Once and for all: post-encoding neural mechanisms promoting rapid episodic memory formation in humans

Xiongbo Wu

UNIVERSITAT DE BARCELONA

Once and for all: post-encoding neural mechanisms promoting rapid episodic memory formation in humans

Xiongbo Wu

Submitted for the degree of Doctor in Psychology
Doctoral Program in Brain, Cognition and Behaviour

Cognition and Brain Plasticity Unit
Department of Cognition, Development and Educational Psychology

University of Barcelona

PhD Supervisor: Dr. Lluís Fuentemilla Garriga

Barcelona, 2022

Acknowledgements

It is never easy to make a decision that you foresee its long-term impact on your life yet know little about in what way. In 2015, I made up my mind to come to Barcelona after graduating from university in search of a possible training trajectory to become a researcher. Looking back, I'm still impressed by the simple motivation of my interests in brain science, yet strong enough to push myself to make such a decision without fear or hesitation. After years of dedication, I'm finally pushing to the finish line of the first yet utmost milestone on this way. This page is for all of you who accompanied me all the way through. From the bottom of my heart, you have made a substantial impact on all aspects of my life, even more than you could imagine.

First and most importantly, I would like to thank my supervisor Lluís Fuentemilla, the man who made all this possible. Thank you for being such a wonderful mentor in my life. You have guided me into the world of science, trusted me when even myself had doubts, cheered for me when I made the tiniest progress and comforted me when I felt worried and stressed. Thank you for giving me so many strengths, strengths that will help me to carry on the fights in my future scientific career.

Together with Lluís, all my dearest people in the DMF research group, Alexis, Ludo, Joan, Marta…... I want to give you my heartful gratitude, for being such a fantastic team to work with, talk to, and laugh with together. I have always felt lucky to have you around me, present or virtual, during this whole journey. I will always recall these beautiful memories from time to time.

I want to thank David Cucurell and Joan Rodríguez for always being patient with me, guiding me through all kinds of technical problems or administrative issues, and always being ready to help.

And, of course, the rest of the Brainvitge unit. It is always a pity for me that I have not been able to spend more time with you as I would like. I will cherish the moments we shared together, and I'm sure I will see you in the near future.

Thank you, my dearest friends, who I’m so lucky to get to know since the Master program in 2017: Akke, Alex, Berta, Gemma, Guillem, Joan, Lídia, and Marta. We were so enthusiastic about life, science and fun. After all these years, I’m so glad to see our friendship grows steady and strong. It feels so warm to have you by my side.

Thank you, my beloved Otaku people, Eunyoung, David, Stefanie, Yosuke, Chiara and Stefanie. Thank you for caring so much about me, giving me your time and love when I need them the most, and constantly reminding me of the happiness and joy in life.

I also want to thank the Pascual family, thank you for having me as a part of the big family. Thank you for offering me such a welcoming and warm place to live, a place that I will always call home.

There are two particular people I want to express my gratitude, Min Jie and Zeyu. Thank you for bringing so much light into my life and showing me happiness in its simplest yet purest form. I’m so proud to see how we learn and grow together, let’s keep it up!

I want to give my deepest gratitude to my parents and grandma. You have supported me from the very beginning. Not a single second you have doubted me. There would not be any chance for me to come this far without your most profound and unconditional love. I miss you.

Finally, I want to dedicate this last paragraph to my beloved girlfriend, Yuelu He. Thank you, my best friend, for understanding me the most, for facing together with me all forms of challenges in life, for simply being you. No words can represent how grateful and fortunate I feel to have you by my side. So, cheers! For all the time we shared together and all the time yet to come.

Table of Contents

List of abbreviations

ANOVA	Analysis of Variance
CA	Cornu Ammonis
CLS	Complementary Learning System
D	Distance value (Study 2)
DG	Dentate gyrus
EEG	Electroencephalography
ERP	Event-related potential
EST	Event segmentation theory
F	F Statistic
FEF	Frontal eye field
H.M.	Henry Molaison
HMM	Hidden Markov Model
HPC	Hippocampus
Hz	Hertz
iEEG	Intracranial electroencephalography
fMRI	Functional Magnetic Resonance Imaging
M/EEG	Magneto and Encephalography
MEG	Magnetoencephalography
mPFC	Medial Prefrontal cortex
ms.	Milliseconds
MTL	Medial temporal lobe
MVPA	Multivariate pattern analysis
NC	Neocortex
NMDA	N-methyl-D-aspartate
LDA	Linear Discriminant Analysis
LMM	Linear Mixed Effects Model
LS	Posterior lateral sulcus

LTP	Long term potentiation
p	p value
PFC	Prefrontal cortex
PVT	Psychomotor vigilance task
r	Pearson’s rho
RSA	Representational Similarity Analysis
s.	Seconds
SD	Standard deviation
SEM	Standard error of the mean
STS	Superior temporal sulcus
SWRs	Sharp wave ripples
SWS	Slow-wave sleep
t	t statistic
TCM	Temporal Context Model
TPJ	Temporal-parietal junction
WM	Working memory
z	z score
µV	Microvolts

List of figures

Abstract

Deciphering the mechanisms of human declarative memory is one of the main goals of neuroscience. Theoretical models such as Standard Consolidation Theory (Alvarez and Squire, 1994) and the Multiple Trace Theory (Nadel et al., 2000) have led the memory research field for the last two decades. However, with recent advances in research, the highly dynamic nature of memory systems seems to be oversimplified by these classic theoretical models, especially in the context of answering how memories may be formed as the experience unfolds and why specific episodes after only one-time exposure can be remembered over a long period but others not.

It has been reported in recent memory studies the existence of an immediate post-stimulus activity in the hippocampus following the detection of an event boundary and its predictive role in subsequent memory performance (Ben-Yakov and Dudai, 2011). It is well accepted that encoding of cohesive episodes requires integration of information that is likely to engage offline processes, which in rodents are mainly supported by mechanisms that enable replay of the sequential reactivation of hippocampal place cells that represent previously experienced trajectories. According to the latest evidence from human studies, this brain activity time-locked to the offset of the event boundary might reflect a rapid memory reactivation of the just-encoded sequence episode as such memory reactivation showed high neural similarity with neural activity elicited during the previous episode encoding phase (Sols et al. 2017; Silva et al., 2019).

Here, aiming to advance the understanding of why, when, and how immediate post-encoding memory reactivation in humans may be associated with memory formation, we conducted three EEG studies to explore its defining neural features that support the rapid transformation from experience to memory representation once the end of the event is conceived, a mechanism that we believe might contain the key to understand the human memory systems.

In Study 1, we first explored the specific post-encoding neural signature that promotes the rapid - ‘one-shot’ - learning of an unfolding episodic event. We speculated that such a process could be achieved by reactivating the neural patterns elicited during initial encoding to stabilise the memory trace beyond the past learning experience. Using a sequential encoding paradigm, we identified the neural trace of memory reactivation of novel episodic picture sequences elicited particularly at the episode offset after the event sequences were completed. And importantly, the degree of the elicited neural reactivation was predictive of memory recollection of an encoded episode. Moreover, memory reactivation was only present at the offset of sequences that could be perceived as meaningful episodes.

We then took a further step to explore the representational nature of this episodic immediate offset-locked neural activity supporting the recollection of the just encoded experience. In Study 2, participants encoded trial-unique combinations of face-object-scene picture triplet sequences to be subsequently recalled in a test. We successfully replicated the elicitation of the post-encoding neural reactivation mechanism at the event offset period and its functional role in promoting the event episode memory formation. Then using a pattern classifier, we discovered that rapid offset-locked signals triggered reactivation of a detailed representation of the elements of a just-encoded experience to be bound as a cohesive episodic memory in the long term.

Finally, we explored whether and how the congruence of an upcoming element within an unfolding event would influence rapid memory reactivation and shape long-term memory. In Study 3, participants encoded episodic naturalistic sequences depicting everyday activities followed by a target object image. We found that the perceived incongruence triggered stronger neural reactivation of the preceding event sequences, though they were less accurate to be remembered. In two follow-up behavioural experiments, we further found that the increase in accuracy in recognition memory for the congruent object was also accompanied by a decreased detailed item representation but a stronger association between the object and the preceding context.

Our results from the three studies suggested that the post-encoding reactivation represents a crucial mechanism for promoting the rapid formation of unique and meaningful episodes; its elicitation triggers the reinstatement of a detailed representation of experience-specific episodic elements to be bound as a cohesive memory trace; and it relies on a flexible mechanism sensitive to the top-down expectancy driven by context congruence.

Altogether, these findings shed light on the neural mechanisms that support the rapid learning of novel episodic events that unfold over time and advance the understanding of immediate post-encoding memory reactivation in humans that serves to transform experiences into long-term memory representations.

Resumen

Descifrar los mecanismos de la memoria declarativa humana es uno de los principales objetivos de la neurociencia. Modelos teóricos como la Teoría de la Consolidación Estándar (Standard Consolidation Theory, Álvarez y Squire, 1994) y la Teoría de la Huella de Memoria Múltiple (Multiple Trace Theory, Nadel et al., 2000) han liderado el campo de investigación de la memoria durante las últimas dos décadas. Sin embargo, los avances recientes en la investigación sugieren que estos modelos teóricos clásicos parecen estar simplificando en exceso la naturaleza altamente dinámica de los sistemas de memoria. Especialmente al tratar de responder cómo se pueden formar los recuerdos a medida que se desarrolla la experiencia, así como por qué ciertos episodios pueden ser recordados durante un largo período después de una sola exposición, pero otros no.

En estudios de memoria recientes se ha detectado la existencia de una actividad post-estímulo en el hipocampo. Dicha actividad aparece inmediatamente tras la detección de un límite de evento jugando un papel predictivo en el rendimiento de la memoria posterior (Ben-Yakov y Dudai, 2011). Es comúnmente aceptado que la codificación de episodios cohesivos requiere la integración de información que probablemente involucre procesos *offline*. En roedores, estos procesos están apoyados principalmente por mecanismos que permiten la reproducción de la reactivación secuencial de las células de lugar del hipocampo, representando trayectorias previamente experimentadas. De acuerdo con la evidencia más reciente en estudios con humanos, esta actividad cerebral anclada en la ventana temporal que sigue al límite del evento podría reflejar una rápida reactivación de la memoria del episodio recién codificado. Este fenómeno se daría a cabo ya que dicha reactivación de la memoria muestra una gran similitud con la actividad neural provocada durante la fase de codificación del episodio anterior. (Sols et al. 2017; Silva et al., 2019).

Con el objetivo de avanzar el conocimiento de por qué, cuándo y cómo la reactivación de la memoria posterior a la codificación en humanos puede estar asociada con la formación de la memoria, llevamos a cabo tres estudios de EEG para

explorar sus características neurales definitorias que apoyan la rápida transformación de la experiencia a la representación de la memoria una vez concebido el final del evento, el mecanismo que creemos que podría contener la clave para comprender los sistemas de la memoria humana.

En el Estudio 1, primero exploramos la propiedad neural posterior a la codificación que promueve el aprendizaje rápido -"de una sola vez"- de un evento episódico en desarrollo. Especulamos que dicho proceso podría lograrse reactivando los patrones neurales provocados durante la codificación inicial para estabilizar la traza de la memoria más allá de la experiencia de aprendizaje. Usando un paradigma de codificación secuencial, identificamos el rastro neural de la reactivación de la memoria de nuevas secuencias de imágenes episódicas provocadas particularmente en la ventana de tiempo posterior a la finalización de las secuencias de eventos. Sin embargo, lo más importante fue que el grado de reactivación neural provocada predijo el posterior recuerdo de la memoria de un episodio codificado. Además, la reactivación de la memoria solo estuvo presente al final de aquellas secuencias que podrían percibirse como episodios con sentido.

Luego dimos un paso más para explorar la naturaleza representacional de esta actividad neural anclada al límite de evento que apoya el recuerdo de la experiencia recién codificada. En el Estudio 2, los participantes codificaron combinaciones únicas de secuencias de tríadas formadas por imágenes de rostros-objetos-escenas para recordarlas posteriormente en una prueba. Reproducimos con éxito la elicitación del mecanismo de reactivación neural posterior a la codificación seguido al límite del evento y su papel funcional en la promoción de la formación de la memoria del episodio. Luego, utilizando un clasificador de patrones, descubrimos que las señales rápidas ancladas al límite del evento desencadenaron la reactivación de una representación detallada de los elementos de una experiencia recién codificada para unirse como una memoria episódica cohesiva a largo plazo.

