UNIVERSITY OF CAMBRIDGE

Neural Event Segmentation in Mild Cognitive Impairment: Hippocampal Boundary Attenuation and Altered Cortical Event Dynamics

Marcello Pepe

Magdalene College

August 2025

This thesis is submitted for the degree of

Master of Philosophy

Preface

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specifically indicated in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee. Generative AI tools were used for language and clarity editing.

Acknowledgements

I am grateful to Rik Henson for insightful supervision and for sharpening my rigor - continually asking why each step is necessary and what question it answers - while remaining consistently considerate and generous with his help. I thank Petar Raykov for patient and perceptive day-to-day guidance and Linda Geerligs for helpful collaboration and advice. I am indebted to David Nesbitt for collecting the dataset used here as part of his PhD. I also acknowledge the MRC Cognition and Brain Sciences Unit for institutional support and the Cam-CAN team for providing control data. Finally, my thanks go to the participants who made this research possible.

Abstract

Everyday memory failures in ageing and mild cognitive impairment (MCI) may arise from disrupted event segmentation, i.e., failures to appropriately parse continuous experience into discrete episodes. In this thesis, we focus on two complementary systems: the hippocampus, which shows selective, boundary-evoked responses at event offsets (Ben-Yakov & Henson, 2018) that decline with age (Reagh et al., 2020); and a posterior-medial cortical hub (angular gyrus) that maintains stable situation models over long timescales (Ranganath & Ritchey, 2012; Baldassano et al., 2017; Geerligs et al., 2022). Leveraging naturalistic fMRI (an 8-min Hitchcock film) in 54 older adults (18 with MCI), we examined how each of these systems differs in MCI.

In Chapter 3, using independently-defined event boundaries, the MCI group showed a reduced boundary response relative to controls in posterior hippocampus (pHPC). This deficit persisted after accounting for hippocampal volume and for V1 activity at boundaries, suggesting that it provides complementary functional information. A whole-brain analysis revealed additional boundary-response deficits across posterior-medial network hubs (e.g., PCC/precuneus, angular gyrus), indicating a network-level boundary-processing impairment.

Chapter 4 evaluated whether MCI patients express a distinct cortical event structure, or largely share the control structure but implement it with altered timing/strength. This used a method called Greedy State Boundary Search (GSBS) to identify changes in multivoxel patterns in angular gyrus (and V1). We show that patients largely share the Control group's event structure, but implement it with subtle, systematic timing shifts and weaker boundary transitions. In contrast, GSBS failed to identify stable multivoxel states in pHPC for either group, consistent with its role in transient, boundary-locked dynamics, rather than sustained state representations.

Taken together, a degraded cortical situation model and an attenuated hippocampal "save" signal may provide a neural account of the fragmented recall seen in MCI, indicating that a breakdown in event segmentation is a key contributor to everyday episodic memory impairment. To our knowledge, this study provides the first evidence of neural event segmentation impairments in MCI.

Table of Contents

1 Introduction

1.1 Episodic Memory Failure in Ageing and MCI

Episodic memory is the fundamental cognitive system that allows us to mentally travel back in time to re-experience past events (Tulving, 1983, 2002). This capacity is central for navigating daily life, from mundane tasks like cooking to complex social interactions like following a conversation or a television programme (Zacks et al., 2006). Normal cognitive ageing is characterized by declines in processing speed, attentional control, and, most prominently, episodic memory (Park & Reuter-Lorenz, 2009; Balota et al., 2000). These deficits are amplified in Mild Cognitive Impairment (MCI), a critical transitional state between healthy ageing and dementia, particularly Alzheimer's Disease (AD) (Petersen, 2000; Flicker et al., 1991). MCI is defined as a subjective memory complaint corroborated by objective memory impairment, while general cognitive function and daily living activities remain largely preserved (Petersen et al., 1999). MCI affects 12-18 % of adults over 60, and around 10-15 % of those affected develop dementia each year. Furthermore, one-third of individuals whose MCI stems from Alzheimer's pathology convert within five years (Alzheimer's Association, 2025). The MCI stage is therefore important to study, because it provides a critical window for detecting early functional brain changes that underlie the irreversible deficits seen in AD, at a time when patients are still largely able to carry out everyday activities.

In both healthy ageing and AD, the most noticeable deficits are with respect to recollecting very recent, everyday experiences, more so than word-finding or remote autobiographical recall (Balota et al., 2000; Greene et al., 1996). Accordingly, ageing and AD seem to disrupt the adaptive encoding of ongoing activity (Balota et al., 2000; Hodges, 2000). Failure to register events in the recent past makes it harder to keep track of what is currently happening. In line with this, dementia patients often lose orientation in time and space (Giannakopoulos et al., 2000; Robert et al., 2003). One hypothesis is that these impairments stem from a failure to appropriately parse the continuous stream of experience into meaningful, discrete events, a process known as event segmentation (Zacks & Tversky, 2001). By investigating how the neural mechanisms of event segmentation are compromised in MCI, this thesis aims to improve our understanding of cognitive decline, providing a potential target for future diagnostic and therapeutic strategies.

1.2 Event Segmentation Theory (EST): A Cognitive Framework

Event Segmentation Theory (EST) is the dominant theoretical framework for explaining this process of event segmentation (Zacks et al., 2007; Zacks & Swallow, 2007). In short, EST proposes that the brain actively maintains a working model of the current situation (termed event model) that represents what is happening now. This moment-to-moment model is informed by stored event schemas (sometimes called scripts), which are abstract knowledge structures about how a given class of events typically unfolds (Schank & Abelson, 1977; Radvansky & Zacks, 2014). By filling schemas with the perceptual details of the ongoing situation, the event model generates predictions about what will happen next. When incoming information violates these predictions, a spike in prediction error triggers an update of the event model. This is the moment a new event is perceived to begin and marks an event boundary. Behavioural evidence for EST is growing: people show high agreement on where event boundaries occur in a continuous stream (Zacks et al., 2006). This segmentation process also has measurable consequences for memory: information is bound more strongly within an event than across events (Ezzyat & Davachi, 2011; DuBrow & Davachi, 2013). Furthermore, information from a previous event becomes less accessible in working memory immediately after a boundary is crossed, as if the mind has ‘closed a file’ on the completed event (Swallow et al., 2009; Radvansky & Zacks, 2017).

1.3 The Neural Basis of Event Segmentation

The neural implementation of event segmentation has been studied extensively. Converging evidence indicates that event segmentation relies on at least two partially dissociable yet mutually dependent neural subsystems. Boundary-evoked hippocampal signals often coincide with abrupt pattern shifts in the posterior medial network (PMN) and with spikes in functional connectivity between the two systems (Baldassano et al., 2017; Barnett et al., 2024; Liu et al., 2022). This has been framed as transient indexing of events (hippocampus) and stable/sustained situation modelling (PMN) (Chen & Bornstein, 2024; Cooper et al., 2021).

Hippocampal Boundary-Evoked Activity

More specifically, fMRI studies show transient increases in hippocampal activity precisely when observers mark event boundaries (Ben-Yakov & Henson, 2018). This

boundary-evoked hippocampal activity has been interpreted as a 'save' signal that reflects the encoding of a compressed representation of the just-ended event. Evidence for this comes from studies showing that film clips that are later remembered, compared to those that are later forgotten, are associated with greater hippocampal activity at the end of (but not throughout) the encoding of those clips (Ben-Yakov & Dudai, 2011). Additionally, Swallow et al. (2011) observed hippocampal involvement during memory retrieval across event boundaries, but not during retrieval occurring within an ongoing event. Extending this link, Baldassano et al. (2017) showed that a larger hippocampal BOLD response at the end of a high-level event was associated with more extensive reinstatement of the event's multivoxel pattern during later free recall, directly tying the boundary burst to episodic storage. Critically, Barnett et al. (2024) demonstrate that stronger functional coupling between the hippocampus and the PMN at the event offset predicts richer, more detailed recollection of the preceding scene after a two-day delay. This indicates that boundary-locked hippocampal-cortical interaction is a key mechanism for stabilizing episodic memories over time.

Stable Event Models in the PMN

In contrast to the transient boundary-evoked hippocampal activity, cortical activity is thought to maintain stable situation models, updating only when a boundary is reached. The PMN, a network encompassing core Default Mode Network (DMN) regions like the angular gyrus (AG) and posterior cingulate/precuneus (PCC), is theorized to be the primary location these stable states. The PMN is specialized for constructing and maintaining coherent models of ongoing situations, and is highly sensitive to the structure of those situations (Ranganath & Ritchey, 2012; Ritchey, Libby & Ranganath, 2015). Furthermore, Baldassano et al. (2017) have shown that the PMN plays a pivotal role in driving the hippocampal activity at event boundaries. The neural correlates of this process can be examined using data-driven state detection methods like Hidden-Markov Models (HMM) and the more recent Greedy State Boundary Search (GSBS) algorithm, which identify moments of abrupt shifts in multivoxel activity patterns without relying on external annotations (Baldassano et al., 2017; Geerligs et al., 2021). The neural states corresponding to stable multivariate patterns are organized in a partially nested temporal hierarchy, with durations progressively lengthening from short states in sensory regions like the primary visual cortex (V1) to long, integrative states in high-order association cortices (Baldassano et al., 2017; Geerligs et al., 2022). These data-driven boundaries identified in high-level regions like the AG and the posterior medial cortex (PMC) align well with subjectively perceived event boundaries, underscoring their validity

(Baldassano et al., 2017). More recent whole-brain analyses confirm that regions like the PCC and medial prefrontal cortex (mPFC) sit at the top of this neural hierarchy, and that boundaries with larger pattern shifts (i.e., higher boundary strength) are more likely to align with subjectively perceived boundaries (Geerligs et al., 2022).

1.4 Deficits in Event Processing in Ageing and MCI

The event-segmentation account and evidence for its neural implementation outlined in the sections above provide a useful framework for understanding the episodic memory deficits characteristic of cognitive ageing and dementia. A growing body of evidence suggests that the ability to segment experience into meaningful events declines with age and is further compromised in clinical populations.

Behaviourally, event segmentation becomes less consistent and more idiosyncratic in older adults. When prompted to mark event boundaries in a movie, older adults show significantly less agreement with group norms compared to younger adults (Zacks et al., 2006; Kurby & Zacks, 2011). The degree to which an individual's segmentation aligns with the norm predicts their subsequent memory for event content, even when controlling for general cognitive abilities like working memory, processing speed and executive function (Sargent et al., 2013). This suggests that normative segmentation is a fundamental and adaptive cognitive skill. This deficit is further exacerbated in clinical populations, with individuals with mild dementia of the Alzheimer's type showing even less segmentation agreement compared to healthy older adults (Zacks et al., 2006). Beyond simply misplacing boundaries, older adults also show a deficit in perceiving the hierarchical structure of events. They show poorer alignment between coarse-grained events (e.g., "washing hands") and the fine-grained sub-events (e.g., "putting soap on hands") that comprise them (Kurby & Zacks, 2011). Furthermore, in paradigms testing event integration, both individuals with MCI and individuals with AD showed impairments in arranging scrambled sub-events to form coherent macro-events (Kokje et al., 2021). Critically, the MCI group's performance was nearly identical to that of the AD group, suggesting that these event processing deficits emerge early in the course of the disease. This further underscores the importance of studying the MCI stage for detecting early functional brain changes associated with cognitive decline.

Neural findings corroborate the observed behavioural impairments in event segmentation. A large-scale lifespan study by Reagh et al. (2020) on a subset of the Cam-CAN dataset revealed that boundary-evoked BOLD responses decline with age specifically in

the posterior hippocampus (pHPC) and PMN regions like the AG and PMC, with no significant age-related changes observed in the anterior hippocampus or V1. Furthermore, the magnitude of this pHPC boundary response predicted performance on a separate, general measure of narrative memory, establishing its value as a marker of episodic memory ability.

The cortical systems maintaining stable event models also show age-related dysfunction. In a study of healthy ageing, Geerligs and Campbell (2018) found that Inter-Subject Correlation (ISC) during movie viewing - a proxy for shared information processing - significantly declines with age in the DMN, PMN and medial temporal lobe (MTL). This indicates more idiosyncratic neural responses during in older adults during movie viewing. This pattern of idiosyncratic neural processing is exacerbated in MCI. Raykov (2019) found that individuals with MCI exhibited significantly reduced ISC while they watched short clips compared to both healthy age-matched controls and individuals with subjective memory complaints. In line with the findings in healthy ageing, this deficit was localized in key nodes of the DMN and PMN, including the PMC, ventromedial prefrontal cortex (vmPFC), and anterior temporal poles. Importantly, these idiosyncratic neural responses in the MCI group were accompanied by significantly worse subsequent memory for the videos. This strongly suggests that the event processing deficits observed in healthy ageing become even more pronounced in MCI, positioning them as a critical target for investigation.

These functional deficits are grounded in the known pathophysiology of AD. The disease process classically begins in the medial temporal lobes, causing significant hippocampal atrophy and dysfunction even in the MCI stage (Pennanen et al., 2004; Du et al., 2004). Furthermore, a study by Jones et al. (2015) found that dysfunction within the PMN, specifically its posterior regions, emerges early in the disease process. Critically, this network dysfunction precedes the accumulation of amyloid plaques. Given the central role of the hippocampus in the encoding of events and the PMN in maintaining event models, this provides a strong biological rationale to expect that event segmentation processes will be compromised even in the early clinical stage of MCI, and therefore any deficits have potential as early, functionally-grounded markers of Alzheimer's-related pathology.

1.5 Naturalistic Stimuli

Traditional memory research has historically relied on simple, controlled stimuli like lists of words or static pictures. While these reductionist laboratory paradigms are valuable for

isolating specific cognitive processes, they fail to capture the continuous, multimodal and structured nature of real-world experience. Since the core deficit in MCI seems to be real-word episodic memory failures, which we hypothesize stem from impaired event segmentation, we must use paradigms that better approximate the complexity of everyday life.

Naturalistic stimuli, like movies or narrative stories, are an ideal candidate for this. They provide a controlled, yet ecologically-valid method for studying the continuous processing of complex events. They present a rich and dynamic stream of information that unfolds over time, requiring the brain to actively segment and integrate events just like in daily life. Crucially, naturalistic narrative stimuli are ideal for engaging the cognitive functions and brain networks involved in event segmentation and known to be affected in MCI. For example, temporal scrambling studies show that processing in regions such as the AG and PMC is dependent on the narrative structure being intact (Hasson et al., 2008; Honey et al., 2012; Lerner et al., 2011). Furthermore, MCI patients can show problems in discourse comprehension (Drummond et al., 2015; see also Fraser et al., 2019). More broadly, the comprehension of naturalistic stimuli requires a host of processes beyond just memory, such as prolonged sustained attention (Naci et al., 2014), executive functioning (Cannizzaro et al., 2013), and switching between encoding and retrieval (Hasson et al., 2015). While the predominant difficulty in MCI is episodic memory, attentional/executive weaknesses often emerge under higher task load (Kirova et al., 2015). Because naturalistic narratives engage episodic integration, while also drawing on attention and control, they provide a sensitive stress-test of precisely the systems most relevant to everyday failures in MCI. Finally, narrative comprehension draws on prior knowledge in the form of schemas. In healthy ageing, schematic knowledge is often preserved (Umanath & Marsh, 2014), but in MCI the picture is mixed: patients can show degraded action scripts and semantic weaknesses (e.g., naming/lexical access), while other aspects of semantic memory appear preserved (e.g., autobiographical facts, low-effort decisions) (Jekel et al., 2015; Joubert et al., 2021; Taler et al., 2020; Murphy et al., 2008), supported by compensatory recruitment in default-mode/posterior-temporo-parietal regions (Woodard et al., 2009). Therefore, schema/script problems in MCI might reflect difficulty maintaining and updating the current situation model during continuous narratives, rather than wholesale semantic loss. Together, these points show that naturalistic narrative paradigms are particularly sensitive for revealing cognitive impairments in MCI.

1.6 Thesis Outline

The goal of this thesis is to provide a characterization of the neural mechanisms that underlie event cognition deficits in MCI. To achieve this, it combines two complementary analytical approaches that probe different aspects of the system involved in event cognition and address distinct knowledge gaps.

The first gap concerns the hippocampal boundary activation in MCI. Healthy ageing is associated with a blunted pHPC boundary response (Reagh et al., 2020), but it is an open question whether this response is further degraded in MCI. Our investigation in Chapter 3 specifically targets the pHPC, as this is the region that showed the most robust decline in boundary-evoked activity with age (Reagh et al., 2020). Using subjective event boundaries defined by an independent sample, Chapter 3 asks: 1) is the magnitude of boundary-evoked activity in the pHPC significantly more diminished in MCI patients compared to age-matched healthy controls, 2) is any such attenuated activity simply a consequence of structural atrophy in the pHPC (i.e., less brain tissue to activate), and 3) is this putative hippocampal deficit behaviourally meaningful, as reflected in its relationship with a general measure of episodic memory?

The analysis in Chapter 3 is activation-based. It measures the magnitude of a transient hippocampal response to externally-defined boundaries. While this addresses a potential encoding deficit, it cannot determine if the deficit lies in the very structure of the neural representations between these boundaries. It assumes MCI patients are trying to segment at the same moments but simply produce a weaker signal. This points to the second critical knowledge gap: what if the core deficit is not just a weak response to boundaries, but a failure to construct a coherent event structure in the first place?

To investigate this, Chapter 4 pivots to a multivariate, pattern-based approach that derives boundaries directly from neural signals. We focus on the AG, a PMN hub that integrates information over long timescales (Hasson et al., 2008; Lerner et al., 2011). Its central role in maintaining stable event models makes it a good cortical site to probe for deficits in the integrity of event structure (Baldassano et al., 2017). Using a data-driven event segmentation method (GSBS) (Geerligs et al., 2021), we formalise two competing accounts of the MCI deficit: (i) a “coherent alternative” event structure (patients segment differently from Controls, but consistently as a group), and (ii) a “shared” structure (patients segment like

Controls). We adjudicate among these by combining two complementary tests: a boundary-overlap analysis quantifying temporal correspondence between Control- and MCI-derived boundary templates, and a boundary template-preference test that assesses whether each group's data align more strongly with the Control template, the MCI template, or neither. To characterise how distinctly the AG transitions between states, we use a multivariate measure of neural state boundary strength - the degree of pattern change - and compare this between groups. Finally, to integrate findings across chapters, we apply the pattern-based stable state model to the hippocampus (pHPC), to test the fundamental assumption that cortex maintains stable event states while the hippocampus does not.

2 General Methods

2.1 Participants

A total of 54 older adults were included in the present study: 36 cognitively healthy controls (15 female; mean age = 72.5 years, SD = 9.11) and 18 patients with MCI (5 female; mean age = 71.7 years, SD = 10.67). The groups did not significantly differ in age ($t(29.72) = 0.28$, $p = .778$). A Chi-squared test of independence also showed no significant difference in sex distribution between the control and MCI groups ($\chi^2 = 0.486$, $p = .486$). Across the full sample, participants ranged in age from 53 to 90 years and were native or bilingual speakers of English with the capacity to provide informed consent.

The control group was drawn from the population-based Cambridge Centre for Ageing and Neuroscience (Cam-CAN) cohort (Shafto et al., 2014). Controls were designated as cognitively healthy based on their performance on two widely used screening tools: the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), a brief 30-point questionnaire assessing global cognitive function including orientation, attention, and memory; and the Addenbrooke's Cognitive Examination–Revised (ACE-R; Mioshi et al., 2006), a more comprehensive 100-point assessment that provides a detailed profile across five domains: attention/orientation, memory, verbal fluency, language, and visuospatial skills. To be included as Controls, participants had to score ≥ 26/30 on the MMSE, > 88/100 on the ACE-R total, and > 24/26 on the ACE-R memory subscale.

In contrast, the MCI group was recruited from specialist memory clinics in Cambridge. All patients satisfied the established Petersen diagnostic criteria for MCI (Petersen et al., 1999), which require: (1) a subjective memory complaint, (2) objective evidence of memory impairment, (3) largely preserved general cognitive function, (4) intact activities of daily living, and (5) the absence of dementia. The clinical diagnosis was corroborated, where available, by cerebrospinal-fluid or amyloid-PET biomarkers (15/18 patients had biomarker data: 8 positive, 7 negative).

All participants completed the same three-session, "Cam-CAN Frail" protocol, and only those who completed the movie-viewing fMRI paradigm in Session 3 were included in the present study. General exclusion criteria mirrored those of the original Cam-CAN study, encompassing contraindications to MRI or MEG, major psychiatric or neurological disorders

(with the exception of the MCI diagnosis in the patient group), insufficient English proficiency, and inability to complete the testing sessions. Ethical approval for the Cam-CAN Frail arm was granted by the East of England-Cambridge Central Research Ethics Committee (10/H0308/50), and additional approval for patient recruitment and assessment was obtained from the East of England-Essex Research Ethics Committee (07/H0307/64). All procedures conformed to the Declaration of Helsinki, and written informed consent was obtained from every participant prior to participation.

2.2 Stimulus

Inside the scanner participants passively watched a shortened black-and-white edit of Alfred Hitchcock's *Bang! You're Dead*. This stimulus has been used in previous studies (Hasson et al., 2010; Shafto et al., 2014) and is known to elicit robust, synchronized brain activity across individuals (Geerligs & Campbell, 2018). The full (~25-minute) episode was condensed to 8 minutes to accommodate time constraints in the scanner while preserving the core narrative arc. No overt task was given; participants were instructed simply to attend to the film.

