET)

Participants performed a memory paradigm based on eye movements (Fig. 15). During the encoding phase, participants were simply asked to 'enjoy' viewing a set of color images without any other explicit instructions (i.e., they were not informed they were being given a memory task). The images were selected for their positive aesthetic appeal. During the recognition phase, the participants viewed a modified version of the same set of images in the same order with either an item removed ("removed condition") or an item added ("added condition"). Images were selected from an open access database of images from Pixabay and Pexel and edited by a medical illustrator using Adobe Photoshop. Images were selected to be interesting and with varying degrees of complexity.

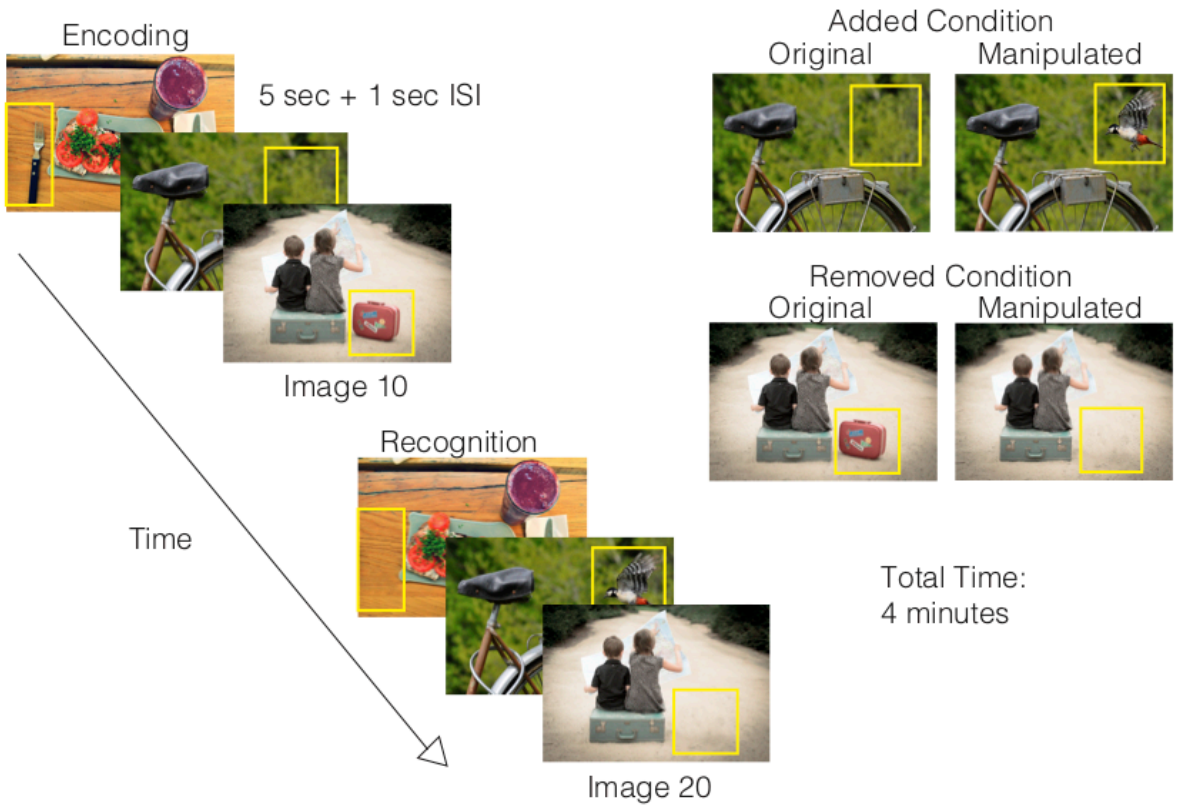

Figure 15. Visuospatial Memory Eyetracking Test (VisMET)

Participants were asked to view a set of images for 5 seconds with a 1 second inter-stimulus interval each during the encoding phase. During the recognition phase, participants viewed the same set of images with either one item removed (removed condition) or one item added (added condition). The manipulated regions used to quantify memory performance are indicated by the yellow box, which was not visible during viewing. The final test parameters consisted of the presentation of *two* sets of 10 original-manipulated pairs (7 with removed condition and 3 with added condition) with a delay of 1 minute in between the original and manipulated presentations. The entire task took 4 minutes.

Images were presented on a 24'' monitor 26'' away from the screen at a visual angle of 27 x 20 degrees. Each image was shown for 5 seconds, with a white fixation cross appearing for 1 second

between images. Each set consisted of original images followed by the presentation of slightly manipulated images, i.e. object added or removed. Performance was initially assessed over a range of image numbers and delay periods with a minimum of 10 images and a maximum of 20 images. Presenting 10-20 images resulted in a delay of 1-2 minutes between the original and manipulated images. The final test parameters consisted of the presentation of *two* sets of 10 original-manipulated pairs (7 with removed condition and 3 with added condition) with a delay of 1 minute in between the original and manipulated presentations. Images with 2-3 objects or focal points for the removed condition and 3-5 objects for the added condition were selected as optimal for assessing memory performance. We also only added or removed items in non-central locations to minimize the impact of delayed eye movements from the fixation cross at the center of the calibration screen preceding each image.

3.2.3 Eye movement detection

The locations of an individual's focus on the screen were estimated using an EyeTribe Infrared Scanner which sampled at 30 Hz. The scanner was attached to the bottom panel of a computer screen that was mounted to the wall using an adjustable arm to allow adaptation to participants of different heights. This hardware comprises a linear array of infrared LEDs that illuminate the eye and allow for the capture of pupil and corneal reflection using a near-infrared sensitive camera. The rotation of the eye was determined by the relative positions of the corneal and pupillary reflections. At the start of each session, participants performed a 9-point calibration procedure in order to convert eye rotations into a set of gaze positions relative to the screen. A small number of participants were excluded from the analysis (8.3%) who could not calibrate to the eye tracker or did not make any attempts to view the images.

3.2.4 Fixation Detection

For each participant, raw eye movement data was extracted using the EyeTribe Software and the data was analyzed off-line using custom scripts in MATLAB (MathWorks, Natick, MA) and Python. Raw gaze positions were converted into a set of fixations using a dispersion-based algorithm[101]. Each fixation was defined as a point of gaze continually remaining on the screen within 2 degrees of visual angle for a period of 100 ms or more.

3.2.5 Measurement of Visual Exploration

We developed methods to quantify visual exploration for each participant by measuring the viewing of the unmanipulated object during the first presentation (Fig. 16A). To quantify viewing of unmanipulated objects during the encoding phase, we identified the location of each object to be removed or added by drawing a rectangle, defined by the x,y coordinates of its four corners, around each object. This region was identified as the 'critical region' (average size: 8.8 x 7.5 visual angles). The number of fixations and the percentage of time viewing the critical region was calculated for each original image presentation. The percentage of time spent in the critical region was averaged across all original image presentations for each subject (Metric 1). We also calculated the percentage of all original image presentations with at least one fixation in the critical region for each subject (Metric 2). The calculated metrics varied depending on the specific images presented to each participant. To correct for this variation, each participant's metrics were normalized by the mean and standard deviation of the metrics derived from the viewing behavior of the control population (# subjects > 25 for each set) viewing the same images. These metrics were compared between healthy controls and memory-impaired participants (MCI/AD) using an unpaired student t-test. Of note, the images for the added condition during the original presentation

contained critical regions without any objects (Fig. 16). Therefore, we expected these eye movement-based metrics to be nearly zero for the added condition in the first viewing of the images.

3.2.6 Measurement of VisMET Performance

A similar approach was used to quantify viewing of the manipulated objects during the recognition phase. A rectangular critical region was drawn outlining the location of the removed or added object. The number of fixations and the percentage of viewing time within the critical region was calculated for each manipulated image presentation. The metrics for the added and removed conditions were derived separately for each participant. For each participant, we calculated the average percentage of time spent in the critical region across all manipulated images (Metric 1). We also calculated the percentage of manipulated images with at least one fixation in the critical region for each participant, with separate measurements for the removed and added objects (Metric 2). As before, these metrics were normalized by the mean and standard deviation of the healthy controls viewing the same image sets. These metrics of memory performance were compared across age groups and disease categories using an unpaired student t-test. Although both metrics were calculated for the added and removed conditions, we focused on Metric 1 for the added condition since spending at least one fixation in the critical region did not necessarily constitute successful memory. We focused on Metric 2 for the removed condition since the viewing time of many of the manipulated images was zero.

VisMET offers the option to vary the task difficulty, and as a result, track memory performance at different degrees of severity of impairment. The difficulty for remembering an image was defined based on the performance of healthy persons. The higher the percentage of healthy people that

viewed the manipulated critical region, the lower the difficulty for that image. Formally, difficulty for an image containing a removed object was the percentage of healthy people that spent at least one fixation in the critical region. The difficulty for an image containing an added object was the percentage of time spent in the critical region. Images were binned into categories based on their difficulty (0-25, 25-50, 50-75, and 75-100 percent of healthy controls showing the response) and performance was compared between healthy and memory-impaired participants for each of these bins. Of note, bins 50-75 and 75-100 were not analyzed for the added condition as no images had an average viewing time greater than 50% within the critical region.

3.2.7 Logistic Regression Models

Using a leave one out cross validation procedure, we quantified whether visuospatial memory performance on the task could be used as a screening tool for measuring cognitive impairment and predicting a diagnosis of MCI and AD. Two logistic regression classifiers were trained, each using a combination of three features: viewing time in the critical regions (added condition), percentage of critical regions viewed (removed condition), and age. The output of the models predicted the likelihood of accurately predicting performance on a standard measure of cognitive impairment (MoCA $\leq$ 23 or MoCA > 23) and disease status (healthy control or MCI/AD), respectively. The performance of the model was assessed using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. To conduct this analysis, participants needed to view images for both the added and removed condition. This analysis could only be completed with data from participants who viewed image sets that included both types of manipulations (added and removed). Thus, only the subset of participants (n= 126) that received the final version of the task were analyzed (Table 2).

Table 2. Demographics of Controls, MCI, and AD Participants in Final Version

	Controls	MCI	AD
Number	77	27	22
Age	64.5 ± 7.5	69.5 ± 9.5	76.0 ± 7.0
MoCA	26.7 ± 2.0	21.3 ± 4.0	13.5 ± 5

3.3 Results

3.3.1 Visual Exploration in Healthy Aging, Mild Cognitive Impairment, and Alzheimer's Disease during Encoding

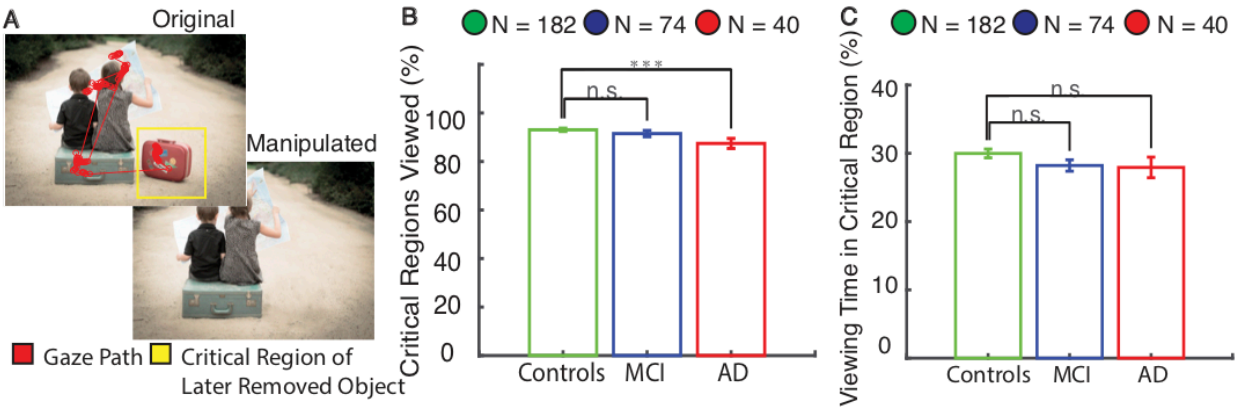

Figure 16. Visual Exploration of Later Removed Objects during Encoding Phase

(A) Participants viewed images during the encoding phase containing an object that was removed in the future during the recognition phase as indicated by the yellow critical region. (B) Healthy controls, MCI, and AD participants fixated on approximately 90% of the subsequently removed objects. The MCI group viewed the same percentage of critical regions as healthy controls ($p > 0.05$, unpaired t-test) while the AD group viewed slightly fewer critical regions than healthy controls ($p < 0.001$, unpaired t-test) C) Healthy, MCI, and AD participants all spent roughly 30% of the time viewing the critical regions, with no significant differences across groups. Asterisks in each panel indicate significant differences in performance as shown (*0.05, **0.01, ***0.001, ****0.0001; unpaired t-test).