Finalmente, exploramos si y cómo la congruencia de un elemento próximo dentro de un evento en desarrollo influiría en la reactivación rápida de la memoria y formaría la memoria a largo plazo. En el Estudio 3, los participantes codificaron secuencias episódicas naturalísticas que representan actividades cotidianas, seguidas de una imagen de un objeto. Descubrimos que la incongruencia percibida desencadenó una reactivación neural más fuerte de las secuencias de eventos anteriores, aunque la probabilidad de ser recordadas era menor. En dos experimentos conductuales de seguimiento, encontramos además que el aumento en la precisión en la memoria de reconocimiento para el objeto congruente también estuvo acompañado por una representación menos detallada del elemento, pero una asociación más fuerte entre el objeto y el contexto anterior.

Nuestros resultados de los tres estudios sugirieron que la reactivación posterior a la codificación representa un mecanismo crucial para promover la formación rápida de episodios únicos y con significado; su elicitación desencadena el restablecimiento de una representación detallada de elementos episódicos específicos de la experiencia para unirlos como una traza de memoria cohesivo; y se basa en un mecanismo flexible y sensible a la expectativa de control arriba-abajo (top-down control) impulsada por la congruencia del contexto.

En conjunto, estos hallazgos ayudan a esclarecer los mecanismos neurales que apoyan el aprendizaje rápido de nuevos eventos episódicos que se desarrollan con el tiempo y contribuyen al avance del conocimiento de la reactivación inmediata de la memoria posterior a la codificación en humanos que sirve para transformar experiencias en representaciones de memoria a largo plazo.

Chapter 1. Introduction

1.1. Historical background of memory studies

Human memory is a broad term. Even limiting its definition to a specific cognitive process, memory still seems to be a somewhat subjective concept that refers to a particular ability of living entities. One can intuitively describe memory as the feeling of being immersed in the past events of one´s own, the vivid mental image of one´s dearest friends, or even the capacity to remember a certain number of new words in a short time when learning a new language. Despite the discrepancy in one´s own vision to the question ‘What is memory?’, there is a converging point on agreeing that all those subjective feelings that are termed as part of memory ability represent the core part of being a human. Throughout history, many great minds have been amazed by the flexible yet complex nature of memory and pondered the essence of memory as a subject of study. The way they framed the questions and the approaches they adopted gradually contributed to parsing the generic thinking about the term ‘Memory’ into more fine-grained multidisciplinary lines of research with distinct conceptual and methodological perspectives.

One of the earliest reflections on memory can be traced back to Aristotle’s book *On Memory and Reminiscence*. From a philosophical perspective, he defined memory as, different from perception or conception, a state or affection of the two conditioned by a lapse of time. He believed that the mind is like an unscribed tablet (*blank slate*) with memory analogous to the written marks on it. Much later in the 18th century, English philosopher David Hartley built on the concept of the mind as a ‘blank slate’ and further extended it to the idea of linking the phenomena of perceptions to their physical entity in the brain. In his proposal, perceptions are caused by the vibration of small particles in the nerves. These variations fainted and left similar traces in the brain, which generates the sensation of memory. The connection between physiological and psychological facts emerged. However, it was only until the end of the 19th century that Ebbinghaus first showed that it was possible to study memory experimentally (Ebbinghaus, 1885). His method of simplifying experimental situations was later advanced with more sophisticated control of factors and conditions, leading to a direct interpretation of the association between stimuli and response. Building on the advances in behavioural methods, with later development of Gestalt psychology and Neurology, memory studies, among other mental operations, had gradually shifted to be studied as an internalised operation which requires a deep understanding of its biological foundation.

Donald Hebb, possibly inspired by the neuron morphology discovered by Ramón y Cajal, proposed in 1949 that learning is supported biologically by long-term synaptic changes between neurons. Hebb´s model (Hebb, 1949), recited widely in nowadays´ textbooks as "cells that fire together wire together", for the first time bridged the psychology and biology and opened up a new perspective to bridge memory, as elusive as might be the term at the time, to a neural activity with tangible traces in the brain.

Theoretical breakthroughs in science are generally initiated from advances in methods development and new technology. In the field of memory studies, there were specific historical points where the emergence of new technologies overturned the course of the discipline and laid down the tool base to make the brain mechanisms ‘observable’. For example, in 1929, German psychiatrist Hans Berger first introduced a new technique for measuring electrical current in the human brain, which he named electroencephalogram recordings. After nearly 100 years, electroencephalography (EEG) remains one of the most effective imaging techniques widely used by neuroscientists. It is a non-invasive technique that directly measures the brain's activity that results from neuronal firing by recording the summed electrical activity of many neurons at the scalp. Its high temporal resolution makes it an optimal tool to infer underlying neural mechanisms under experimental control on a millisecond-by-millisecond basis.

The scientific field of cognitive neuroscience got its name in the late 1970s. Leveraged by the new theoretical and methodological approaches, it targeted directly to unveil the physiological foundations behind the complex cognitive mechanisms. As one of the central topics in the nascent field, memory studies started to focus on the neural substrates underlining distinct memory processes. The modern era of memory research commenced.

1.2. Episodic Memory

In 1953, a 27-year-old young man with the initials HM underwent a neural surgery in an attempt to alleviate the severe seizure he had suffered since childhood. After successful resection of the bilateral Medial Temporal Lobe (MTL), the seizures got controlled. However, there was an unexpected and devastating outcome of the operation. He became severely amnesic. During the rest of his life, he lost the ability to generate new memories. The case of patient HM, with the real name Henry Molaison, revealed after his death in 2008, revolutionised the science of memory with valuable insights that allowed countless upcoming studies to explore the distinct segments of memory processes and the neural circuits underlie them. Suzanne Corkin, the leading scientist investigating the case, wrote in her book in memories of Henry Molaison:

> Henry was famous, but did not know it. His striking condition had made him the subject of scientific research and public fascination. For decades, I received requests from the media to interview and videotape him. Each time I told him how special he was, he could momentarily grasp, but not retain, what I had said. […] He would have been proud to know how much his tragedy has benefited science and medicine. (Corkin, 2013. p xv-xvi).

HM´s deficits in memory are centred in *declarative memory*, the ability to state what they know and remember explicitly. However, he showed intact capacity in *non-*

declarative memory, the ability to acquire new motor skills or perform the learnt ones with no recollection of the learning experiences (Scoville & Milner, 1957; Cohen & Squire, 1980). Further, the specificity of HM´s impairment was prominent even within the range of declarative memory. He could somehow recall details from events experienced remotely in the past and retrieve remote semantic memory. The deficit was especially evident when facing the new learning experiences.

The concept of episodic memory was expounded later by Tulving in 1972. He emphasised the critical distinction between knowing and remembering. While the former refers to the semantic knowledge or facts about the world, the latter represents memory about personally experienced unique episodes (Tulving, 1972). His taxonomic distinction between episodic and semantic memory laid the conceptual foundation for subsequent memory studies on the two parallel yet dependent information processing systems. The concept carried over gradually in later studies to subordinate stages of processing, from how memories were acquired (encoding), organised and stored (consolidation) to how memories can be afterwards recollected (retrieval).

Hallmarks of episodic memory can be intuitively conceived from a subjective point of view. As vivid as they might feel, they are however not a literal record of one’s experiences (Conway, 2009). Conway summarised systematically nine properties of episodic memory:

1. Contain summary records of sensory-perceptual-conceptual-affective processing.
2. Retain patterns of activation/inhibition over long periods.
3. Often represented in the form of (visual) images.
4. They always have a perspective (field or observer).
5. Represent short time slices of experience.
6. They are represented on a temporal dimension roughly in order of occurrence.
7. They are subject to rapid forgetting.
8. They make autobiographical remembering specific.
9. They are recollectively experienced when accessed. (Conway, 2009)

As we can see, although these properties resonate easily as they can be intuitively deduced at a conceptual level, questions such as: how is episodic memory represented in the brain? Or, how can it be acquired, stored and retrieved? may not have a trivial answer. Exploring these questions requires a deep understanding of the underlying mechanisms, structurally and functionally, to which the methodological approaches are not always straightforward enough to build a direct link. Since the first study by Ebbinghaus (1885) with a simple form of stimuli such as lists of words or pictures, memory studies have been expanding the scope of focus with more flexible experimental designs to pinpoint the targeted process. In terms of episodic memory, studies have explored more sophisticated stimuli and test measures to restore ecological validity. And thanks to the rapid advances in neurophysiological and neuroimaging techniques, behavioural outcomes under new episodic memory measures can be attributed to their neural underpinnings. Together, they have helped deconstruct the complex episodic memory formation processes into finer-grained structural and functional components. However, accumulating empirical evidence gradually showed that the memory system is somewhat flexible and adaptive to the richness of highly dynamic spatial-temporal features contained in the events we experience daily. It came to be noticed that when exploring the human episodic memory, the aspect of ‘episodic’ might weigh just as much as that of ‘memory’. In other words, studies need to take into account that an event or an episode is not only comprised of a piling up of external stimuli, such that the formation of episodic memory can be treated as a simple accumulation of information processing at different stages. But instead, the properties of memory activities need to be further explored within the context of unfolding experience, under the constraints of its intrinsic dynamic nature. Understanding the mechanisms of these activities might be the only path to unveil the fundamental yet elusive question: how does an episode become an episodic memory, and what are the relevant mechanisms that determine whether or not an episode could be remembered in the long term.

1.3. Event

Let´s take a moment to mentally simulate a regular day for Mr. D. He wakes up in the early morning like usual. After a quick breakfast at home, he heads for the office by metro. His workplace is a half-hour travel distance from home, and he usually browses the phone to spend his time. His job consists of providing after-sales service for the company's products, which requires him to attend to the incoming calls most of the time. He has lunch with some colleagues in his group in the restaurant under the company building they used to go to, chat over work-related topics and novelty in each other´s life. He continues the work over the afternoon until all things on the agenda clear out. After leaving work, he goes to catch up with one of his old college friends in a bar. He walks back home after a cup of drink. As tedious as the day might be, there are still several moments that he can vividly remember when asked about by others or ponder on by himself. The memory of those moments might stick with him for a considerable amount of time or even for the rest of his life. However, when recalling these moments, he might be surprised by the fact that he appears to 'jump in' one scene by another with no apparent feeling of transition in the middle. He might remember walking on the road, sitting in the office, or drinking in the bar. Still, those scenes are represented in a discreet manner with a minimised sensation of continuity as they should have been since the experience of the physical world contains no break in time. As a matter of fact, even by describing it from a third-person perspective, the narrative itself turns out to be segmented, in a similar form of self-reported recall. As we can see, this partitioning sensation of activity consists of one of the most intriguing phenomena when thinking about memory: there exists a transition from the continuous day-to-day experience to the discrete memory pieces or segments represented in our brains.

1.3.1. An event and its context

How do we segment the unfolding experience into event segments, and when does this transition happen? To answer these questions, the first step is to determine what an event consists of. Following the formal definition of the term 'event' by Zacks and Tversky: "a segment of time at a given location that is conceived by an observer to have a beginning and an end" (Zacks and Tversky, 2001, p. 17), the event represents a specific chunk of time embedded in the continuous stream of life with a various temporal length ranging from seconds to hours and days or even longer. Events might be directed to a specific goal (e.g., writing an article; or playing football) or simply display natural occurrences (e.g., an earthquake) (Zack et al., 2007). However, under the context of memory studies, the events we focus on here will be centred on those that contain cognitive activities anchored in the first-person´s perspective.

A typical event contains various features that are usually associated with each other. The list could include an animated agent, often human, a place or series of sites where the event takes place and objects. The link connecting these features creates a common context for one specific event. However, even under restrictive laboratory settings, context remains a vague and elusive term in memory research, as it could refer to anything represented in the participant´s brain during the experiment (Manning et al., 2015). Smith and Vela (2001) reviewed studies of incidental environmental context-dependent memory in humans and found reliable context effects generated by various types of experimental settings, ranging from features of the stimuli such as the colour and the spatial information to the external environment or even thoughts about encoded items. Therefore, instead of defining context by the type of information, Howard and Kahana (2002) focused specifically on the time scale of information representation. The *Temporal Context Model* (TCM) was thus proposed to describe how temporal flow is embedded in the context and how it may alter the memory performance at the behavioural level.