2.3 Cognitive Measures

Episodic memory was assessed using the Logical Memory subtest of the Wechsler Memory Scale (WMS; e.g., Wechsler, 2009) to provide the primary behavioural measure for this study. In this test, two short stories are read aloud to the participant, who is then asked to recall them immediately (Immediate Recall). After a delay of approximately 20-30 minutes, the participant is asked to recall the stories again (Delayed Recall). To create a single, robust index of episodic memory performance, raw scores from the immediate and delayed recall conditions were combined into a composite score using a principal component analysis.

2.4 MRI Data Acquisition

MRI data were collected on a 3T Siemens Prisma (32-channel head coil) at the MRC CBU in Cambridge. Structural data were acquired using a T1-weighted MPRAGE sequence with 1mm^3 isotropic resolution. Functional data were acquired using a T2*-weighted gradient echo-planar imaging (EPI) sequence (TR = 1.12 s, TE = 30 ms, flip angle = 74°, Field of view = 192 × 192 × 120 mm, matrix size = 64 × 64 × 60, multiband acceleration = 4, phase-

encoding direction = anterior → posterior). Each functional volume contained 60 interleaved axial slices in LAS orientation with 2 mm isotropic voxels, and 433 volumes were collected. Slice-timing information was recorded for retrospective correction and the first non-steady-state volume was automatically discarded. Field-maps for distortion correction were not available at the time of analysis.

2.5 Preprocessing

All images were processed with fMRIPrep 24.1.1 (Esteban et al., 2018, 2019; RRID: SCR_016216), a Nipype 1.8.6-based tool (Gorgolewski et al., 2011).

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity using N4BiasFieldCorrection (ANTs 2.5.3), skull-stripped using the antsBrainExtraction.sh workflow, and used as the anatomical reference. The resulting brain-extracted T1w was segmented into cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) using fast (FSL). Cortical surfaces were reconstructed with recon-all (FreeSurfer 7.3.2). The T1w reference was non-linearly normalized to the MNI152NLin2009cAsym standard space using antsRegistration (ANTs).

Functional data preprocessing

Reference volumes were generated and used to estimate head-motion parameters with mcflirt (FSL). The BOLD reference was then co-registered to the T1w reference using boundary-based registration (bbregister, FreeSurfer). The head-motion correction and co-registration transforms were concatenated and applied in a single resampling step into standard space using cubic B-spline interpolation.

A comprehensive set of confound time series was calculated for subsequent analyses. This standard set of confounds, used across all analyses, included the six rigid-body motion parameters, their first temporal derivatives, and the mean time series from white matter (WM) and cerebrospinal fluid (CSF). We then applied a high-pass temporal filter with a cutoff of 0.0078 Hz (128-second period) to remove low-frequency signal drifts. Finally, functional images were spatially smoothed using an isotropic Gaussian kernel of 6.0 mm Full-Width at

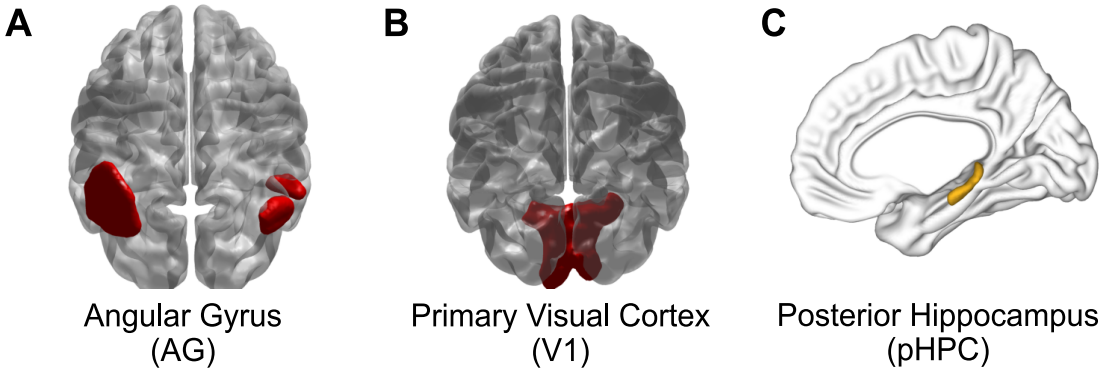
Half-Maximum (FWHM). A data-driven optimization that motivated including WM/CSF regression and spatial smoothing is presented in the Appendix (section 7.1).

2.6 Region of Interest Definition

This thesis investigated neural activity within four a priori regions of interest (ROI), selected to test specific hypotheses about event processing across different levels of the cognitive and neural hierarchy. These regions were the pHPC, the AG, and the primary visual cortex (V1) (see Figure 1). The pHPC was selected as the primary ROI for Chapter 3, motivated by prior findings that boundary-evoked activity is specific to this subregion (Reagh et al., 2020) and its established role in processing contextual event information.

Figure 1

ROIs Used in Subsequent Analyses



Note. Panel (a) displays the AG mask, panel (b) the V1 mask, and panel (c) pHPC mask. Definition of each ROI is explained within each chapter. Colors are arbitrary and do not convey statistical values.

The analysis in Chapter 4 focused on event segmentation in the cortex. The AG, a key hub of the DMN, was chosen as the primary high-level cortical region due to its established role in integrating information over longer timescales and processing event structure (Baldassano et al., 2017).

The V1 was included in both chapters to serve as a consistent low-level sensory control region. By analyzing V1 in parallel with higher-order regions like the pHPC and AG, we can interpret whether observed group differences or segmentation patterns are specific to

high-level associative cortex or reflect more domain-general changes that are also present in early sensory processing.

To suit the different analytical goals of the subsequent chapters, two different ROI masking strategies were employed. Chapter 3 used subject-specific masks to maximize anatomical precision for individual-level modelling, while Chapter 4 used a single common mask to ensure valid group-level comparisons. The specific details of each procedure are described in the methods sections of their respective chapters.

3 Boundary-Evoked Activity Analysis

3.1 Introduction

A growing body of work shows that the hippocampus responds with a burst of activity at neural and subjective event boundaries, and importantly, the magnitude of this neural signal directly predicts subsequent memory performance (Baldassano et al., 2017; Ben-Yakov & Henson, 2018; Ben-Yakov & Dudai, 2011; Reagh et al., 2020). This mechanism appears to be vulnerable to age-related decline, with older adults showing significantly blunted pHPC responses at boundaries compared to the young (Reagh et al., 2020). It remains a critical open question whether this boundary-evoked activity is further degraded in MCI. Given that MCI is defined by hippocampal pathology more advanced than in healthy ageing (Pennanen et al., 2004) and is characterized by general hippocampal hypoactivation during memory tasks (Wang et al., 2016), there is a strong basis to hypothesize a specific deficit in this boundary-encoding process.

Using fMRI data acquired during a naturalistic movie stimulus, this chapter therefore aims to answer three primary questions: (1) Is boundary-evoked neural activity in the pHPC diminished in individuals with MCI compared to healthy controls? (2) Is any reduction in boundary activity in the MCI group simply because of smaller hippocampal volumes (less grey-matter)? (3) Is the magnitude of this activity linked to a general measure of episodic memory, and does this brain-behaviour relationship differ between the groups?

3.2 Methods

3.2.1 ROI Definition

As outlined in Chapter 2, the analyses in this chapter focused on the pHPC and V1. To maximize anatomical precision for the individual-level GLM, this analysis required ROIs that were maximally representative of each participant's unique anatomy within the shared MNI space. Therefore, ROIs were defined in MNI152NLin2009cAsym standard space using a subject-specific gray matter masking procedure. For each participant, a binary gray matter mask was generated by thresholding their T1w-derived gray matter probability map at a value of 0.3. This individual mask was then intersected with anatomical or functional atlas templates.

For the pHPC, a bilateral hippocampal mask was adapted from the segmentations of a previous study (Ritchey et al., 2015). The pHPC mask comprised the combined hippocampal body and tail. The hippocampal segmentations from Ritchey et al. were supplied on a 1 mm grid. For every participant they were therefore aligned to that participant's 2 mm grey-matter map with nearest-neighbour interpolation before the intersection step. After intersection, the mean (±SD) number of voxels of the resulting pHPC ROI in the control group was 427 ± 19 and 383 ± 43 voxels in the MCI group. Independent samples t-tests confirmed that this difference was significant ($t = 5.24$, $p < .001$), consistent with hippocampal atrophy in AD.

The V1 ROI was defined using the Schaefer et al. (2018) 500-parcel, 17-network functional atlas by combining all parcels of the visual network. This template underwent the same subject-specific masking procedure. The mean (±SD) number of voxels of the resulting ROIs was 2432 ± 198 for the control group and 2438 ± 201 for the MCI group. Unlike the pHPC, these V1 volumes did not differ significantly between groups ($t = -0.12$, $p = 0.91$), suggesting that volume loss was specific to medial temporal lobe regions.

3.2.2 Defining Event Boundaries and Within-Event Timepoints

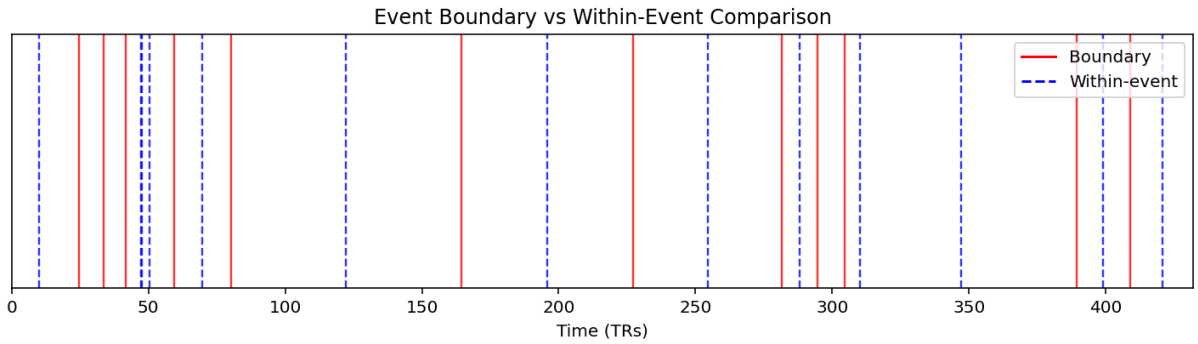
To identify the neural response to event boundaries, we used event boundary annotations derived from the data of Ben-Yakov and Henson (2018). In their study, 16 observers pressed a key whenever they perceived an event to end. The authors created an initial set of consensual boundaries by correcting for reaction times and clustering nearby responses. Building on this, we applied the stricter filtering criteria used by Reagh et al. (2020) to isolate boundaries with the highest level of agreement. Specifically, we selected boundaries from the Ben-Yakov & Henson data that were endorsed by at least half of the observers (8 of 16) within a five-second window, while ensuring a minimum spacing of 6 s between boundaries. This majority-vote rule ensures that every retained moment reflects an event change recognized by most viewers and resulted in a final set of 12 boundaries for our analysis, identical to the set used in the Reagh et al. (2020) study.

To isolate neural activity specifically related to event segmentation, we also sampled an equal number of 'within-event' time-points that lay at least six seconds away from any boundary (Figure 2). These time points were designed to capture brain activity associated with the general processing of the continuous audiovisual stimulus that is not unique to event boundaries. By directly contrasting boundary-evoked activity against within-event activity (boundary > within-event), we can minimize potential MCI-related confounds in overall task

engagement or global shifts in baseline BOLD signal. This approach aims to increase confidence that group differences are specifically time-locked to perceived event boundaries.

Figure 2

Event Boundary Versus Within-Event Time Points Across the Movie Run



Note. Red solid vertical lines denote event boundaries identified in the Cam-CAN movie task, and blue dashed vertical lines mark within-event control time points selected to be at least 6 s (≈ 5 TRs) from any boundary. The horizontal axis is expressed in scanner repetitions and spans the full length of the run. Because the boundary and within-event locations are fixed for every participant, the pattern is identical across subjects. The legend (upper right) reiterates the color coding.

3.2.3 Statistical Analyses

All statistical analyses were conducted using Python 3.7 and its scientific libraries (e.g., Nilearn, Statsmodels, Pingouin). The significance level for all hypothesis tests was set at $\alpha = .05$. The analysis pipeline consisted of two core stages: first-level (subject-level) modelling of individual fMRI data, followed by second-level (group-level) statistical comparisons.

Group Differences in Boundary-Evoked Activity

First, for each participant, a General Linear Model (GLM) was fit to the pre-processed fMRI timeseries at each voxel. This model aimed to isolate the neural signature of processing an event boundary from more general activity related to viewing the movie. Separate regressors were therefore created for the boundary and within-event timepoints, which were modeled as impulse responses (delta functions) convolved with a canonical double-gamma

hemodynamic response function (HRF). Additional confound regressors were added, as detailed in Section 2.5. The primary *T*-contrast of interest was boundary > within-event. The resulting contrast estimates were averaged across voxels within our two ROIs: pHPC and V1.

These contrast values for each participant and ROI were fit by a second, group-level GLM, designed to contrast boundary-related activation between the Control and MCI groups. A participant was excluded only if they were identified as an outlier on both mean Framewise Displacement (FD) and the boundary > within-event contrast values, exceeding 1.5 times the interquartile range (IQR) above the third quartile or below the first quartile. The rationale was to remove only those data points where an extreme neural outcome was most likely attributable to a data quality issue (i.e., motion). This procedure resulted in no participants being excluded.

Our primary hypothesis corresponded to a one-tailed *T*-contrast, testing whether boundary-related activation was greater in the Control than Patient groups. As a sanity check, we also performed one-sample *T*-tests within each group separately, to confirm that boundary-related activations were significantly greater than zero. To test the robustness of any group differences, however, we added progressively more covariates to the GLM (corresponding to an Analysis of Covariance, ANCOVA), as expanded below.

Our first GLM included covariates for age and mean FD (i.e., pHPC_Activation ~ Group + Age + mean_FD). Although mean age did not differ significantly between groups, it was included to reduce residual error due to its known effects on brain activation. Mean FD was also included to control for motion, particularly as it was found to be higher in the MCI group than the Control group (see Results).

To test our second research question - whether group differences in neural activity are driven by atrophy - we included each participant's pHPC volume as a covariate. This volume was defined as the number of voxels within their individual pHPC mask that exceeded the 0.3 grey matter probability threshold.

In a final GLM for pHPC, we added a further covariate of V1 boundary activation for each participant. This was to test whether the hippocampal group difference was independent of any differences between boundary and non-boundaries in low-level visual processing.

To assess the anatomical specificity of our pHPC findings, we applied our first GLM to the V1 control region (i.e., V1_Activation ~ Group + Age + mean_FD). As there was no strong directional hypothesis for this control region, the group effect was assessed using a two-tailed *T*-contrast.

Whole-Brain Confirmatory Analysis

To further assess the anatomical specificity of our findings, we conducted an exploratory whole-brain, group-level analysis. This analysis employed the same first-level contrast maps (boundary > within-event) but performed a *T*-contrast between Control and MCI groups for each voxel separately. This group-level contrast was tested in the context of the primary GLM above, i.e., controlling for Age and mean FD only. The analysis was restricted to ~330,000 voxels within a group-level mask comprising all voxels with data in every participant. To correct for multiple comparisons, we applied False Discovery Rate (FDR) correction at $q < 0.025$, given that we tested for significant effects in both directions.

Brain-Behaviour Analysis: Relationship with Episodic Memory

To investigate the link between brain activity and general memory, we first created a composite "trait memory" score via principal component analysis (PCA) on the raw scores from the immediate and delayed recall conditions of the WMS Logical Memory test. The first principal component was used as the single dependent variable for subsequent analyses.

We then used a multiple linear regression analysis to assess whether activity in the pHPC ROI explained unique variance in memory performance within each group (after accounting for mean group differences), and whether the strength of this brain-behaviour relationship differed between the Control and MCI groups. The primary model was specified as: Memory ~ Group + Activation + Age + Group * Activation. Although this addresses a primary research question, we used a two-tailed test in the absence of a strong directional prediction for the brain-behaviour relationship or its potential modulation by group. A follow-up model including pHPC volume was planned to test the influence of atrophy, conditional on finding a significant brain-behaviour relationship.

Confirmatory Finite Impulse Response (FIR) Analysis

To further investigate the nature of group differences in boundary-evoked activity, a second analysis was conducted using a Finite Impulse Response (FIR) basis set instead of a

canonical HRF within each subject's fMRI timeseries model. The FIR set consisted of 14 consecutive time bins of 1s duration, covering the window from –2 to +12 seconds relative to event onset. This approach allowed us to investigate whether any group differences were due to a weaker response, a delayed response, or a fundamentally different evoked response shape. If the response in the MCI group are just a down-scaled (weaker) version of that in the Control group, for example, then we can be more confident that the differences found with the canonical HRF reflect neural differences, rather than differences in the neurovascular coupling (i.e., HRF shape) between groups.

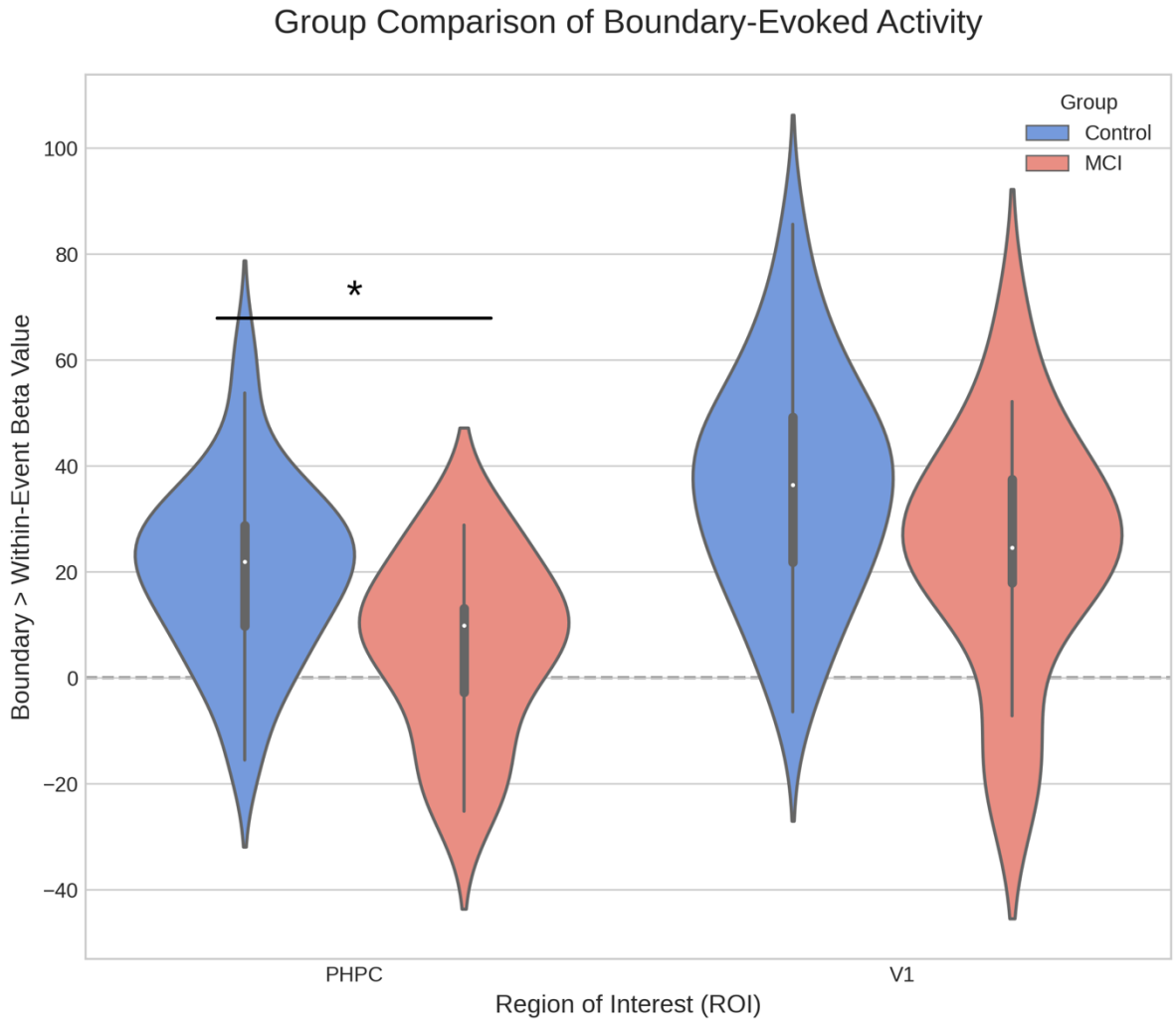
To assess differences in the HRF, the boundary - within difference score was calculated for each FIR time bin. These difference scores were then submitted to a mixed-design ANOVA, with Time (14 levels) as the within-subjects factor and Group (Control vs. MCI) as the between-subjects factor. This allowed us to test for a main effect of Group on the overall response amplitude, and a Group × Time interaction, which would suggest a difference in response shape or timing. As the assumption of sphericity is often violated in time-series data, we applied a Greenhouse-Geisser correction to the degrees of freedom for all effects involving the Time factor.

3.3 Results

The final sample for analysis comprised 36 cognitively healthy controls and 18 patients with MCI. The MCI group tended to move more during the scan, as evidenced by significantly higher FD than controls ($t(23.44) = –2.18$, $p = .040$). Given that motion correction of fMRI data is rarely perfect, this finding reinforced the importance of using mean FD values as a covariate at the group level. A preliminary one-sample t-test confirmed that, as expected, activation in the pHPC was significantly greater for boundaries than non-boundaries within the Control group ($t(35) = 7.20$, $p < .001$, one-tailed), validating that our paradigm successfully elicits the expected neural response in healthy individuals. In contrast, this effect did not reach significance in the MCI group ($t(17) = 1.51$, $p = .074$, one-tailed), a null result that should be interpreted with caution given the smaller patient sample. As illustrated in Figure 3, the distributions of raw boundary activations across pHPC and V1 highlight this divergence; the critical test of the group difference is addressed next.