The inability to shift attentional resources may lead to inadequate viewing of the to-be manipulated item, which could confound later assessments of visuospatial memory performance. For these reasons, we first evaluated baseline attentional shifts in eye movement during image viewing to ensure that controls and memory-impaired participants all had adequate opportunity to encode the images.

The control, MCI, and AD participant groups all made at least one fixation within the critical regions containing a later removed object for nearly 90% of the images. The MCI group viewed the same percentage of critical regions as healthy controls ($p > 0.05$, unpaired t-test) while the AD group viewed slightly fewer critical regions than healthy controls ($p < 0.001$, unpaired t-test) (Fig. 16B). In the encoding phase, we also compared the average viewing time in the critical region containing a later removed object. Healthy controls spent 30 ± 0.6% of the viewing time in the critical regions, with similar amounts for MCI (28 ± 0.8 %, $p > 0.05$, unpaired t-test) and AD (28 ± 1.5%, $p > 0.05$, unpaired t-test) populations (Fig. 16C). In summary, controls and memory-impaired participants viewed nearly 90% of the critical regions containing a later removed object, spending roughly 30% of the viewing time in the critical regions. These results indicate that the subject groups had similar viewing behavior with adequate opportunity to encode the majority of the images, and that any differences in visuospatial memory were unlikely due to differences in viewing behavior or attention.

The encoding phase also included a set of images in the first presentation containing "empty" regions where items were subsequently added for the recognition phase. All three populations fixated on less than 5% of these empty critical regions with a viewing time of less than 1%. Therefore, the viewing behavior during the added condition for the groups also did not suggest any major differences that would confound later assessments of visuospatial memory.

3.3.2 VisMET Performance in Healthy Aging

The visuospatial memory paradigm in this study required memory for a complex set of associations between objects and locations and was assessed passively using eye movements rather than explicitly. We sought to evaluate whether visuospatial memory showed age-related declines in performance and how these differences compared to age-related declines in verbal free recall performance. We hypothesized that healthy older participants would show impairments in discerning the manipulated objects compared to healthy younger participants. To test this hypothesis, we compared the percent viewing time and the percentage of trials with at least one fixation in the critical region across three age groups [50-59, 60-69, 70+]. We compared these metrics separately for the added and removed conditions.

The group of 50-59 year-old individuals performed better on the memory task than the 60-69 and 70+ age groups. The 50-59 age group fixated within 58 ± 3% of the critical regions with a removed object, whereas this percentage dropped for the older controls (45 ± 3% for the 60-69 year olds, and 46 ± 3% for the 70+ year olds) (Fig. 17A). These results were significant when comparing the percentage of the critical regions viewed by the 50-59 age group to the 60-69 ($p < 0.001$, unpaired t-test) and the 70+ age groups ($p < 0.01$, unpaired t-test). There was no difference in performance between the 60-69 and 70+ age groups. Similar age-related declines in memory were observed when using viewing time within critical regions as a performance metric (Fig. 17B). The 50-59, 60-69, and 70+ age groups spent 12 ± 0.9%, 9.3 ± 0.8%, and 10 ± 1% of the time viewing the critical regions with a removed object (Fig. 17B). The 50-59 age group spent significantly more time in the critical regions compared to the 60-69 ($p < 0.01$, unpaired t-test) and the 70+ ($p < 0.05$, unpaired t-test) age groups. We did not find such differences for the added condition ($p > 0.05$,

unpaired t-test; data not shown). These results indicate that healthy adults aged 50-59 show better memory performance for the removed condition compared to healthy adults over the age of 60.

To compare these results to a commonly used neuropsychological measure of memory, we assessed age-related differences in delayed free recall performance using the Free Cued and Selective Reminding Test (FCSRT). In this population, the 50-59, 60-69, and 70+ age groups remembered 77 ± 1%, 76 ± 1%, and 70+1.6% of the words. The 50-59 age group did not show any significant differences in verbal free recall compared to the 60-69 age group ($p > 0.05$, unpaired t-test; Fig. 17C). Rather, age-related decline in free recall performance became apparent later in the 70+ age group ($p < 0.001$, unpaired t-test). Collectively, these results suggest visuospatial memory performance for the removed condition provides a means to detect age-related memory decline earlier than the FCSRT.

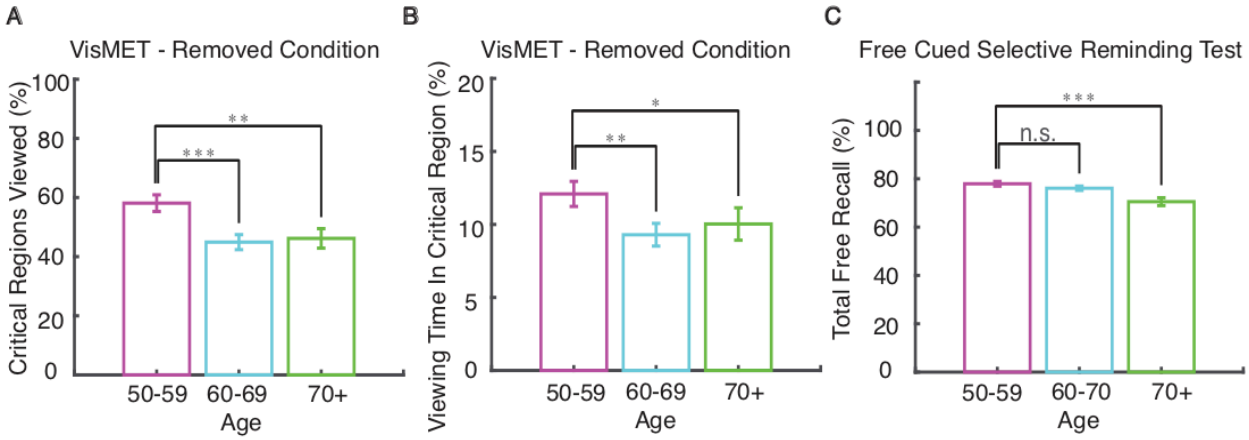

Figure 17. Age-related Changed in VisMET Performance.

A) Younger participants (50-60) viewed more of the critical regions containing removed objects compared to older participants (60-70, 70+). (B) Younger participants (50-60) spent a greater percentage of viewing time in the critical regions containing the removed object compared to older participants (60-70, 70+). (C) For comparison, memory scores on the free recall portion of the FCSRT are shown. Asterisks in each panel indicate significant differences in performance as shown (*0.05, **0.01, ***0.001, ****0.0001; unpaired t-test).

3.3.3 VisMET Is impaired in Mild Cognitive Impairment and Alzheimer's Disease

We compared visuospatial memory performance among healthy, MCI, and AD populations separately for the removed and added conditions. We first quantified viewing of the removed objects during the recognition phase (Fig. 15). The critical region for the removed objects was an empty location and as a result, the percentage of critical regions in which a visual fixation was recorded was a strong indicator that the removed object had been successfully remembered (Fig. 18A). Control participants fixated on nearly twice as many of the critical regions compared to MCI ($p < 10^{-8}$, unpaired t-test) and AD ($p < 10^{-10}$, unpaired t-test) populations (Fig. 18B). We also quantified viewing time within the critical regions for the removed condition. Control participants spent a significantly greater percentage of the viewing time in the critical region compared to MCI ($p < 0.001$, unpaired t-test) and AD populations ($p < 10^{-4}$, unpaired t-test).

We varied the difficulty among the manipulated images to determine if this would alter task performance (see Methods). We hypothesized that the more difficult images, operationally defined by the percentage of healthy controls with viewing behavior indicative of intact recall, to remember would best distinguish controls and memory-impaired subjects. The largest differences between control and MCI/AD participants were observed for images that were recognized as different by only 25-50% of the control participants ($p <10^{-5}$, unpaired t-test; $p < 10^{-8}$, unpaired t-test). Notably, the easiest images (operationally defined as viewing behavior indicative that they were "remembered" by 75 to 100 percent of the participants) could only distinguish healthy controls from AD but not from MCI ($p > 0.05$, unpaired t-test). Thus, by controlling the difficulty of manipulated images, we were able to create a potential method to track memory performance across varying degrees of memory impairment.

Next, we determined whether viewing of the added object in the delayed presentation was also different among healthy, MCI, and AD groups. The critical region for the added condition contained an item that was not present during the encoding phase (Fig. 18D). Control participants with intact memory spent a greater percentage of time viewing the critical region containing the added object ($p < 0.01$, unpaired t-test) (Fig. 18E) and viewed a greater percentage of the critical regions compared to MCI populations ($p < 0.01$, unpaired t-test). There was a similar relationship between the control and AD populations ($p < 10^{-4}$, unpaired t-test; $p < 10^{-9}$, unpaired t-test). We also assessed the impact of image difficulty for the added condition (see Methods). Although both bins were able to separate healthy controls from memory-impaired populations, manipulated images with 25-50% viewing time in the critical region showed the most significant differences between the healthy controls and MCI ($p < 10^{-4}$, unpaired t-test) and AD populations ($p < 10^{-4}$, unpaired t-test) (Fig. 18F).

Although we expected that such differences in viewing would be due to lack of recognition of the changes in the images viewed previously, we also evaluated potential effects of eye movement-related differences in fixation accuracy across the participant groups. To control for possible differences in fixation accuracy, we quantified the average distance to the fixation cues during the first and second half of the task in visual angles. We found decreases in fixation accuracy between the first and second half of the tasks in the x direction (1.35 ± 0.05 and 1.92 ± 0.07 visual angles; unpaired t-test, < 0.05) but not the y direction (1.30 ± 0.05 and 1.14 ± 0.07 visual angles; unpaired t-test, $p > 0.05$). Importantly, there were no differences in fixation accuracy in MCI and AD participants compared to healthy controls for the first or second half of the experiment (unpaired

t-test, $p>0.05$). We also did not observe any significant correlation between a participant's performance and the distance to the fixation dot ($r=-0.03$, $p>0.05$). From these results, the variation in performance for each group is unlikely due to the changes in fixation accuracy throughout the experiment but rather, differences in memory of the removed objects when comparing the controls and memory-impaired subjects with MCI or AD.