TCM poses two different yet interconnected layers of computational elements. The *item layer* refers to the specific encoded stimuli, and the *context layer* represents the series of encountered stimuli over the temporal flow. A simplified example was described in Manning et al. (2015) in the case of learning a list of words (Figure 1.1). The encoding of the word ‘dog’ is represented both at the item layer and context layer. Fast activation of representation at the item layer becomes associated with the current state of the context layer. At the same time, the context layer gets updated with the new input item. However, the context layer carries information of all learnt items, and computationally speaking, it is formed by a running average of all previous states. This relatively slow update in the context layer thus creates an evolving context drift over time, continuously interacting with new inputs from the item layer.

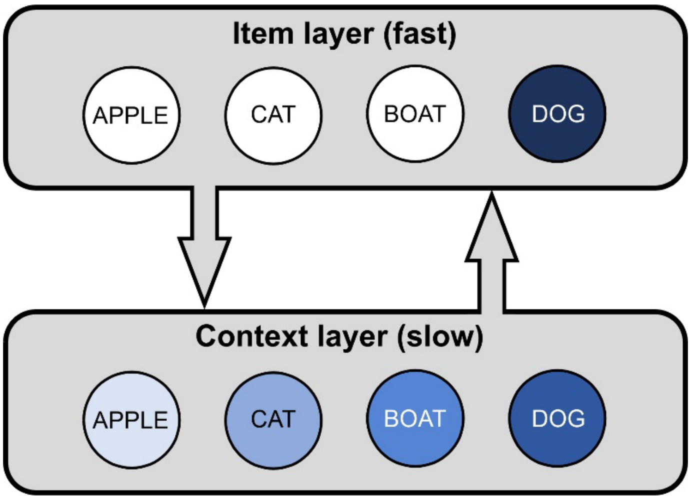

Figure 1.1. The Temporal Context Model (TCM). The Temporal Context Model (TCM). The item layer contains the sensory representations of the words to be encoded (marked as circles). When an item is studied (in the current example, the fourth word 'dog' in the list), the specific item representation is activated (coloured in deep blue) while the rest items are silent (coloured in white). The representations in the context layer reflect a running average of item-layer activation, with the darkness of the colour standing for the weight of each representation. The current context state contains activation representations of all studied words in the list with a graded increase in strength over proximity. The arrows indicate an interaction between the two layers. (Figure adapted from Manning et al., 2015)

Based on TCM, two well-known behavioural phenomena in memory research, namely the *recency effect* (Murdock, 1962) and the *contiguity effect* (Kahana, 1996), can be explained by the temporal position of encoded information at the context layer. The recency effect refers to cognitive bias in which those items, ideas, or arguments that came last are remembered more clearly than those that came at the beginning. TCM attributes the recency effect to the fact that the context at the end of item encoding assimilates more to the context of a test. Thus, the items associated with the end of the context can be more easily retrieved than those learnt at the beginning. The contiguity effect describes the stimuli initially encoded close in time become associated with each other. Later on, they are more likely to be activated together during recall. According to TCM, the contiguity effect rises when retrieving one item activates its contextual feature, probing the memory for more items associated with the same context.

1.3.2. Brain representing event context

An essential step in evaluating the validity of a theoretical model for a specific cognitive system consists in identifying the neural mechanisms underlying the proposed processes. Many recent studies, leveraged by neuroimaging techniques, have been devoted to tracing the neural signatures supporting how the event is perceived and segmented from the stream of stimuli, aiming ultimately at revealing the underlined neural processes of the formation of memory from continuous experiences. To approach these questions, many studies focused their first step on exploring how the contextual information is represented in the brain and what neural mechanism tracks its changes over time. Studies have especially pinpointed the functional involvement of the Prefrontal cortex (PFC) as well as several regions in MTL (Especially the hippocampus (HPC)) in such processes.

Prefrontal cortex

As reviewed by Polyn and Kahana (2008), several properties make PFC a suitable candidate for tacking the representation of contextual information: neural activity patterns were found stable in PFC (Postle, 2006); it has been shown to be functionally involved in the formation of context (Brave et al., 2001); also, it was shown to possess the ability to selectively update activity patterns by the task demands (Frank et al., 2001). Furthermore, it has been proposed a systematic organisation of representations across diverse areas in PFC for flexible information integration and generation of novel predictions (O’Reilly, 2010), thereby suggesting its functional role in providing a cognitive map of task space (Saez et al. 2015; Wilson et al. 2014). For example, Jenkins and Ranganath (2010) reported trial-to-trial changes in multivoxel activation patterns of rostrolateral PFC predicted successful encoding of temporal context, suggesting its direct involvement in the representation of time-varying contextual states in a manner similar to that proposed by TCM. Moreover, concordant with the drifting nature of encoding context, a recent study in Macaques (Cowley et al., 2020) showed that a natural drift in perceptual behaviour, such as fluctuations in pupil diameter and hit rate, is accompanied by a slow drift in the spiking rate of neurons in multiple brain areas, including V4 and PFC. The slow drift in neuronal activity did not modify the sensory stimulus encoding itself. Instead, it represented the internal states that contribute to the decision-making process.

On the other side, converging evidence was reported from lesion studies where damage in PFC disrupted episodic memory measured by tests of recognition and recall (Wheeler et al., 1995; Nolde et al., 1998), as well as the recollection of context in which memory was initially encoded (Duarte et al., 2005). Altogether, these shreds of evidence imply the functional involvement of PFC in context representation and tracing facing the encoding of the unfolding environment.

Hippocampus

Hippocampus has been widely proposed to be linked to context-dependent learning, but theories diverged in its specific functional implication in the process (Fuhs and Touretzky, 2007; Hasselmo and Eichenbaum, 2005; Hirsh, 1974; Howard et al., 2005; Jarrard, 1993; Nadel, 1995; Nadel and MacDonald, 1980; O’Keefe and Nadel, 1978; Rudy and O’Reilly, 1999; Staresina, and Davachi 2008). The common putative role of the hippocampus lies in its crucial function in binding elements of our experience with contextual information (Davachi, 2006; Diana et al., 2007; Eichenbaum et al., 2007; Ranganath, 2010). Early work by O’Keefe and Nadel (1978) first linked the neural activity in MTL to spatial context representation. In their model, the core function of the hippocampus is to encode and maintain the spatial maps of the environment. This view has been updated over the years. Nowadays, the hippocampus has mainly been proposed to have a prominent role in relational learning, including integrating stimulus elements with a corresponding context (Cohen and Eichenbaum, 1995; Eichenbaum, 2004). Furthermore, studies have shown that the hippocampus also can be sensitive to context drifting. For example, evidence from rodents demonstrated that neural patterns in the hippocampus could evolve gradually over time. In a study conducted by Manns and colleagues (2007), hippocampal activity drifted over time in correspondence with the odours across lists, even when there was no requirement to remember the odours from one list to the next. Other studies also showed that hippocampal ensembles gradually changed neural patterns during spatial navigation tasks, even in the absence of explicit spatial memory tasks (Hyman et al., 2012; Mankin et al., 2012). Such neural drift in the hippocampus thus suggests a neural representation for temporal information embedded in the context, enabling a contextual separation over a larger timescale.

1.4. From continuous experience to episodes

As discussed, TCM focuses on time-stamping the drift of an event context and linking the variance in behavioural memory performance to the temporal evolution in a context. However, what is the driving factors that lead the context to drift? According to TCM, context drift, instead of a random process, is caused by changes in features associated with encoded stimuli. Such changes can be induced by the external environment. A simple example could be walking into another room, the changes in the brightness, the colour of the wall paint, the layout of furniture etc., all together inform the perceptual systems that it is time to update novel information to the yet upholding context representation. However, the drift of a context can also be driven by the intrinsic flow of mental states under stable external features. One could sit in the office the whole day in a changeless environment while spending the morning focusing on work and the afternoon eager to go home. The gradual drift of internal state may lead to conceiving the daily experience separately as a productive morning mode and an irritable afternoon mode. As we can see here, context drift is a relatively common outcome driven by a variety form of causes. Instead of being limited to a passive or automatic process, it can be internally generated to adapt sensitively to one´s top-down goals (Dubrow et al., 2017).

1.4.1. Changes in context have an impact on memory

There is a common phenomenon that many people report to have encountered in their daily life: sometimes, right after we get into another room, generally at the moment of passing through the door, we suddenly forget the initial purpose of going to that room. It seems that the sudden change of environment somehow interrupts the maintenance of our goal-directed memory. And instead of a gradual shift in environmental features, this ‘passing through the door’ moment represents more of an abrupt change of context rather than a relatively slow drift. In 2006, Radvansky and Copeland (2006) first explored this phenomenon in laboratory settings. They investigated the ability of people to retrieve information about objects while moving them through rooms in a virtual space. The results showed that information about the associated object was less accessible when there was a spatial shift (i.e., change of rooms), which was indicated by a longer reaction time to the memory probe of the object and a higher error rate. The authors thus concluded the findings on the disruption of processing caused by spatial shifts. Indeed, if the memory system is highly susceptible to the context, as both theoretical models (Howard and Kahana, 2002) and empirical findings (O' Reilly and Rudy, 2001) suggested, it should not be surprising the fluctuation in contextual information may have an impact on the memory encoding of all associated items in that context. Such implications should be directly reflected at the behavioural level. Several upcoming studies further confirmed that the impaired memory caused by context drift could be replicated using various types of shifts in contextual features beyond spatial cues. A study by Swallow and colleagues (Swallow et al., 2009) used video clips to induce the contextual information for memory encoding of the containing objects. The transition from one video clip to another created a shift in context. They found that *event boundaries*, the moments when the situation is inferred to have changed, altered the accessibility of recently encountered objects, which can be reflected by a poorer ability to recognise its perceptual features. Ezzyat and Davachi replicated the effect with narratives containing context shifts marked by temporal event boundary cues in the form of text (Ezzyat and Davachi, 2011). They found that participants' long-term associative memory for information spanned across adjacent events was lower than their memory for information within the same event.

Accumulating evidence showed that the encoding of stimuli is sensitive to their context, and changes in the context can induce a direct disturbance in memory-related performance. However, changes in the context do not always happen passively and gradually, as proposed by TCM. As seen in the studies mentioned above, context can shift abruptly and may be sensitive to top-down goals. Also, changes in the context can be induced by the manipulation of various types of stimulus (e.g., text, audio-visual, spatial), depending on the particular task environment. Altogether, it suggests that contextual drift is an active process (Dubrow et al., 2017). And more importantly, diverging in its driving factors, contextual shift concords with the same ethological consequence: we conceive an end of the previous event at the same time the next one commences, with an emerged boundary in time cutting off the string of the continuous stream of mental state. So, under the continuously unfolding context, what actually leads our brain to decide the moment of 'closing' on the current context representation, wrapping up the past event as a whole, and start tracking the upcoming event?

1.4.2. Event boundaries and prediction error

As the experience unfolds, different neural states following the gradual drift in the context could lead identical stimuli to be perceived differently at different times (Polyn et al., 2009; Howard and Kahana, 2002). This process may further elicit the formation of event boundaries that the memory system may use to parse the continuous ongoing stream of experience into episodic segments (Polyn et al., 2009). Zacks and colleagues attributed the detection of event boundaries as transient errors in predictions, which refers to a mismatch between the inferred or expected upcoming situation and the actual outcome (Zacks et al., 2007, 2011; Reynolds et al., 2007). This prediction error may guide the process of segmenting continuous experience into meaningful chunks of events which construct the basis for memory formation. The *Event segmentation theory* (EST) (Figure 1.2) proposed a hierarchical information-processing system to account for how the brain deals with the transition from continuous experience encoding to discrete event representations. The core process of such transition relies on a 'gating mechanism' that tracks the stability of contextual consistency and guides the re-organisation of context representations when updates are necessary. Such updates can happen in either a bottom-up or top-down manner. When the influence of the pathway from sensory input and perceptual representations to the event models increases as a prediction error of the current environment appears, the gating mechanism responds to it by resetting the context representation and leading the event model to a new state. In a similar way, top-down processing, driven by the input from *event schemata*, which are described as semantic representations extracted from previous experience, can also guide the update of the event model.