Figure 3

Group Differences in Boundary-Evoked Activity Across ROIs



Note. Violin plots overlaid with boxplots showing the distribution of boundary-evoked activity (raw boundary > within-event beta values) for the Control and MCI groups across pHPC and V1. Each violin displays the probability density of beta values, with embedded indicators showing the median (white dot), interquartile range (thick dark line), and data range (thin lines). The asterisk indicates a statistically significant group difference in pHPC (one-tailed $p < 0.05$). The MCI group showed significantly reduced boundary-evoked activation in pHPC compared to controls, while no significant group difference was observed in the control region V1, supporting the hypothesis of hippocampal-specific deficits in event boundary processing in cognitive impairment.

3.3.1 Group Differences in Boundary-Evoked Activity

Within the initial GLM that adjusted for Age and mean FD, there were no significant effects of the covariates Age (p = .399) or mean FD (p = .944). Nonetheless, a one-tailed T-contrast showed a significant effect of group (β = –14.61, t(50) = –2.82, p = .004), with greater boundary-activation in the Control group, as expected.

To confirm that the group difference was not simply a consequence of less gray-matter in the MCI group, we added ROI volume as a further covariate. There was no significant effect of pHPC volume on boundary activation (t(49) = –0.81, p = .422). Nonetheless, the group difference remained significant (β = –17.75, t(49) = –2.73, p = .004), suggesting it is not a consequence of hippocampal atrophy in the MCI group.

Finally, we added V1 boundary activation as another covariate to control for visual differences between boundaries and non-boundaries. There was a significant positive relationship between V1 and pHPC activation (t(49) = 2.30, p = .026). Nonetheless, the hippocampal group difference remained significant (β = –11.83, t(49) = –2.31, p = .013), suggesting that it is not a simple effect of visual changes at boundaries, to the extent that such changes are captured by V1 activation.

Within the GLM for the V1 control region that adjusted for Age and mean FD, a two-tailed T-contrast showed no significant effect of group (t(50) = –1.72, p = .092). The lack of a group difference in this sensory area suggests the functional deficit is specific to higher-order regions like the hippocampus.

3.3.2 Whole-Brain Analysis

The whole-brain analysis revealed widespread group differences in boundary-evoked activity. Testing both directions with FDR correction at $q < 0.025$ for each, we found 68,545 voxels showing greater activation in Controls than MCI patients, while only 11,811 voxels showed the opposite pattern (MCI > Control), as shown in Figure 4. Consistent with our ROI findings, hippocampal regions showed robust Control > MCI group differences, with the strongest effects at peak voxels in pHPC (right posterior peak: t = 9.34; left posterior peak: t = 7.38).

Beyond the hippocampus, significant clusters emerged throughout regions associated with the PMN. The largest cluster (497,568 mm³) encompassed PCC and precuneus, with a

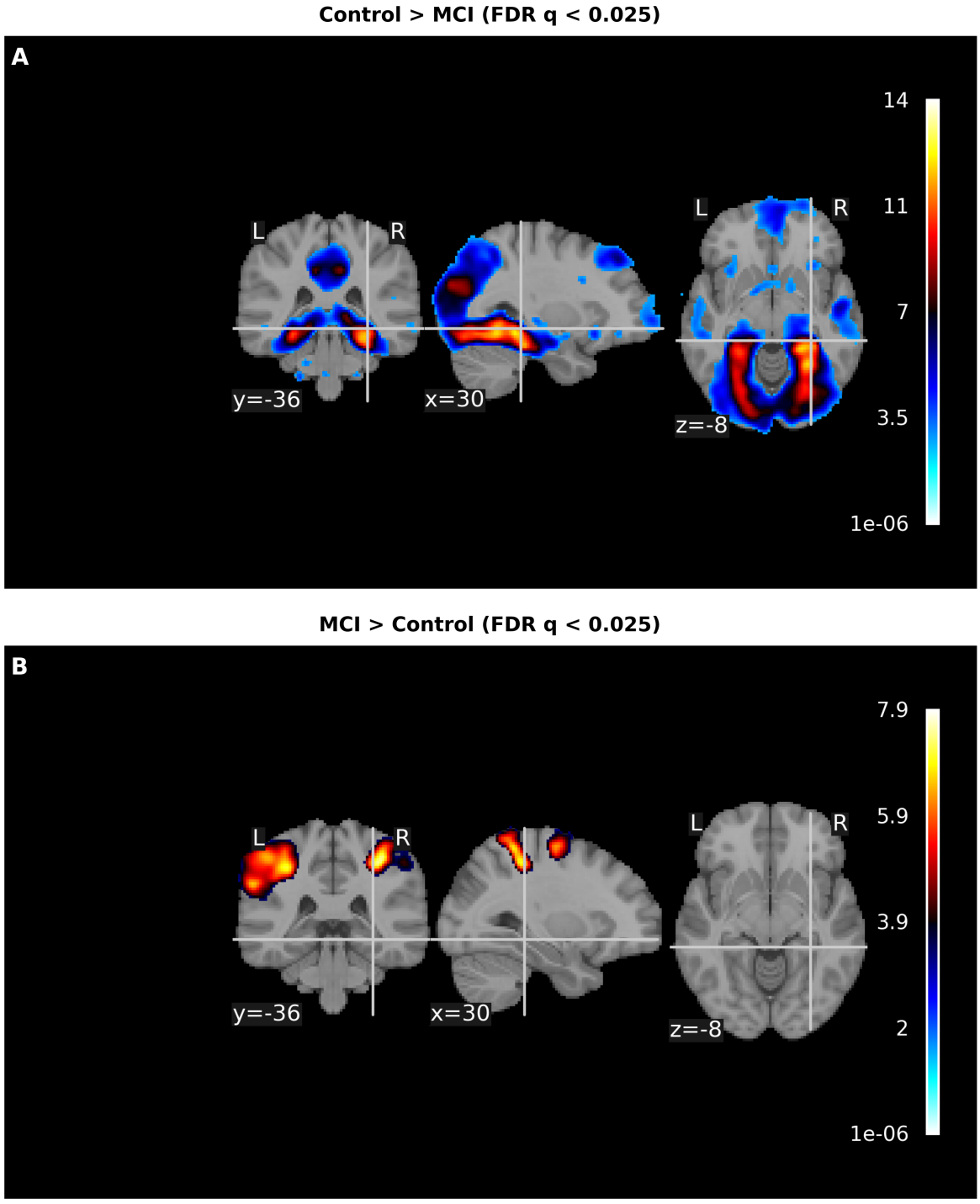
peak *t*-value of 14.02. Additional significant clusters were identified in lateral temporal cortex, AG, and medial prefrontal regions. This widespread pattern suggests the boundary processing deficit in MCI is not isolated to the hippocampus but involves the broader network for event cognition and episodic memory.

In the opposite contrast (MCI > Control), we found a markedly smaller set of 11,811 voxels organized into 17 clusters. These regions were predominantly located in sensorimotor cortex, including bilateral pre- and postcentral gyri (peak $t = 7.89$), superior parietal regions (peak $t = 7.50$), and supplementary motor areas (peak $t = 7.40$).

Notably, several clusters within visual cortex also survived FDR correction in the Control > MCI contrast. In our ROI analysis, V1 showed no significant group difference ($p = .092$). This discrepancy likely reflects the difference between examining peak voxels in the whole-brain analysis versus averaging across all voxels in our ROI analysis, with the latter providing a more conservative and representative measure of regional activity.

Figure 4

Whole-brain boundary-evoked activation differences between Control and MCI groups



Note. (A) Control > MCI: Orthogonal MNI-152 slices display voxels where control participants showed significantly greater boundary-related BOLD signal than individuals with MCI. Warm colors represent higher activation in controls. (B) MCI > Control: Orthogonal MNI-152 slices display voxels where individuals with MCI showed significantly greater boundary-related BOLD signal than control participants. Warm colors represent higher activation in MCI. Both contrasts were thresholded at False Discovery Rate $q < .025$ (one-tailed) after controlling for age and mean FD. Crosshairs mark x = 30, y = –36, z = –8.

3.3.3 Relationship Between pHPC Activity and Memory

Due to missing data on the memory task, one MCI patient was excluded from this analysis, resulting in a final sample of 53 participants (36 Controls, 17 MCI). The PCA showed that the first component explained 95.1% of the variance in the immediate and delayed memory measure.

The multiple regression model predicting this trait memory score revealed a highly significant main effect of Group ($\beta = –10.83$, $t(48) = –4.94$, $p < .001$), confirming, as expected, that the MCI group had significantly poorer memory performance than controls. However, there was no significant main effect of pHPC boundary activation on memory across the sample ($p = .802$), nor was there a significant interaction between Group and pHPC activation ($p = .985$), suggesting that the relationship between boundary-related activation in the pHPC and this general measure of trait memory does not differ between Controls and MCI (see Figure 5).

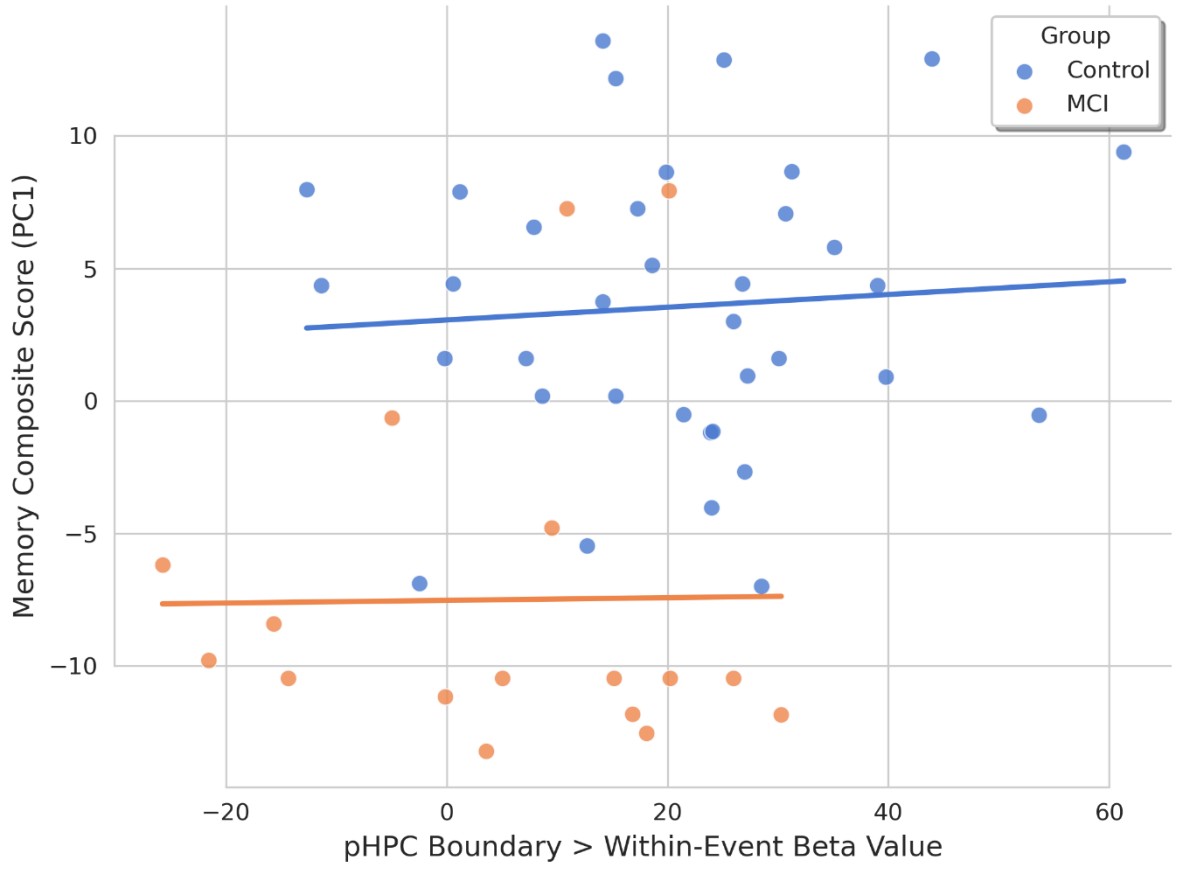
Given that Reagh et al. (2020) found a correlation between pHPC activation and episodic memory within the same CamCAN paradigm - albeit in a much larger, independent sample, we tested a simpler model (Memory ~ Activation + Age) on the Control group only, thereby avoiding the statistical confound of mixing patient and control data. The regression model within the Control showed no relationship between pHPC activation and memory ($\beta = 0.018$, $t(33) = 0.31$, $p = .760$).

This null result provides two key insights. First, our data do not replicate the brain-behaviour relationship reported by Reagh et al., likely due to limited statistical power. Second, and more importantly, it confirms that any correlation observed when collapsing across our clinical groups would be a statistical artifact driven by group status, not a genuine link between pHPC activation and memory within individuals. Therefore, we conclude that in our sample, only

clinical group status, and not pHPC activation, is a significant predictor of memory performance.

Figure 5

Relationship Between pHPC Boundary Activity and Memory Performance by Group



Note. Scatter plot illustrating the relationship between boundary-evoked activity in the pHPC and the composite memory score. Each point represents an individual participant, colored by their clinical group (Control: $n = 36$; MCI: $n = 17$). Solid lines represent the linear regression line of best fit for each group separately. The plot visually demonstrates the main effect of group on memory performance (a vertical separation between the groups) and the non-significant interaction (parallel slopes of the regression lines).

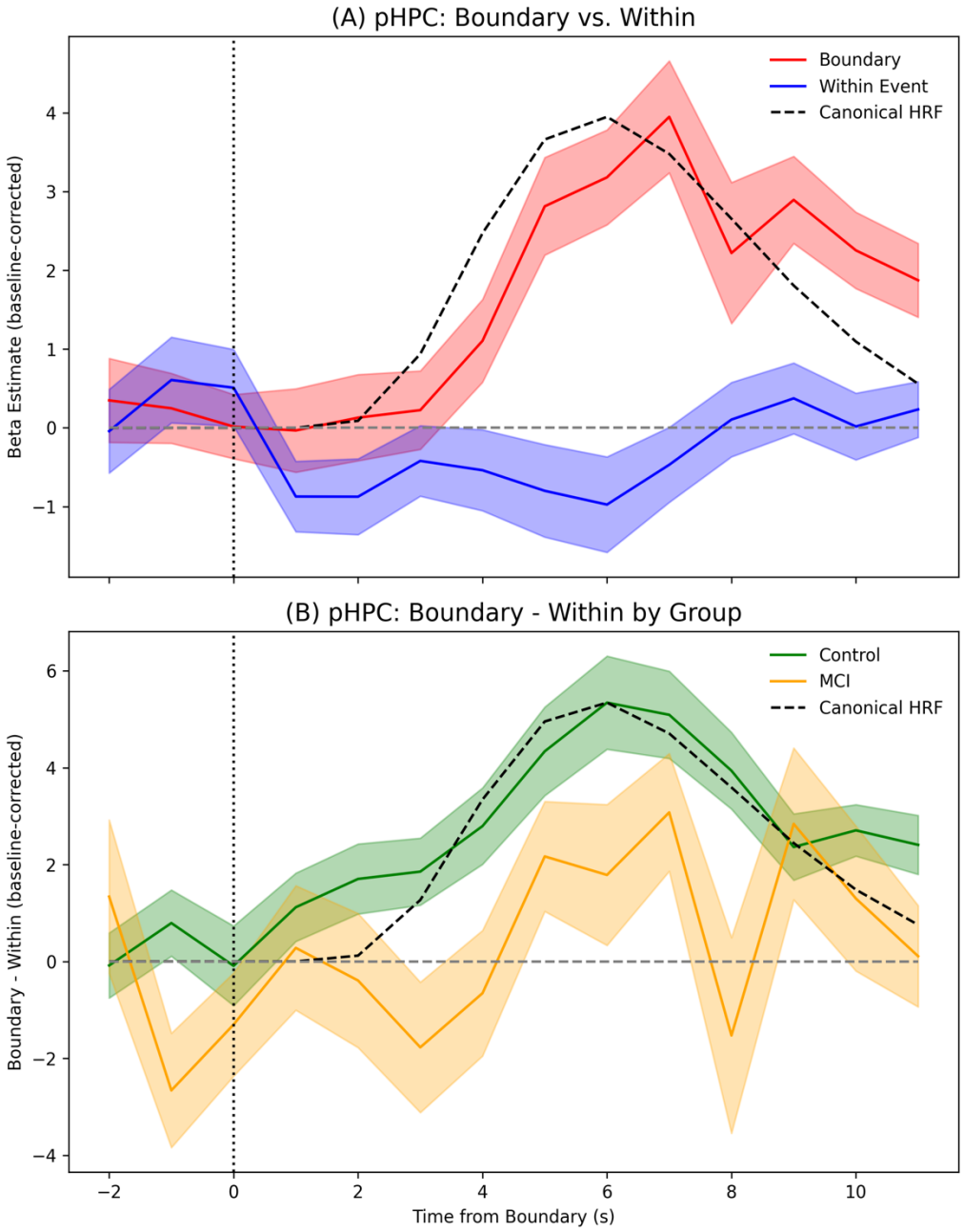
3.3.4 Confirmatory FIR Analysis

To validate our use of the canonical HRF in the primary analysis and further investigate the temporal dynamics of the boundary-evoked response, we performed a mixed-design ANOVA on the FIR beta estimates from the pHPC. Visual inspection of the resulting time courses supported the use of a canonical HRF in our main analysis, as shown in Figure 6, where the FIR-estimated responses correspond well with the canonical HRF. Specifically, the observed response peaks at a latency of approximately 5-7 seconds in both groups, which is consistent with the assumed timing of the canonical model. The presence of this robust, time-varying signal was statistically confirmed by a significant main effect of Time ($F(8.07, 419.7) = 6.29$, $p < .001$, Greenhouse-Geisser corrected).

Crucially, this analysis also replicated our primary finding. A significant main effect of Group ($F(1, 52) = 9.04$, $p = .004$) demonstrated that the overall magnitude of the boundary-specific signal was significantly weaker in the MCI group compared to Controls. This effect is visually represented in Figure 6B by the clear vertical separation between the two groups' response curves. After applying the necessary Greenhouse-Geisser correction, the Group × Time interaction was not significant ($F(8.07, 419.7) = 1.85$, $p = .066$). This indicates that the overall shape of the boundary-evoked response over time did not reliably differ between the groups, suggesting the deficit reflects a uniform reduction in signal strength rather than a change in response timing.

Figure 6

Finite Impulse Response (FIR) Time Courses of Boundary-Evoked Activity in the pHPC



Note. Panel A shows the mean BOLD signal (± SEM) evoked at event boundaries ("Boundary," red) compared with within-event time points ("Within Event," blue) collapsed across groups. Panel B displays the Boundary – Within difference for the Control group (green) and the mild-cognitive-impairment (MCI) group (orange). Time courses are baseline-corrected. Shaded bands represent ± 1 SEM across participants. The dashed black curve in each panel is the canonical SPM HRF scaled to the peak amplitude of the plotted data and aligned so that its time-zero sample coincides with the boundary (vertical dotted line). Positive values reflect greater activity at the boundary (or boundary minus within) relative to the pre-boundary baseline (horizontal dashed line). Time is expressed in seconds relative to the boundary; the FIR model spans –2 s to +11 s.

3.4 Discussion

MCI is Associated with a Blunted Hippocampal Boundary Response

Our findings in cognitively healthy controls confirm that the hippocampus is strongly engaged at perceived event changes, showing robust boundary-evoked BOLD activity. Importantly, our data replicate the boundary-evoked pHPC activity reported by Reagh et al. (2020) in an independent cohort of cognitively healthy older adults, thereby validating their findings in a distinct control sample. In contrast, when analyzed in isolation, the MCI group did not show a statistically significant boundary response ($p = .074$), a finding that is likely attributable to lower statistical power from the smaller patient sample ($n = 18$) rather than a complete absence of a signal. More importantly, the amplitude of the MCI response was significantly smaller than that of the Control group. In general, this is consistent with prior work supporting the hippocampus's role in event segmentation (Ben-Yakov & Henson, 2018; Reagh et al., 2020).

Crucially, the smaller hippocampal boundary response in the MCI group was not merely a reflection of hippocampal atrophy, since it remained significant even after statistically controlling for individual differences in pHPC volume. This suggests that the blunted boundary response in MCI is a functional deficit, providing information that is complementary to structural biomarkers. Furthermore, the group difference in the pHPC also persisted after accounting for concurrent activity in V1. Though our ROI analysis of V1 showed no significant group difference, there was a trend towards less boundary activation in the MCI Group. The fact that group effect in pHPC persisted after controlling for this shared

variance confirms that the hippocampal deficit is not merely a downstream consequence of altered low-level sensory processing. Nonetheless, the significant effect of V1 on pHPC activation in this model demonstrates that some portion of the pHPC's boundary response is related to visual differences at boundaries.

Finally, our FIR analysis suggested that the pHPC deficit in MCI did not simply owe to a delayed or dispersed response due to different haemodynamics. The MCI group showed an HRF that was simply a down-scaled version of that in the Control group, with similar peak latency, suggesting that it was simply caused by less neural activity.

Interpreting the Absence of a Brain-Behaviour Relationship

According to EST, the brain resets its current event model when a spike in prediction error signals an event boundary (Zacks et al., 2007). This reset appears to be implemented by the hippocampus through the brief burst of activity locked to the boundary. Ben-Yakov and Henson (2018) described its function as a "film editor" that responds preferentially to event transitions. Moreover, this hippocampal activity is believed to reflect registration of the preceding event into long-term memory as a single, bound representation (Ben-Yakov & Dudai, 2011; Richmond & Zacks, 2017). This segmentation process therefore affects not just immediate perception, but also the subsequent organisation of memory, ensuring that elements within an event are bound together more cohesively than elements across different events (Ezzyat & Davachi, 2011; DuBrow & Davachi, 2013).