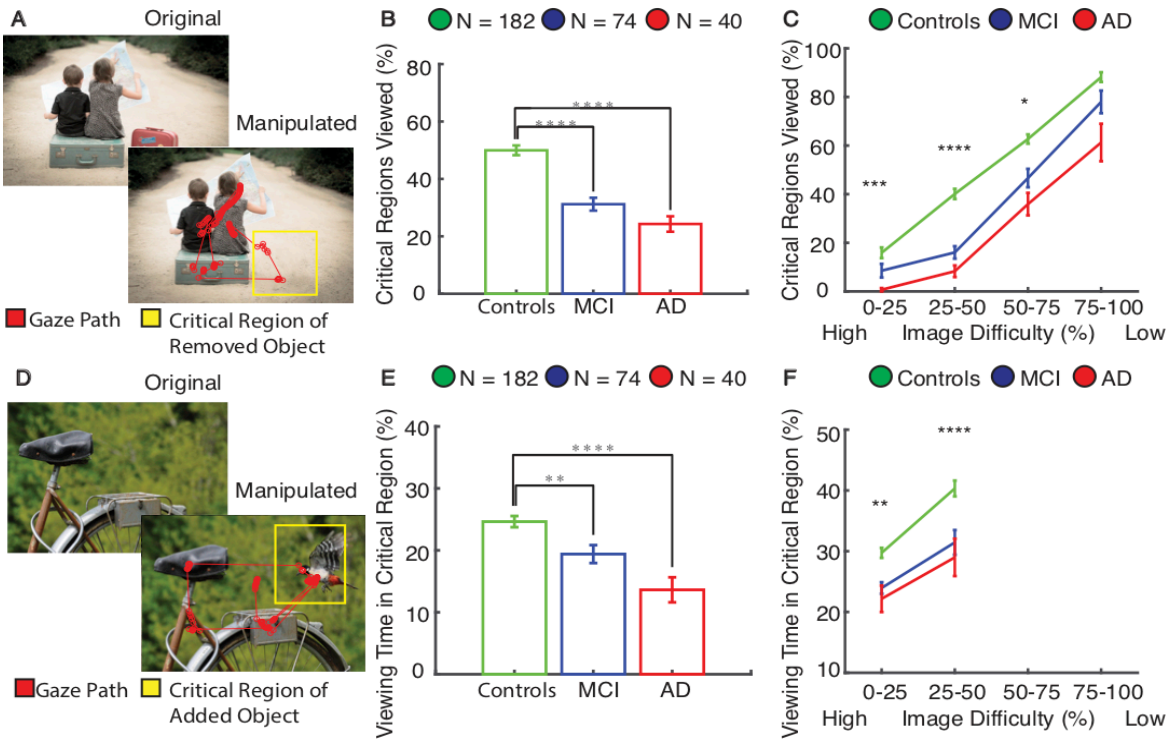

Figure 18. VisMET Performance in Mild Cognitive Impairment and Alzheimer's Disease

Participants viewed images with either an object removed (A) or added (D) as indicated by the yellow critical regions, which was invisible to the viewer. (B) Subjects with MCI and AD showed impaired visuospatial memory performance (removed condition) compared to controls. (C) Control subjects viewed a greater percentage of the critical regions compared to AD participants regardless of the extent of difficulty between the original and manipulated presentations. The less difficult images better distinguished healthy and MCI individuals. Asterisks indicate significant differences in performance between healthy controls and MCI (*0.05, **0.01, ***0.001, ****0.0001; unpaired t-test). (E) Memory performance for the added condition was impaired (i.e., less time viewing the added object) in the MCI and AD populations compared to controls. (F) Manipulated images with high difficulty showed the most significant differences in performance between healthy and MCI participants as indicated by the asterisks. Viewing times for any of the

added objects did not exceed 50% and as a result, difficulty could not be measured at higher viewing times.

3.3.4 VisMET as a screening tool for cognitive impairment and disease status

Visuospatial memory performance showed robust differences between healthy controls and memory-impaired participants. The reliability of these differences suggest that visuospatial memory performance may be used as a screening tool. To evaluate VisMET as a screening tool for cognitive impairment, we trained a logistic regression classifier using a combination of three features: viewing time in the critical regions (added), percentage of critical regions viewed (removed), and age. The output of the models estimated the likelihood of cognitive impairment as measured by the Montreal Cognitive Assessment, a widely-used screening tool used to test a range of cognitive domains including memory, attention, executive function, and language. The performance of the model was assessed using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve of a leave one out cross validation analysis.

When training the models to predict cognitive impairment (MoCA $\leq$ 23), we found VisMET performance was able to achieve an AUC of 0.85 compared to an AUC of 0.71 and 0.56 when using age and education alone. This model was able to achieve a sensitivity and specificity of 0.83 and 0.74, respectively, using a cutoff probability of 0.64 (Fig. 19A). To further evaluate VisMET, we identified the specific cognitive domains that may be assessed by the task. We correlated visuospatial memory performance to other neuropsychological assessments given to the study population, regressing out age and education and correcting for multiple comparisons. We found robust correlations with the CERAD Word List Delayed Recall, Benson Complex Figure Delayed Recall and to a lesser extent, verbal and category fluency tests (Table 3). Based on these results,

VisMET performance offers a sensitive screening method to identify cognitive impairment, particularly for the memory domain.

Table 3. Correlations between VisMET Performance and Other Neuropsychological Instruments

	Removed Condition N=114	Added Condition N=65
General Cognition		
MoCA	0.32 ($p<10^{-3}$)	0.20
Memory		
CERAD Delayed Recall	0.37 ($p<10^{-4}$)	0.33 ($p<0.01$)
Benson Delayed Recall	0.41 ($p<10^{-5}$)	0.20
Attention		
Trails A Time	-0.07	0.13
Forward Span	0.17	0.19
Backward Span	0.07	-0.05
Executive Function		
Trails B Time	-0.16	0.08
Verbal Fluency (FAS)	0.21	0.13
Category Fluency (Animals)	0.21 ($p<0.001$)	0.17

We next aimed to determine the sensitivity of VisMET performance in estimating disease status. We trained a logistic regression classifier with the same three features as before, but instead the output of these models was the diagnostic classification of healthy controls, MCI or AD. Memory performance estimated MCI/AD status with an AUC of 0.85 compared to 0.73 and 0.58 when using age and education alone (Fig. 19B). Using all of the features, the model achieved a sensitivity and specificity of 0.85 and 0.75 with a cutoff probability of 0.63.

We further explored the relationship between VisMET performance and disease status. Each participant's raw performance (% of critical regions viewed and % viewing time) was normalized using the mean and standard deviation of healthy controls viewing the set of images. We then plotted each participant's performance on this 2D feature space (Fig. 19C). The lower left quadrant

of this plane indicated below average performance for both the added and removed condition and as expected, nearly all of the MCI/AD participants were in this lower left quadrant. Interestingly, a significant portion of healthy controls' performance fell in this same quadrant, with memory performance on VisMET similar to those with MCI and AD. Together, these results suggest that VisMET can be used as a sensitive tool for separating healthy controls from MCI/AD and to identify a population of healthy controls with an AD-like memory profile.

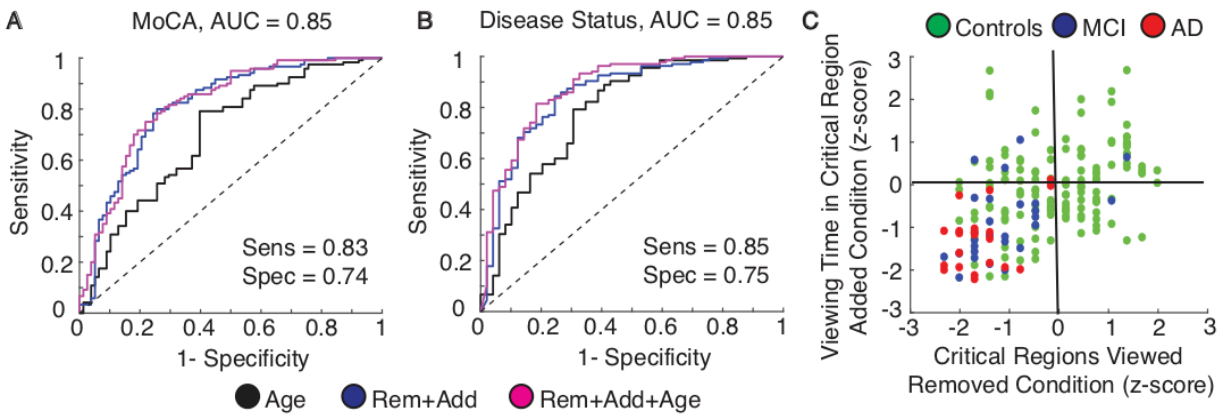

Figure 19. VisMET performance predicts cognitive impairment and disease status

(A) Viewing of the removed and added objects during the recognition phase could accurately predict performance on the Montreal Cognitive Assessment (MoCA ≤ 23 or MoCA > 23), a standard measure of cognitive impairment. (B) Viewing of the removed and added objects could separate those clinically diagnosed with MCI/AD from healthy controls. (C) Memory performance was visualized on a two-dimensional plane representing performance. Most MCI/AD participants fell within the lower left quadrant of the plane indicating below average performance for both the added and removed condition. Healthy participants within this quadrant exhibited a memory profile indicative of AD.

3.4 Discussion

The aim of this study was to develop an easily administered, enjoyable, and sensitive paradigm for detecting mild memory deficits, and assess its performance in a large group of healthy controls as well as a population of patients with memory impairment. To this end, 296 participants were presented with a memory paradigm in which we used eye movements to infer memory. Using eye tracking as an index of memory, we found that memory performance on the task is both age-related and different between healthy and MCI/AD participants. Performance was also dependent on the difficulty of the original and manipulated images, which allows for the task to be sensitive across a broad range of memory abilities. A multivariate model of memory performance on the task predicted cognitive impairment and AD status with high sensitivity and identified a subpopulation of healthy controls with relatively weak performance on the task.

A few studies have examined whether memory can be measured using eye movements and to determine whether memory measured using eye movements depends on the hippocampus[66–68]. In these studies, participants were presented with a series of images and cued with a question about the relationships between the objects within the scene [66]. These images were followed by another set of images that were either novel, repeated or manipulated. Participants spent more time viewing the manipulated regions only when they were unable to verbally report the manipulation had occurred. In these early studies, cueing the participants towards the manipulation could have influenced later assessments of memory. Later studies replicated this paradigm *without* cueing the participants toward the manipulation during the first presentation[67,68], and found that increased viewing of the manipulated region only occurred when participants were aware of the manipulation. Moreover, the viewing time and explicit identification of the manipulation was

reduced in a small group of amnestic patients with non-specific damage to the medial temporal lobe. Although the role of cueing and delay in eye-movement based memory need to be carefully examined in future studies [135,136], these initial studies provided evidence for the use of eye movements as an indicator of memory dysfunction.

To extend these findings, we evaluated the ability of eye movements to predict memory impairment in normal aging and found an age-related decline in memory performance. Although a number of studies have shown memory decline with age [137–139], the pattern of memory decline is unclear. Most cross-sectional aging studies measure memory decline using verbal learning paradigms and show a linear decrease in episodic memory function starting in the 20's and progressing through the rest of adult life [140,141]. In contrast, longitudinal studies of aging show a very different pattern, demonstrating a decline in memory performance only after the age of 60-65 [140,141]. A potential reason for this difference stems from the varying levels of education attainment for the different age groups in cross-sectional studies. When education is controlled in cross-sectional studies, declines in memory appear after the age of 60 [137]. More recently, the effect of aging on memory has been investigated using mnemonic discrimination paradigms [142,143]. These paradigms find an entirely different trend than previous cross-sectional studies, namely that memory begins to decline around age 20 but plateaus in the 60-90 range. We also found a similar trend with performance on visuospatial memory starting to plateau after 60 years of age. To clarify the effects of aging on these different types of memory, longitudinal studies need to be conducted administering these paradigms in the same cohort of patients to allow for proper comparison.

Visuospatial memory performance was also evaluated in a group of participants with memory impairment due to AD. Memory impairment due to AD has typically been established using verbal learning tests such as the Rey Auditory Verbal Learning Test and the Free Cued Selective Reminding Test. Formal neuropsychological tests of memory are often resource intensive, requiring trained personnel and considerable amount of time to administer. Moreover, the explicit responses and awareness of performance deficits often leads to frustration or even distress for impaired subjects – sometimes to the point of discontinuing the task or declining future assessments. Even for symptomatic individuals, memory is often not assessed because of the resource demands of these tests in a clinical setting. Compared to other paradigms, memory was assessed passively using eye movements rather than instructing the participants or requiring explicit memory judgements. Using eye movements as an index of memory offers a number of practical benefits compared to explicit forms of retrieval [70–72]. Without the collection of explicit instructions or behavioral responses, we were able to assess memory passively in only a few minutes. Anecdotally, the paradigm appears to be strikingly less distressing and frustrating to both research participants and clinical patient populations than traditional neuropsychological tasks and also avoids issues related to task comprehension and explicit memory judgements. Conventional tasks are also subject to differences in effort, literacy, cultural variation, and decision-making capacity which can confound the measures of memory. Although the current study did not formally address the ability of this task to overcome these limitations, these potential advantages were important considerations in the development of the task and remain to be investigated.

Cognitive impairment in AD and related disorders has typically been established using cognitive screening tools such as the MoCA or MMSE. These tests often suffer from floor and ceiling effects

due to their inability to change task difficulty for the specific population being tested. To address this issue, we created a means to change the difficulty of the task. In doing so, we were able to create a set of images that could potentially track the transition from MCI to AD but also ones that could track the transition from healthy to MCI. Even without adjusting the difficulty of the items, we found that visuospatial memory performance was highly sensitive for predicting cognitive impairment and disease status. The task also showed a large variation in performance among healthy controls. One possible explanation is that successful recognition of the manipulation can occur even when participants do not view the manipulation. Alternatively, we speculate that healthy controls performing similarly to memory-impaired subjects may be at higher risk for AD, although this needs to be studied carefully using a longitudinal design, and with biomarkers for preclinical AD. These findings come in support of recent work suggesting that separating similar visual images rely on the same circuits affected in the preclinical stages of AD [143–145].