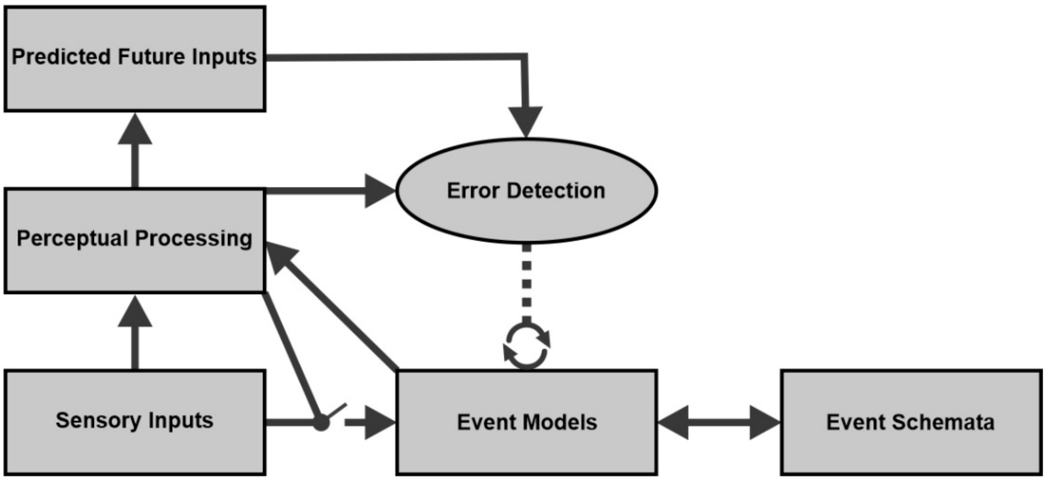

Figure 1.2. Schematic depiction of the event segmentation model. The arrows indicate the flow of information. The entrance of information from sensory inputs to event models is gated. Once there is a mismatch between predicted inputs and actual sensory input, a prediction error is generated. The detection of error lead event models to reset and get updated with new sensory inputs. Event schemata, the semantic representations extracted from experience, can also modulate event models by top-down control. (Figure adapted from Zacks et al., 2007).

Reynolds and colleagues (2007) brought up five features of the hypothesised relationship between prediction error and event perception:

1. Prediction error is greater during event boundaries relative to time points within an event.
2. Stable contextual information improves the prediction of event sequences.
3. A network can use gating signals occurring at event boundaries to learn and update internal context representations that reflect event knowledge.
4. A network can use gating signals based on prediction error to reliably update internal context representations that carry event information and facilitate subsequent prediction.
5. A network can use gating signals based on prediction error to learn internal context representations for events.

These features highlighted the bottom-up influence in event perception particularly, outlining how a mismatch in perceptual input can appropriately update event representations, and brought valuable insights for empirical studies. Also, event models in EST are considered a working memory representation, supported at the neural level by simple transit activation with no permanent changes in synaptic weights. At the systems level, the temporary property of EST assimilates the concept of *episodic buffer* proposed by Baddeley (2000). In doing so, the model emphasised a working mechanism that keeps track of the processing of information from various sources, a prerequisite stage for memory formation in the long term.

Neural trace supporting the process from event segmentation to event update was reported in a recent study, where prediction errors were shown to disrupt hippocampal representations and lead to an update of episodic memory representations (Sinclair et al., 2021). While participants watched familiar video clips, the authors inserted interruptions before the expected endings to generate mnemonic prediction errors. Data from fMRI showed that the hippocampus elicited stable activity patterns as video narratives unfolded, suggesting a neural tracking of the ongoing contextual information. However, prediction errors disrupted these sustained representations and, at the same time, promoted the memory strength, both immediately and in the long-term (i.e., after a 1-day delay), which was reflected by the number of recalled details in the video. The author thus concluded that prediction error could guide memory update and prompt the hippocampus to stop tracing the ongoing prediction.

Similar to EST, a recent theory attributes the detection of prediction error to being driven by latent cause inference (Gershman and Niv, 2010, 2012; Gershman et al., 2010). According to the theory, people constantly infer the state of the world (i.e., the latent or hidden cause) based on observation statistics. In an example of a student sitting in the classroom, the external stimuli (e.g., desks, blackboard, the teacher, etc.) are associated with the inferred latent cause (e.g., an ongoing course). Discrepancy rises when the current sensory input no longer fits with the inferred situation (e.g., some students come in with birthday cake and candles in hand). In turn, this can generate large prediction errors urging the system to infer the new latent cause (e.g., a surprise party). Consequently, tracing the previous experience stops, and a new state commences with a distinct underlying latent cause.

However, an event segmentation mechanism gated only by perceptually driven prediction errors is suboptimal since the perceptual changes in the environment are not always relevant or meaningful (O'Reilly, 2013). In line with this reasoning, a recent study showed that the boundary-induced memory effects appeared when the environment was stable, but such effects disappeared when changes frequently occurred (Siefke et al., 2019). These results indicated that the detection of prediction error alone might not be sufficient to drive event segmentation since its influence on the inference about the situation can be flexibly adjusted to the stationarity of the environment (Behrens et al., 2007, Nassar et al., 2010, 2012). In addition, the partition of continuous experience into smaller temporal units does not always happen due to external changes in the environment. For example, it has been shown that event representation can be generated and shaped following community structure (i.e., clusters of stimuli), where transition probabilities between stimuli (the basis of uncertainty and surprise) are uniform (Schapiro et al., 2013). It is possible, then, that event segmentation can be goal-directed or inference-based (Hard et al., 2011). This brings up the necessity to explore event segmentation as a dynamic mechanism, taking into account that boundaries can be conceived even when external changes are predictable (Shin and DuBrow, 2021).

1.4.3. Event segmentation in the brain

How does the neuroimaging data support the event segmentation model? The question can be further parsed into two subparts. The first consists of exploring ‘event’ representation: how the brain represents the hierarchical and structured properties of an unfolding event. And the second focuses on understanding the ‘segmentation’ mechanism.

Research using continuous streams of stimuli (movies, videoclips) has been recently adopted, given its advantages in preserving the temporal dynamics of the stimuli with abundant details and the experimental control (i.e., comparable across participants). A study conducted by Hasson and colleagues (2004) reported a shared (i.e., across-participants) large extent of the human cortex stereotypically responsive to naturalistic audio-visual stimuli, suggesting an intrinsic cortical specialisation sensitive to the processing of the context of a realistic environment. The fine-organised hierarchy of such cortical specialisation was further explored in a later fMRI study (Hasson et al., 2008). By examining the neural activity while participants viewed silent movie clips, the authors found that early visual regions (primary visual cortex and the motion-sensitive area MT+) responded consistently for short timescales information processing regardless of disruptions in temporal structure, whereas higher-association regions (e.g., superior temporal sulcus (STS), precuneus, posterior lateral sulcus (LS), the temporal-parietal junction (TPJ), and frontal eye field (FEF)) were associated to integrative processes that took long period.

To examine the neural underpinnings induced by the detection of event boundaries, Zacks et al. (2001) applied an fMRI scan to participants while viewing movies of everyday activities (e.g., making a bed, doing the dishes, etc.). Movies were watched twice. Participants viewed it passively the first time, and the second time, they were requested to press a button to indicate where they felt an event boundary occurred. An increase in neural activity at these event boundaries was found in the posterior cortex, including the occipital/temporal junction and Brodmann's areas. Similar increases at event boundaries in these regions were also present during passive viewing. At the same time, the magnitude of increase in these regions was greater for coarse-event boundaries (larger temporal units) than for fine-event boundaries (smaller temporal units). Altogether, the results suggested the existence of a specific neural response at event boundaries, sensitive also to the temporal hierarchy of the event structure.

Exploration of the two subparts of the question, namely how we first trace and then segment events, converged in a recent study by Baldassano and colleagues (2017). Using an fMRI data-driven approach, signals elicited during stimuli encoding were submitted to the Hidden Markov Model (HMM) to infer the current processing state. Such a state represents a limited time window where the neural pattern of the region remains relatively stable, suggesting the processing of the same piece of information. With changes in contextual information over time, pattern alternation in the fMRI signatures then leads the model to automatically identify event boundaries where the neural stability gets interrupted and updated. Based on the outputs of the HMM, the results first demonstrated that the topology of whole-brain neural activity for event processing showed a hierarchical property, with a shorter state in timescale at sensory regions and more extended events in high-order areas (including angular gyrus and posterior medial cortex), as well as a nested boundary structure from low to high processing layer (i.e., boundary a given layer was present at lower layer but not vice versa). In addition, these fMRI-inferred long event boundaries are closely related to human-annotated event boundaries. The results all suggested a brain-wide neural trace sensitive to parsing the ongoing experience into event segments.

Altogether, both theoretical models and empirical neuroimaging data exploring the event segmentation mechanism converged on proposing that the segmentation of ongoing experience is an efficient way to process the online information simultaneously across different hierarchical levels; at the same time, it can also help to reconstruct the perceived information for better long-term storage with the strengthened association. In line with this argumentation, further studies showed that items within and across event boundaries elicited distinct physiological and behavioural patterns: items within the events were strongly connected in memory (Speer and Zacks, 2005; Ezzyat and Davachi, 2011), preserved better the temporal order of episodic sequences (DuBrow and Davachi, 2013, 2014, 2016; Horner et al., 2016; Heusser et al., 2018), and also elicited similar neural patterns when compared to items across events (Chen et al. 2017).

Thus, the extensive physiological consequences observed at event boundaries do not only reflect the outcomes of the processing of ongoing experiences but also imply that the event segmentation mechanism has a functional role in organising memory for past experiences, contributing to the transformation from episode to memory. In other words, the event segmentation mechanism may not only be relevant for the way we perceive the world but also for how it is represented and persisted in the long term.

1.5. Transforming a perceived event into an episodic memory unit

As discussed in the previous section, human brain is actively engaged in tracing and segmenting continuous experience into meaningful units. Recent literature suggests this segmentation process also has a functional role in forming solid memory traces for segmented events for an extended period. To understand how this transformation is conducted, say, from a newly segmented episodic unit to a long-lasting memory representation, we should first explore the biological basis of memory storage in the brain and then the neural mechanisms engaged in strengthening the newly formed storage after the initial encoding experience.

1.5.1. Neurobiological models of episodic memory

The idea of linking memory with a physical entity in the brain originates from Semon in 1904, who introduced the term “engram” to describe the neural substrate supporting memory storage (Semon, 1923). This concept emphasises the possible existence of a group of neurons that encode and store specific experiences (i.e., episodic memory) (Tonegawa et al., 2015). Although it is commonly agreed that a memory trace implicates widely distributed brain circuits with no simple form of point-to-point unit-based storage, searching for an engram remains an exciting research topic.

Converging evidence from rodents and humans showed the critical role of the hippocampal formation in episodic memory (Scoville and Milner, 1957; Tulving, 1983; Nadel and Moscovitch, 1997; Aggleton and Brown, 1999; Lisman, 1999; Redish, 1999). At the anatomical level, hippocampal formation includes the hippocampus (CA fields), subiculum, dentate gyrus and the surrounding regions, including the entorhinal cortex, presubiculum, and parasubiculum (Andersen et al., 2006). One key structural feature of the hippocampus consists of its complex neuron circuits distinctive from regions of neocortical whose connections are largely reciprocal (Felleman and Van Essen, 1991). Certain parts of connections in the hippocampus are however unidirectional. For example, the perforant path, the primary hippocampal input source, consists of a one-way projection from the entorhinal cortex to the dentate gyrus (DG). Moreover, within the hippocampus, granule cells in the dentate gyrus project through mossy fibres to CA3 pyramidal cells, which the latter further projects to the CA1 regions via Shaffer collaterals. All these projections constrain a one-directional flow of processing with no backward projection. The neocortical input into the hippocampus gets back to the entorhinal cortex from direct and indirect projection (intermediated by the subiculum) from CA1 after the aforementioned intrinsic circuit. Such unique structural morphology in the hippocampus, known as the tri-synaptic pathway, highlights the hippocampal circuitry formed by DG (synapse 1) to CA3 (synapse 2) and then to CA1 (synapse 3) (Andersen et al., 1971), lays as the biological basis supports the functional architecture of hippocampus network responsible for learning and memory.