The blunted pHPC boundary response observed in the MCI group suggests a failure of this snapshot mechanism. Without an effective, boundary-locked "save" command, newly incoming information may overwrite incompletely stored episodes. This would lead to the kind of fragmented narrative recall that is a key characteristic of MCI.

Ideally this hypothesis could be tested with data on recall of the movie itself. Unfortunately, such memory data were not acquired in this cohort. Instead, we related pHPC activation to performance on an independent episodic memory test, which required immediate and delayed recall of a verbal story. This was the same test used by Reagh et al. (2020) to find a correlation between pHPC boundary activity and memory in their much larger CamCAN sample. Reagh and colleagues reported this brain-behaviour relationship to be statistically stable across the healthy lifespan. However, grounded in the principle that MCI-related pathology can disrupt such functional coupling, we tested for a Group × Activation interaction in the prediction of episodic memory ability.

Our analysis revealed a clear main effect of Group on memory performance, with the MCI group recalling less, as expected. However, there was no main effect of pHPC activation, nor an interaction between this activation and Group. A focused post-hoc test within our healthy control group also failed to find a direct brain-behaviour relationship, yielding a negligible observed effect size for pHPC activation ($f^2 \approx 0.003$). In other words, we failed to replicate Reagh et al.'s finding that pHPC boundary activation during a movie predicts a healthy participant's general memory ability. This failure to replicate is explained by limited statistical power. Our control group ($n = 36$) was a fraction of the original Reagh et al. sample ($N = 546$), giving us only 23.3% power to detect the effect they reported ($f^2 \approx 0.046$). Future investigations would require a substantially larger sample, estimated at $N \approx 175$, to be adequately powered for Reagh et al's effect size.

Anatomical Specificity and the PMN

The anatomical distribution of our whole-brain findings provides important insights into the network-level nature of boundary processing deficits in MCI. The pHPC is a key component of the PMN, which processes contextual information and is highly sensitive to the structure of ongoing events (Ranganath & Ritchey, 2012; Ritchey, Libby & Ranganath, 2015). Multiple studies have shown that the PMN plays a pivotal role in driving the hippocampal activity at event boundaries (Barnett et al., 2024; Baldassano et al., 2017), making it particularly relevant to our findings.

Our whole-brain analysis revealed extensive group differences across the PMN, with a peak cluster centred in the PCC, suggesting this hub may be particularly vulnerable to disruption in MCI. This is consistent with Alzheimer's pathophysiology, in which PMN dysfunction emerges early (Jones et al., 2016). The breadth of the effect (68,545 voxels showing Control > MCI) indicates that boundary-processing deficits in MCI are not subtle or localized but widespread, pointing to a network-level disruption that likely reflects a fundamental breakdown in the neural mechanisms that segment continuous experience into discrete, memorable events.

To interpret our opposite finding of greater boundary-related increases in the MCI than Control group in sensorimotor cortex, we first considered whether residual head motion could explain this pattern. Even after standard motion controls (including covarying mean FD), residual motion-related confounds can linger in fMRI data. However, the fMRI signatures most predictive of head motion tend to localize to cerebellar Crus II and medial default-mode

regions, rather than sensorimotor cortex (Tomasi & Volkow, 2023). Consistent with this, in motor paradigms where motion is highest, sensorimotor activation does not reliably scale with head-motion metrics, and can even be uncorrelated with motion in patients, arguing against a purely motion-artifact account (Rumshiskaya et al., 2013). As an alternative explanation, we propose that in Controls, boundary-evoked signals in hippocampal/PMN transiently inhibit sensorimotor cortex, pausing ongoing action representations while the event model is updated. In MCI, these boundary signals are weaker. Consequently, sensorimotor cortex is less inhibited at those same moments, producing a relatively larger boundary > within-event response in sensorimotor cortex in MCI.

Importantly, these network-level findings, while informative about disease pathology, do not diminish the theoretical importance of understanding how the hippocampus with its unique role in marking boundaries contributes to event memory deficits. Our focused ROI approach allowed us to test specific hypotheses about hippocampal function while controlling for relevant confounds such as atrophy and visual processing. The whole-brain findings complement rather than replace this targeted investigation, revealing the broader context within which hippocampal dysfunction occurs. This raises a critical question for future work regarding the distinct functional roles of boundary-evoked activity across the PMN. For example, future studies should test whether activity in other key nodes, such as the PCC, have the same impact on subsequent event memory as the hippocampal response.

Limitations and Future Directions

Several limitations of this chapter should be acknowledged. First, our analysis relied on event boundaries defined by an independent healthy sample. This complicates the interpretation of our main finding: the reduced pHPC signal in MCI could reflect impaired neural encoding at these shared transitions, or it could indicate that patients' brains are segmenting the narrative at fundamentally different moments. Chapter 4 of this thesis will address this ambiguity by using data-driven event segmentation to investigate whether MCI patients exhibit a different pattern of neural event segmentation.

Further limitations relate to the scope of our analysis. Our primary comparisons were restricted to the pHPC as our primary ROI of interest and the V1 as a control region. However, the whole-brain analysis revealed that boundary-related group differences extended well beyond the hippocampus to encompass much of the PMN. The widespread nature of these findings suggests that our ROI approach may have underestimated the full extent of

boundary processing disruption in MCI. Future studies might benefit from network-based approaches that can capture the coordinated dysfunction across these distributed regions. Furthermore, as discussed previously, our modest sample size (n = 18 MCI) limited our statistical power to detect small effects, particularly for the brain-behaviour interaction. Similarly, the use of a general trait test of memory, rather than one specific to the movie's content, likely limited our statistical power to detect an effect. Finally, the study's cross-sectional design prevents us from determining whether reduced boundary-evoked pHPC activation is a predictive marker for the rate of future cognitive decline.

Despite these limitations, our findings establish a key functional deficit in the pHPC and have important clinical implications. A challenge for any potential functional biomarker is its ability to outperform established measures like hippocampal volume and CSF/PET, especially when such measures contribute to the diagnostic criteria used to define the patient groups in the first place. The value of a functional biomarker, therefore, is not necessarily to outperform these metrics but to provide uniquely informative signal above and beyond them. Our key finding that the group difference in pHPC activation remained significant even after statistically controlling for hippocampal volume shows that this functional measure provides information complementary to structural atrophy. While promising, it is unlikely that an fMRI-based measure would replace standard diagnostic biomarkers for AD like CSF or PET in a clinical setting, given its practical expense and complexity. Instead, its value may lie in providing a useful marker for the synaptic-level dysfunction that may precede widespread atrophy.

4 Event Segmentation Analysis

4.1 Introduction

Chapter 3 revealed a core functional deficit in MCI: a blunted pHPC response to event boundaries. However, those analyses focused on the hippocampal activity evoked by boundaries defined externally from subjective reports. This leaves a critical question unanswered: is the MCI deficit rooted in the neural implementation of the event boundaries themselves? Here we provide a complementary perspective by shifting the focus from boundary-evoked hippocampal activity to multivoxel patterns in AG, using a data-driven, pattern-based approach that derives boundaries directly from neural signals.

Within this framework, we test two mutually exclusive hypotheses for the MCI deficit: (i) a “coherent alternative” model, in which patients agree on a boundary template that is systematically different from controls, i.e., consistent group-level timing differences; and (ii) a “shared” model, in which patients and controls express the same underlying event structure. Orthogonal to these “structural models”, we also consider an “implementation fidelity” model: even when the structure is shared, transitions may be blurred (weaker within-participant pre/post pattern contrast).

To adjudicate among these scenarios, we apply data-driven segmentation in AG and V1 and combine timing-sensitive tests. First, boundary overlap quantifies temporal correspondence between Control- and MCI-derived templates; a substantial overlap is inconsistent with a largely distinct "coherent-alternative" template, though it does not preclude small systematic timing differences. Second, a boundary template-preference test asks which template better indexes each group’s transitions. Its logic is as follows: (a) if MCI patients systematically prefer their own template, this supports the "coherent-alternative" model (consistent timing differences, possibly small); (b) if MCI patients systematically prefer the Control template, this is consistent with a shared structure in which the Control-derived template is the cleaner estimate due to higher *N*; (c) indifference at unequal *N* can arise from opposing forces (own-alignment vs Control precision). We therefore perform an additional *N*-matched split-half analysis: under the "shared" model (no systematic timing difference), *N*-matching predicts indifference; under the "coherent-alternative" model (even with small differences), it predicts an own-template advantage in MCI.

To characterize implementation fidelity, we estimate a mean boundary-strength metric. Because strength depends both on the clarity of within-participant transitions and on any residual timing mismatch between group-average template times and an individual's true transitions, this metric is ambiguous in isolation. Accordingly, we interpret boundary strength alongside the overlap and template-preference tests that speak directly to timing.

Finally, our investigation operates on the assumption that cortex and hippocampus play different dynamic roles - the cortex maintaining stable states and the hippocampus marking transient changes. To synthesise our findings with Chapter 3, we will also test whether the stable-state segmentation model applies to hippocampal activity, providing a comprehensive picture for interpreting any cortical deficits we uncover.

4.2 Methods

4.2.1 Data Preparation: ROI Definition and Voxel Selection

ROI Definition

This chapter's analyses required single, common ROIs for the AG, V1, and the hippocampal subregions to ensure that group comparisons could not be attributed to anatomical differences in voxel selection. To construct a common space that is nonetheless sensitive to potential systematic anatomical differences between the cohorts (such as cortical atrophy in the MCI group), a multi-step procedure was employed in MNI152NLin2009cAsym standard space. First, group-specific gray matter masks were created by averaging individual probability maps within each clinical group (Control and MCI), smoothing the average with a 3mm FWHM Gaussian kernel, and thresholding at a probability of 0.3. This step ensures that the typical anatomy of each group is represented. The final ROI for each cortical region was then generated by creating an intersection of the atlas-based template, the control group's gray matter mask, and the MCI group's gray matter mask. This ensures that any voxel included in the final analysis is considered valid gray matter in both groups. This common masking procedure was applied to all three ROIs. For the AG and V1, templates were defined using the Schaefer et al. (2018) atlas, with the AG ROI combining parcels from the Default Mode Network. The pHPC ROI was defined using the bilateral posterior hippocampal mask adapted from Ritchey et al. (2015). The final common masks encompassed 2042 voxels for the AG, 2150 for V1, and 460 voxels for pHPC.

Voxel Selection

Within the common anatomical ROIs defined for both groups, a reliability-based voxel selection was performed to identify the most stimulus-synchronized voxels. Each voxel's time series was *z*-scored and the inter-subject correlation (ISC) was computed across the entire sample. The two hundred voxels with the highest ISC were retained. The ISC range for the top 200 voxels was 0.0839 - 0.1355 in AG, 0.1182 - 0.1552 in V1, and 0.0045 - 0.0194 in pHPC. Concentrating subsequent analyses on these stimulus-synchronised voxels improves the signal-to-noise ratio for state segmentation and, because the identical voxel set is applied to every brain, guarantees that any difference observed between groups cannot be attributed to differences in voxel sampling. A possible objection is that pooling across cohorts might attenuate genuine group differences by discarding voxels that are reliable only in one group. This trade-off is acknowledged; the present choice favours anatomical comparability over maximal sensitivity. By the same token, any group effect that still emerges after this conservative filtering can be regarded as especially robust.

Preprocessing Optimizations

To ensure the input data was optimally prepared for identifying event-related neural dynamics, we conducted an exploratory analysis comparing several preprocessing pipelines (see Appendix for full details). We systematically evaluated the effects of an extended denoising protocol (including spatial smoothing) and functional hyperalignment. This revealed a critical trade-off: while functional hyperalignment dramatically increased voxel-level signal synchrony, it paradoxically blurred the sharp transitions between distinct neural states. Conversely, the extended denoising protocol markedly enhanced the block-like structure critical for segmentation. Given the importance of preserving the sharpness of these neural transitions and to ensure comparable preprocessing pipelines across chapter 3 and 4, this pipeline was selected for all analyses reported in this thesis and is equivalent to the preprocessing/denoising pipeline described in Chapter 2, Section 2.5.

4.2.2 Generating Neural Event Boundary Templates using GSBS

Group averaging

The voxel-filtered and extensively denoised data were then averaged within each cohort at every timepoint to create group-average time series for the control and MCI groups.

Group averaging greatly boosts signal-to-noise by cancelling subject-specific fluctuations and retaining variance that is locked to the movie. Geerligs et al. (2021) show that GSBS produces stable state counts once ≈17 participants are averaged (with single-subject input of the same length, noise drives the algorithm to over-segment and to return erratic estimates of the optimal number of states). Guided by those findings, we refrained from fitting GSBS at the individual level and relied instead on cohort averages to obtain reliable neural boundary estimates.

Greedy State Boundary Search

Neural event boundaries were detected using an improved version of the GSBS algorithm. GSBS is a data-driven method that partitions a multivariate timeseries into a succession of non-recurring, temporally contiguous states (Geerligs et al., 2021). Though generally applicable, it was specifically developed for identifying state transitions in fMRI data, where the fundamental assumption is that each state is characterized by a relatively stable spatial activity pattern across voxels. The algorithm operates by iteratively finding the boundary locations that best segment the data according to this principle. It begins with the simplest possible model, in which the entire fMRI run constitutes a single state. From there, it greedily adds boundaries to produce progressively more fine-grained models.

The core of the algorithm is an iterative search. At each iteration, the algorithm aims to split an existing state to achieve the largest improvement in model fit. The fit is defined as the mean Pearson correlation between the activity pattern of each timepoint and the averaged pattern of its provisionally assigned state (Geerligs et al., 2021). At each iteration, GSBS chooses between (i) splitting a state (add one boundary) or (ii) inserting a substate (add two boundaries), selecting the option that best improves model fit. This prevents late detection of certain strong transitions and overestimation of state count (Geerligs et al., 2022).

To decide between these two placement strategies at each iteration, the algorithm uses the *t*-distance metric. The *t*-distance quantifies the separation between two distributions: the distribution of pattern correlations for all pairs of timepoints within the same state, and the distribution of correlations for pairs of timepoints in consecutive states. This separation is expressed as a single *t*-statistic (Geerligs et al., 2021). At each step, the algorithm calculates the potential improvement in *t*-distance from adding either one or two boundaries and selects the option that yields the higher value. This makes the iterative search more robust by allowing it to identify either simple transitions or entire encapsulated states as needed.

Because GSBS adds boundaries sequentially (not jointly optimized), a fine-tuning step is applied after each iteration: boundaries are re-examined boundaries from weakest to strongest and shifted by ±1 TR if this increases mean within-state correlation. This prevents early boundaries from being locked in suboptimal positions (Geerligs et al., 2022).

Finally, the iterative process of placing and fine-tuning boundaries is not run to find a single solution, but is repeated to generate a complete set of boundary locations for every possible number of states, k, up to a predefined maximum (in this analysis, $T/2$, where T is the number of timepoints). The final step is to select the single, optimal state count from all these candidate models. This selection is again performed using the t-distance. The t-distance is calculated for the final segmentation of every k, and the optimal model is chosen as the one with the global maximum t-distance. This criterion proved markedly less susceptible to haemodynamic autocorrelation than earlier metrics such as Within-Across correlation (WAC) or log-likelihood (Geerligs et al., 2021). The algorithm therefore produces a single, optimal set of boundaries for the timeseries.

4.2.3 Temporal Alignment of Event Boundaries

Absolute Boundary Overlap

We first performed a descriptive analysis to directly quantify the temporal alignment of event boundaries. This allowed us to test whether both groups were tracking a shared event structure. The analysis was conducted on the unconstrained (i.e., not k-constrained) group templates from the AG and V1. Second, to characterize the nature of these neural boundaries, we quantify their temporal alignment with the set of externally-defined subjective boundaries used in Chapter 3. For this, we shifted the subjective boundary times by +5 s to account for the hemodynamic response before applying the overlap procedure.

The core metric for this comparison was the absolute overlap, which measures the proportion of co-occurring boundaries corrected for the number of overlaps expected by chance (Geerligs et al., 2022). This calculation allowed for a temporal tolerance of ±1 TR to account for minor timing variations. First, we identified the number of actual co-occurrences, defined as the count of boundaries in the group with fewer boundaries (the "target" group) that fell within this ±1 TR window of any boundary in the other ("reference") group. Next, we calculated the number of co-occurrences expected by chance, based on the probability of a target boundary randomly falling within the union of ±1 TR windows around all reference

boundaries (i.e., a 3-TR effective window per boundary, accounting for window overlaps). The final absolute overlap score was then calculated as the number of actual co-occurrences minus the expected number, normalized by the maximum possible co-occurrences above chance (i.e., the total number of target boundaries minus the expected number). A score of 1 indicates perfect overlap, 0 indicates overlap at the level of chance, and negative values indicate less overlap than expected by chance.

Statistical Evaluation

To assess statistical significance for this overlap analysis, a permutation test (1000 iterations) was performed. For each iteration, the order of state durations for each group's template was shuffled. This preserved the number of boundaries and the distribution of state lengths while randomizing their temporal positions. When comparing to subjective boundaries, the subjective series was held fixed and only the neural series was permuted. Because our a-priori hypothesis was directional (neither the coherent alternative model hypothesis nor the shared-but-weaker model hypothesis predict anti-alignment), we used a one-tailed test: the *p*-value was computed as the proportion of permuted overlap scores that were greater than or equal to the empirically observed score. Significance was assessed against an alpha of 0.05.

4.2.4 Boundary Strength Analysis: Template Preference and Implementation Clarity

To determine whether individuals with MCI segment continuous experience using a boundary template that differs systematically from Controls, or instead share the same underlying event structure, and to assess implementation fidelity, we conducted two related analyses based on boundary-strength. First, we generated group-specific event templates with GSBS, then super-imposed those templates on each participant's time-series to derive a boundary-strength score at every template boundary. These scores underpin (i) the template-preference test, which probes whether the two groups favour different timing templates, and (ii) a between-group comparison of average strength, which gauges how distinct the neural transitions are once each group is evaluated against its own template.

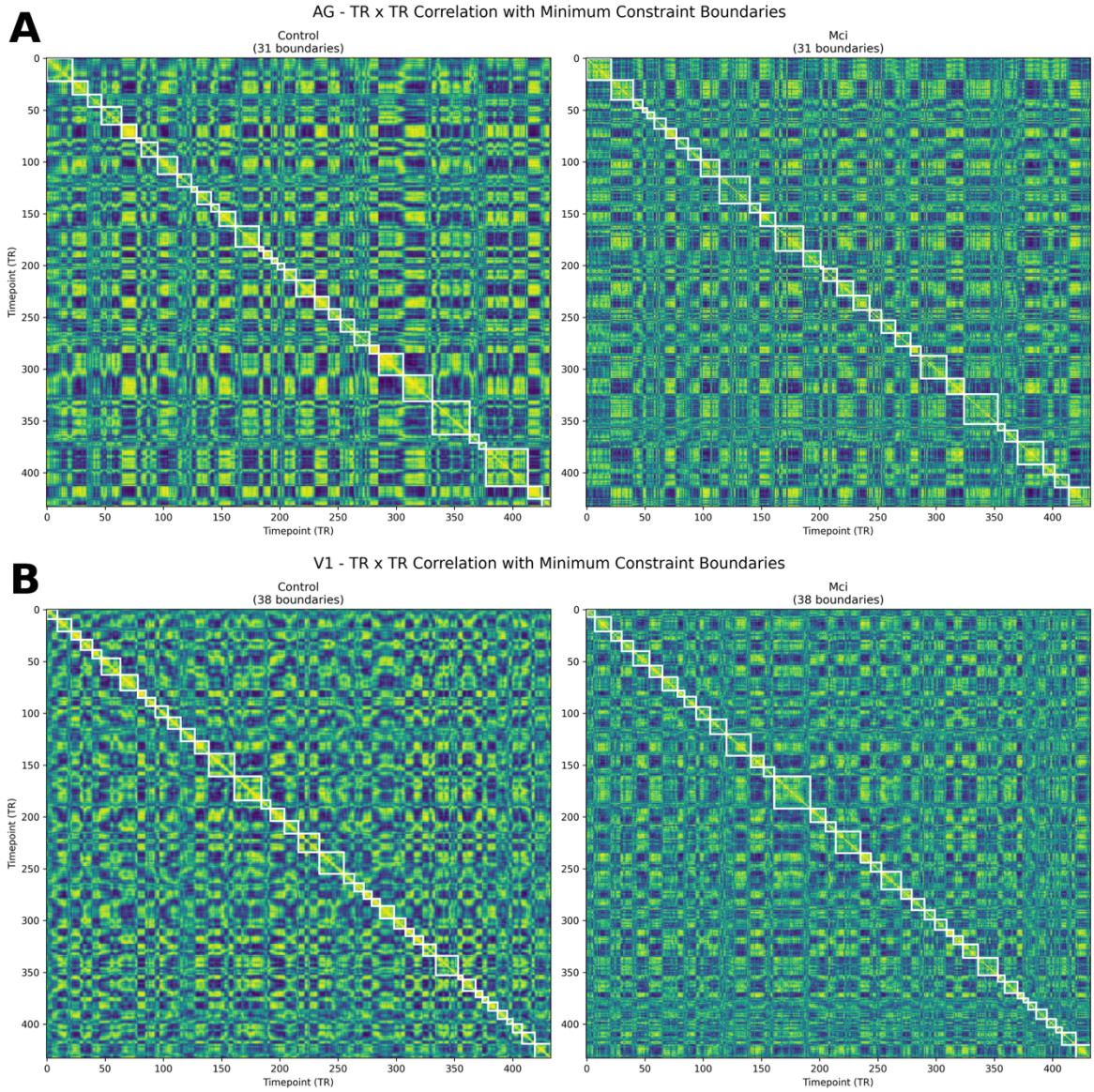
Generating Constrained Boundary Templates

A critical methodological challenge arose in the first step of this procedure. A preliminary analysis in which the optimal number of boundaries (*k*) was estimated from a full

GSBS run for each group revealed a potential confound. While the groups had similar boundary counts in the AG (k = 32 for Control group; k = 31 for MCI group), the Control group template produced substantially more boundaries in V1 (k = 59) than the MCI group template (k = 38). This discrepancy makes it impossible to disentangle group differences in the temporal timing of events from differences in the number of boundaries. Therefore, to isolate the effects of boundary timing from the confounding effect of boundary count, we explicitly constrained k to be identical across all group templates. We used the optimal k identified for the MCI group, which we call the MCI-k constraint (k = 31 in AG; k = 38 in V1; see Figure 7). This a priori choice was critical for two reasons. First, it provides the most conservative test of implementation strength by preventing the MCI group from being penalized for "over-segmenting" their data to match the higher granularity of controls. Second, it strengthens our test for an altered segmentation structure by building the MCI template at its own natural granularity, whereby we give it the best possible chance to reveal a coherent, alternative event pattern. A failure of the MCI group to prefer this optimally-derived template provides stronger, more definitive evidence against a coherent alternative model.