Our data provides further support for the use of eye movements to measure objective memory impairment. In previous work, eye-tracking was used to assess novel object recognition memory using the visual paired comparison task [73,74]. The amount of time viewing a novel image when placed side by side with an old image could differentiate healthy controls from memory-impaired participants. To extend these findings, we developed a task that assesses memory for the relationship between an object and its location (visuospatial memory). Compared to previous methods, we developed a method to assess memory capacity across a broad range of memory impairment, and using machine learning techniques we could estimate cognitive impairment and disease status with high sensitivity. An open question remains whether memory based on eye movements can predict future memory decline earlier than standard verbal memory assessments.

Based on the role of the entorhinal-hippocampal circuit in visuospatial memory, we hypothesize visuospatial memory to be an earlier predictor of entorhinal-hippocampal degeneration, as occurs in early AD, compared to current assessments and therefore predict earlier memory decline than standard memory assessments. Nonetheless, this visuospatial memory paradigm based on eye movements offer a passive, sensitive, and efficient memory paradigm capable of detecting objective memory impairment and predicting disease and cognitive status.

Chapter 4: Deep convolutional neural networks and transfer learning for measuring cognitive impairment using eye-tracking in a distributed tablet-based environment

4.1 Introduction

In the previous chapter, we developed the Visuospatial Memory Eye-tracking Test (VisMET), an eye-movement based memory paradigm that provides a passive, efficient, and sensitive assessment of cognitive impairment in Alzheimer's disease [146]. During this task, participants are presented with a set of images followed by the same set of images with an object removed or added (Fig. 20). The amount of time viewing these manipulations could differentiate healthy participants from those with MCI and AD. A major limitation of this platform for administration of VisMET is that it requires the use of an external eye tracker, which is often expensive and widely unavailable. As a result, large scale screening for memory loss remains difficult. Convolutional neural networks have recently been used to demonstrate eye-tracking on the mobile phone and tablet. These networks are trained with the crops of both the eyes and the face and have enabled gaze estimation without the need for an external eye tracker. Whether the accuracy of these networks is sufficient to assess visuospatial memory performance based on eye movements remains unknown.

We developed an iPad based version of VisMET that utilizes iTracker, a convolutional neural network (CNN) architecture used to track eye movements on Apple devices [147]. We trained this architecture using a public data set of over 1.5 million images of human faces unrelated to the study population [147], and then updated the network using a transfer learning paradigm using approximately 90,000 images collected from research and clinical populations. We administered

a four-minute memory test to 250 healthy controls (MoCA > 24) and cognitively-impaired participants (MoCA ≤ 24) recruited from the Emory Healthy Brain Study (EHBS) and the Goizueta Alzheimer's Disease Research Center (ADRC) at Emory. We compared memory performance between these two populations and found performance could be used as a screening tool for identifying cognitive impairment. This tablet-based version of VisMET has potential to enable large scale screening of memory loss.

4.2 Materials and Methods

4.2.1 Participants

A total of 552 participants were recruited from the EHBS and ADRC and were administered the VisMET using either the tablet or the EyeTribe (Table 4). All participants received detailed evaluations that included neuropsychological testing, although the specific batteries varied. The MoCA was used as a common test to evaluate cognitive performance across the EHBS and ADRC populations. A MoCA greater than 24 was considered normal, and MoCA less than or equal to 24 was indicative of cognitive impairment [148]. The research participants from the EHBS were further classified as healthy controls, MCI, or AD by an interdisciplinary team of research neurologists, geriatricians, neuropsychologists, and clinical participants from the ADRC received a clinical diagnosis by a sub-specialty trained cognitive neurologist. Participants recruited from the EHBS primarily consisted of healthy controls whereas participants recruited from the ADRC consisted of participants at different stages of MCI and AD. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation.

Table 4 Demographics of Controls, MCI, and AD Participants

	Tablet	EyeTribe
Number	250	302
Age	72.3 ± 9.1	64.0 ± 8.1
Gender (M/F)	111/140	115/187
Race (C/AA/As/Oth)	199/41/8/2	270/20/4/4
Education	16.4 ± 2.3	16.2 ± 2.5
MoCA	20.0 ± 7.1	25.0 ± 4.8
EHBS/ADRC	83/167	240/62

*C=Caucasian, AA = African American, As = Asian, Oth - Other

4.2.2 Mobile Device Data Capture

VisMET was presented to participants using iPad Air 9.7" tablets with a viewable screen area of 154 x 203 mm and a resolution of 1536 x 2048 pixels. Each iPad was running at least iOS 10, capturing a series of frames from the 'selfie' camera at a resolution of 720p and sampling rate of 30 Hz. The selfie camera has an aperture of f/2.2 and a focal length of 31 mm.

4.2.3 Calibration Procedure

Each participant was seated in front of an iPad that was attached to a stand in portrait orientation. A silhouette of a face appeared on the iPad screen to center the participant's face, resulting in an approximate distance of 350 mm between the iPad and the participant's eyes. The calibration procedure sequentially presented 16 random numbers and letters, 35 x 35 pixels in size, at random locations on the screen. Participants were instructed to say the letters and numbers aloud to increase their attention and assess their understanding. Characters were presented for 1.3 seconds without delay, long enough for participants to find and say the character aloud. The locations of the characters served as labels to train the CNN and support vector regression (SVR) models for gaze estimation (Fig. 20). We removed frames collected during the first 250 ms of each character to

reduce periods during which participants were transitioning from one character to the next. We also removed frames collected from the first four calibration dots to accommodate learning the calibration procedure.

4.2.4 Visuospatial Memory Eye-tracking Task (VisMET)

VisMET was administered to assess memory passively using eye movements rather than explicit memory judgments. As MoCA was delivered to both populations within clinic and research sites, we used MoCA as the standard of cognitive performance for this paper. The only instruction given to participants was to enjoy the images. The task begins with the presentation of 20 images of scenes for a duration of five seconds, followed by a slightly modified set of images with either one object added (six trials) or removed (14 trials) from each image (Fig. 20). The delay between the first and second presentation of each image was approximately one minute. The images were presented on the top half of the iPad screen since eye-tracking accuracy has been shown to decrease as a function of distance from the iPad camera [147]. The added condition was removed from further analysis due to the limited number of total trials. The images were selected to contain 2-5 objects based on visual inspection. This selection was performed prior to the collection of eye tracking and provided a method for filtering out images that would likely lead to a diffuse pattern in viewing.

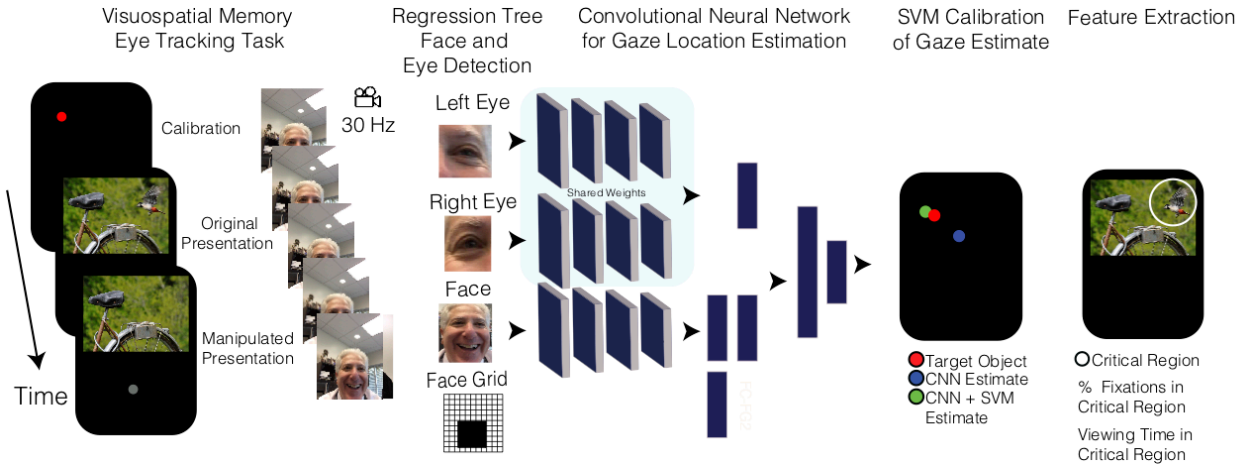

Figure 20. Distributed tablet-based environment for measuring memory peformance

Participants were presented with a calibration procedure followed by VisMET on iPad devices. The task consisted of the presentation of a set of images followed by the same set of images with an object added or removed. While the task was running, the mobile application sampled images of a participant's face at 30 Hz using the built-in 'selfie' camera. Upon task completion, the images were uploaded to the cloud for face and eye detection, and then used to train a CNN to estimate gaze. For each subject, a support vector regression (SVR) was trained using the gaze estimations from the CNN as input. Gaze estimations from the SVR were then used to extract memory features.

4.2.5 Regression Tree Face and Eye Detection

The CNN required a face crop, left eye crop, right eye crop, and face grid vector to estimate eye gaze locations. In previous work, crops were detected using built-in face and eye detection libraries on Apple mobile devices but led to a significant amount of data loss (39%). To prevent such data loss, we calculated bounding boxes for crops using the *dlib* library[149] through OpenCV's Python interface [150], an implementation that uses an ensemble of regression trees for facial feature alignment [151]. If three or more consecutive frames crossed an eye-aspect ratio of 0.1, the frames were omitted [152]. Frames without detected faces or eyes were also removed from future analysis. The mean percentage of valid frames per subject was 93%.

4.2.6 Convolutional Neural Network Training Datasets

We used two separate datasets for training the CNN: MIT's GazeCapture and Emory's ADRC set. MIT's dataset consisted of 1450 people and approximately 1.5 million frames. To improve the accuracy of gaze estimates, significant variability was introduced when creating MIT's data-set. By crowdsourcing administration, participants were administered the task in a wide variety of settings and lighting conditions with 75% of the participants collected through Amazon Mechanical Turk, 16% were collected from UGA's campus, and the 9% collected using app store downloads. Participants were also asked to move continuously throughout the task and encouraged to change the orientation of the device to further increase the variability in the dataset [147]. Emory's dataset consisted of 250 people and approximately 90,000 frames. Key differences exist between MIT's and Emory's datasets: (1) MIT's was comprised of camera frames collected from both Apple mobile phone and tablet devices. (2) Face and eye crops for the MIT set were generated

from Apple's built-in face and eye detection libraries rather than the offline OpenCV python interface described above. The labels of the MIT set were the location of characters using a different calibration procedure [147].

4.2.7 Convolutional Neural Network and Support Vector Regression for Tablet-Based Gaze Estimation

Method 1: Method 1 involved training a CNN with 90,000 frames collected during the calibration phase from a total of 250 Emory participants using a five-fold cross validation procedure, where the set of images used to train the CNN is different than the set of images used to extract participant memory performance. For each fold, we randomly divided the participants into train, validation, and test splits consisting of 125, 62, and 63 participants, respectively. The features for the CNN consisted of a face crop, left eye crop, right eye crop, and face grid vector extracted from each frame (Fig. 20). The labels of the network were the locations of the character during the calibration procedure. The CNN model was implemented in PyTorch and trained for 35 epochs with a batch size of 16, a weight decay of 0.0001, and momentum of 0.9. A global learning rate of 0.0001 was used and decayed by a factor of 10 every five epochs. The model with the lowest validation error was used to generate gaze estimations. These gaze estimations were used to evaluate the average Euclidean distance to the target character for the train, validation, and test participants.

Method 2: Method 2 involved training the network using approximately 1.5 million frames from the GazeCapture dataset [147]. This dataset consisted of 1450 participants with an 80%, 10%, 10% train, validation, and test split. For Method 2, the CNN was first trained with the inputs and labels from the MIT dataset using the same hyperparameters and training procedures as Method 1. This

pre-trained model was then retrained using inputs and labels from the Emory dataset as described in Method 1.

Method 3: Method 3 expanded on Method 1 by adding a support vector regression (SVR) layer for each participant separately. The features for the SVR were the gaze estimations from the final layer of the CNN from Method 1 while the labels were the location of the characters. These features and labels were separated into a train, validation, and test split of 50%, 25% and 25%. For each participant, we used the estimates from the test set to calculate the average Euclidean distance to the target character. We report this average error of the test set (Table 2.2) for the train, validation, and test participants from Method 1.