One of the most influential models of memory systems, the *Hippocampus indexing theory*, states that the functional role of the hippocampus is to store a map or index of cortical regions activated by the encoding experience (Teyler and DiScenna, 1986; Teyler and Rudy, 2007). Such index represents the stored spatiotemporally patterns of an experience in the hippocampus that indicates the location as well as the temporal sequencing of the cortical neuronal ensembles that encoded the initial experience (Figure 1.3a). Therefore, the formation of a memory trace relies, in its essence, on establishing such a link between certain ensembles of hippocampal neurons and a specific cortical activity pattern (Figure 1.3b). Once the memory trace is established, partial activation of the cortical regions is sufficient to trigger the activation of the corresponding hippocampal neuron ensembles (Figure 1.3c), which, in turn, activates ('completes') the rest of the cortical neurons that encoded the entire memory trace (Figure 1.3d). Reactivation of these cortical regions in the originally encoded temporal sequence thus resembles the initial encoding experience. Biologically, a memory index represents an elicited neocortical-to-hippocampal neural pathways that can be established via Long-Term Potentiation (LTP). It is prone to be strengthened by repetitive reactivations or to gradually decay over time, leading to remembering or forgetting at the behavioural level (Teyler and DiScenna, 1986).

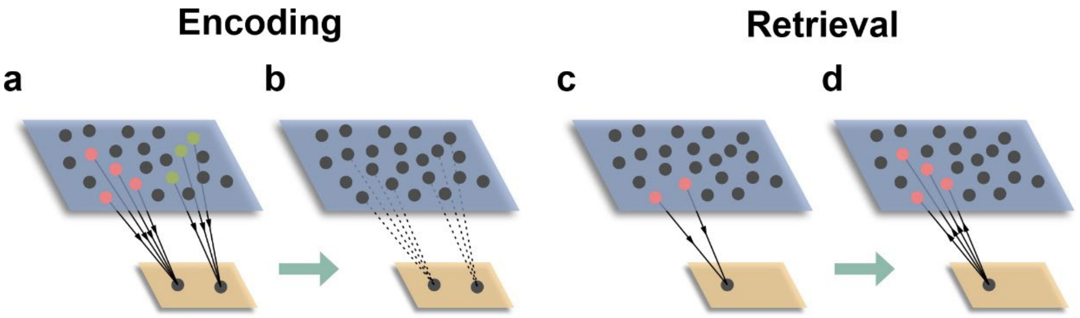

Figure 1.3. Schematic representation of Hippocampus indexing theory. The blue panels represent the cortical regions, and the yellow panels represent the hippocampus. **(a)**, Different encoding experiences can be represented at the cortical level by specific spatial-temporal neuronal activity patterns, and particular hippocampal neuron ensembles can register such patterns in the form of an index. **(b)**, Memory trace is represented by hippocampal neurons indexing the cortical neuronal activity pattern. **(c)**, Activation of partial neurons from the memory trace triggers its index in the hippocampus. **(d)**, The back projections from the hippocampus index to cortical regions enable retrieval of the whole memory representation. (Figure adapted from Teyler and Rudy, 2007)

The concept of 'indexing' emphasises the coordinating role of the hippocampus with no intervention or modification to the intracortical connections. This requires the hippocampus to be able to integrate cross-modal information from the neocortex (NC) and to rapidly accomplish the associative process. The physiological property of the hippocampus has been shown to meet up with such functional demands: CA3 is heavily innervated by collaterals of their own axons (Ishizuka et al., 1990; Li et al., 1994), so the activation of a small group of neurons can trigger population activities (Miles and Wong, 1983; Fujisawa et al., 2006). Such recurrent connections have been proposed to support an associative synaptic network generally linked to processes of completion and recall of stored information (McNaughton and Morris, 1987). Consequently, a partial cue of a particular episode leads to the activation of the index that further triggers the activity pattern in the rest of the engaged neocortical regions, completing the reconstruction of the retrieval of the whole episode; Also, NMDA receptors, which play a crucial role in LTP and memory (Morris, 2003), abundantly exist at associational synapses in CA3 regions (Jonas et al., 1993). They are primarily implied in paired-associate learning (Rajji et al., 2006), recall of associative memory (Nakazawa et al., 2002), and, most importantly, in the rapid acquisition and storage of information even from a novel one-time experience (Nakazawa et al., 2003).

Leveraging the distinctive physiological features in the hippocampus and NC as well as their reciprocal connections, the *Complementary Learning System* (CLS) (McClelland et al., 1995) proposed a complementary relationship between the two systems. According to CLS, the neocortical system relies on slow learning of distributed representations to infer statistical regularities of the environment, while the hippocampal system can rapidly encode the specifics of individual experiences to construct pattern-separated representations (McClelland et al., 1995; O'Reilly and Norman, 2002; O’Reilly et al., 2014; Kumaran et al., 2016). Memory-related behaviour relies on the joint contribution of the two systems with varying weight assigned to the extracted associative pairs (i.e., knowledge about the facts) and individual context-based experience (i.e., incidental episodic memories) (Kumaran et al., 2016). However, challenges to the dichotomic model arose recently from empirical studies. The original proposal of a slow-learning neocortical system emphasises its advantage in generating statistical regularities and is thus resistant to catastrophic interference. Recent studies have shown that it is not necessarily the case when newly encoded information is congruent with the existing knowledge structure (Tse et al., 2007, 2011; Brodt et al., 2018), suggesting a more flexible system memory consolidation mechanism than initially hypothesised. At the same time, the hippocampus has been proposed to support generalisation achieved by its recurrent activation (Eichenbaum, 2004; Zeithamova et al., 2012; Kumaran and McClelland, 2012), expanding its originally envisaged role to inferential reasoning across individual existing memories. Together, it highlights the necessity to revise the dynamic nature of the memory formation mechanism with flexible experimental paradigms.

1.5.2. Memory consolidation after learning

Like an old Chinese proverb says: “Awkward at first but skilful later on”, it’s only logical to think that a one-time exposure to the new information may not be enough to establish a stable memory trace. Newly formed memory must go through a particular process to be stored in a more consolidated format. The original concept of *consolidation* is generally credited to German experimental psychologists Georg Elias Müller and his student Alfons Pilzecker. They proposed that learning does not permit an instantaneous and long-lasting memory formation, which instead takes time to be fixed. As a result, memory remains vulnerable and prone to changes for a certain period after acquisition (Lechner et al., 1999). The claim is consistent with evidence from patients as well as healthy participants where such post-encoding vulnerability of memory can extend over the course of weeks or even years (Duncan 1949; Russel and Nathan 1946; Squire et al. 1975), such that premorbid memory loss centred majorly in recently acquired information while sparing the ones learnt remotely (Hodges, 1994; Squire and Alvarez, 1995), a phenomenon denoted as *temporally graded retrograde amnesia*. The term *consolidation* has been constantly refined over the years. Nowadays, it is generally described as the progressive post-encoding stabilisation of long-term memory (Dudai, 2004). It contains two types of processes with respective neurobiological underpinnings, namely *synaptic consolidation* and *system consolidation*. The former refers to a relatively rapid process that takes place within the first minutes to hours after encoding experience, which is underlined at the molecular and cellular level by changes at local nodes of neurons. The latter represents the reorganisation of memory representation over a long period during which the involving neural circuits at the encoding stage spread to new locations in the brain, establishing a new dependence network for that memory trace.

Inspired firstly by retrograde amnesia patients, accumulative empirical evidence concluded that the memory consolidation process relies on the critical role of MTL (especially the hippocamps) and NC. However, divergence appeared when it came to modelling the functional dynamics of the involved brain circuits to explain the changes in memory organisation in the brain over time. An influential model, the *Standard Consolidation Theory* (Zola-Morgan and Squire, 1990; Squire and Alvarez, 1995; Milner et al., 1998), posited HPC´s involvement in episodic memory formation is limited to the initial stage of learning and from beyond one week after, memory can be gradually ´transferred´ to NC. After the establishment of cortical-cortical connections, memory became hippocampus-independent for permanent storage (Dudai, 2004). Later empirical findings challenged the model by showing HPC´s ubiquitous involvement in remote autobiographical memory retrieval (Nadel and Moscovitch, 1997). The *Multiple Trace Theory* (Nadel and Moscovitch, 1997; Nadel et al., 2000) was then proposed to stress the dependency of episodic memory on HPC even in the long term. The model, which was built on hippocampal function offered by the *Cognitive Map Theory* (O'Keefe & Nadel, 1978) and the *Component Process Model* (Moscovitch, 1992), further described the retrieval of contextual details embedded in the episodic memory, though it might diminish over time, will always reply on hippocampal complex (Nadel et al., 2007). Both models posit that NC gradually integrates knowledge representation (e.g., over days and weeks) while HPC rapidly acquires new episodes online.

The aforementioned classic models of episodic memory formation stressed the existence of the post-learning consolidation process that permits the encoded experience to be transformed into a long-persistent memory representation. However, at that time, there was still a paucity of understanding when and how the transformation process was completed at the neural level (Frankland and Bontempi, 2005). Studies in the last decade have made a substantial contribution toward this direction.

1.5.3. Memory consolidation via neural reactivation

A crucial step for memory formation consists in understating how an experience-induced mental representation becomes a stable neural trace. A widely held view is that part of the transformation process is achieved by synaptic modifications to the neural circuits via recurrent reactivations (Dudai et al., 2015). Memory reactivation refers to the repeated activation of the original neural pattern beyond its direct experience. The primary role of neural reactivation consists in stabilising the newly formed synaptic strength to be resistant to future modification or interference. In the current section, we will focus on the offline memory consolidation process via neural reactivation at distinct post-learning temporal stages, from sleep to awake period at resting, and finally to the recently discovered immediate moment after learning.

1.5.3.1. Neural reactivation during sleep

The benefit of sleep for memory enhancement has been long recognised. One of the earliest studies by Jenkins and Dallenbach (1924) showed that recall performance of an encoded list of syllables was better after time that included asleep than awake. However, it was not until recently that neural reactivation was hypothesised as a crucial process underlying such behavioural outcomes (Diekelmann and Born, 2010; Dudai, 2012; Lewis and Durrant, 2011). It has been proposed that sleep contains the optimal time window where neural activation patterns induced during the daytime get revoked. Such reactivation can alter structural or chemical changes at synapses. Thus certain newly established synaptic connections become strengthened/re-organised while others are weakened/eliminated, leading to a subsequent memory boost or loss at the behavioural level. To confirm such a casual effect relationship, empirical studies have explored the neural patterns that get reactivated and wherein the brain such reactivation takes place.

In rodents, inspired by the cognitive map theory developed by O’Keefe and Nadel (1978), early studies explored the trace of neural reactivation in the hippocampus during slow-wave sleep (SWS) after performing simple spatial navigation tasks (e.g., maze exploration). Leveraging the possibility of direct invasive recordings of neuronal activity, neural reactivation of a memory trace has been detected in hippocampal cells in several manners, including measuring the changes in firing rates of individual cells (Pavlides and Winson, 1989), correlating firing patterns of multiple cells between encoding and sleep stage (Wilson and McNaughton, 1994), as well as evaluating the amount of explained variance in the correlation patterns of firing during sleep (Kudrimoti et al., 1999). In humans, studies bridging neural reactivation and sleep consolidation used data from MEG/EEG (Piantoni et al., 2015; Schönauer et al., 2017; Schreiner et al., 2021), fMRI (Bergmann et al., 2012; Sterpenich et al., 2021) and intracranial recordings with epileptic patients (Zhang et al., 2018). These studies focused on registering the neural activity of population neurons in various locations in the brain during the sleep stage. It was used to compare with stimulus-specific patterns elicited during the encoding stage. Neural reactivation can then be inferred based on pattern classification or similarity between neural patterns extracted from the two states. For example, Schreiner et al. (2021) showed that category-specific multivariate neural patterns of scalp EEG elicited during associative learning tasks were reactivated at the post-learning nap stage and that such reactivation was time-locked to the up-states of slow oscillations and sleep spindles, a neural signature of synaptic plasticity underlying the formation of memories at cortical networks (for a review, see Klinzing et al., 2019). The reactivation strength further predicted the level of subsequent memory performance, indicating a direct functional link between sleep neural reactivation and memory consolidation at the behavioural level.