Figure 7

Constrained GSBS state structure in the AG and V1



Note. The upper row (Panel A) shows the Control and MCI TR × TR pattern-correlation matrices for the AG; the lower row (Panel B) shows the same for V1. White squares mark the 31 (AG) or 38 (V1) boundary positions derived with the MCI-*k* constraint - the template that serves as the basis for all subsequent overlap and boundary-strength analyses. The block-diagonal pattern illustrates the stable neural states identified by GSBS.

To test whether conclusions depend on segmentation granularity, we repeated all analyses at a (somewhat) finer model order using the Control-optimal *k*. This robustness

check allowed us to investigate whether any potential deficit is amplified by increased segmentation demand or, alternatively, whether forcing a higher granularity could reveal a latent, fine-grained structure and thereby reduce any group difference.

The templates were generated by selecting the specific boundary solution from the full GSBS output that matched the desired *k*. We generated two types of these templates, each tailored for a different analytical purpose. First, for analyses assessing an individual participant's fit to their own group's event structure, we used a leave-one-out (LOO) procedure to ensure non-circularity. This involved re-estimating the group template for each participant while excluding that participant's own data. Second, for analyses requiring a single, stable representation of each group's structure (e.g., when comparing the two groups directly), we used the full group-average template, created by averaging across all participants within a group to provide the most robust estimate of their overall segmentation pattern.

Strength Score Calculation

Once a template was generated, either via LOO or using the full data, its boundary vector was superimposed on an individual participant's fMRI time series. For each boundary, b_i, in the template, two adjacent neural states were defined. The state preceding the boundary comprised all timepoints from the previous boundary, b_{i-1} (inclusive), to the current boundary, b_i (exclusive). The subsequent state began at the current boundary, b_i (inclusive), and continued to the next one, b_{i+1} (exclusive). We then calculated the mean spatial pattern for each of these two states and computed their Pearson correlation, r_i. This similarity coefficient was transformed into a dissimilarity value, or strength score, using the formula $s_i = 1 - r_i$. This ensures that $s_i = 0$ indicates no neural pattern shift across the boundary, while larger positive values indicate sharper transitions (0 = no change; 2 = maximal change when $r = -1$). The final boundary-strength score for that participant-template pair was the arithmetic mean of these individual scores.

A potential caveat of this score is that boundary strength can be artificially inflated for very short states, as the means are estimated from noisier, less stable patterns. To rule out the possibility that unequal state lengths might spuriously drive group differences, we compared the resulting state-duration distributions between the group templates for each constraint condition. This comparison confirmed that, under the k-constraint, the resulting distributions of state lengths were indeed well-matched between groups. In both the AG and V1, the

Control and MCI templates exhibited virtually identical means, medians, and spreads (e.g., mean difference < 0.4 TR). Because the potential relationship between duration and strength is therefore balanced, variable state length cannot account for any group effects on boundary strength.

Statistical evaluation

The primary family of hypotheses consisted of three tests (two within-group tests of preference, one between-group test of strength) conducted in our primary cortical ROI, the AG. These were therefore assessed against a Bonferroni-corrected alpha level of $\alpha = .05/3 = .0167$. The same family of tests was then performed in V1 to assess anatomical specificity.

Within-Group Tests of Template Preference

To adjudicate between a coherent-alternative model and a shared model, we compared, within each group, boundary strength computed against the own-group template (estimated with LOO) versus the other-group template. Templates serve only as time indices: they partition a participant's time series, and strength is the pre/post pattern change drawn solely from that participant's data. The LOO procedure prevents inflation from circularity when evaluating the own-group template. The decisive outcomes are as follows: if MCI reliably show higher strength on their own template than on the Control template, this supports a coherent-alternative model, i.e., MCI-specific systematic timing differences; if MCI prefer the Control template, this is consistent with a shared structure in which the Control template is the cleaner estimate due to higher *N*; indifference at unequal *N* can arise from opposing forces (own-alignment vs Control precision) and is thus ambiguous. Accordingly, we also performed an *N*-matched split-half analysis as the decisive test: under the shared model (no systematic timing difference), *N*-matching predicts indifference; under the coherent-alternative model, it predicts an own-template advantage in MCI.

For the Control group, we compared each participant's strength on their LOO Control template with their strength on the full-group MCI template using a one-tailed paired-samples *t*-test, reflecting the expectation that Controls share a common timing model best indexed by the Control template. For the MCI group, we compared strength on the LOO MCI template with strength on the full-group Control template using a two-tailed paired-samples *t*-test, because both a coherent-alternative outcome (own-template advantage) and a shared-structure outcome (Control-template advantage due to precision) are theoretically informative.

Between-Group Test of Implementation Strength

We also compared the mean boundary strength produced by each cohort when evaluated on its own full-group template, using a two-tailed independent-samples t-test. This metric predominantly reflects implementation fidelity (the intrinsic clarity of within-participant transitions), while acknowledging that it may also capture some template-participant timing mismatch. Scoring each cohort on its own template reduces (but does not eliminate) timing-related penalties, because group-average timings still smooth over participant-specific transitions. Therefore, we interpret the strength gap in conjunction with the timing-sensitive tests (overlap and template preference).

4.2.5 Contrasting Regional Dynamics: Applying GSBS to the Hippocampus

To directly contrast the functional dynamics of the cortex with those of the hippocampus, we performed a final analysis applying the GSBS algorithm to the time series from the pHPC. The core purpose was to test if the stable-state model, which proved effective in the cortex, is equally applicable to a region with putatively different, more transient dynamics. This direct comparison serves to synthesize our findings from Chapter 3 by assessing whether the hippocampus maintains stable states between boundaries. We used an exploratory, qualitative framework: similarity to the cortical pattern (comparable optimal *k*, elbow in *t*-distance, block-like TR×TR structure) would be read as consistent with stable states in pHPC, whereas preference for very high k with a monotonic t-distance and weak block structure would be read as limited compatibility with a stable-state model. All data preparation steps and GSBS parameters were identical to those used in the primary cortical analyses.

4.3 Results

4.3.1 Temporal Alignment

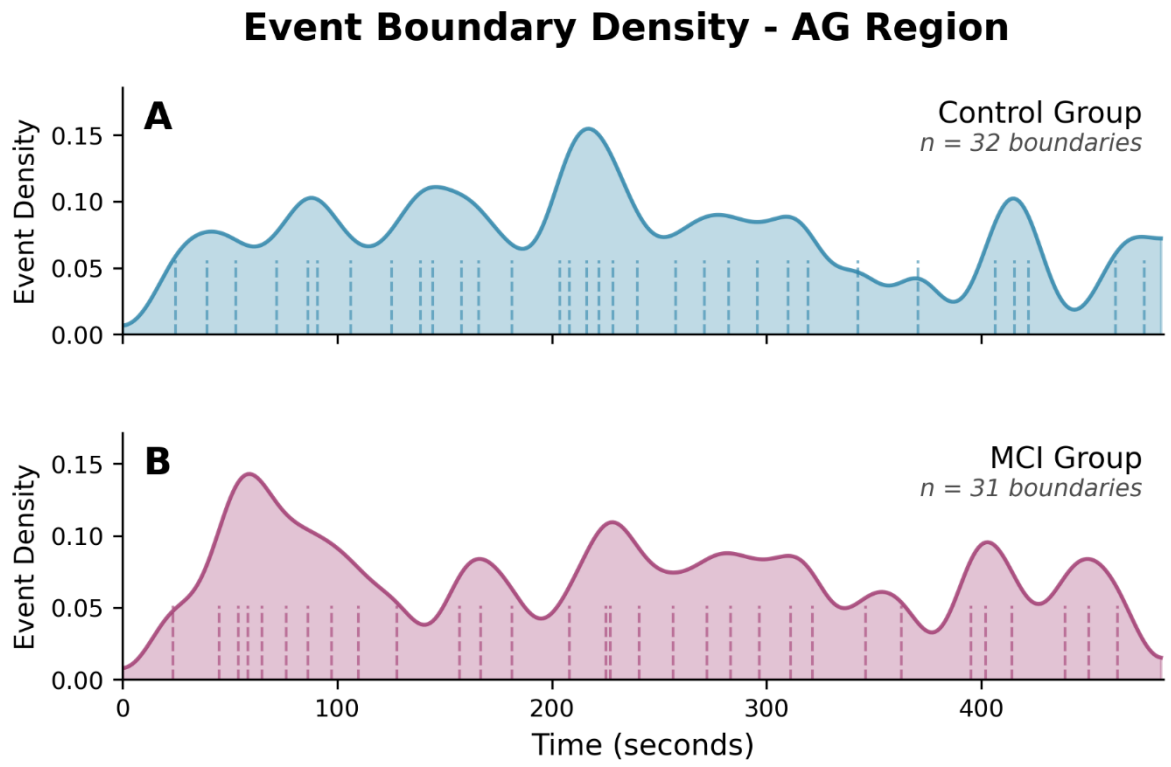
The boundary-overlap analysis showed that the two groups shared a substantial subset of boundary time-points. In both the AG and V1, we observed a significant temporal overlap between the unconstrained neural templates of the two groups (AG: absolute overlap = 0.38, $p < .001$; V1: absolute overlap = 0.51, $p < .001$). This significant overlap is inconsistent with a fully distinct MCI template, but does not rule out systematic timing differences, since the overlap is far from perfect. Figure 8 illustrates these temporal alignment patterns, showing

the smoothed boundary activity across the movie timeline for both groups. The visualization reveals that while boundaries cluster at similar temporal regions (explaining the significant overlap), the groups also show distinct segmentation patterns at other movie moments, consistent with the overlap being partial.

In contrast, we observed selective correspondence between the neural and the subjective boundaries. In AG, neither group showed significant alignment (Control: absolute overlap = 0.143, $p = .285$; MCI: 0.048, $p = .508$). In V1, the Control template showed robust above-chance alignment (absolute overlap = 0.718, $p < .001$), whereas the MCI template showed a positive but non-significant trend (absolute overlap = 0.321, $p = .082$). Thus, low-level visual state boundaries in Controls align with perceived event boundaries, while AG does not, and MCI shows weaker/non-significant alignment.

Figure 8

Temporal Distribution of Event Boundaries Across Movie Viewing for Control and MCI Groups in the AG Region



Note. Smoothed boundary activity showing the temporal distribution of neural event states throughout the movie stimulus for (A) control group and (B) MCI group. Vertical dashed lines indicate individual boundary locations, while the continuous curve represents Gaussian-smoothed boundary activity (σ = 10 TRs = 11.2 s). Peaks indicate temporal regions where boundaries are densely placed, while valleys represent periods of stable neural states. The alignment of peaks and valleys between groups reflects their shared event segmentation structure, consistent with the significant temporal overlap (absolute overlap = 0.38) observed in the quantitative analysis. The control group identified 32 boundaries compared to 31 boundaries in the MCI group. Time is presented in seconds, with total movie duration of 484.96 seconds (433 TRs × 1.12 s/TR).

4.3.2 Boundary Strength

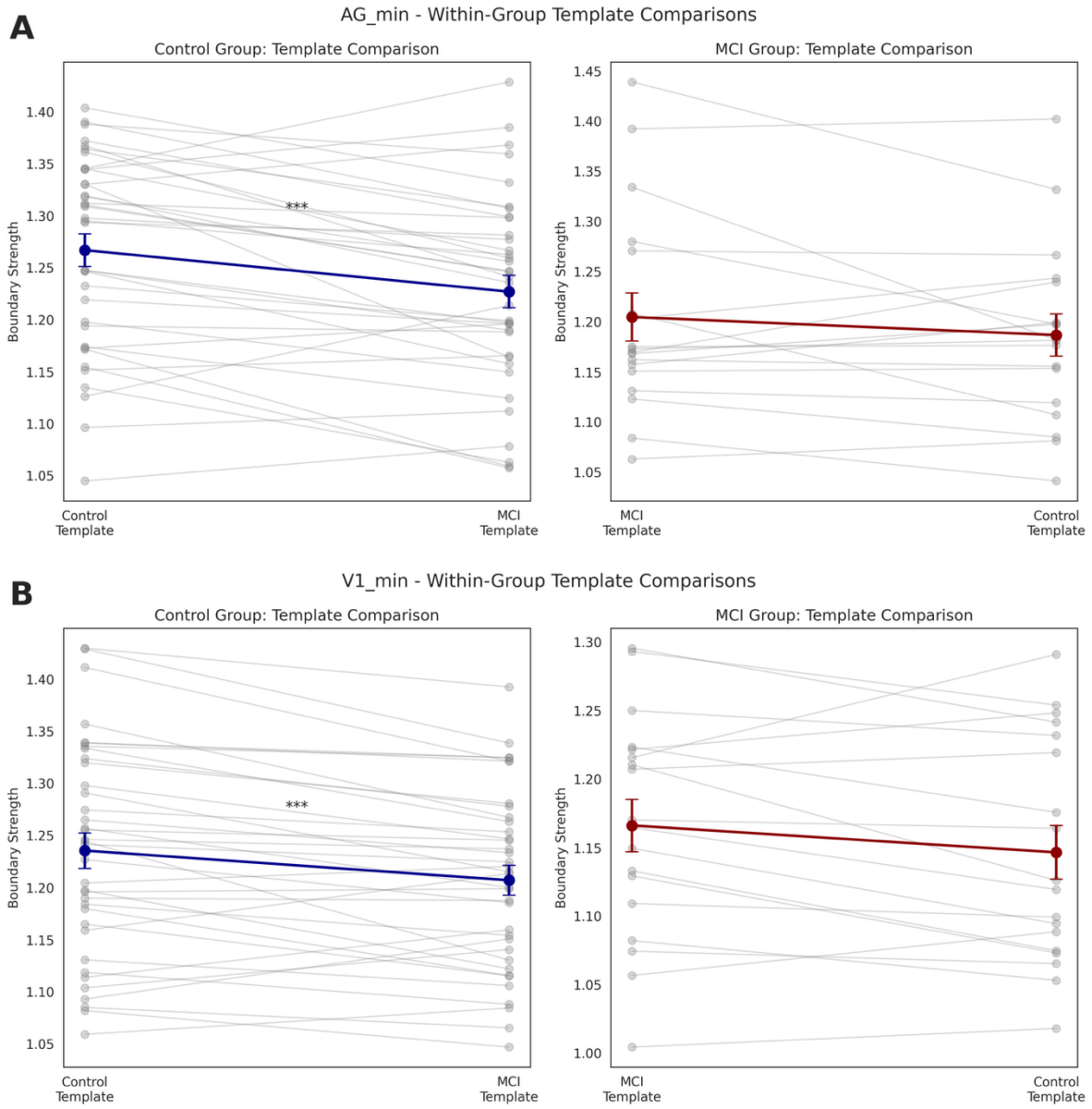
Having established significant - but not complete - temporal overlap, we next performed the boundary-strength analysis to evaluate which event template each cohort prefers and how distinct their neural transitions are.

Angular Gyrus

First, to test for an altered event structure, we assessed whether each group showed a preferential fit to their own template. A paired-samples t-test confirmed that the Control group's brain activity showed a significantly stronger fit to the Control template than to the MCI template ($t(35) = 4.24$, $p < .001$, $d_z \approx 0.71$) (see Figure 9, Panel A). In contrast, a symmetrical test on the MCI group revealed no such preference; their boundary strength scores did not differ significantly between their own MCI template and the Control template ($t(17) = 1.30$, $p = .212$, $d_z \approx 0.31$). The between-group comparison of mean boundary strength (each group evaluated on its own template), revealed lower scores in MCI ($t(52) = 2.23$, $p = .030$, $d = 0.64$), although this did not survive the Bonferroni-corrected threshold ($\alpha = .0167$).

Figure 9

Within-group template comparisons in the AG and V1



Note. Panel A shows the boundary-strength trajectories for the Control and MCI groups in AG when each participant is evaluated on both the Control and MCI templates. Panel B presents the corresponding trajectories in V1. Each grey line represents a single participant; coloured symbols and error bars depict the group mean ± 1 SE. Asterisks denote the result of the paired-samples preference test (*** $p < .001$).

V1

The results showed the same asymmetric pattern for template preference observed in the AG. The Control group demonstrated a robust preference for their own template ($t(35) = 4.10$, $p < .001$, $d_z \approx 0.68$), while the MCI group showed no statistically significant preference for theirs ($t(17) = 2.06$, $p = .055$, $d_z \approx \approx 0.49$) (see Figure 9, Panel B). The between-group comparison of mean boundary strength provided clear evidence of a deficit. Mean boundary strength was significantly lower in MCI and passed the Bonferroni threshold ($t(52) = 2.51$, $p = .015$, $d = 0.73$).

When repeating the analyses under the Control-optimal k, all effects persisted and the group strength gap increased in both ROIs (AG $p = .016$; V1 $p = .006$), with the Control template still preferred by Controls and no preference in MCI. This confirms that our results are not an artefact of a specific *k* value.

4.3.3 Robustness to template size: *N*-matched split-half analysis

One concern about the MCI template-preference result is that unequal template precision could mask a true Control template advantage in MCI. The Control template is estimated from more participants, which yields cleaner boundary times. Conversely, the MCI template might reflect MCI-specific timing differences. These two influences could, in principle, oppose each other to produce the indifference for either template that was observed in MCI, even if patients and controls share the same event structure.

Therefore, to remove unequal precision as an explanation, we matched group size. We randomly split the Control group into two halves of $n = 18/18$ to match the MCI sample size, rebuilt half-templates in each ROI, and re-ran the preference tests, repeating this procedure for 200 random splits per ROI. Specifically, for each Control subject, we compared own-half LOO against the other-half full template. For each MCI subject, we compared MCI LOO against the mean of the two Control-half full templates. (For a complete description of procedures, see Appendix, section 7.2)

With unequal group sizes (as in the original template preference analysis), a shared-structure + noise-only account predicts a Control advantage (because Control template precision is higher due to larger *N*), so indifference argues against a shared-structure + noise-only account. After *N*-matching, the same account predicts no preference (both templates equally precise). Any own-template advantage at matched *N* indicates genuine MCI-specific systematic timing differences.

Results and Interpretation

In Controls, Half-A vs Half-B produced near-zero differences in both regions (V1 mean $\Delta \approx -0.0001$; 95% CI [−0.0020, 0.0019]; AG mean $\Delta \approx -0.0024$; 95% CI [−0.0047, −0.0001]). The small negative bias in AG is expected from the asymmetry (own-half LOO uses $n = 17$, other-half full uses $n = 18$) and indicates the halves are effectively interchangeable. In MCI, across 200 random splits in both AG and V1, patients now showed an own-template advantage. If anything, residual precision still favors Control (MCI LOO vs the mean of two Control-half full templates), yet Δ remained positive: V1 mean $\Delta = +0.0038$ (95% CI [0.0024, 0.0052]); AG mean $\Delta = +0.0171$ (95% CI [0.0159, 0.0184]).

If MCI indifference were solely due to a noisier MCI template, *N*-matching should have yielded indifference (i.e., no bias toward either template). Instead, the MCI group now showed a reliable own-template advantage at matched *N* in both ROIs. This closes the "unequal precision" loophole and supports a broadly shared event structure with small but systematic timing differences in MCI, rather than a purely precision-driven artefact or a fully distinct alternative structure.