Method 4: Rather than using the CNN described in Method 1 to generate features for the SVR, Method 4 used gaze estimates from the final layer of the CNN from Method 2 to generate features.

4.2.8 Re-Calibration of Gaze Estimations Between Successive Images

Between each set of images, a red fixation dot was presented in the center of the viewing frame to reset a participant's gaze to the center. For each fixation dot, we calculated the median vertical and horizontal distance between the location of the fixation dot and the gaze estimations generated from the SVR. We shifted the estimations from the SVR for a particular image by the vertical and horizontal distances calculated from the preceding fixation dot. We performed these adjustments to account for any translations in gaze estimations throughout the task.

4.2.9 EyeTribe-Based Gaze Estimation

Using an EyeTribe Infrared Scanner (Copenhagen, Denmark), participant gaze locations were estimated at a sampling rate of 30 Hz. This eye-tracking scanner uses a linear array of infrared LEDs to illuminate the eye and allow for the capture of pupil and corneal reflection. The rotation of the eye was determined by the relative positions of the corneal and pupillary reflections. The scanner was attached to the bottom of a computer monitor that was mounted to the wall using an adjustable arm to accommodate participants of different heights. At the start of each session, participants completed the calibration procedure to convert eye rotations into a set of gaze positions relative to the screen. More information can be found in [153].

4.2.10 Feature Extraction

After gaze positions were estimated for each subject using the methods described above, the gaze estimations were converted into a set of fixations using a dispersion-based algorithm [101]. Each fixation was defined as a point of gaze continually remaining on the screen within 2 degrees of visual angle for a period of 100 ms or more. Data were analyzed off-line using custom scripts written in MATLAB and Python.

Measurement of Visual Exploration: We developed methods to quantify visual exploration for each participant by measuring the viewing of the unmanipulated object during the first presentation of the image (Fig. 21A). To quantify viewing of unmanipulated objects during the learning phase, we identified the location of each object to be removed (”critical region”) by drawing an ellipse around the object (defined by the x,y coordinates of the center, major, and minor axis). The number of fixations within the critical region and the percentage of time viewing the critical region was calculated for each original image presentation. The percentage of time spent in the critical region

was averaged across all original image presentations for each subject (Metric 1). We also calculated the percentage of all original image presentations with at least one fixation in the critical region for each subject (Metric 2).

Measurement of Memory Performance: We developed a similar approach to quantify viewing of the region containing the manipulated objects as a measure of memory [146]. An ellipse was drawn outlining the location of the removed object. The number of fixations and the percentage of viewing time within the critical region was calculated for each manipulated image presentation. For each participant, we calculated the average percentage of time spent in the critical region across all manipulated images (Metric 1). We also calculated the percentage of manipulated images with at least one fixation in the critical region for each participant (Metric 2).

4.2.11 Logistic Regression Models for Detection of Cognitive Impairment

We quantified whether memory performance on the task could serve as a screening tool for cognitive impairment. The test participants from Methods 3 and 4 were further divided into a train and testing set using an 80/20 split. Logistic regression classifiers were trained using either age (as a baseline) or the percentage of removed objects viewed. The output of the model estimated the performance on a standard measure of cognitive impairment (MoCA $\leq$ 24 or MoCA > 24). Using a five-fold cross validation procedure, the performance of the model was assessed using the area under the receiver operating characteristic curve (AUROC).

4.3 Results

4.3.1 Performance Evaluation of Tablet-Based Methods for Gaze Estimation

Table 5 shows the average distance between the gaze estimate and the target character for each of the tablet-based methods. Implementing SVR (Methods 3 and 4) reduced testing error substantially compared to methods using the CNN alone (Methods 1 and 2). We also observed a significant improvement when training the network with both datasets rather than the Emory dataset alone (Method 2 vs 1 and Method 4 vs 3). The method implementing both of these approaches (Method 4) outperformed all other methods with a testing error of 2.72 cm, which was comparable to the 2.58 cm and 2.12 cm errors from previous studies [147]. These results replicated the feasibility of performing eye-tracking on the tablet, showing improved performance with support vector regression, and extended previous studies by implementing a transfer learning approach for gaze estimation.

Table 5. Performance of Tablet-Based Methods of Gaze Estimation

Method	Model	Dataset	Train Error (cm)	Valid Error (cm)	Test Error (cm)
1	CNN	Emory	4.90 ± 0.18	4.83 ± 0.26	4.90 ± 0.15
2	CNN	MIT+Emory	3.87 ± 0.20	3.75 ± 0.16	3.89 ± 0.30
3	CNN+SVR	Emory	3.20 ± 0.06	3.20 ± 0.08	3.30 ± 0.06
4	CNN+SVR	MIT+Emory	2.71 ± 0.12	2.68 ± 0.09	2.72 ± 0.15

4.3.2 Assessment of Visual Exploration on the Tablet and EyeTribe

Figure 4.2 shows viewing of the critical region during the initial presentation of images for healthy controls and cognitively-impaired participants. The differences in viewing between the populations was first evaluated using the EyeTribe, a proprietary commercial eye-tracker. Healthy controls viewed 94% of the later removed objects, slightly more than the 87% viewed by

cognitively impaired participants ($p < 10^{-5,}$ unpaired t-test; Fig. 21C). This relative change of 8% was preserved when using the tablet to generate gaze estimations ($p < 0.05$, unpaired t-test; Fig. 21C). We also compared the average viewing time of the later removed objects between healthy controls and cognitively-impaired patients (Fig. 21B). Both EyeTribe and tablet-based methods produced relative changes in viewing time of 6-8% between healthy controls and cognitively-impaired participants ($p < 0.05$, unpaired t-test; Fig. 21B). From these results, we conclude that initial viewing of later removed objects is similar for both populations on either platform.

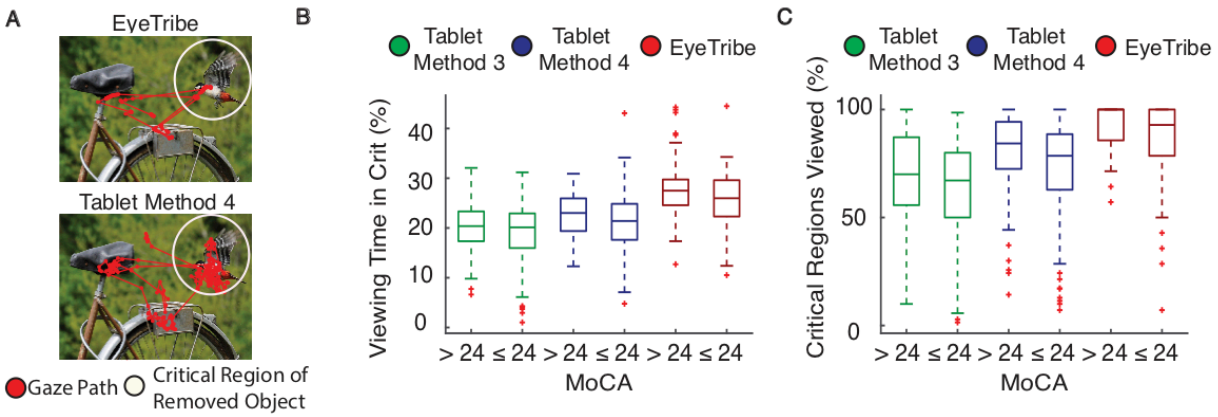

Figure 21. Visual exploration of later removed objects using a tablet and EyeTribe for gaze estimation

(A) Participants viewed images containing an object that was removed later as indicated by the white critical region (invisible to the viewer). The red line illustrates gaze path estimations using the EyeTribe and tablet. (B) Healthy controls (MoCA > 24) and cognitively impaired participants (MoCA <= 24) spent comparable amounts of time in the critical region across all platforms (6-8% relative change). Tablet based-methods 3 and 4 produced estimates that reduced viewing time in the critical region by 26% and 18% compared to the EyeTribe. (C) Healthy controls viewed a similar number of critical regions compared to cognitively-impaired participants across all platforms (8% relative change). Tablet based-methods 3 and 4 produced estimates that reduced the percentage of critical regions viewed by 27% and 15% compared to the EyeTribe.

We hypothesized the differences in viewing estimates between the tablet and EyeTribe were primarily driven by the accuracy of the gaze estimates. In support of this hypothesis, we observed that the differences in viewing time and viewed objects between the platforms were reduced to 7% and 11% after removing participants with a calibration error greater than 2cm. We also found tablet-based method 4 produced viewing estimates more comparable to the EyeTribe than tablet-based method 3 (Fig. 21), most likely due to its higher accuracy (Table 5). From these observations, we conclude that the differences in viewing between the platforms are primarily driven by the accuracy of gaze estimates. These results also suggest that the calibration errors in Table 5 extend beyond calibration and generalize to the task.

4.3.3 Assessment of Cognitive Impairment on the Tablet and EyeTribe

Figure 4.3 shows the differences in viewing of the manipulated regions between cognitively-impaired participants and healthy controls. Using the EyeTribe, control participants viewed 54% of the removed objects, nearly twice as many viewed by cognitively impaired participants ($p < 10^{-14}$, unpaired t-test). Tablet-based methods 3 and 4 produced relative increases of 24% and 29% when comparing to healthy controls to cognitive impaired participants, respectively ($p < 10^{-6}$, unpaired t-test; $p < 10^{-9}$, unpaired t-test). When using only participant data with a calibration error less than 2cm, healthy controls viewed nearly twice as many of the critical regions as cognitively impaired participants, comparable to differences produced by the EyeTribe. We also compared average viewing time between control and cognitively impaired participants on both platforms. Both the tablet methods and the EyeTribe showed significant increases in viewing time for healthy controls compared to cognitively impaired participants ($p < 10^{-4}$, unpaired t-test; $p < 10^{-7}$, unpaired t-test; $p < 10^{-10}$, unpaired t-test). In summary, we find both platforms are capable of distinguishing control and cognitively impaired participants and produce comparable viewing of the manipulated

regions, especially when the calibration errors are low.

The differences in cognitively impaired participants and controls during the recognition phase could be purely driven by fewer fixations, especially considering the lower viewing time for objects during the learning phase. We found a small but reliable decrease in the average number of fixations for cognitively impaired participants (11.0 ± 0.2) compared to healthy controls (11.8 ± 0.3) ($p < 0.05$, unpaired t-test). To ensure these viewing differences were not the primary driver of the differences between the two populations, we first calculated the correlation between MoCA and VisMET performance and found this correlation was 0.32 ± .01 across the testing folds. For comparison, we performed the same correlation partialling out the average number of fixations made by the participants during the recognition phase and found the correlation was 0.29 ± 01. These results indicate that the differences between the two groups during the recognition phase are largely driven by the memory of the manipulated objects.

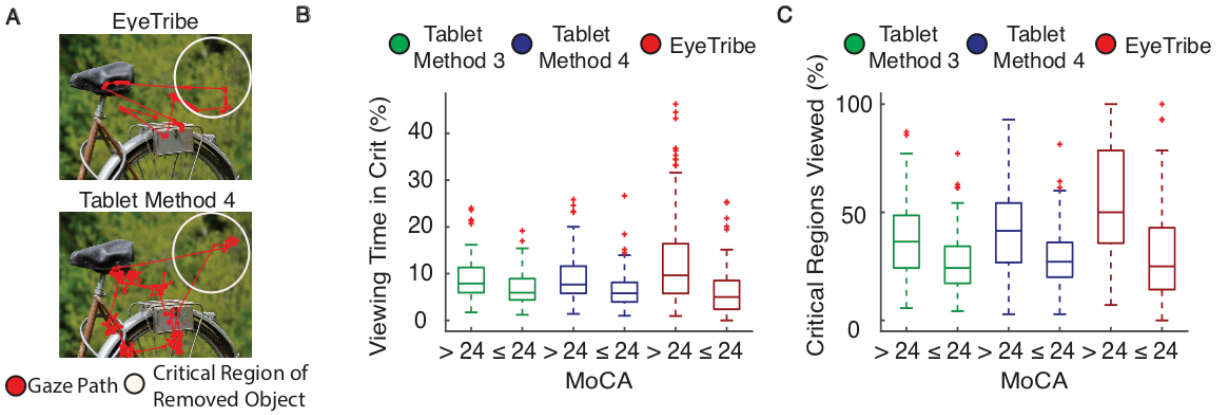

Figure 22. Memory of removed objects using a tablet and EyeTribe for gaze estimation

(A) Participants viewed images where an object was removed as indicated by the white critical region (invisible to the viewer). The red line illustrates gaze path estimates for EyeTribe and tablet-based methods. (B) Healthy controls (MoCA >24) spent a greater percentage of their time in the critical regions compared to cognitively impaired participants (MoCA <= 24) across platforms.(C) When using the EyeTribe, control participants (MoCA > 24) viewed twice as many of the removed objects compared to cognitively impaired participants. Tablet-based methods 3 and 4 produced relative increases of 24% and 29% when comparing healthy controls to cognitively impaired participants.