Another important aspect of neural reactivation during sleep is that its trace has been identified, mostly in rodent studies, across distributed brain areas and not limited to the hippocampus. The regions range from cortical (Peyrache et al., 2009; Wilber et al., 2017) to several subcortical brain areas, including the entorhinal cortex (O'Neill et al., 2017), ventral striatum (Lansink et al., 2009) and the Ventral Tegmental Area (Valdés et al., 2015). Though hippocampal reactivation has been shown to occur in coordination with neuronal firing in these areas (Ji and Wilson, 2017; Lansink et al., 2009; Qin et al., 1997), it has also been reported the occurrence of neural reactivation independent of hippocampal neuronal activity, for example in entorhinal cortex (O'Neill et al., 2017).

1.5.3.2. Neural reactivation at wakefulness

The consolidation process may not be limited to sleep states. Using multiple single-unit recordings of hippocampal neuronal activity, Foster and Wilson (2006) found evidence for neural reactivation in rodents during awake periods immediately after the spatial navigation experience. Such reactivation entails a sequential increase in firing rates of hippocampal *place cells* that indexed the initial position encoding in space. Similarly, Diba and Buzsáki (2007) showed sequential reactivation of hippocampal cells in rodents at the beginning and end of a running lap session, respectively, in a forward and a reverse replay manner. The directionality of the sequential reactivation has been proposed to be behaviourally relevant. A reverse reactivation occurs typically at the end of a trajectory when rats reach the reward point, suggesting a mechanism to reinforce the link between newly acquired memory traces with positive feedback. Forward reactivation instead occurs preferentially before the start of spatial navigation, indicating a preparation mechanism for trajectory planning (Carr et al., 2011). Moreover, neural reactivation of past experiences has also been found to be elicited frequently during resting periods while awake (Karlsson and Frank, 2009), suggesting a link between just encoded information and consolidation (Figure 1.4). Also, it has been reported that neural reactivation occurs in a time-comprised format by a factor of 5 to 20 (O'Neill et al., 2010; Ji and Wilson, 2007). These time-compressed replays stand for the fact that while the firing sequence of the neurons encoding the initial experience is preserved (either in a forward or backward manner), the firing rate of these neurons increases and the pattern persists for a shorter duration, as found during sleep stage (Euston et al., 2007) as well as wakefulness (Foster and Wilson, 2006; Diba and Buzsáki, 2007).

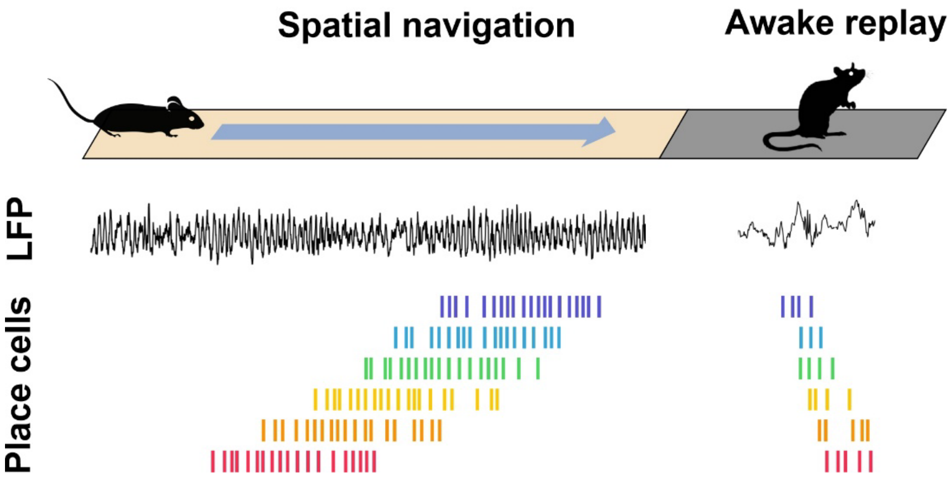

Figure 1.4. Awake replay in rodents. Local field potential (LTP) comprises neuronal activities of the place cells. During spatial navigation, place cells sequentially activate to encode the current position of the rodent. After the completion of the trajectory, a sequential neuronal pattern can be later reactivated reversely during the resting period while the rodent stays awake. Such reactivation happens in a temporally-compressed manner and is functionally linked to promoting the learnt experience (i.e., the memory of the trajectory) in the long term. (Figure adapted from Carr et al., 2011)

In humans, it has been found that the learning experience can induce an *offline activity* during the post-learning rest. Such activity is related to task performance in the form of a re-emergence of brain activity elicited during the learning phase (Peigneux et al., 2006). Other studies showed a post-learning increase in brain connectivity between task-relevant brain regions and the hippocampus (Tambini et al., 2010). These studies have been interpreted in line with the rodent findings, by which rapid memory consolidation processed may be induced right after a learning experience in humans too. More direct evidence came from a study by Staresina and colleagues (2013). Applying pattern similarity analysis on fMRI signal showed spontaneous neural reactivation during the delay period after the associative learning task in both the entorhinal and retrosplenial cortex. Such neural reactivation was shown to be event-specific, and the frequency of its occurrence during the delay period determined later successful recollection. The converging evidence suggests a plausible sign for memory consolidation during wakefulness after the initial learning experience. This can be functionally achieved by reactivating the neural pattern elicited during the learning phase.

1.5.3.3. Rapid post-encoding neural reactivation

As mentioned in Section 4, the brain actively segments the continuous experience into smaller chunks, guided by the detection of event boundaries and the event segmentation mechanism also contributes to how we organise the past experiences and transform them into episodic memory representations (Baldassano et al., 2017; Chen et al., 2017; DuBrow and Davachi, 2013, 2014, 2016; Horner et al., 2016; Heusser et al., 2018). A question yet to be explored is to delineate the exact neural mechanism engaged in promoting the memory strength in the long term once an event boundary is detected.

A direct link between immediate post-encoding neural activity and memory promotion was reported by Ben-Yakov and Dudai (2011). Using fMRI, they found increased neural activity in participants time-locked to the offset of short narrative audio-visual movie clips. The magnitude of the neural activity is predictive of the subsequent memory of the gist of the movie clips, and such memory-predictive activity engages several brain regions, most prominently the bilateral hippocampus and the bilateral caudate nucleus. Previous studies mainly had attributed the post-encoding activity, especially in MTL, to a carry-over effect from working memory (WM) maintenance (Ranganath and D’Esposito, 2001; Schon et al., 2004; Nichols et al., 2006; Olsen et al., 2009), it was the first time a direct link between offline neural processes with episodic encoding in humans was supported. The finding paved the way for upcoming human neuroimaging studies examining memory formation during a continuous stream of stimuli, such as naturalistic video clips, and results have gradually shown a memory-related post-encoding increase in a distributed network of brain regions (Ben-Yakov et al., 2013; Ben-Yakov and Henson, 2018; Baldassano et al., 2017). These neural activities time-locked the offset of learning experience have been proposed to be the signature of the early step or the initiation of the memory consolidation process (Dudai et al., 2015).

With the implementation of multivariate pattern analysis, a previous study conducted by our lab (Sols et al., 2017) found a rapid memory reactivation mechanism preferentially elicited during the context shifts (i.e., event boundary) of the event sequence encoding. Such memory reactivation is supported by a similar neural activation pattern of EEG signals resembling the encoding experience, with the degree of similarity predicting participants' ability to later remember associative information from within the prior event. This neural reactivation mechanism at the event boundary and its functional implication in memory formation was later replicated by another study using a naturalistic stimulus (Silva et al., 2019). In this study, the neural pattern at event transition during movie watching, operationalised as transition points in the encoding time whereby one episode ends, and a new one starts, was found to assimilate the neural pattern elicited during the encoding of the preceding but not the upcoming scene, and the strength of the indicated neural reactivation correlated with the recall memory of the scene in a subsequent test. These shreds of new evidence suggest that immediate offline reactivation plays a vital role in stabilising memory beyond initial learning processes. It can be elicited as a neural concomitant to the detection of boundary during the segmentation process of a long continuous stream of experience to promote the memory trace for the just-encoded episode in the long term.

Building on these empirical findings, a question emerges: why would an immediate reactivation mechanism following an event boundary be necessary for memory formation? More specifically, what is the function of such a phenomenon that needs to be revoked at the exact time point when an episode is completed? The possible underlying reasons could include the following: It can strengthen the memory trace for the previous episode after only one-shot exposure; it may indicate an early transformation process for long-term memory representation; also, it can be especially advantageous to avoid future interference to the upcoming unfolding experience, while the just ended episode can also be stored more cohesively and accurately. However, to date, there are scarce empirical studies in humans exploring immediate post-learning neural reactivation, so questions regarding *when* this neural mechanism occurs, *how* it can be triggered and *what* is its representational nature and functional implication in episodic memory formation remain elusive.

Chapter 2. Research aims

The main goal of the current thesis is to advance the understanding of why, when, and how immediate post-encoding memory reactivation in humans may be associated to memory formation. In what follows, we argument the objectives in more detail:

Objective 1: To examine the role of post-encoding neural reactivation for 'one-shot' episodic memory formation (Study 1)

Much work in rodents and in humans has provided evidence that offline reactivation plays an important role in stabilising memory beyond initial learning processes (Diba and Buzsáki, 2007; Karlsson and Frank, 2009; Carr et al., 2011). However, previous studies in humans only looked at reactivation after blocks of episodic learning (Staresina et al., 2013) or at transition boundary within continuous stimulus encoding (Baldassano et al., 2017; Silva et al., 2019), rather than after completion of a single episode. Thus, it remains unclear whether memory reactivation is crucial for the rapid - 'one-shot' - learning of an unfolding episodic event or merely a neural concomitant to event segmenting process. Here, we ask whether the reactivation of an episode may preferentially occur post-encoding but only when an individual perceives a meaningful event to be completed, i.e., once the unfolding experience can be fully integrated as a memory representation. We specifically explore "when" the brain binds the continuous experience input into a cohesive episodic memory trace and “how” the brain undergoes this rapid transformation to promote the memory trace for long term.

Objective 2: To explore the representational nature of post-encoding memory reactivation (Study 2)

As a continuation from Study 1, we take a further step to explore the representational nature of this offset-locked neural activity. Having shown that it serves to selectively transform experiences into long-term memory representations via reactivating the neural pattern elicited during encoding, we further explore which sort of information is reactivated to the immediate offset periods: a neural reinstatement/maintenance of trial-specific encoding experience or/and an integrated representation of the encoded information to overcome the representational gaps between disparate perceptual representations.

Leveraging the episodic encoding paradigm formed by trial-unique face-object-scene picture triplet sequences and multivariate pattern analysis, we attempt to address this issue by examining whether episodic immediate offset-locked neural signature relies on the reactivation of an accurate representation of the elements and/or the generic categorical representations of a just-encoded episode and, whether this is associated to recollection from long term.

Objective 3: To investigate the event representation update triggered by contextual incongruency (Study 3)

Event segmentation model proposed that when the input stimulus no longer fits the current event representation, the detection of this prediction error triggers an update process to reset the event models (Zacks et al., 2007, 2011). Empirical studies have highlighted the rapid neural reactivation as the neural concomitant to promote the preceding events when a mismatch expectancy is detected within an unfolding episode (Sols et al., 2017; Silva et al., 2019). However, it is not clear how the two distinct mechanisms driven by the prediction error, namely the disruption of the preceding neural state and the update process, are coordinated in time. In Study 3, we specifically explore whether and how congruence of an upcoming element within an unfolding event influences rapid memory reactivation and its impact in long term memory.

We tackle this question by manipulating the item encoding following episodic naturalistic sequences depicting everyday activities. We explore the extent to which the degree of congruence of the boundary item to the preceding episodic sequences may interact with the neural reactivation, and also, how the elicitation of this memory reactivation of the preceding episodic sequences may affect the encoding of the current object as well as its link with previous contextual information.

Chapter 3. General methods

In the current chapter, the primary analytical approaches will be described in brief. We included the methods shared across three studies, with specific adjustments to the particular experimental conditions. Please note that all methods used in each study will be further explained in detail in the corresponding chapter. The current chapter provides an overview of the general analytical approaches used in the present thesis to illustrate the underlying logic that bridges the experimental designs to our primary objectives of the studies.