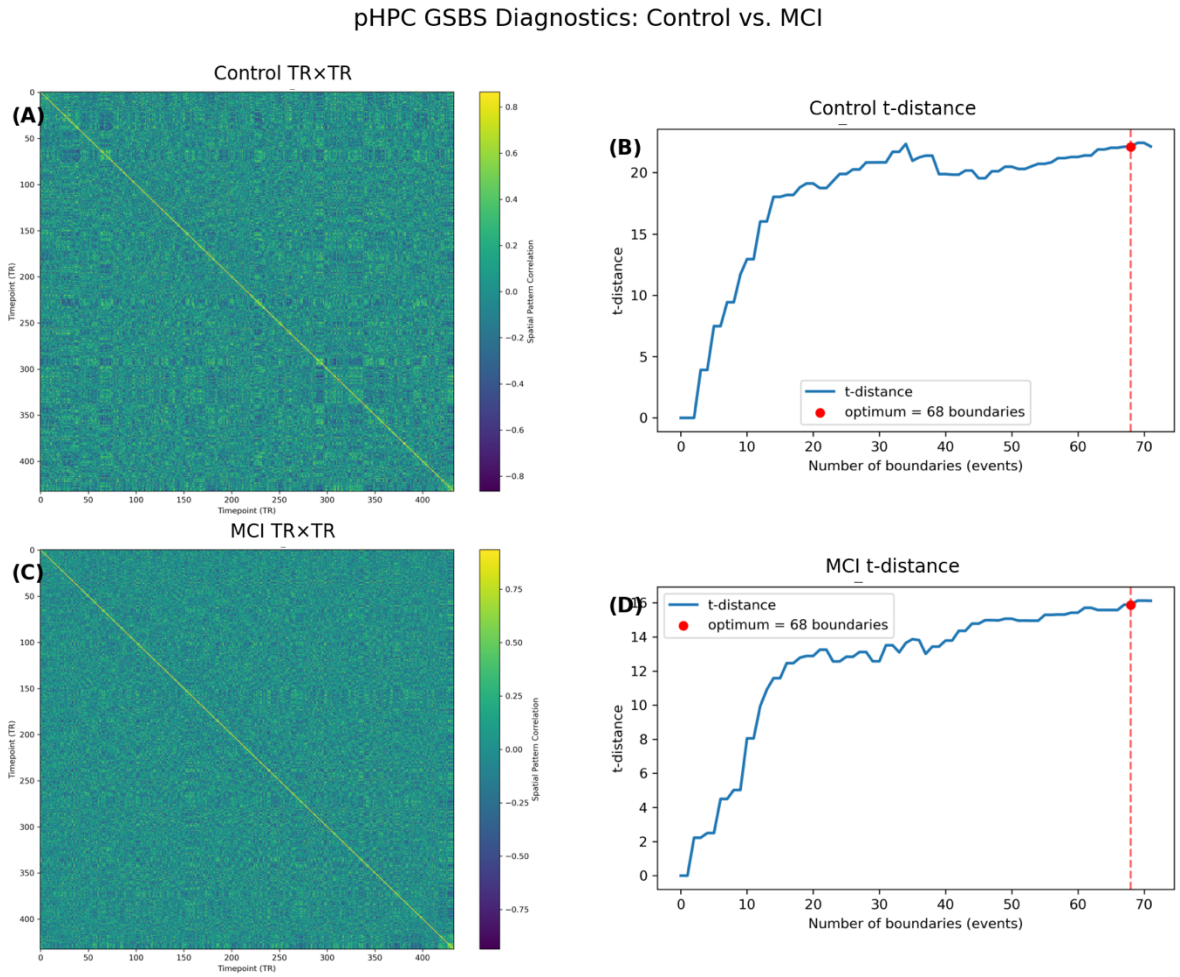
4.3.4 Absence of Stable States in the Hippocampus

In stark contrast to the clear state structure observed in AG and V1, GSBS failed to converge on a parsimonious solution in pHPC. Across both participant groups, the algorithm consistently determined that the optimal number of boundaries was very close to the maximum allowed ($k = 216$). This outcome persisted even when the maximum was constrained to a much lower value ($k = 70$) to rule out a simple ceiling artifact. The *t*-distance curve rose monotonically with no local maximum (Figure 10 B/D).

Visual inspection of the timepoint-by-timepoint correlation matrices confirmed this. Unlike the distinct block-like structure that characterizes stable states visible in the cortical regions, the hippocampal matrices were unstructured, (Figure 10 A/C). Crucially, this pattern held for both the Control and MCI groups, suggesting that the absence of sustained, stable states is a fundamental property of hippocampal processing during this task, not a disease-related phenomenon.

Figure 10

Hippocampal GSBS Diagnostics for Controls and MCI



Note. Panel (A) shows the time-point-by-time-point (TR × TR) spatial-pattern correlation matrix for the control group; brighter off-diagonal blocks would indicate stable neural states, but none are evident. Panel (B) plots the corresponding t-distance curve, with the red dot marking the algorithm-selected optimum of 68 boundaries, close to the maximum allowed number of boundaries (k = 70). Panels (C) and (D) repeat the diagnostics for the MCI group.

4.4 Discussion

The present findings allow us to draw a clear answer to the question posed at the outset of this chapter. The data do not support a fully distinct "coherent alternative" patient template. Instead they point to a broadly shared event structure with small, systematic timing differences, alongside weaker implementation. Support for this conclusion comes from three

converging observations. (1) Temporal overlap: although not complete, a sizeable fraction of boundaries was common to both cohorts (absolute overlaps ≈ 0.4-0.5), indicating that patients and controls recognise many of the same event transitions while allowing for modest timing differences. (2) Template preference: Controls preferred their own template; MCI showed no preference under unequal-*N*, but an *N*-matched split-half (equalizing precision) revealed a small, consistent own-template advantage in MCI. (3) Average neural state boundary strength: when each cohort was evaluated on its own template, transitions were weaker in MCI (significant in V1 and directionally similar in AG). Given the timing evidence above and scoring on each group's own template, this strength gap is most parsimoniously interpreted as reduced implementation fidelity (blurred/less distinct transitions). Taken together, these results favour a model of shared structure with small, systematic timing differences and weaker implementation in MCI, rather than a fully distinct alternative template.

Interpreting the MCI Own-Template Advantage

How should the MCI group's own-template advantage be interpreted? Prior work reports reduced intersubject segmentation agreement (i.e., greater heterogeneity) in ageing (e.g., Zacks et al., 2006). However, such agreement metrics are broad: they are sensitive to both systematic group shifts in timing and non-systematic heterogeneity (e.g., dispersion, omissions/insertions) and therefore do not separate these effects. A dispersion-only account (greater spread of individual boundary times in MCI but the same typical timing as Controls) struggles to produce the own-template advantage we observe under *N*-matching with LOO. With matched *N*, both templates target the same central times. Larger dispersion in MCI merely makes the estimation of boundary times from finite data less precise (group-average transitions are blurrier), which predicts indifference or even a Control-template advantage, not an MCI advantage. Importantly, in our split-half analysis, MCI were scored on MCI-LOO ($n = 17$) against the mean of two Control half-templates (each $n = 18$, then averaged). That tiny precision asymmetry actually favours Control, so the MCI own-template advantage is, if anything, conservative. Mechanistically, even if dispersion occasionally misplaces a boundary by a few TRs or splits a true transition into two nearby boundaries (which would displace a weaker boundary elsewhere because GSBS is greedy and *k* is fixed), these effects cancel across participants and cannot produce a systematic own-template advantage. Therefore, the most plausible interpretation of the observed advantage is structured timing differences in MCI, potentially arising from small boundary-wise shifts and/or dominant subclusters within the group, rather than broader noise/heterogeneity.

In principle, the lower mean neural boundary strength in MCI could reflect weaker implementation and/or timing mismatch/heterogeneity (if template times miss an individual's true switches). Two considerations point toward weaker implementation as the primary driver. First, scoring each cohort on its own template reduces penalties from between-group timing differences. Second, because a dispersion-only account cannot explain the MCI own-template advantage at *N*-matched, residual timing mismatch is unlikely to be the dominant cause of the strength gap. Nonetheless, mismatch is not eliminated by own-template scoring. Within-group heterogeneity and small systematic shifts can still depress strength. Accordingly, we interpret the gap most parsimoniously as weaker implementation in MCI (less distinct pre/post pattern contrast), while acknowledging that residual timing mismatch may contribute secondarily.

Methodological Constraints on Interpretation

Our choice to fit GSBS on group-averaged data boosts signal-to-noise but sacrifices the individual specificity ideally needed to directly measure between-participant dispersion in timing. In principle, fitting GSBS to each participant's data would allow a direct estimate of individual boundary timing and strength. However, this approach faces significant challenges. The original validation of the GSBS algorithm demonstrated that single-subject fMRI data is often too noisy for reliable state estimation, leading to very low reliability and a tendency for the algorithm to overestimate the number of boundaries as it fits to noise (Geerligs et al., 2021). Consequently, group-averaging was recommended as a standard procedure to boost the signal-to-noise ratio and isolate the shared, stimulus-driven neural response. More recent work, however, argues that this apparent "noise" may in fact contain meaningful, idiosyncratic neural signals (Wilford et al., 2025). Future work using more extensive denoising (e.g., multi-echo denoising) and then estimating individual-level GSBS boundaries will be important to quantify heterogeneity directly.

The voxel sampling procedure itself (Section 4.2.1) likely shaped these results in two opposing ways. For each ROI, we retained only the 200 voxels whose time-courses were most stimulus-synchronized across the entire mixed cohort. This conservative approach makes the group differences in boundary strength and template preference particularly hard to detect, as it discards voxels that were highly reliable in one group but not the other, i.e., the very signals that might best distinguish the cohorts. Notably, the boundary-strength gap (especially in V1) and the MCI own-template advantage under *N*-matched splits nonetheless survived this filtering, indicating that these effects are robust to conservative voxel selection. Conversely,

using a common voxel set likely inflates temporal overlap between group templates by construction, so our estimate of shared timing is probably an upper bound of the true overlap.

Limited Overlap Between Neural and Subjective Boundaries

The near-null overlap between our AG/V1 templates and the 12 subjective boundaries warrants careful interpretation and does not invalidate the subjective boundaries from Chapter 3. Instead, the two boundary types likely reflect different levels of the neural hierarchy. The subjective boundaries, derived from group consensus, likely represent the highest-level event structure that most participants, including those with MCI, would consciously agree upon. Evidence shows that neural state durations lengthen progressively from early sensory cortex through the AG to posterior-medial default-mode hubs (Baldassano et al., 2017). Indeed, selectivity analyses show hippocampus and PCC are uniquely sensitive to boundary salience, consistent with a posterior-medial hierarchy at coarse event scales (Ben-Yakov & Henson, 2018). Our coarse behavioural set (12 events spaced ~40 s apart) likely taps this tier, whereas AG/V1 transitions capture sub-event updates within those larger chunks. Consistent with our limited neural-subjective overlap, Henderson et al. (2025) using individual-level EEG GSBS reported no reliable alignment between neural state boundaries and subjective event boundaries, despite robust between-subject alignment of neural states. They further showed that boundary distinctiveness, not boundary-behaviour alignment, predicts memory, reinforcing the view that neural transitions can be finer-grained and functionally meaningful even when they miss coarse, consensual boundaries.

We observed a significant neural-subjective overlap only in V1 for the Control group. Notably, Reagh et al. (2020), who used the same movie, reported that most consensus boundaries (9/12) coincided with spatial-context shifts. Similarly showed that Ben-Yakov & Henson, using the same movie, showed two relevant points: (i) early visual responses (including V1) rise at many consensus boundaries and covary positively with boundary salience, but (ii) boundary-salience effects remain robust only in hippocampus/PCC after controlling for many perceptual features (including V1/A1 responses). In other words, subjective boundaries often coincide with strong sensory transients, but selectivity for boundaries lives in higher-level regions. This pattern fits our findings: the overlap between V1 neural boundaries in Controls and subjective boundaries likely reflects those sensory transients at a subset of boundaries, not that early visual cortex tracks subjective structure

more than higher-level cortex. AG state changes, likely reflecting conceptual transitions, need not align with these coarse markers.

Importantly, the null overlap we find does not contradict robust evidence that hippocampal univariate activity is both sensitive and specific to subjective event boundaries (Ben-Yakov & Henson, 2018). Furthermore, AG pattern transitions only elicit hippocampal increases when they coincide with subjective boundaries. This supports a region × signal-type dissociation: multivoxel state transitions in cortex can be finer-grained and partially idiosyncratic, while hippocampal univariate transients register salient, consensual boundaries. A useful future check would be to test AG univariate responses to subjective boundaries (and/or the alignment between AG pattern transitions and hippocampal univariate peaks) to further bridge these literatures.

The Hippocampus: Transient Dynamics Instead of Stable States

Perhaps one of the most informative findings from this chapter came from the contrast between the clear state structures in the cortex and GSBS's failure to identify any such structure in the hippocampus. This finding is in line with our initial voxel selection results, which revealed substantially lower ISC in the hippocampus (0.0045-0.0194) compared to the cortex (e.g., AG: 0.0839-0.1355), indicating that hippocampal activity is less consistently time-locked to the stimulus across individuals. This finds empirical support in the work of Brunec et al. (2018), who demonstrated that the pHPC is characterized by intrinsically dynamic neural signals. Specifically, they found that the pHPC exhibits both low inter-voxel similarity (implying complex, non-redundant coding) and low temporal autocorrelation (implying rapidly changing patterns), particularly during demanding tasks. Taken together, these characteristics suggest that while the cortex maintains stable representations of ongoing event content, the hippocampus performs a different, more dynamic role.

Two non-exclusive interpretations can account for its dynamic profile. First, the hippocampus may act as a transient event detector (e.g., Ben-Yakov & Henson, 2018), with neural ensembles that fire at a boundary and then rapidly reconfigure, leaving no sustained pattern correlation between time-points. Second, it may generate sparse pointer-like codes (e.g., Teyler & Rudy, 2007) - an index or "barcode" that refers back to the event content stored in cortex. Either mechanism would yield low pattern similarity and would not be well captured by a pattern-based stable-state model. GSBS detects boundaries by searching for stretches of relatively stable multivoxel activity. So if a region delivers only brief boundary-

locked bursts or extremely sparse pointer codes, there is simply too little sustained similarity for the algorithm to latch onto. This is a known failure mode of the algorithm; when a true state structure is swamped by noise, the estimated number of states tends to climb toward the analyst-imposed ceiling (Geerligs et al., 2021). This is precisely the failure mode observed here, with the hippocampus defaulting to kmax. The flat TR × TR correlation matrices reinforce this interpretation: they show no block-like structure, indicating insufficient within-state coherence for GSBS to detect stable states.

Crucially, this lack of stable states, when combined with the findings from Chapter 3, supports a model of hippocampal function as a transient event detector. The boundary-locked bursts identified in Chapter 3 and the absence of sustained states found here jointly indicate brief resets that mark boundaries, rather than persistent models of the events themselves.

5 General Discussion

5.1 Main Findings

We asked three linked questions: (i) whether boundary-evoked pHPC activity is reduced in MCI beyond effects of atrophy and low-level visual differences, and whether offset magnitude relates to episodic memory; (ii) whether patients show a fundamentally different cortical event structure or a largely shared structure with altered timing and strength; and (iii) whether the hippocampus, like cortex, maintains stable multivoxel states between boundaries or instead operates in a transient, boundary-locked mode. Our data point to clear answers. In Chapter 3, pHPC responses at event offsets were reliably smaller in MCI than in Controls, even after adjusting for age, motion, hippocampal volume, and V1 boundary activity; the FIR results indicated a scaled (not delayed) response. Groups differed in episodic memory as expected (MCI < Controls), but pHPC offset magnitude did not predict the general memory score across or within groups. In Chapter 4, cortical segmentation in AG and V1 was broadly shared across groups, with MCI showing lower mean neural boundary strength and small, systematic shifts in boundary timing. By contrast, GSBS did not recover stable multivoxel states in pHPC for either group, consistent with a transient detector/encoder rather than sustained state maintenance.

5.1.1 Situating Chapter 3

Attenuated boundary-locked pHPC activity in MCI aligns with work showing that hippocampal offsets index the just-ended event for episodic storage (Ben-Yakov & Dudai, 2011; Ben-Yakov & Henson, 2018) and that this signal weakens with age (Reagh et al., 2020). Crucially, because the effect remained after controlling for hippocampal volume, it reflects functional impairment not reducible to atrophy and thus offers information complementary to structural biomarkers. The effect also persisted after controlling for low-level visual boundary responses (V1), and the FIR pattern points to reduced neural gain rather than delayed timing, consistent with broader reports of hippocampal dysfunction in MCI/AD (Pennanen et al., 2004; Wang et al., 2016). Whole-brain analyses showed boundary-locked reductions across PMN/DMN hubs, in line with early PMN vulnerability along the Alzheimer's continuum (Jones et al., 2016). Behaviourally, while the expected memory gap between MCI and Controls was present, pHPC offset magnitude did not track a trait-level narrative memory score. This likely reflects limited power and a mismatch between a

stimulus-locked boundary metric and a general, trait-level memory measure. The absence of a pHPC offset-memory correlation should therefore not be over-interpreted.

5.1.2 Situating Chapter 4

Controls and MCI appear to broadly share a common event structure, as indicated by substantial but incomplete boundary overlap. However, the MCI group shows (i) a small, systematic timing shift and (ii) lower mean boundary strength. This refines reports of less normative segmentation and impaired event integration in ageing/MCI by specifying that the group difference reflects typical timing drift plus weaker transition implementation, not a wholesale reorganisation and not a dispersion-only (idiosyncrasy-only) deficit (Zacks et al., 2006; Kokje et al., 2022). Idiosyncratic variability may co-occur, but the own-template advantage argues against a dispersion-only account (see Chapter 4 Discussion). Parallel evidence for lower mean boundary strength is limited; our result therefore extends the literature alongside work showing temporal dedifferentiation (longer states and weaker boundaries) with age (Lugtmeijer, Oetringer, Geerligs, & Campbell, 2025) and work linking stronger boundary-related change to better recall irrespective of age (Henderson, Oetringer, Geerligs, & Campbell, 2025). Finally, localisation of timing and strength effects within the AG, a PMN hub, situates this "shared but shifted and weaker" event implementation within established Alzheimer's systems biology (Jones et al., 2016).

5.2 Integrating Accounts

5.2.1 Cortico-Hippocampal Loop

Converging evidence on event segmentation points to complementary dynamics that may work as a loop: cortex maintains relatively stable situation models, while the hippocampus issues brief, boundary-locked "save" signals that register the just-ended event into episodic memory. Our data are consistent with this division. Chapter 3 indicates that pHPC shows a brief, boundary-locked response that is reduced in MCI. Chapter 4 shows that GSBS did not recover stable states in pHPC, as expected if the hippocampus transiently reconfigures at boundaries rather than sustaining multivoxel states. In contrast, AG and V1 did exhibit clear state structure. Within this framework, we propose a cortico-hippocampal boundary loop. First, weaker cortical transitions may reduce the boundary evidence delivered to the hippocampus, yielding lower boundary-locked activity (the scaled deficit we observe in Chapter 3). Critically, a reduced hippocampal tag could reflect impaired encoding of the just-

ended event. In turn, weaker hippocampal tags may undermine later cortical reinstatement and predictive use of prior events, which could make subsequent cortical updates less decisive (and possibly slightly mistimed), i.e., the deficits we observe in Chapter 4. This is consistent with work showing that cortical pattern shifts at event offsets couple with hippocampal activity and predict later reinstatement, and that hippocampus-PMN coupling at offsets relates to subsequent memory detail (Baldassano et al., 2017; Barnett et al., 2024). Furthermore, human ECoG supports both directions: information flow from auditory cortex to hippocampus occurs near event boundaries (earlier and more robust on a second exposure), and hippocampus-to-cortex flow precedes moments of predictive recall (Michelmann et al., 2021).

5.2.2 MCI Timing Shifts as an Explanation of the pHPC Boundary Deficit

An alternative (not mutually exclusive) account for the Chapter 3 pHPC boundary deficit is timing dilution: if MCI participants' perceived boundary times are slightly shifted (and possibly more idiosyncratic) relative to Controls, averaging responses at healthy-defined times would place fewer MCI participants "at boundary", reducing the group-mean pHPC boundary response even if patients respond normally at their own boundary times. This is consistent with our Chapter 4 finding that small timing differences exist at the neural level, and with prior work showing less normative and more idiosyncratic segmentation in ageing and early AD/MCI and the subjective/behavioural level (Zacks et al., 2006; Kurby & Zacks, 2011).

In practice, both the cortico-hippocampal mechanism and the timing shift may contribute to the effects we observe. A next step would be to estimate boundaries at the individual level and re-analyse hippocampal offsets using each participant's own cortical (and/or subjective) boundaries. If the pHPC attenuation persists under participant-specific alignment, that would indicate a genuine impairment in the hippocampal boundary response; if it shrinks substantially, misalignment likely plays a larger role. Furthermore, one could then test the integrative cortex-hippocampus account more directly by asking whether an individual's cortical boundary strength predicts pHPC offset magnitude (cortex→pHPC) and whether hippocampal offset strength relates to subsequent stabilisation/organisation of cortical event states (pHPC→cortex) on a boundary-to-boundary basis.

5.3 Clinical Translation

5.3.1 Biomarker Potential and Feasibility

Our strongest translational signal is that the pHPC boundary-offset deficit remained significant after covarying hippocampal volume (and V1 boundary activation) in Chapter 3, i.e., it carries unique functional information beyond atrophy. Coupled with prior work showing that pHPC boundary activity relates to narrative memory (Reagh et al., 2020) and that boundary-linked cortical/hippocampal dynamics support reinstatement (Baldassano et al., 2017), this positions the pHPC boundary response as a plausible functional biomarker of AD, to complement structural MRI and CSF/PET. The segmentation phenotype from Chapter 4 (small timing shift / weaker boundary strength) could also serve as a marker if recovered at the individual level, given links between sharper neural boundaries and better memory (Henderson et al., 2025).

Any engaging naturalistic stimulus can elicit boundary-locked signals (the 8-min film we use is a proven option), but fMRI remains costly and logistically complex, so these measures are best suited as additions in specialist settings for risk stratification and longitudinal tracking, rather than routine screening. An important next step is to determine disease specificity. Per the NIA-AA framework, amyloid status classifies MCI by Alzheimer's biology: Aβ-positive (Aβ+) indicates amyloid deposition on CSF/PET and places individuals in the Alzheimer's continuum, while Aβ-negative (Aβ−) lacks this signature (Jack et al., 2018). Future studies should stratify MCI by amyloid status. If pHPC boundary response deficits (and neural boundary-precision deficits) are larger in Aβ+ than Aβ− MCI and potentially also track PMN integrity and subsequent decline, that would support AD-relevant utility. Comparable effects across Aβ+ and Aβ− would indicate a broader frailty marker. This question fits with evidence that PMN dysfunction may emerge very early, even before measurable amyloid, suggesting potential value as an early systems-level marker (Jones et al., 2015). Unfortunately, the numbers for each type of CSF biomarker in the present study were simply too small to investigate further.