4.3.4 Tablet Administration of VisMET as a Screening Tool for Cognitive Impairment

Viewing of the manipulated regions showed robust population differences between cognitively-impaired participants and healthy controls using the tablet. Based on these results, we hypothesized that viewing of the manipulated region could be used as a screening tool for cognitive impairment. Figure 4.4 shows the extent to which viewing of the manipulated region can separate healthy controls from cognitively-impaired participants. When using the EyeTribe gaze tracking hardware, viewing of the manipulated region could estimate cognitive impairment (MoCA ≤ 24) with an AUC of 0.76 compared to an AUC of 0.66 and 0.70 when using tablet-based methods 3 and 4, respectively (Fig. 23A). Furthermore, we compared performance to age, a well-established risk factor for cognitive impairment. Using age as an estimate of cognitive impairment resulted in AUCs of 0.68 and 0.63 for desktop and tablet-based administrations, respectively. Overall these results suggest that VisMET outperforms age as an estimator of cognitive impairment.

Finally, we quantified testing AUC for tablet-based method 4 after removing participants that did not reach a specific calibration error threshold (Fig. 23), indicating the need for re-calibration. By requiring a calibration error of 2 cm, we were able to achieve a testing AUC of 0.76 for the mobile application(Fig 23B,C). These results demonstrate that the mobile version of VisMET can provide equivalent accuracy of cognitive impairment estimates compared to a commercial eye-tracker.

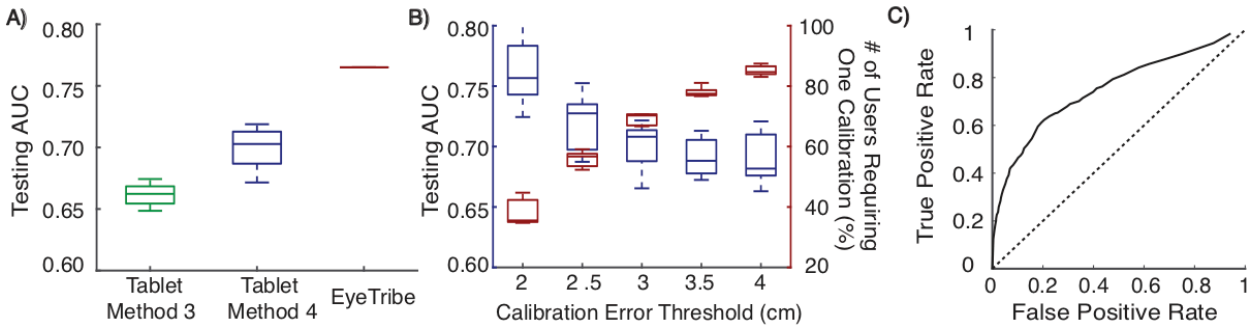

Figure 23. VisMET estimates cognitive impairment on the tablet and EyeTribe

(A) Viewing of the removed objects could accurately estimate performance on the Montreal Cognitive Assessment (MoCA <= 24 and MoCA >24), a standard measure of cognitive impairment. Tablet-based method 3, tablet-based method 4, and the EyeTribe estimated MoCA performance with an AUC of 0.66, 0.70, and 0.76, respectively. (B) Participants with a calibration error less than 2 centimeters achieved an AUC of 0.76 equivalent to the performance achieved using the EyeTribe. (C) Average ROC Curve across testing folds of estimating cognitive impairment in participants with a calibration error less than 2 cm.

4.4 Discussion

Currently, a key clinical criterion for the diagnosis of MCI is objective cognitive impairment, primarily detected using general cognitive screenings such as the MoCA and Mini-Mental Status Exam. These conventional memory tests typically require trained personnel, a considerable amount of time to administer, and are often underused in the clinic as individuals are intimidated by perceived poor performance on such tests. To address these concerns, the novel contribution of this work was the development of a mobile version of VisMET that performs eye-tracking using the standard front-facing camera of an iPad by leveraging a deep convolutional neural network together with a transfer learning approach. Experimental validation using 250 participants from the ADRC at Emory confirms the ability to estimate cognitive impairment within a clinical setting with an AUC of 0.76, equivalent to the accuracy of desktop-based eye-trackers. This study provides an easily administered, sensitive paradigm that can track cognitive impairment at multiple clinical centers.

The primary measure of cognitive impairment in this study was the amount of time participants spent viewing the manipulated region which has previously shown to correlate most with memory performance compared to performance on other cognitive domains. Memory formation has been shown to rely on the entorhinal-hippocampal circuit in the brain, the initial site of cortical pathology in AD. For this reason, memory performance may precede measures of cognitive impairment and predict cognitive decline in AD earlier than the MoCA. Certain images in the task may be more or less successful in estimating cognitive impairment than others, and also show variability in identifying cognitive impairment during different stages of AD. Detailed investigation of this hypothesis remains a future direction of this work.

The purpose of this study was to validate a mobile version of VisMET that could be used to standardize assessment of cognitive performance at multiple clinical centers under the testing conditions developed in this study. However, generalization of the trained CNN beyond the conditions in this study cannot be assured. In particular, if the iPad is handheld, lighting conditions are different, or if more than one face is on the camera, we expect a drop in performance. A more challenging test of the developed algorithm would be to deliver the paradigm remotely. This is likely to require the retraining of the CNN using frames collected from a wide variety of real-world settings, with a range of lighting and background motion changes. If accurate memory performance can be obtained in such a way, we could potentially use this approach with a cloud-based pipeline for widespread screening of cognitive impairment. Moreover, the ease of administration would be conducive for longitudinal assessments to potentially track cognitive impairment over time.

Chapter 5: Future Directions and Conclusions

5.1 Summary

In the third chapter, we developed the Visuospatial Memory Eyetracking Test (VisMET) and administered this paradigm to healthy controls and participants with MCI and AD. We found that VisMET performance was age-related and also showed significant impairment in MCI and AD participants compared to healthy controls. Performance was also dependent on the difficulty of the original and manipulated images, which allows for the task to be sensitive across a broad range of memory abilities. A multivariate model of memory performance on the task predicted cognitive impairment and AD status with high sensitivity and identified a subpopulation of healthy controls with relatively weak performance on the task. Based on the results from this study, VisMET

offered a passive, sensitive, and efficient memory paradigm capable of detecting objective memory impairment and predicting disease and cognitive status.

We sought to overcome the reliance on an external desktop-based eye tracker, which is often expensive and widely unavailable. To increase the availability of VisMET, we developed a tablet-based version of VisMET that utilized iTracker, a deep convolutional neural network capable of tracking eye movements on the iPad. We implemented a transfer learning protocol and performed a subject-specific SVM in order to improve the accuracy of eye-tracking on the tablet. Experimental validation using 250 participants from the ADRC at Emory confirms the ability to estimate cognitive impairment within a clinical setting with an AUC of 0.76, equivalent to the accuracy of standalone eye-trackers. This study provided an easily administered, sensitive paradigm that can track cognitive impairment at multiple clinical centers.

We also investigated the neurophysiological interactions between the neocortex and MTL that underlie visuospatial memory performance in humans. We administered a visuospatial memory task that requires participants to indicate whether an image was manipulated or repeated while recording intracranial EEG in visual association areas and the MTL. Successfully recognizing this manipulation therefore requires participants to not only retrieve the past visual experience but to also then compare the retrieved memory with the present image. We found that recognizing manipulated images, and therefore successfully identifying a difference between past and present experience, evoked a narrowband high frequency prediction error signal in visual association cortex that then propagated towards the MTL. During successful recognition of these manipulations, this error signal was also accompanied by elevated low frequency coherence between the neocortex and MTL. Together, the results provide a direct account of how violations

of the expectations set by the episodic memory of a previous experience are encoded in neocortical-MTL circuits.

5.2 Future Directions

5.2.1 Visuospatial Memory Performance in Healthy Participants

In the third and fourth chapters, we found significant variability in visuospatial memory performance in healthy participants. We expect this variability could arise from a number of different reasons. One potential explanation for this variability may be the different viewing strategies used by the participants during the recognition of a manipulation. We found that participants often did not fixate on the critical region of the removed object even though they were successfully able to locate the removed object when explicitly asked to. There may be a group of healthy participants who do not fixate on the critical region but successfully recognize the manipulation, which would explain the low VisMET performance in healthy controls. A second potential explanation is that the variability in VisMET performance arises from AD pathological processes. VisMET performance may provide a sensitive measure of subtle cognitive impairment that tracks the extent of neurodegeneration and AD pathology within the MTL in healthy controls.

Future work could quantify the relationship between VisMET performance and the presence of AD pathology, especially within healthy controls. Recent advances in molecular profiling technology have enabled identification of the proteomic landscape in order to develop unbiased, data-driven network models of AD [154,155]. Molecular networks for synaptic injury, neuroinflammation, and other pathophysiological mechanisms have been strongly linked to cognitive trajectory and the hallmark pathologies, even in preclinical stages of disease. These

molecular networks may prove useful in identifying if AD explains any of the variance in performance in healthy controls. To ensure the variability in performance arises from AD pathology, images could be selected that best explain the variability in AD pathology in healthy participants. Our preliminary studies indicate certain images are more useful than others in explaining cognitive impairment in AD. Similar studies could be conducted to determine if certain images better estimate the variance in AD pathology in healthy controls.

The development of VisMET for remote delivery may aid in screening for images that best explain variance in AD pathology in healthy controls. This is likely to require the retraining of the CNN using frames collected from a wide variety of real-world settings, with a range of lighting and background motion changes. If accurate memory performance can be obtained in such a way, we could potentially use this approach with a cloud-based pipeline in order to screen for images that best explain the variance in AD pathology of healthy controls. Moreover, the ease of administration would potentially reduce variability in performance within an individual due to the ability to administer VisMET multiple times. Remote delivery of VisMET may be conducive for longitudinal assessments to potentially track cognitive impairment over time.

5.2.2 Neurophysiological Mechanisms of Visuospatial Memory Recognition

Recognition memory within the MTL was largely driven by differences in 80-120 Hz power rather slow gamma activity observed in the rodent MTL [126]. A potential explanation for this discrepancy is that many of the MTL electrodes in this study were located in entorhinal cortex and parahippocampal gyrus. Therefore 80-120 Hz activity may reflect neocortical input into rather than activity within hippocampus. This neocortical input may induce slow gamma oscillations within hippocampal circuits and enable the selective encoding and retrieval of the representations

associated with unexpected input. An alternative explanation for the different frequency profiles in rodents and humans are the intrinsic differences in the frequencies mediating memory-related interactions within the brain for the two species [156,157]. Evidence for narrowband slow gamma activity within human hippocampal circuits has been limited. Rather, gamma activity within human hippocampal circuits has largely been characterized by broadband activity with preferences for frequencies ranges outside of the slow gamma band [158]. These studies have primarily been conducted using larger recording electrodes, averaged activity across many trials, and used imprecise methods of hippocampal localization. Determining the human analog of slow gamma oscillations may prove useful in the future as these oscillations have been shown to promote recognition memory, rescue memory deficits in AD models, and reduce AD pathology via microglia activation within rodent hippocampal circuits [112,114,159,160]. Future studies need to be conducted to determine the interactions of these gamma frequencies with precise recordings of the MTL and whether the causal induction of these frequencies can promote recognition memory and reduce AD pathology in humans.