3.1. EEG data analysis

In each study, we extracted the epochs for each encoding image. The duration of epochs varied as a function of the presentation length defined in each study´s paradigm (e.g., 2500 ms in Study 1). All epochs were baseline corrected to the pre-stimulus interval (-100 to 0 ms). Epochs for the offset period following each encoding sequence were extracted in each study. In Study 1 and 2, the offset epochs corresponded to the post-episode offset signals presentation after each series. In Study 3, the offset epochs were defined as the 500 ms of object presentation at the end of each encoding trial, and the 2000 ms of fixation cross followed immediately. In each study, offset epochs were baseline corrected to the time interval (-100 to 0 ms) before its onset. Epochs with maximum absolute amplitude over 100μV were discarded for further analysis. And for later analysis, in each study, all the epochs were smoothed by averaging data via a moving window of 100 ms (excluding the baseline period) and then down-sampled by a factor of 5.

3.2. Representational Similarity Analysis (RSA)

RSA was applied in all three studies. After smoothing and down-sampling the epochs of interest, RSA was performed at the individual level and included spatial features (i.e., scalp voltages from all the electrodes) (Silva et al., 2019). The degree of similarity between two included EEG spatial patterns was calculated and registered as the Pearson correlation coefficients, which are insensitive to the absolute amplitude and variance of the EEG response.

To evaluate the mechanism of post-encoding neural reactivation, which can be indicated by a re-emergence of image encoding neural pattern elicited again at the offset period. In each study, RSA was conducted at a single-trial level between the EEG signal of each encoding image (e.g., In Study 1, the image at the 1st, 2nd, 3rd, 4th, 5th and 6th position in a sequence) with EEG signal of the corresponding offset period. Point-to-point correlation values were then calculated, resulting in a 2D similarity matrix, where the x-axis represented the offset time points, and the y-axis represented the encoding time points (e.g., In Study 1, the size of the 2D similarity matrix was 250×400, representing 2500 ms of picture encoding x 4000 ms of post-episode offset). The output 2D matrix represents the overall degree of neural pattern similarity between EEG elicited for each pair of encoding images and its corresponding sequence offset.

The behaviourally relevant differences in the similarity value between conditions (e.g., high versus low memory condition in Study 1) were assessed with paired *t*-test (two-tailed) and corrected with cluster-level statistics (Section 3) and also evaluated through linear mixed-effect model regression analysis (Section 4)

3.3. Non-parametric Cluster-based Permutation Test

To account for RSA differences between conditions, we employed a nonparametric statistical method (Maris & Oostenveld, 2007), which identified clusters of significant points on the resulting 2D similarity matrix and corrected for multiple comparison based on cluster-level randomisation testing. Statistics were computed on values between conditions for each time point, and adjacent points in the 2D matrix that exceeded the preset significance threshold ($p < 0.05$, two-tailed) were selected and grouped as a cluster. Within each of the identified clusters, the cluster-level statistics took the sum of the statistics of all included time points. This procedure was then repeated 1000 times. For each time, labels across conditions were randomly shuffled. Cluster-level statistics with the highest absolute value for each permutation were registered to construct a distribution under the null hypothesis. Finally, the nonparametric statistical test was calculated by the proportion of permuted test statistics that exceeded the actual observed cluster-level statistics.

3.4. Linear Mixed-effect Model Regression analysis (LMM)

To explore the direct link between the neural measures (e.g., neural similarity value) and the behavioural measures (e.g., memory performance) on a trial basis, we implemented the linear mixed-effect model regression analysis on the pattern similarity between encoding and offset (Study 1, 2, and 3) as well as ERPs elicited during offset period (Study 3).

The formula for LMM varied across studies. However, the model was constructed following a similar structure and procedure. In the three studies, the neural measure value (e.g., similarity value or ERP amplitude) was included in the model as the independent variable. And following the temporal structure of the resulting neural value, the regression model was conducted independently on the value of each time point (e.g., similarity value on each time point of the 2D similarity matrix). In line with the specific paradigm of the study, behavioural measures were introduced into the model as the fixed effect factors. For example, in Study 1, the fixed effect factors included the number of items correctly recalled, the index of an item´s order in each sequence, and the coherence rating provided by the participant to each sequence at encoding. The subject was included in the model as the grouping variable in all three studies, with random intercept and a fixed slope for each fixed effect variable. Finally, the statistical significance for each fixed effect variable was Bonferroni corrected with a thresholded alpha level (i.e., 0.05 divided by the total numbers of regression analyses).

Chapter 4. Study 1

Post-encoding reactivation promotes one-shot learning of episodes in humans

Please Note: This study has been posted on BioRxiv.
bioRxiv. 2021.04.13.439658; doi: https://doi.org/10.1101/2021.04.13.439658

4.1. Summary

Prior animal and human work have shown that post-encoding reinstatement plays an important role in organising the temporal sequence of unfolding episodes in memory. Here, we investigated whether post-encoding reinstatement serves to promote the encoding of ‘one-shot’ episodic learning beyond the temporal structure in humans. In experiment 1, participants encoded sequences of pictures depicting unique and meaningful episodic-like events. We used representational similarity analysis on scalp electroencephalography recordings during encoding and found evidence of rapid picture elicited EEG patterns reinstatement at episodic offset (around 500 ms post-episode). Memory reinstatement was not observed between successive elements within an episode and the degree of memory reinstatement at episodic offset predicted later recall for that episode. In experiment 2, participants encoded a shuffled version of the picture sequences from experiment 1, rendering each episode meaningless to the participant but temporally structured as in experiment 1, and we found no evidence of memory reinstatement at episodic offset. These results suggest that post-encoding memory reinstatement is akin to the rapid formation of unique and meaningful episodes that unfold over time.

4.2. Introduction

In episodic encoding, an experienced event is rapidly transformed into a memory trace that has the potential to be consciously recollected at long-term (Tulving, 1983). Prior research has largely focused on examining how the brain contributes to successful encoding of individual trial information, such as single images (Paller and Wagner, 2002) or single item-context associations (Davachi, 2006). However, in natural settings, an episode is better characterized by a collection of successive elements that become contextually meaningful as they unfold over time. To be accessible for future retrieval, these elements have to be associatively linked into a bound memory trace. Discerning “when” the brain binds the continuous experience input into a cohesive episodic memory trace and “how” the brain undergoes this rapid transformation is essential to understand memory formation.

Human neuroimaging studies examining memory formation during a continuous stream of stimuli, such as naturalistic video clips, have shown that a distributed network of brain regions comprising the hippocampus and neocortex increased activity at the end of an event (Ben-Yakov et al., 2013; Ben-Yakov and Henson, 2018; Baldassano et al., 2017). This event offset brain signal has been shown to reflect a binding operation of the just encoded event elements into a specific spatio-temporal context (Ritchey and Cooper, 2020), which aligns well with the notion that the hippocampus is crucial for binding elements of our experience with contextual information (Davachi, 2006; Diana et al., 2007; Eichenbaum et al., 2007; Ranganath, 2010). In real life, episodic encoding relies on the possibility to form a coherent memory trace that integrates the temporally evolving sequence of elements into a meaningful context, so that if there is a shift in contextual information this is perceived as the end of one episode and the beginning of another (Zacks et al., 2007). These episodic boundaries are thought to support the segmentation of the continuous experience into discrete episodes (Zacks et al., 2007) and their detection has a direct impact on how events are organised into meaningful units in long-term memory (Kurby and Zacks, 2008; Radvansky, 2012; Ezzyat and Davachi, 2011; DuBrow and Davachi, 2013, 2014).

Work in rodents has provided evidence that memory replay at event offset plays a critical role in stabilising a temporal memory organisation beyond initial learning processes (Foster and Wilson, 2006; Diba and Buzsáki, 2007; Karlsson and Frank, 2008; Carr et al., 2011). In humans, rapid event offset memory reinstatement has been shown to be induced at the detection of context shifts during encoding of sequences of episodes and to predict their temporal order memory accuracy of the encoded sequential episodes in a later memory test (Sols et al., 2018; Silva et al., 2019). These findings suggest that the reactivation of an event contemporaneously with the experience of a subsequent adjacent event could theoretically result in the co-activation of the past and present events, promoting the binding of sequential events in their temporal order. In the real world, however, the recall of encoded episodes does not always depend on maintaining the exact order of the sequential representations, which can be fragile in many situations. Instead, it has been shown that when individuals are asked to recall episodes encoded in naturalistic conditions, they structured the recall along the causal (Brownstein & Read, 2007), semantics (van

Kesteren et al., 2013; Baldassano et al., 2018) and the relations between the elements embedded in the episodes (Lee and Chen, 2021). In fact, psychological models of event comprehension have emphasized that as the experience unfolds, memory is carved by how people construct a coherent model of a situation, which consists of agents and objects, semantic and spatiotemporal contexts, and the relations between them (Radvansky and Zacks, 2011). Notably, neuroimaging studies using video clips involving multiple consecutive episodes have started to find evidence that brain activity is naturally structured into events organised along these representational dimensions (e.g., Baldassano et al., 2017; Bird, 2020; Reagh and Ranganath, 2021; Lee and Chen, 2021; Heusser et al., 2021). Thus, if memory reactivation at episodic offset is an important neural signature of the rapid - 'one-shot' - learning of an unfolding realistic episode in humans, it is important to clarify whether it is concomitant to the encoding of episodes that included sequence of elements depicting coherent relations between them.

To address the issue, we designed a task that required participants to encode and later recall sequences of pictures depicting unique episodic-like events followed by a delay period with no stimulus. We used representational similarity analysis of scalp electroencephalography (EEG) recordings during encoding and found evidence for memory reactivation of the sequence elements of the episode after encoding, i.e., at the offset of the episode, and the degree of memory reinstatement at the offset predicted later memory recall for the specific episode. Memory reinstatement was not observed between successive elements within an episode, indicating that memory reactivation was specifically induced once participants perceived the unfolding episode to be completed. In a separate experiment, we also found that offset memory reinstatement was not present when participants encoded sequences of pictures that were not perceived as meaningful episodes. These results suggest that rapid memory reinstatement at episodic offset may be a neural signature engaged to integrate elements of the unfolding experience into a coherent memory for the just encoded episode.

4.3. Materials and Methods

4.3.1. Participants

Participants were native Spanish speakers who were recruited for pay (10€/h). All participants had normal or corrected-to-normal vision and reported no history of medical, neurological or psychiatric disorders. Twenty-five participants (17 females, age range 18-29) were recruited for Experiment 1. In addition, twenty-eight new participants (15 females, age range 19-35) were recruited for Experiment 2. The sample size is based on similar studies in the literature (Ben-Yakov et al., 2013; Sols et al., 2018; Silva et al., 2019). Informed consent was obtained from participants in accordance with procedures approved by the Ethics Committee of the University of Barcelona.

4.3.2. Experimental Procedure

Both Experiments (Figure 4.1) consisted of an encoding and a retrieval phase, separated by a 10-15 mins-break in the middle. Task timing and visual stimulus presentation were under the control of commercially available software e-Prime 2.0 (Psychology Software Tools).

The encoding phase of Experiment 1 included 100 different sequences, each consisting of 6 pictures that each was presented to participants only once. All pictures depicted emotionally neutral events and were controlled for saliency. All sequences of pictures described day life routine circumstances such as go shopping, reading a book at home or frying an egg in the kitchen. The pictures of each of the series were related to each other as in a story presented chronologically ordered. The order of pictures within each sequence was the same across participants but the order of the sequence presentation was counterbalanced across participants. Prior to the Encoding phase, participants were taught to attend to the picture series for a subsequent cue-recalled test yet not to engage in active rehearsal, especially when the blue fixation cross at the end of the series appeared. They were informed beforehand about the format of the subsequent test, entailing cued-recall of the images of each series. Participants were encouraged to memorise as many pictures as possible in a narrative form. The experiment only began when the examiner made sure that the participants fully understood the task. During the task, participants were encouraged to minimise eye movements and blinking. Each trial began with the presentation of text ‘New Episode’ for 2000 ms, which marked the start of a new sequence. This was followed by a fixation screen with a red asterisk lasting 2500 ms. Each picture was then presented sequentially on a white screen for 2500 ms and followed by a 1500 ms black fixation cross. Immediately after the presentation of the last picture in each series, a blue asterisk was presented on the screen, indicating a post-episode offset period of 4000 ms during which participants were previously instructed to avoid rehearsing the just-encoded picture sequence. The asterisk remained visible on the screen during the offset period. Immediately after the offset period, participants were asked to provide a subjective rating of coherence for the just encoded sequence. A rating scale ranged from 1 to 4, where 1 stood for ‘not coherent’ and 4 stood for ‘very coherent’. The next trial began after a fixed time interval of 2000 ms.