5.3.2 Can We Improve Memory?

Our findings may also inform how memory could be improved. A recent review by Smith and Zacks (2024) discusses why interventions requiring active, self-generated segmentation often fail to benefit older adults, proposing several explanations, including a "representation deficit." Our results support to this specific hypothesis. The Chapter 4

findings of a small timing shift and weaker implementation in MCI can be interpreted as a neural-level signature of such a representational deficit. An active segmentation task requires participants to rely on these internal event models to identify boundaries. If the neural representations of those event structures are degraded and less distinct, the task is unlikely to yield a memory benefit.

By contrast, external cueing, i.e., editing the material to flag likely event boundaries, has been shown to improve memory in older adults, especially when cues align with boundaries (Gold et al., 2017). Our findings may also offer a mechanistic account for the efficacy of these external cueing interventions. As outlined above, the less distinct state transitions observed in the cortex may provide an impoverished or ambiguous input signal to the hippocampus, culminating in the blunted mean boundary-evoked response we observed in the pHPC. External cues, as used by Gold et al. (2017), may be effective precisely because they bypass this failing mechanism. Rather than relying on a subtle, top-down cortical signal to indicate a boundary, an external cue provides a salient bottom-up sensory event. This exogenous marker could drive a state update, triggering the residual encoding capacity of the hippocampus, which the cortex appears no longer able to reliably elicit on its own. In this view, our findings provide a neural explanation for why interventions that provide external structure may be more beneficial for MCI than those that demand the internal generation of that structure.

5.4 Modulators of Cortical Updates in MCI

While the primary deficits in MCI lie in episodic memory - and, as argued here, in event segmentation - semantic, attentional, and executive weaknesses are common, variable across individuals and tend to emerge under higher task load. These factors could therefore modulate the strength and timing of cortical event updates and contribute (at least to some extent) to the results we find.

Event Schemas

Event schemas provide a top-down scaffold that stabilizes and predicts ongoing narratives. In naturalistic narrative fMRI, default-mode hubs (especially mPFC, PMC, and SFG) carry schema-specific event patterns that generalize across stories, modalities, and people (Baldassano, Hasson, & Norman, 2018). Critically, disrupting temporal order weakens mPFC responses, and shared context increases alignment in PCC/mPFC/AG, while

perspective cues nudge boundary placement toward normative timings (Baldassano, Hasson, & Norman, 2018; Nguyen et al., 2019; Yeshurun et al., 2017; Newberry & Bailey, 2019). If schema access or deployment is impaired in MCI, situation-model updates at event transitions are likely to resolve less decisively and less normatively, which could yield the small timing shifts in boundary placement and weaker boundary implementation we observe. This is consistent with degraded everyday action knowledge in dementia (Roll et al., 2019).

Reduced Inhibition

Another plausible contributor is reduced inhibition: insufficient "reset" at transitions allows residual activity to bleed into the next state, weakening boundary strength. Converging evidence from magnetic resonance spectroscopy (MRS) shows decreased GABA in MCI (e.g., ACC/PCC and sensory cortices), and reviews implicate GABAergic dysfunction in MCI/AD, consistent with a weaker inhibitory "reset" at transitions (i.e., residual activity bleeding into the next state) (Fu et al., 2023; Zuppichini et al., 2025; Jiménez-Balado & Eich, 2021). Importantly, incomplete suppression could also produce slight temporal lags in settling into the next state, aligning with small timing shifts, while simultaneously reducing the pre/post pattern contrast (lower strength).

Attentional Control

Deficits in attentional control may also contribute to the segmentation impairments we report. MCI often entails weaknesses in sustained/divided attention and executive control (Saunders & Summers, 2011; Kirova et al., 2015; Gordon et al., 2015), and attention modulates inter-participant neural alignment during narratives (Ki et al., 2016). The literature supports a task- and load-dependent view of attention in ageing and MCI. During movie-watching, intersubject neural synchrony declines with age and covaries with attentional control (higher fluid intelligence, lower RT variability), implying that top-down control helps the brain follow a complex storyline (Campbell et al., 2015). In MCI, reviews and task studies point to heterogeneous but reliable weaknesses in executive/attentional control, with divided attention and higher-load vigilance showing earlier or larger effects than simple vigilance, and performance worsening as demands increase (Kirova et al., 2015). Under standard event-segmentation accounts, maintaining and updating a situation model requires sustained control to keep the current model active and to detect when predictions fail. Diminished control would naturally yield less distinct neural states. Taken together, we view attentional control as a context-dependent moderator of segmentation under narrative load, not a sufficient explanation on its own.

5.5 Strengths and Limitations

Naturalistic Stimuli

A key strength of our approach is its ecological validity. Engaging narratives let us probe how the brain processes continuous, task-free experience, much closer to daily life than trial-wise tasks. Movies naturally recruit event segmentation mechanisms and a constellation of systems known to weaken in cognitive decline (PMN/DMN integration over long timescales, hippocampal offset responses, sustained attention, executive control, schema use, etc.), and the data distribution better matches real-world inputs, improving the odds that findings generalize beyond the scanner. At the same time, naturalistic designs afford less control and manipulation of variables than traditional experiments. Multiple features co-occur at boundaries, causal inference is harder, and effects can be stimulus-bound (e.g., genre, cultural familiarity, script knowledge). In addition, passive viewing may not fully capture the demands of active, goal-directed behaviour, where maintaining goals and preparing actions can shape segmentation. Future work should therefore test whether MCI shows similar or distinct segmentation deficits during active, goal-directed contexts (e.g., task switching or goal maintenance paradigms) compared with passive narrative viewing.

Stimulus-feature controls

Naturalistic designs may raise concerns about low-level confounds. Reagh et al. (who used the same 8-minute Hitchcock film) explicitly modeled a high-frequency visual regressor (edge-pixel proportion per TR) alongside the boundary > within-event contrast. They found that low-level visual change did not account for PMN boundary responses or their age effects, and visual cortex showed boundary responses without an age relationship. Many (9/12) boundaries coincided with spatial-context shifts, but segmentation can also arise from prediction error even without large sensory changes (Zacks et al., 2010). In our study, we also contrasted boundary > within-event and covaried V1 boundary activation for the hippocampus and the pHPC group effect persisted. Because both groups viewed the identical stimulus, unmodeled low-level variance is largely shared, so not including detailed low-level visual regressors is unlikely to create spurious group differences. However, it is still possible that groups differ in neural sensitivity to the same low-level features.

Noise Considerations and Estimation Strategy

Our 8-min movie yields only 12 boundaries, limiting within-subject precision. For the GLM in Chapter 3, more boundaries would increase samples for the boundary > within-event contrast, lowering beta variance. With respect to neural state detection, Geerligs et al. (2021) report that single-participant state detection can be noisy and that across-participant averaging improves reliability (relevant to Chapter 4). But this choice comes with trade-offs: (i) it conflates weaker within-participant transitions with any potential between-subject timing heterogeneity, making the MCI strength gap ambiguous in isolation; and (ii) it penalizes individuals whose true boundaries are offset from the group template, which may disproportionately affect the potentially more heterogeneous MCI group. Consequently, group-level fitting limits immediate individual-level biomarker utility.

To support individual-level estimation, averaging across within-subject repeats would stabilize estimates, but re-watching can alter neural responses. A better alternative is multi-echo EPI with ME-ICA denoising to remove non-BOLD components (Kundu et al., 2012; Geerligs et al., 2021). This would allow us to quantify per-person timing variance as a disease-relevant phenotype and to test whether an individual's neural boundary strength and timing predict their hippocampal offsets and memory on a moment-by-moment basis.

Against these challenges, the thesis already incorporates several SNR-positive choices: a boundary-locked contrast (boundary vs within) to cancel shared variance; an optimised fMRIPrep-based pipeline (WM/CSF regression, motion + derivatives, high-pass at 0.0078 Hz, 6 mm smoothing; see Appendix); inclusion of mean FD as a group-level covariate; V1 boundary activation as a covariate in hippocampal models; FIR checks showing a scaled (not shifted) pHPC response in MCI; conservative voxel selection (top-ISC voxels) within common ROIs; and group-average GSBS with leave-one-out templates. Together, these choices reduce noise-driven false positives and, if anything, make the observed effects harder, not easier, to detect.

5.6 Conclusion

This thesis asked whether the episodic memory deficits characteristic of MCI/AD can be understood as a failure to parse continuous experience into discrete events. Our findings support this view, pointing to complementary impairments within the cortico-hippocampal system for event segmentation. In the pHPC, boundary-evoked responses were reliably attenuated even after controlling for atrophy and low-level visual change. Concurrently, we

found that, while individuals with MCI share a broadly similar cortical event structure with controls, it is implemented less distinctly and with subtle timing shifts. This functional dissociation was reinforced by the observation that the hippocampus acts as a transient boundary detector, failing to show the stable event states characteristic of the cortex. Taken together, a degraded cortical situation model and an attenuated hippocampal “save” signal may provide a neural account of the fragmented recall seen in MCI, indicating that a breakdown in event segmentation is a key contributor to their everyday episodic memory impairment.

6 References

Alzheimer's Association. (2025). 2025 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *21*(5), e1-e188.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709–721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *The Journal of Neuroscience*, *38*(45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395–409). Oxford University Press.

Barnett, A. J., Nguyen, M., Spargo, J., Yadav, R., Cohn-Sheehy, B. I., & Ranganath, C. (2024). Hippocampal-cortical interactions during event boundaries support retention of complex narrative events. *Neuron*, *112*(2), 319–330.e7. https://doi.org/10.1016/j.neuron.2023.10.010

Ben-Yakov, A., & Dudai, Y. (2011). Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *The Journal of Neuroscience*, *31*(24), 9032–9042. https://doi.org/10.1523/JNEUROSCI.0702-11.2011

Ben-Yakov, A., & Henson, R. N. (2018). The hippocampal film editor: Sensitivity and specificity to event boundaries in continuous experience. *The Journal of Neuroscience*, *38*(47), 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018

bogpetre. (2021, August 26). Searchlight hyperalignment inappropriately scales values (#626) [Issue]. In PyMVPA/PyMVPA. GitHub. https://github.com/PyMVPA/PyMVPA/issues/626

Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z. X., Grady, C., Rosenbaum, R. S., Winocur, G., Barense, M. D., & Moscovitch, M. (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Current Biology*, *28*(13), 2129–2135.e6. https://doi.org/10.1016/j.cub.2018.05.016

Campbell, K. L., Shafto, M. A., Wright, P., Tsvetanov, K. A., Geerligs, L., Cusack, R., Cam-CAN, & Tyler, L. K. (2015). Idiosyncratic responding during movie-watching predicted by age differences in attentional control. *Neurobiology of aging*, *36*(11), 3045–3055. https://doi.org/10.1016/j.neurobiolaging.2015.07.028

Cannizzaro, M. S., & Coelho, C. A. (2013). Analysis of narrative discourse structure as an ecologically relevant measure of executive function in adults. *Journal of Psycholinguistic Research*, *42*(6), 527–549. https://doi.org/10.1007/s10936-012-9231-5

Chen, J., & Bornstein, A. M. (2024). The causal structure and computational value of narratives. *Trends in Cognitive Sciences*, *28*(8), 769–781. https://doi.org/10.1016/j.tics.2024.04.003

Cooper, R. A., Kurkela, K. A., Davis, S. W., & Ritchey, M. (2021). Mapping the organization and dynamics of the posterior medial network during movie watching. *NeuroImage*, *236*, Article 118075. https://doi.org/10.1016/j.neuroimage.2021.118075

Drummond, C., Coutinho, G., Fonseca, R. P., Assunção, N., Teldeschi, A., de Oliveira-Souza, R., Moll, J., Tovar-Moll, F., & Mattos, P. (2015). Deficits in narrative discourse elicited by visual stimuli are already present in patients with mild cognitive impairment. *Frontiers in Aging Neuroscience*, *7*, Article 96. https://doi.org/10.3389/fnagi.2015.00096

Du, A. T., Schuff, N., Kramer, J. H., Ganzer, S., Zhu, X. P., Jagust, W. J., Miller, B. L., Reed, B. R., Mungas, D., Yaffe, K., Chui, H. C., & Weiner, M. W. (2004). Higher atrophy rate of entorhinal cortex than hippocampus in AD. *Neurology*, *62*(3), 422–427. https://doi.org/10.1212/01.wnl.0000106462.72282.90

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–1286. https://doi.org/10.1037/a0034024

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., et al. (2018). *fMRIPrep* (Version 24.1.1) [Software]. Zenodo. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, *22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Flicker, C., Ferris, S. H., & Reisberg, B. (1991). Mild cognitive impairment in the elderly: predictors of dementia. *Neurology*, *41*(7), 1006–1009. https://doi.org/10.1212/wnl.41.7.1006

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fraser, K. C., Lundholm Fors, K., Eckerström, M., Öhman, F., & Kokkinakis, D. (2019). Predicting MCI status from multimodal language data using cascaded classifiers. *Frontiers in Aging Neuroscience*, *11*, Article 205. https://doi.org/10.3389/fnagi.2019.00205

Fu, X., Qin, M., Liu, X., Cheng, L., Zhang, L., Zhang, X., Lei, Y., Zhou, Q., Sun, P., Lin, L., Su, Y., & Wang, J. (2023). Decreased GABA levels of the anterior and posterior cingulate cortex are associated with executive dysfunction in mild cognitive impairment. *Frontiers in Neuroscience*, *17*, 1220122. https://doi.org/10.3389/fnins.2023.1220122

Geerligs, L., Cam-CAN, & Campbell, K. L. (2018). Age-related differences in information processing during movie watching. *Neurobiology of Aging*, *72*, 106–120. https://doi.org/10.1016/j.neurobiolaging.2018.07.025

Geerligs, L., Gözükara, D., Oetringer, D., Campbell, K. L., van Gerven, M., & Güçlü, U. (2022). A partially nested cortical hierarchy of neural states underlies event segmentation in the human brain. *eLife*, *11*, Article e77430. https://doi.org/10.7554/eLife.77430

Geerligs, L., van Gerven, M., & Güçlü, U. (2021). Detecting neural state transitions underlying event segmentation. *NeuroImage*, *236*, Article 118085. https://doi.org/10.1016/j.neuroimage.2021.118085

Giannakopoulos, P., Gold, G., Duc, M., Michel, J. P., Hof, P. R., & Bouras, C. (2000). Neural substrates of spatial and temporal disorientation in Alzheimer's disease. *Acta Neuropathologica*, *100*(2), 189–195. https://doi.org/10.1007/s004019900166

Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, *2*(1), Article 1. https://doi.org/10.1186/s41235-016-0043-2

Gordon, B. A., Zacks, J. M., Blazey, T., Benzinger, T. L., Morris, J. C., Fagan, A. M., Holtzman, D. M., & Balota, D. A. (2015). Task-evoked fMRI changes in attention networks are associated with preclinical Alzheimer's disease biomarkers. *Neurobiology of Aging*, *36*(5), 1771–1779. https://doi.org/10.1016/j.neurobiolaging.2015.01.019

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, Article 13. https://doi.org/10.3389/fninf.2011.00013

Greene, J. D., Baddeley, A. D., & Hodges, J. R. (1996). Analysis of the episodic memory deficit in early Alzheimer's disease: Evidence from the doors and people test. *Neuropsychologia*, *34*(6), 537–551. https://doi.org/10.1016/0028-3932(95)00151-4

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral cortex (New York, N.Y. : 1991)*, *26*(6), 2919–2934. https://doi.org/10.1093/cercor/bhw068

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *The Journal of Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Henderson, S. E., Oetringer, D., Geerligs, L., & Campbell, K. L. (2025). Neural state changes during movie watching relate to episodic memory in younger and older adults. *Cerebral Cortex*, *35*(5), Article bhaf114. https://doi.org/10.1093/cercor/bhaf114

Hodges, J. R. (2000). Memory in the dementias. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 441–459). Oxford University Press.

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., Doyle, W. K., Rubin, N., Heeger, D. J., & Hasson, U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*, *76*(2), 423–434. https://doi.org/10.1016/j.neuron.2012.08.011

Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E.,

Snyder, H. M., Sperling, R., ... (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jekel, K., Damian, M., Wattmo, C., Hausner, L., Bullock, R., Connelly, P. J., Dubois, B., Eriksdotter, M., Ewers, M., Graessel, E., Kramberger, M. G., Law, E., Mecocci, P., Molinuevo, J. L., Nygård, L., Olde-Rikkert, M. G., Orgogozo, J. M., Pasquier, F., Peres, K., Salmon, E., … Frölich, L. (2015). Mild cognitive impairment and deficits in instrumental activities of daily living: A systematic review. *Alzheimer's Research & Therapy*, *7*(1), Article 17. https://doi.org/10.1186/s13195-015-0099-0

Jiménez-Balado, J., & Eich, T. S. (2021). GABAergic dysfunction, neural network hyperactivity and memory impairments in human aging and Alzheimer's disease. *Seminars in Cell & Developmental Biology*, *116*, 146–159. https://doi.org/10.1016/j.semcdb.2021.01.005

Jones, D. T., Knopman, D. S., Gunter, J. L., Graff-Radford, J., Vemuri, P., Boeve, B. F., Petersen, R. C., Weiner, M. W., Jack, C. R., Jr., & Alzheimer's Disease Neuroimaging Initiative. (2016). Cascading network failure across the Alzheimer's disease spectrum. *Brain: A Journal of Neurology*, *139*(2), 547–562. https://doi.org/10.1093/brain/awv338

Joubert, S., Gardy, L., Didic, M., Rouleau, I., & Barbeau, E. J. (2021). A Meta-Analysis of Semantic Memory in Mild Cognitive Impairment. *Neuropsychology review*, *31*(2), 221–232. https://doi.org/10.1007/s11065-020-09453-5

Ki, J. J., Kelly, S. P., & Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *The Journal of Neuroscience*, *36*(10), 3092–3101. https://doi.org/10.1523/JNEUROSCI.2942-15.2016

Kirova, A. M., Bays, R. B., & Lagalwar, S. (2015). Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *BioMed Research International*, *2015*, Article 748212. https://doi.org/10.1155/2015/748212

Kokje, E., Gerwien, J., & von Stutterheim, C. (2022). Macro-event recognition in healthy ageing, Alzheimer's disease, and mild cognitive impairment. *Journal of Neuropsychology*, *16*(2), 306–323. https://doi.org/10.1111/jnp.12271

Kundu, P., Inati, S. J., Evans, J. W., Luh, W. M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, *60*(3), 1759–1770. https://doi.org/10.1016/j.neuroimage.2011.12.028

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition*, *39*(1), 75–91. https://doi.org/10.3758/s13421-010-0027-2

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Liu, W., Shi, Y., Cousins, J. N., Kohn, N., & Fernández, G. (2022). Hippocampal-medial prefrontal event segmentation and integration contribute to episodic memory formation. *Cerebral Cortex*, *32*(5), 949–969. https://doi.org/10.1093/cercor/bhab258

Lugtmeijer, S., Oetringer, D., Geerligs, L., & Campbell, K. L. (2025). *Temporal dedifferentiation of neural states with age during naturalistic viewing* [Preprint]. bioRxiv. https://doi.org/10.1101/2025.01.26.634954

Michelmann, S., Price, A. R., Aubrey, B., Strauss, C. K., Doyle, W. K., Friedman, D., Dugan, P. C., Devinsky, O., Devore, S., Flinker, A., Hasson, U., & Norman, K. A. (2021). Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. *Nature communications*, *12*(1), 5394. https://doi.org/10.1038/s41467-021-25376-y

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, *21*(11), 1078–1085. https://doi.org/10.1002/gps.1610

Murphy, K. J., Troyer, A. K., Levine, B., & Moscovitch, M. (2008). Episodic, but not semantic, autobiographical memory is reduced in amnestic mild cognitive impairment. *Neuropsychologia*, *46*(13), 3116–3123. https://doi.org/10.1016/j.neuropsychologia.2008.07.004

Naci, L., Cusack, R., Anello, M., & Owen, A. M. (2014). A common neural code for similar conscious experiences in different individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(39), 14277–14282. https://doi.org/10.1073/pnas.1407007111

Newberry, K. M., & Bailey, H. R. (2019). Does semantic knowledge influence event segmentation and recall of text? *Memory & Cognition*, *47*(7), 1173–1187. https://doi.org/10.3758/s13421-019-00926-4

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, *184*, 161–170. https://doi.org/10.1016/j.neuroimage.2018.09.010

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology*, *60*, 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hänninen, T., Laakso, M. P., Hallikainen, M., Vanhanen, M., Nissinen, A., Helkala, E. L., Vainio, P., Vanninen, R., Partanen, K., & Soininen, H. (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of Aging*, *25*(3), 303–310. https://doi.org/10.1016/S0197-4580(03)00084-8

Petersen, R. C. (2000). Aging, mild cognitive impairment, and Alzheimer's disease. *Neurologic Clinics*, *18*(4), 789–806. https://doi.org/10.1016/s0733-8619(05)70226-7

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*(3), 303–308. https://doi.org/10.1001/archneur.56.3.303

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133–140. https://doi.org/10.1016/j.cobeha.2017.08.006

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713–726. https://doi.org/10.1038/nrn3338

Raykov, P. (2019). *Neurobiology of specific and general prior knowledge* [Doctoral dissertation, University of Sussex].