5.3 Conclusions

Our data indicate neocortical-MTL interactions are associated with the successful visuospatial recognition memory and performance on this task is impaired in participants with Alzheimer's disease. In Chapters 2 we show the recognition of visuospatial manipulations relies on a distributed set of electrophysiological interactions between the visuospatial association areas and the MTL, the site of early AD pathology. In Chapter 3 and 4, eye movement recordings from both the desktop and tablet show that participants with MCI and AD spend significantly less time viewing

visuospatial manipulations. Importantly, the performance on this task showed significant variability across healthy controls. These studies suggest that visuospatial memory performance may become impaired early in the disease process and identify electrophysiological targets to improve visuospatial memory performance in humans. Future studies investigating how the molecular mechanisms underlying AD manifest in electrophysiological dysfunction and visuospatial memory impairment could provide novel diagnostics and therapeutics for alleviating the AD crisis.

1. Jack, C. R., Jr. *et al*. Evidence for Ordering of Alzheimer Disease Biomarkers. *Arch. Neurol.* (2011) doi:10.1001/archneurol.2011.183.
2. Serrano-Pozo, A., Frosch, M. P., Masliah, E. & Hyman, B. T. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* (2011) doi:10.1101/cshperspect.a006189.
3. Glenner, G. G. & Wong, C. W. Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* (1984) doi:10.1016/S0006-291X(84)80190-4.
4. Selkoe, D. J. Alzheimer's disease. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
5. Sherrington, R. *et al*. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* (1995) doi:10.1038/375754a0.
6. Scheuner, D. *et al*. Secreted amyloid β-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat. Med.* (1996) doi:10.1038/nm0896-864.
7. Goate, A. *et al*. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704–706 (1991).
8. Levy-Lahad, E. *et al*. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* (1995) doi:10.1126/science.7638622.
9. Hardy, J. A. & Higgins, G. A. Alzheimer's disease: The amyloid cascade hypothesis. *Science* (1992) doi:10.1126/science.1566067.
10. Jack, C. R. *et al*. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia* (2018) doi:10.1016/j.jalz.2018.02.018.
11. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Tredici, K. Staging of Alzheimer

disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol*. **112**, 389–404 (2006).

12. Braak, H. & Braak, E. Staging of alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* **16**, 271–278 (1995).
13. Braak, H., Thal, D. R., Ghebremedhin, E. & Del Tredici, K. Stages of the Pathologic Process in Alzheimer Disease: Age Categories From 1 to 100 Years. *J. Neuropathol. Exp. Neurol*. **70**, 960–969 (2011).
14. Nelson, P. T. *et al*. Correlation of alzheimer disease neuropathologic changes with cognitive status: A review of the literature. *Journal of Neuropathology and Experimental Neurology* (2012) doi:10.1097/NEN.0b013e31825018f7.
15. Jack, C. R. *et al*. Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *The Lancet Neurology* (2013) doi:10.1016/S1474-4422(12)70291-0.
16. Nelson, P. T. *et al*. Alzheimer's disease is not 'brain aging': Neuropathological, genetic, and epidemiological human studies. *Acta Neuropathologica* (2011) doi:10.1007/s00401-011-0826-y.
17. Jack, C. R. *et al*. Associations of Amyloid, Tau, and Neurodegeneration Biomarker Profiles with Rates of Memory Decline among Individuals Without Dementia. *JAMA - Journal of the American Medical Association* (2019) doi:10.1001/jama.2019.7437.
18. McKhann, G. M. *et al*. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia* (2011) doi:10.1016/j.jalz.2011.03.005.

19. Nasreddine, Z. S. *et al*. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
20. Petersen, R. C. *et al*. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201–209 (2010).
21. Goldstein, F. C., Milloy, A., Loring, D. W. & for the Alzheimer's Disease Neuroimaging Initiative. Incremental Validity of Montreal Cognitive Assessment Index Scores in Mild Cognitive Impairment and Alzheimer Disease. *Dement. Geriatr. Cogn. Disord.* **45**, 49–55 (2018).
22. Rey, A. L'examen psychologique dans les cas d'encephalopathie traumatique. *Arch. Psychol.* 286–340 (1941).
23. Grober, E. & Buschke, H. Genuine memory deficits in dementia. *Dev. Neuropsychol.* **3**, 13–36 (1987).
24. Lemos, R., Duro, D., Simões, M. R. & Santana, I. The free and cued selective reminding test distinguishes frontotemporal dementia from Alzheimer's disease. *Arch. Clin. Neuropsychol.* **29**, 670–679 (2014).
25. Grober, E., Sanders, A. E., Hall, C. & Lipton, R. B. Free and cued selective reminding identifies very mild dementia in primary care. *Alzheimer Dis. Assoc. Disord.* **24**, 284–290 (2010).
26. Grober, E., Lipton, R. B., Hall, C. & Crystal, H. Memory impairment on free and cued selective reminding predicts dementia. *Neurology* **54**, 827–832 (2000).
27. Ivanoiu, A. *et al*. Memory evaluation with a new cued recall test in patients with mild cognitive impairment and Alzheimer's disease. *J. Neurol.* **252**, 47–55 (2005).
28. Chang, Y. L. *et al*. Brain substrates of learning and retention in mild cognitive impairment

diagnosis and progression to Alzheimer's disease. *Neuropsychologia* **48**, 1237–1247 (2010).

29. Jansen, W. J. *et al*. Prevalence of cerebral amyloid pathology in persons without dementia: A meta-analysis. *JAMA - Journal of the American Medical Association* (2015) doi:10.1001/jama.2015.4668.
30. Ricci, M., Graef, S., Blundo, C. & Miller, L. A. Using the Rey Auditory Verbal Learning Test (RAVLT) to differentiate Alzheimer's dementia and behavioural variant fronto-temporal dementia. *Clin. Neuropsychol.* **26**, 926–941 (2012).
31. Suzuki, W. A. & Amaral, D. G. Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J. Neurosci.* (1994) doi:10.1523/jneurosci.14-03-01856.1994.
32. Squire, L. R., Stark, C. E. L. & Clark, R. E. The medial temporal lobe. *Annu. Rev. Neurosci.* **27**, 279–306 (2004).
33. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
34. Mishkin, M. Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature* (1978) doi:10.1038/273297a0.
35. Meunier, M., Hadfield, W., Bachevalier, J. & Murray, E. A. Effects of rhinal cortex lesions combined with hippocampectomy on visual recognition memory in rhesus monkeys. *J. Neurophysiol.* **75**, 1190–1205 (1996).
36. Mishkin, M. & Delacour, J. An analysis of short-term visual memory in the monkey. *J. Exp. Psychol. Anim. Behav. Process.* **1**, 326–334 (1975).
37. Mahut, H., Zola-Morgan, S. & Moss, M. Hippocampal resections impair associative learning and recognition memory in the monkey. *J. Neurosci.* **2**, 1214–1220 (1982).

38. Murray, E. A. & Mishkin, M. Severe tactual as well as visual memory deficits follow combined removal of the amygdala and hippocampus in monkeys. *J. Neurosci.* **4**, 2565–2580 (1984).

39. Murray, E. A. & Mishkin, M. Visual recognition in monkeys following rhinal cortical ablations combined with either amygdalectomy or hippocampectomy. *J. Neurosci.* **6**, 1991–2003 (1986).

40. Zola-Morgan, S. & Squire, L. R. Medial temporal lesions in monkeys impair memory on a variety of tasks sensitive to human amnesia. *Behav. Neurosci.* **99**, 22–34 (1985).

41. Nemanic, S., Alvarado, M. C. & Bachevalier, J. The hippocampal/parahippocampal regions and recognition memory: insights from visual paired comparison versus object-delayed nonmatching in monkeys. *J. Neurosci.* **24**, 2013–2026 (2004).

42. Mumby, D. G., Glenn, M. J., Nesbitt, C. & Kyriazis, D. A. Dissociation in retrograde memory for object discriminations and object recognition in rats with perirhinal cortex damage. *Behav. Brain Res.* **132**, 215–226 (2002).

43. Norman, G. & Eacott, M. J. Dissociable effects of lesions to the perirhinal cortex and the postrhinal cortex on memory for context and objects in rats. *Behav. Neurosci.* **119**, 557–566 (2005).

44. Winters, B. D., Forwood, S. E., Cowell, R. A., Saksida, L. M. & Bussey, T. J. Double dissociation between the effects of peri-postrhinal cortex and hippocampal lesions on tests of object recognition and spatial memory: heterogeneity of function within the temporal lobe. *Journal of Neuroscience* **24**, 5901–5908 (2004).

45. Zola, S. M. *et al*. Impaired recognition memory in monkeys after damage limited to the hippocampal region. *J. Neurosci.* **20**, 451–463 (2000).

46. Mumby, D. G. Perspectives on object-recognition memory following hippocampal damage: lessons from studies in rats. *Behav. Brain Res.* **127**, 159–181 (2001).

47. Lee, A. C. H. *et al*. Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus* **15**, 782–797 (2005).

48. Eichenbaum, H. Conscious awareness, memory and the hippocampus. *Nat. Neurosci.* **2**, 775–776 (1999).

49. Knierim, J. J., Neunuebel, J. P. & Deshmukh, S. S. Functional correlates of the lateral and medial entorhinal cortex: objects, path integration and localglobal reference frames. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, (2014).

50. Staresina, B. P., Duncan, K. D. & Davachi, L. Perirhinal and Parahippocampal Cortices Differentially Contribute to Later Recollection of Object- and Scene-Related Event Details. *Journal of Neuroscience* **31**, 8739–8747 (2011).

51. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).

52. Markov, N. T. *et al*. Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* **522**, 225–259 (2014).

53. Rolls, E. T. The mechanisms for pattern completion and pattern separation in the hippocampus. *Front. Syst. Neurosci.* **7**, 74 (2013).

54. Kitamura, T. *et al*. Engrams and circuits crucial for systems consolidation of a memory. *Science* vol. 356 73–78 (2017).

55. Josselyn, S. A. & Tonegawa, S. Memory engrams: Recalling the past and imagining the future. *Science* **367**, (2020).

56. Buzsáki, G. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory

and planning. *Hippocampus* **25**, 1073–1188 (2015).

57. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
58. Tonegawa, S., Morrissey, M. D. & Kitamura, T. The role of engram cells in the systems consolidation of memory. *Nat. Rev. Neurosci.* **19**, 485–498 (2018).
59. Norman, Y. *et al.* Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science* **365**, (2019).
60. Sperling, R. A. *et al.* Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia* **7**, 280–292 (2011).
61. von Helmholtz, H. *Handbuch der physiologischen Optik*. (Voss, 1867).
62. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
63. Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).
64. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
65. Keller, G. B. & Mrsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**, 424–435 (2018).
66. Ryan, J., Althoff, R., Whitlow, S. & Cohen, N. Amnesia is a deficit in relational memory. *Psychological science : a journal of the American Psychological Society / APS* **11**, 454–461

(2000).

67. Smith, C. N., Hopkins, R. O. & Squire, L. R. Experience-dependent eye movements, awareness, and hippocampus-dependent memory. *J. Neurosci.* **26**, 11304–11312 (2006).
68. Smith, C. N. & Squire, L. R. Experience-Dependent Eye Movements Reflect Hippocampus-Dependent (Aware) Memory. *Journal of Neuroscience* **28**, 12825–12833 (2008).
69. Smith, C. N. & Squire, L. R. Awareness of what is learned as a characteristic of hippocampus-dependent memory. *Proceedings of the National Academy of Sciences* **115**, 11947–11952 (2018).
70. Freitas Pereira, M. L. G., von Zuben A Camargo, M. Z., Aprahamian, I. & Forlenza, O. V. Eye movement analysis and cognitive processing: Detecting indicators of conversion to Alzheimer's disease. *Neuropsychiatric Disease and Treatment* (2014) doi:10.2147/NDT.S55371.
71. Whitehead, J. C., Gambino, S. A., Richter, J. D. & Ryan, J. D. Focus group reflections on the current and future state of cognitive assessment tools in geriatric health care. *Neuropsychiatr. Dis. Treat.* **11**, 1445–1466 (2015).
72. Whitehead, J. C. *et al*. Portable eyetracking-based assessment of memory decline. *J. Clin. Exp. Neuropsychol.* (2018) doi:10.1080/13803395.2018.1444737.
73. Crutcher, M. D. *et al*. Eye tracking during a visual paired comparison task as a predictor of early dementia. *Am. J. Alzheimers. Dis. Other Demen.* **24**, 258–266 (2009).
74. Zola, S. M., Manzanares, C. M., Clopton, P., Lah, J. J. & Levey, A. I. A behavioral task predicts conversion to mild cognitive impairment and Alzheimer's disease. *Am. J. Alzheimers. Dis. Other Demen.* **28**, 179–184 (2013).
75. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193

(1954).

76. Barlow, H. B. & Others. Possible principles underlying the transformation of sensory messages. *Sensory communication* **1**, 217–234 (1961).
77. Srinivasan, M. V., Laughlin, S. B. & Dubs, A. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* **216**, 427–459 (1982).
78. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
79. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* **2**, 338–347 (1998).
80. Spratling, M. W. A review of predictive coding algorithms. *Brain Cogn.* **112**, 92–97 (2017).
81. Bastos, A. M. *et al*. Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
82. Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
83. Kersten, D. Predictability and redundancy of natural images. *J. Opt. Soc. Am. A* **4**, 2395–2400 (1987).
84. Atick, J. J. & Redlich, A. N. What Does the Retina Know about Natural Scenes? *Neural Comput.* **4**, 196–210 (1992).
85. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
86. Tulving, E. Episodic memory: from mind to brain. *Annu. Rev. Psychol.* **53**, 1–25 (2002).
87. Maviel, T., Durkin, T. P., Menzaghi, F. & Bontempi, B. Sites of neocortical reorganization critical for remote spatial memory. *Science* **305**, 96–99 (2004).

88. Burwell, R. D., Bucci, D. J., Sanborn, M. R. & Jutras, M. J. Perirhinal and postrhinal contributions to remote memory for context. *J. Neurosci.* **24**, 11023–11028 (2004).

89. Cowansage, K. K. *et al*. Direct reactivation of a coherent neocortical memory of context. *Neuron* **84**, 432–441 (2014).

90. Xie, H. *et al*. In vivo imaging of immediate early gene expression reveals layer-specific memory traces in the mammalian brain. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2788–2793 (2014).

91. Brodt, S. *et al*. Fast track to the neocortex: A memory engram in the posterior parietal cortex. *Science* **362**, 1045–1048 (2018).

92. Lesburguères, E. *et al*. Early tagging of cortical networks is required for the formation of enduring associative memory. *Science* **331**, 924–928 (2011).

93. Yaffe, R. B. *et al*. Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18727–18732 (2014).

94. Vaz, A. P., Inati, S. K., Brunel, N. & Zaghloul, K. A. Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. *Science* **363**, 975–978 (2019).

95. Trotta, M. S. *et al*. Surface based electrode localization and standardized regions of interest for intracranial EEG. *Hum. Brain Mapp.* **39**, 709–721 (2018).

96. Desikan, R. S. *et al*. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).

97. Wittig, J. H., Jr, Jang, A. I., Cocjin, J. B., Inati, S. K. & Zaghloul, K. A. Publisher Correction: Attention improves memory by suppressing spiking-neuron activity in the

human anterior temporal lobe. *Nat. Neurosci.* **22**, 143 (2019).

98. Nunez, P. L. & Srinivasan, R. *Electric Fields of the Brain: The Neurophysics of EEG*. (Oxford University Press, 2006).
99. Kovach, C. K. *et al*. Manifestation of ocular-muscle EMG contamination in human intracranial recordings. *Neuroimage* **54**, 213–233 (2011).
100. Staresina, B. P. *et al*. Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nat. Neurosci.* **18**, 1679–1686 (2015).
101. Salvucci, D. D. & Goldberg, J. H. Identifying fixations and saccades in eye-tracking protocols. in *Proceedings of the symposium on Eye tracking research & applications - ETRA '00* 71–78 (2000).
102. Addison, P. S. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, Second Edition*. (CRC Press, 2017).
103. Mitra, P. *Observed Brain Dynamics*. (Oxford University Press, 2007).
104. Cohen, M. X. *Analyzing Neural Time Series Data: Theory and Practice*. (MIT Press, 2014).
105. Zaveri, H. P., Duckrow, R. B. & Spencer, S. S. On the use of bipolar montages for time-series analysis of intracranial electroencephalograms. *Clin. Neurophysiol.* **117**, 2102–2108 (2006).
106. Morris, G., Arkadir, D., Nevet, A., Vaadia, E. & Bergman, H. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* **43**, 133–143 (2004).
107. Steinmetz, P. N. *et al*. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* **404**, 187–190 (2000).

108. Jang, A. I., Wittig, J. H., Jr, Inati, S. K. & Zaghloul, K. A. Human Cortical Neurons in the Anterior Temporal Lobe Reinstate Spiking Activity during Verbal Memory Retrieval. *Curr. Biol.* **27**, 1700–1705.e5 (2017).

109. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

110. Colgin, L. L. *et al*. Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* **462**, 353–357 (2009).

111. Zheng, C., Bieri, K. W., Hwaun, E. & Colgin, L. L. Fast Gamma Rhythms in the Hippocampus Promote Encoding of Novel Object-Place Pairings. *eNeuro* **3**, (2016).

112. Trimper, J. B., Galloway, C. R., Jones, A. C., Mandi, K. & Manns, J. R. Gamma Oscillations in Rat Hippocampal Subregions Dentate Gyrus, CA3, CA1, and Subiculum Underlie Associative Memory Encoding. *Cell Rep.* **21**, 2419–2432 (2017).

113. Zheng, C., Bieri, K. W., Hsiao, Y.-T. & Colgin, L. L. Spatial Sequence Coding Differs during Slow and Fast Gamma Rhythms in the Hippocampus. *Neuron* **89**, 398–408 (2016).

114. Etter, G. *et al*. Optogenetic gamma stimulation rescues memory impairments in an Alzheimer's disease mouse model. *Nat. Commun.* **10**, 5322 (2019).

115. Staresina, B., Wimber, M. & Hanslmayr, S. Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. *Proceedings of the* (2019).

116. Chapeton, J. I., Haque, R., Wittig, J. H., Inati, S. K. & Zaghloul, K. A. Large-Scale Communication in the Human Brain Is Rhythmically Modulated through Alpha Coherence. *Current Biology* vol. 29 2801–2811.e5 (2019).

117. Khodagholy, D., Gelinas, J. N. & Buzsáki, G. Learning-enhanced coupling between ripple

oscillations in association cortices and hippocampus. *Science* **358**, 369–372 (2017).

118. Arnal, L. H. & Giraud, A.-L. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* **16**, 390–398 (2012).

119. Hosoya, T., Baccus, S. A. & Meister, M. Dynamic predictive coding by the retina. *Nature* **436**, 71–77 (2005).

120. Spratling, M. W. Predictive coding as a model of response properties in cortical area V1. *J. Neurosci.* **30**, 3531–3543 (2010).

121. Day, M., Langston, R. & Morris, R. G. M. Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature* **424**, 205–209 (2003).

122. Nakazawa, K. *et al*. Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. *Neuron* **38**, 305–315 (2003).

123. Bieri, K. W., Bobbitt, K. N. & Colgin, L. L. Slow and Fast Gamma Rhythms Coordinate Different Spatial Coding Modes in Hippocampal Place Cells. *Neuron* **82**, 670–681 (2014).

124. Bragin, A. *et al*. Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* **15**, 47–60 (1995).

125. Carr, M. F., Karlsson, M. P. & Frank, L. M. Transient slow gamma synchrony underlies hippocampal memory replay. *Neuron* **75**, 700–713 (2012).

126. Colgin, L. L. & Moser, E. I. Gamma oscillations in the hippocampus. *Physiology* **25**, 319–329 (2010).

127. Colgin, L. L. Do slow and fast gamma rhythms correspond to distinct functional states in the hippocampal network? *Brain Research* vol. 1621 309–315 (2015).

128. Bastos, A. M. *et al*. Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390–401 (2015).

129. Michalareas, G. *et al*. Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron* **89**, 384–397 (2016).

130. Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J. & Desimone, R. Laminar differences in gamma and alpha coherence in the ventral stream. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11262–11267 (2011).

131. van Kerkoerle, T. *et al*. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14332–14341 (2014).

132. Albert, M. S. *et al*. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia* vol. 7 270–279 (2011).

133. Dubois, B. *et al*. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's and Dementia* vol. 12 292–323 (2016).

134. Mortamais, M. *et al*. Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimers. Dement*. 1–25 (2016).

135. Hannula, D. E., Tranel, D. & Cohen, N. J. The Long and the Short of It: Relational Memory Impairments in Amnesia, Even at Short Lags. *Journal of Neuroscience* **26**, 8352–8359 (2006).

136. Ryan, J. D. & Cohen, N. J. Processing and short-term retention of relational information in amnesia. *Neuropsychologia* **42**, 497–511 (2004).

137. Nilsson, L.-G. *et al*. Betula: A Prospective Cohort Study on Memory, Health and Aging. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* **11**, 134–148 (2004).

138. Park, D. C. *et al*. Mediators of Long-Term Memory Performance Across the Life Span. *Psychol. Aging* **II**, 621–637 (1996).

139. Park, D. C. *et al*. Models of visuospatial and verbal memory across the adult life span. *Psychol. Aging* **17**, 299–320 (2002).

140. Hedden, T. & Gabrieli, J. D. E. Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci*. **5**, 87–96 (2004).

141. Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U. & Bäckman, L. Memory aging and brain maintenance. *Trends Cogn. Sci*. **16**, 292–305 (2012).

142. Reagh, Z. M. *et al*. Greater loss of object than spatial mnemonic discrimination in aged adults. *Hippocampus* **26**, 417–422 (2016).

143. Yassa, M. A., Mattfeld, A. T., Stark, S. M. & Stark, C. E. L. Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences* **108**, 8873–8878 (2011).

144. Yassa, M. A., Muftuler, L. T. & Stark, C. E. L. Ultrahigh-resolution microstructural diffusion tensor imaging reveals perforant path degradation in aged humans in vivo. *Proceedings of the National Academy of Sciences* **107**, 12687–12691 (2010).

145. Yassa, M. A. *et al*. High-resolution structural and functional MRI of hippocampal CA3 and dentate gyrus in patients with amnestic Mild Cognitive Impairment. *Neuroimage* **51**, 1242–1252 (2010).

146. Haque, R. U. *et al*. VisMET: a passive, efficient, and sensitive assessment of visuospatial memory in healthy aging, mild cognitive impairment, and Alzheimer's disease. *Learn. Mem*. **26**, 93–100 (2019).

147. Krafka, K. *et al*. *Eye Tracking for Everyone*. http://gazecapture.csail.mit.edu.

148. Borland, E. *et al*. The Montreal Cognitive Assessment: Normative Data from a Large Swedish Population-Based Cohort. *J. Alzheimers. Dis.* **59**, 893–901 (2017).

149. King, D. E. *Dlib-ml: A Machine Learning Toolkit*. vol. 10 1755–1758 http://jmlr.csail.mit.edu/papers/volume10/king09a/king09a.pdf (2009).

150. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

151. Kazemi, V. & Sullivan, J. One millisecond face alignment with an ensemble of regression trees. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1867–1874 (2014).

152. Soukupová, T. *Real-Time Eye Blink Detection using Facial Landmarks*. http://vision.fe.uni-lj.si/cvww2016/proceedings/papers/05.pdf (2016).

153. Ooms, K., Dupont, L., Lapon, L. & Popelka, S. Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups. *J. Eye Mov. Res.* **8**, (2015).

154. Seyfried, N. T. *et al*. A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Systems* (2017) doi:10.1016/j.cels.2016.11.006.

155. Wingo, A. P. *et al*. Large-scale proteomic analysis of human brain identifies proteins associated with cognitive trajectory in advanced age. *Nat. Commun.* (2019) doi:10.1038/s41467-019-09613-z.

156. Jacobs, J. Hippocampal theta oscillations are slower in humans than in rodents: implications for models of spatial navigation and memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130304 (2014).

157. Buzsáki, G., Logothetis, N. & Singer, W. Scaling brain size, keeping timing: evolutionary

preservation of brain rhythms. *Neuron* **80**, 751–764 (2013).

158. Burke, J. F., Ramayya, A. G. & Kahana, M. J. ScienceDirect Human intracranial high-frequency activity during memory processing : neural oscillations or stochastic volatility ? *Curr. Opin. Neurobiol.* **31**, 104–110 (2015).

159. Bass, D. I. & Manns, J. R. Memory-Enhancing Amygdala Stimulation Elicits Gamma Synchrony in the Hippocampus. *Behav. Neurosci.* **129**, 244–256 (2015).

160. Iaccarino, H. F. *et al*. Gamma frequency entrainment attenuates amyloid load and modifies microglia. *Nature* **540**, 230–235 (2016).