Reagh, Z. M., Delarazan, A. I., Garber, A., & Ranganath, C. (2020). Aging alters neural activity at event boundaries in the hippocampus and posterior medial network. *Nature Communications*, *11*(1), Article 3980. https://doi.org/10.1038/s41467-020-17713-4

Reagh, Z. M., & Ranganath, C. (2018). What does the functional organization of cortico-hippocampal networks tell us about the functional organization of memory? *Neuroscience Letters*, *680*, 69–76. https://doi.org/10.1016/j.neulet.2018.04.050

Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: Event models from perception to action. *Trends in Cognitive Sciences*, *21*(12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005

Ritchey, M., Libby, L. A., & Ranganath, C. (2015). Cortico-hippocampal systems involved in memory and cognition: The PMAT framework. *Progress in Brain Research*, *219*, 45–64. https://doi.org/10.1016/bs.pbr.2015.04.001

Ritchey, M., Montchal, M. E., Yonelinas, A. P., & Ranganath, C. (2015). Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. *eLife*, *4*, Article e05025. https://doi.org/10.7554/eLife.05025

Robert, P. H., Schuck, S., Dubois, B., Olié, J. P., Lépine, J. P., Gallarda, T., Goni, S., Troy, S., & Investigators' Group. (2003). Screening for Alzheimer's disease with the short cognitive evaluation battery. *Dementia and Geriatric Cognitive Disorders*, *15*(2), 92–98. https://doi.org/10.1159/000067971

Roll, E. E., Giovannetti, T., Libon, D. J., & Eppig, J. (2019). Everyday task knowledge and everyday function in dementia. *Journal of Neuropsychology*, *13*(1), 96–120. https://doi.org/10.1111/jnp.12135

Rumshiskaya, A. D., Mershina, E. A., & Pechenkova, E. (2013). *Head motion as a potential source of artifacts in fMRI of motor cortex: A comparative study of healthy volunteers and patients with space-occupying lesions of the brain* [Scientific exhibit]. European Congress of Radiology (ECR 2013), Vienna, Austria. https://doi.org/10.1594/ecr2013/C-1975

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Saunders, N. L., & Summers, M. J. (2011). Longitudinal deficits to attention, executive, and working memory in subtypes of mild cognitive impairment. *Neuropsychology*, *25*(2), 237–248. https://doi.org/10.1037/a0021134

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates.

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., Calder, A. J., Marslen-Wilson, W. D., Duncan, J., Dalgleish, T., Henson, R. N., Brayne, C., Matthews, F. E., & Cam-CAN. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*, Article 204. https://doi.org/10.1186/s12883-014-0204-1

Smith, M. E., & Zacks, J. M. (2025). Event segmentation interventions improve memory for naturalistic events. *Current Directions in Psychological Science*. Advance online publication. https://doi.org/10.1177/09637214251350690

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, *23*(5), 1052–1064. https://doi.org/10.1162/jocn.2010.21524

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257. https://doi.org/10.1037/a0015631

Taler, V., Monetta, L., Sheppard, C., & Ohman, A. (2020). Semantic Function in Mild Cognitive Impairment. *Frontiers in psychology*, *10*, 3041. https://doi.org/10.3389/fpsyg.2019.03041

Teyler, T. J., & Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory: Updating the index. *Hippocampus*, *17*(12), 1158–1169. https://doi.org/10.1002/hipo.20350

Tomasi, D., & Volkow, N. D. (2023). Brain motion networks predict head motion during rest- and task-fMRI. *Frontiers in neuroscience*, *17*, 1096232. https://doi.org/10.3389/fnins.2023.1096232

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*, *9*(4), 408–426. https://doi.org/10.1177/1745691614535933

Wang, P., Li, J., Li, H. J., Huo, L., & Li, R. (2016). Mild cognitive impairment is not "mild" at all in altered activation of episodic memory brain networks: Evidence from ALE meta-analysis. *Frontiers in Aging Neuroscience*, *8*, Article 260. https://doi.org/10.3389/fnagi.2016.00260

Wechsler, D. (2009). *Wechsler Memory Scale* (4th ed.). Pearson.

Wilford, R. E., Chen, H., Wharton-Shukster, E., Finn, A. S., & Duncan, K. (2025). Personalized Neural State Segmentation: Validating the Greedy State Boundary Search Algorithm for Individual-level Functional Magnetic Resonance Imaging Data. *Journal of cognitive neuroscience*, 1–24. Advance online publication. https://doi.org/10.1162/jocn_a_02345

Woodard, J. L., Seidenberg, M., Nielson, K. A., Antuono, P., Guidotti, L., Durgerian, S., Zhang, Q., Lancaster, M., Hantke, N., Butts, A., & Rao, S. M. (2009). Semantic memory activation in amnestic mild cognitive impairment. *Brain : a journal of neurology*, *132*(Pt 8), 2068–2078. https://doi.org/10.1093/brain/awp157

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same story, different story. *Psychological Science*, *28*(3), 307–319. https://doi.org/10.1177/0956797616682029

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind–brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*, 168. https://doi.org/10.3389/fnhum.2010.00168

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, *21*(3), 466–482. https://doi.org/10.1037/0882-7974.21.3.466

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zuppichini, M. D., Hamlin, A. M., Zhou, Q., Kim, E., Wyatt, K., Reardon, N., Hampstead, B. M., & Polk, T. A. (2025). GABA levels are significantly reduced in the visual, motor, and auditory cortex of patients with mild cognitive impairment. *Aging and Disease*. Advance online publication. https://doi.org/10.14336/AD.2025.0334

7 Appendix

7.1 Exploratory Preprocessing Optimizations

This section documents the preprocessing experiments conducted to optimize the input data for boundary detection with GSBS. We systematically compared four distinct preprocessing pipelines built from a common starting point. This initial step, which we term the Baseline condition, included standard denoising (regression of 6 head motion parameters and their first temporal derivatives, plus 128s high-pass filtering). We then evaluated the effects of two additional procedures - functional hyperalignment (HA) and an extended denoising (ExtDen) protocol - resulting in a 2x2 comparison of four pipelines: (1) Baseline, (2) Baseline + HA, (3) ExtDen, and (4) ExtDen + HA. Functional hyperalignment introduces a rotational transformation to bring multivariate response trajectories into a common space across participants. The extended denoising pipeline consisted of two further steps: global white matter (WM) and cerebrospinal fluid (CSF) signal regression, plus spatial smoothing.

Our evaluation of these pipelines centred on two key quality metrics. The first was inter-subject correlation (ISC), a measure of shared neural responses across participants. The second was the qualitative clarity of the block structure within ROI-averaged, timepoint-by-timepoint (TR × TR) correlation matrices of activity patterns, as this structure is the primary feature that GSBS is designed to detect.

Investigating Functional Hyperalignment

Conventional anatomical normalization aligns gross cortical landmarks between brains but is insufficient to align the fine-grained, idiosyncratic topographies of neural representations at the voxel level. Because GSBS models the trajectory of multivariate activity patterns, these individual differences can obscure the detection of shared event structure when data are aggregated across a group. Functional hyperalignment addresses this limitation by using the neural data itself to find an optimal transformation that aligns these functional topographies (Haxby et al., 2011; Guntupalli et al., 2016). The algorithm takes as input the time-series of high-dimensional voxel response vectors for each participant and computes a unique rotation matrix (R_s) for each subject (s). When applied, this matrix transforms the subject's data to maximize the similarity of their functional trajectory to a

common, group-level trajectory. This optimization is solved using an iterative Procrustes procedure, which guarantees the transformation contains no scaling or shearing. Reflecting its utility, hyperalignment is increasingly a standard step in fMRI event segmentation studies (e.g., Geerligs et al., 2021).

Procedure and Methodological Rationale for Hyperalignment

Our implementation of this iterative process was performed separately for each participant group using PyMVPA. The alignment was based on data from the top 200 most reliable voxels, identified using a pooled, all-groups ISC analysis (detailed in Section 4.2.1). The procedure began by incrementally building a common model space, starting with a pre-selected reference participant and sequentially aligning and averaging in each subsequent participant. This initial model was then refined over three subsequent iterations; in each step, every participant's original data was re-aligned to the improved group-average trajectory computed in the previous step. A final third pass then realigned every participant once more to the converged common space, producing the transformation ultimately applied in downstream analyses. Throughout, every voxel time-series was z-scored before each Procrustes step to preserve relative variance.

We opted for this ROI-level hyperalignment approach over a searchlight-based method for two primary reasons. First, since our final analysis with GSBS operates on ROI-averaged signals, it is sensitive to the overall pattern change across all voxels in the region, not to distinct sub-patterns within it. Any fine-grained alignment improvements afforded by a searchlight approach would likely be negated by running GSBS at the ROI-level. Second, searchlight hyperalignment can introduce a significant geometric artifact, a scaling issue first documented by Bogpetre (2021). The procedure involves computing transformation matrices for thousands of overlapping searchlight discs and then summing them to create a single transform. This summation process is the source of the artifact: the final transformation for any given voxel is scaled in proportion to the number of searchlights in which it participated. To demonstrate this, the author calculated a "searchlight count" for each vertex on a standard cortical surface and compared this to the diagonal elements of a cumulative transformation matrix from a separate hyperalignment analysis. A clear positive correlation was found, confirming that the magnitude of the transformation is directly related to a voxel's searchlight participation count. This superimposes a gain field across the data that reflects the geometry of the searchlight tiling rather than underlying functional organisation. Our whole-ROI

approach, which computes a single, unaggregated transformation for the entire region, avoids this confound.

During initial implementation, a comparison between two commonly used software packages, nltools and PyMVPA, revealed an important methodological difference. Preliminary analyses conducted with nltools indicated a substantial reduction in voxel-wise variance, with approximately 90% of voxels showing near-zero signal fluctuation after alignment. This effect likely arises from the global Frobenius-norm scaling step applied within each iteration of the Procrustes alignment algorithm. Specifically, this step involves dividing each participant's data matrix (X) by its Frobenius norm, defined as the square root of the sum of squared matrix values ($\sqrt{\Sigma X_{vt}^2}$), which represents the total squared amplitude or energy of the matrix. Because this norm is particularly sensitive to the highest voxel values, participants with intrinsically higher signal variance have a disproportionately large scaling factor, substantially diminishing their voxel-wise fluctuations. PyMVPA avoids this limitation by performing voxel-wise z-scoring individually on each voxel's time-series prior to alignment. This preserves the relative variance structure across voxels while standardizing each voxel's dynamic range. In light of these observations, all subsequent hyperalignment analyses were conducted using PyMVPA.

Impact of Hyperalignment on ISC and Event Structure

To quantify the impact of hyperalignment, we employed two distinct measures of ISC. The first, ROI-level ISC, measures the synchrony of the region's overall activity. It was calculated by first averaging the time-series of all voxels within the ROI for each participant and then computing the mean pairwise correlation between these single, region-averaged time-courses. The second, voxel-wise ISC, directly assesses the success of functional alignment at the finest-grained level. It was calculated on a per-voxel basis by correlating that voxel's time-series across all pairs of participants and then averaging these pairwise correlations; the final metric is the mean of these ISC values across all voxels in the ROI. Using both metrics is crucial because they probe complementary spatial scales. Voxel-wise ISC tells us whether hyperalignment succeeded in bringing fine-grained functional topographies into register, which is the purpose of the rotation. On the other hand, ROI-level ISC captures whether the coarse, region-level trajectory that GSBS will analyse remains synchronised across brains.

When applied to the Baseline data, hyperalignment (Baseline + HA) substantially increased ISC at both scales (AG-controls: voxel-ISC 0.022→0.343; ROI-ISC 0.198→0.392; Table 1). However, this impressive gain in ISC was accompanied by a degradation of the block-diagonal structure in the ROI-averaged TR × TR correlation matrices (Figure 11). It is noteworthy that prior to extended denoising, both the Baseline and the Baseline + HA conditions exhibited poor block-like structure (Figure 11), but the addition of hyperalignment visibly weakened it further. A plausible mechanistic explanation for this dissociation is that hyperalignment, in optimising for moment-to-moment pattern similarity, can inadvertently blur the boundaries between distinct neural events. The algorithm computes a single, global rotation matrix for each participant that is solved across the entire time-series. It is therefore blind to event boundaries. If two adjacent but distinct events recruit a partially overlapping set of neural components, the rotation that best aligns one event will invariably make the other more similar across subjects as well. This "leaking" of similarity across temporal boundaries softens the sharp transitions that define the blocks in the TR × TR matrix.

Investigating Extended Denoising

The second set of optimisations involved applying an extended denoising pipeline to the data, which had already undergone initial preprocessing with fMRIPrep 24.1.1 as well as the standard denoising mentioned previously. This extended procedure consisted of two additional steps, with parameters selected based on common practice rather than systematic tuning. Physiological Nuisance Regression involved regressing the mean signals from WM and CSF (provided by fMRIPrep's confounds file) from the data. Spatial smoothing was then performed by applying a Gaussian smoothing kernel of 6 mm full-width at half-maximum (FWHM), a value chosen for its common use in previous GSBS studies.

The ExtDen pipeline produced the most distinct and well-defined block-diagonal patterns of any pipeline (Figure 11). This improvement was chiefly attributable to spatial smoothing, which enhances the signal-to-noise ratio on the presumption that the states that GSBS detects have a coarser spatial scale than individual voxels. When hyperalignment was subsequently applied to create the ExtDen + HA pipeline, the results were particularly revealing. Both ROI- and voxel-ISC improved relative to ExtDen alone (AG-controls ROI 0.231→0.274; voxel 0.130→0.232) but remained below Baseline + HA (ROI 0.392; voxel 0.343; Table 1). This is likely because hyperalignment achieves its largest gains by aligning fine-grained, high-spatial-frequency variance; spatial smoothing, as a low-pass filter, removes

this very information before the alignment algorithm can use it, thereby setting a lower ceiling on the maximum achievable ISC.

Given that the primary goal was to optimise the input for GSBS, and considering its reliance on clear event structure, the modest ROI-level ISC gain offered by the ExtDen + HA pipeline did not justify the degradation of the block pattern. Therefore, the final preprocessing pipeline selected for the main analysis was the ExtDen pipeline, as it offered the best compromise between respectable signal synchrony and the sharp event boundary structure critical for segmentation.

Table 1

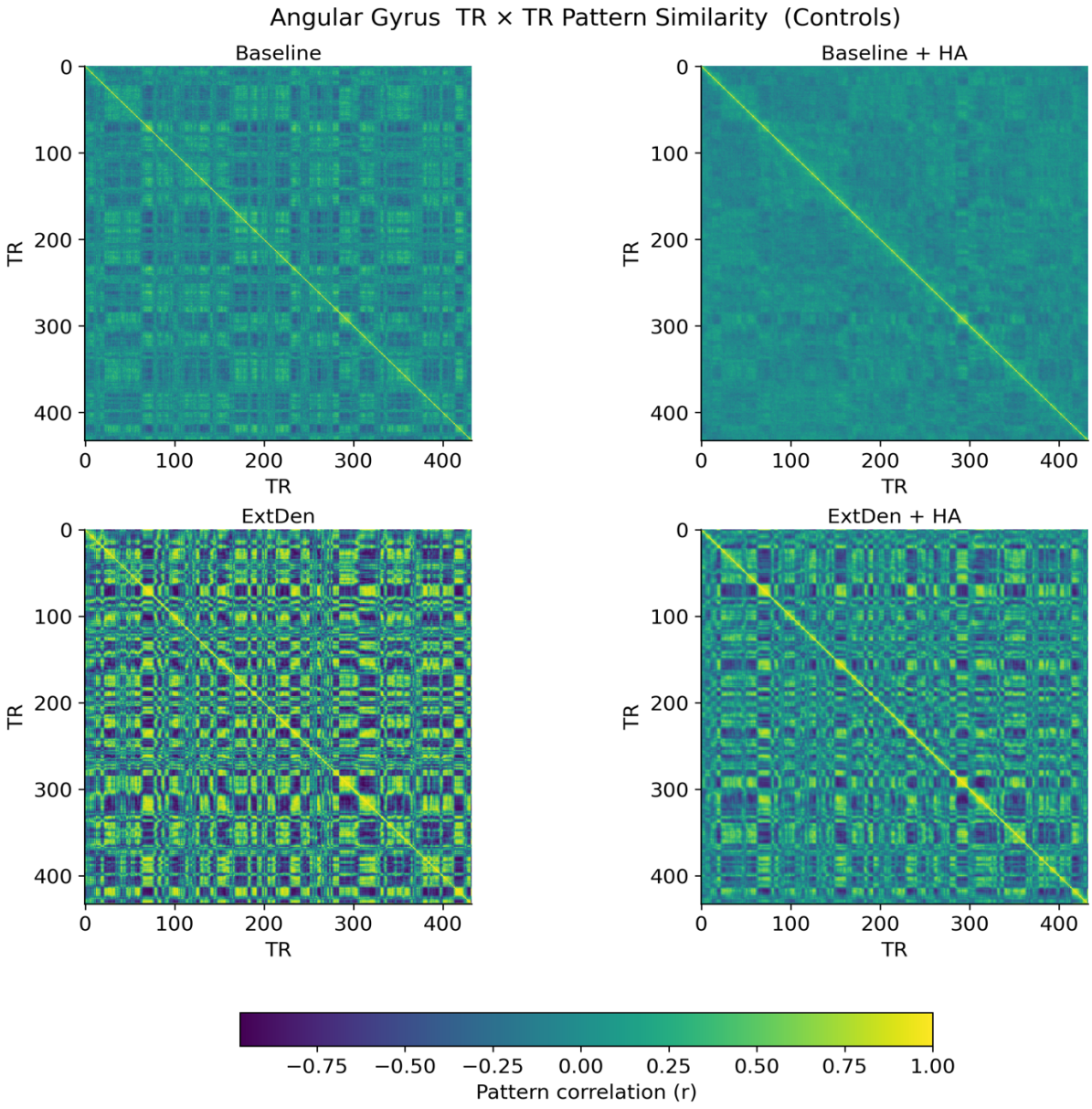
Inter-subject Correlation (ISC) at Each Preprocessing Stage, by Parcel and Group

ROI	Group	ROI-ISC (mean *r*)				Voxel-ISC (mean *r*)			
		Baseline	Baseline + HA	ExtDen	ExtDen + HA	Baseline	Baseline + HA	ExtDen	ExtDen + HA
AG	Control	.198	.392	.231	.274	.022	.343	.130	.232
	MCI	.111	.367	.097	.170	.010	.336	.057	.161
V1	Control	.205	.385	.233	.286	.029	.344	.163	.249
	MCI	.126	.354	.139	.196	.014	.340	.090	.181

Note. ROI-ISC is the Pearson *r* computed on the ROI-averaged time-series for every pair of participants and then averaged; it indexes synchrony in the coarse (parcel-level) trajectory that GSBS segments. Voxel-ISC is the mean across voxels of the pair-wise correlations for each voxel's time-series; it captures how well fine-grained functional topographies are brought into register. All values are averaged across participants and rounded to three decimals. Baseline = standard motion correction and high-pass filtering; Baseline + HA = Baseline plus hyperalignment; ExtDen = Baseline plus white matter/cerebrospinal fluid (WM/CSF) regression and 6 mm smoothing; ExtDen + HA = ExtDen plus hyperalignment.

Figure 11

TR × TR pattern-similarity matrices for the AG (Control group)



Note. Time-point-by-time-point correlation matrices (TR × TR) averaged across the Control group for four preprocessing pipelines: (top-left) Baseline, (top-right) Baseline + HA, (bottom-left) ExtDen, and (bottom-right) ExtDen + HA. Warmer colours indicate higher inter-subject similarity; tick marks on each axis show 100-TR intervals. ExtDen sharpens the block-diagonal structure that reflects putative event boundaries, whereas hyperalignment raises overall correlation but visibly blurs those boundaries.

7.2 N-matched Control split-half analysis

Rationale and decision logic

This analysis tests whether the original MCI template-preference result could be an artefact of unequal template precision (Control template estimated from more participants). Under a shared-map + template-estimation noise account, unequal *N* should favour the cleaner Control template; after matching ***N***, the MCI and Control templates become equally precise, unbiased estimates of the same boundary times, so the predicted MCI preference is ~0 (indifference). Any consistent own-template advantage at matched *N* indicates that the MCI template is systematically closer in time to MCI participants' transitions, i.e., a real timing difference rather than unequal precision.

Methods

All preprocessing, voxel selection, GSBS settings, and scoring followed §4.2. The split-half procedure changed only the way templates were estimated and compared:

1. **Control split-halves (*n* = 18/18 per split).** For each random partition of the Control cohort (*n* = 36) into two non-overlapping halves (Half-A, Half-B; *n* = 18 each), we estimated two full Control templates (one per half) at the MCI-constrained *k*. Within each half, we also computed LOO templates (*N*−1 averaging) to avoid circularity when scoring members of that half. As a sanity/precision check, Control participants were scored twice per split:
   ΔControl = Strength(own-half LOO) − Strength(other-half full).
2. **MCI preference at matched *N*.** For each MCI participant (*n* = 18) we computed strength on their MCI-LOO template (computed once per ROI from the MCI group and reused across splits) and on each Control-half full template from the current split. The Control reference was the mean of the two Control-half scores to stabilize the estimate at matched *N*:
   ΔMCI = Strength(MCI-LOO) − mean{Strength(Control Half-A full), Strength(Control Half-B full)}.
3. **Number of partitions.** We analyzed 200 random splits per ROI

Statistics

For each split, we formed within-subject differences (Controls: own-half LOO − other-half FULL across both halves, yielding ~36 paired differences; MCI: own − mean(Control halves), yielding 18 paired differences) and tested the mean difference against zero with a paired *t*-test. Because participants are reused across random partitions, split-level tests are not independent; we therefore summarized across splits using the mean Δ, its SD and 95% CI (normal approximation). Reporting and interpretation of results appear in Chapter 4.