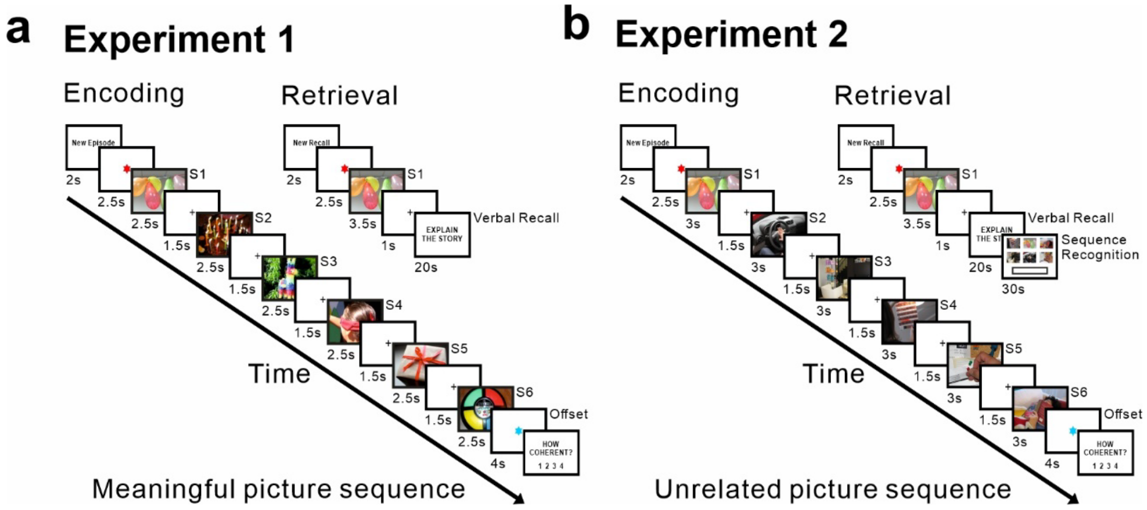

Figure 4.1. The experimental design. **(a)**, Task design in Experiment 1. During encoding, 100 different sequences were presented only once. Each sequence included 6 different pictures that unfold a life-like coherent narrative episode. Each picture was presented 2.5s, followed by a 1.5s fixation cross. After each sequence of images, there was an offset period (4s) during which participants were instructed to avoid rehearsing the just-encoded picture sequence. Participants were asked to provide a subjective rating of episodic coherence to the just encoded sequence at the end of each trial. Retrieval task was conducted 10-15 minutes after encoding. During retrieval, the first picture of each sequence was presented for 3.5 s which was followed fixation cross, and a message prompted at the screen instructing to report the associated episode during encoding. Participants were asked to verbally report within 20 seconds their memory associated episode or to indicate whether no memory came up associated to that picture cue. **(b)**, Task design in Experiment 2. Pictures were shuffled across sequences so that no meaningful story could be constructed after each sequence presentation. 60 shuffled sequence series were selected. The procedure was identical to Experiment 1 except two adjustments for task difficulty: 1. Time duration for the presentation of each

picture during encoding was increased to 3000 ms; 2. After each cued-picture recall task, participants were requested to perform a sequence order recognition task within 30 seconds, during which all 6 pictures from the same sequence series (including the cue picture) were presented on the screen in random positions and participants were asked to type the order of them as the original sequence presented during encoding phase.

For the retrieval phase, each trial started after the presentation of text ‘New Recall’ for 2000 ms on the screen. This was followed by a fixation screen with a red asterisk lasting 2500 ms. After the asterisk, the first picture of one sequence was presented to participant for 3500 ms on the screen serving as a cue to prompt the free verbal recall for the rest of pictures in that sequence. Participants were instructed to start the verbal recall once the text ‘Explain the story’ was presented on the screen following a 1000 ms fixation cross. The verbal recall was limited to 20 seconds and participants could stop the recall when finished by pressing the space bar. The order of the picture cues was randomised before their presentation at the retrieval phase.

In Experiment 2, we shuffled the pictures across sequences from Experiment 1 so that each sequence was formed by 6 pictures from different sequences. Thus, each sequence no longer depicted a meaningful episodic sequence. As in Experiment 1, the order of pictures within each sequence was kept the same across participants but the order of the sequence was randomised between participants. The general experimental settings for the encoding and retrieval phases and the instructions to participants were identical to Experiment 1. However, three adjustments were made: i) the presentation time of each picture during encoding was 3000 ms; ii) the total number of sequences presented to the participants was sixty; and, iii) we added an order recognition task after each cued-picture recall task. During the sequence order recognition task, all 6 pictures from the same sequence (including the cue picture) were presented on the screen in random positions and participants were asked to type the order in which they appeared in the encoding phase. Participants had 30 seconds to type the order of the pictures and they could skip to the next trial when finished by pressing the space bar. These changes were motivated by previous pilot studies with small sample of participants that ensured this number of sequences provided a balanced outcome between picture sequences whose order memory was relatively preserved or not.

4.3.3. EEG Recording

For both Experiments, EEG was recorded using a 31-channel system at a sampling rate of 500 Hz, using a BrainAmp amplifier and Ag/AgCl electrodes mounted in an electrocap (EasyCap) located at 28 standard positions (Fp1/2, Fz, F7/8, F3/4, Fc1/2, Fc5/6, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T7/9, P7/8, O1/2, Oz) and at the left and right mastoids. An electrode placed at the lateral outer canthus of the right eye served as an online reference. EEG was re-referenced offline to the linked mastoids. Vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. Electrode impedances were kept below 3 kΩ. A band-pass filter (0.1 Hz - 20 Hz) was implemented offline before the analysis.

4.3.4. Verbal recall analysis

During the retrieval phase of Experiment 1 & 2, participants were asked to verbally recall in the form of a “narrative” as many pictures as possible within each sequence corresponding to the cue. Free verbal recall of each trial was recorded through an audio recorder and the audio files were later analysed by a native Spanish speaker in the laboratory. Within each retrieval trial, a picture was considered as successfully retrieved when the precise details of the picture were described, or its core feature was mentioned during recall. Memory for each sequence was then quantified by the number of pictures (excluding the cue) correctly recalled.

4.3.5. Sequence order analysis

For Experiment 1, the order of the verbally recalled items in each trial was analysed. A trial was considered in-order when all recalled items followed the same sequential order as the encoding sequence. For Experiment 2, the temporal order memory for picture sequences recognition was compared to the true order of the sequence, and the result for each trial was coded as the maximum number of pictures (including the cue) correctly ordered in a row.

4.3.6. EEG data analysis

For each participant, we first used EEG data of the encoding phase to extract epochs for each item within the sequence, namely an EEG epoch for the 1st, 2nd, 3rd ,4th ,5th and 6th picture. Each epoch had a duration of 2500 ms and was baseline corrected to the pre-stimulus interval (-100 to 0 ms). We then extracted epochs for the post-episode offset signals after each sequence with duration of 4000 ms and baseline corrected to the time interval (-100 to 0 ms) before its onset. Finally, we repeated the procedure to extract epochs for the post-item offset of 1500 ms (with baseline corrected to -100 to 0 ms), that corresponded to the inter-stimulus interval separating each item presentation during episodic sequence encoding phase.

4.3.7. Time-resolved Representational Similarity Analysis (RSA)

For RSA, each EEG epoch data was smoothed by averaging data via a moving window of 100 ms throughout the EEG epoch (excluding the baseline period) and then down-sampled by a factor of 5. RSA was performed at individual level and included spatial features (i.e., scalp voltages from all the 28 electrodes) (Silva et al., 2019). The similarity analysis was calculated using Pearson correlation coefficients, which are insensitive to the absolute amplitude and variance of the EEG response.

For both Experiment 1 & 2, we conducted a trial-based RSA between the EEG signal elicited by each encoding item (1st, 2nd, 3rd ,4th ,5th and 6th) and the EEG signal elicited during immediate post-episode offset. After data smoothing and down-sampling, EEG epoch data for each item encoding contained 250 time points (given the 500 Hz EEG recording sampling rate) covering the 2500 ms of item picture presentation, and EEG data for each post-episode offset contained 400 time points, equivalent to 4000 ms. Point-to-point correlation values were then calculated, resulting in a single trial 2D similarity matrix with the size of 250×400, where the *x*-axis represented the offset time points and the *y*-axis represented the encoding time points. A trial-based RSA was computed between EEG patterns elicited by the encoding of 1st to 4th and 1st to 5th picture and the EEG patterns elicited at the immediate post-4th and post-5th picture ISI interval (i.e., 1500 ms), respectively. More concretely, we conducted RSA between EEG signal elicited at post-stimulus period after the 4th item and EEG patterns triggered during the encoding of each of the preceding 1st, 2nd, 3rd and 4th items. Then we repeated the same RSA but between EEG signal elicited at post-stimulus period after the 5th item and EEG patterns triggered during the encoding of each of the preceding 1st, 2nd, 3rd, 4th and 5th items. This resulted in 9 similarity matrices in total for each participant. The resulting similarity matrices were then averaged which resulted in a single 2D matrix with size of 250×150 (i.e., 2500 ms of picture encoding x 1500 ms of post-stimulus offset or ISI), depicting the overall degree of similarity between EEG patterns elicited during item encoding and post-item offset.

4.3.8. Nonparametric Cluster-Based Permutation Test

To account for RSA differences between conditions, we employed a nonparametric statistical method (Maris & Oostenveld, 2007), which identifies clusters of significant points on the resulting 2D similarity matrix and corrects for multiple comparison based on cluster-level randomisation testing to control the family-wise error rate. Statistics were computed on values between conditions for each time point, and based on adjacency in the 2D matrix, adjacent points that passed the significance threshold ($p < 0.05$, two-tailed) were selected and grouped together as a cluster. The cluster-level statistics were then calculated by summing up the statistics of all time points within in each identified cluster. The procedure was then repeated 1000 times with randomly shuffled labels across conditions to simulate the null hypothesis. For each permutation, the cluster-level statistics with highest absolute value was registered to construct a distribution of the cluster-level statistics under the null hypothesis. The nonparametric statistical test was calculated by the proportion of permuted test statistics that exceeded the true observed cluster-level statistics.

4.3.9. Bayes Factor statistical analysis

To further evaluate the power of the RSA effects that could be observed between conditions, we implemented the Bayes Factor statistical analysis (Kass & Raftery. 1995) in the point-to-point 2D similarity matrix. Bayes Factor was computed using Matlab Toolbox (Bart Krekelberg (2021). bayesFactor.

https://github.com/klabhub/bayesFactor) on RSA values between conditions for each point of the resulting 2D similarity matrix (250×400). A Bayer Factor greater than 10 indicates strong evidence for difference between conditions comparing to the null hypothesis.

4.3.10. Linear-mixed effect model

To investigate the relationship between EEG similarity values and behavioural memory on a trial basis we implemented a Linear Mixed Effect Model (LMM), which accounts for intra- and inter-individual variances. This analysis would also allow scrutinize the extent to which possible differences in EEG similarity results when comparing High and Low accuracy trials observed with a median-split approach described above were independent on the partitioning trial strategy implemented at subject level. Thus, we specified in our LMM the correlation values for one specific point on the resulting 2D similarity matrix as the dependent variable and included the following factors as fixed effect variables: the number of items correctly recalled (which ranged from 0 to 5, without counting the picture cue present at the recall task); the index of an item´s order in each sequence (1st, 2nd, 3rd ,4th ,5th and 6th), and the coherence rating provided by the participant to each sequence at encoding (which ranged from 1 to 4). Participant number was introduced into the model as the grouping variable, with random intercept and a fixed slope for each of the fixed-effect variables. To balance the requirement for computational power and signal to noise ratio, we further smoothed the resulting 2D similarity matrix for each item-offset pair by averaging over a moving window of 200 ms and then down-sampled by the factor of 5, both smoothing and down-sampling were conducted 2-dimensionally across the x and y axes. We applied the model fitting analysis independently for each position on the resulting 2D similarity matrix (50×80), then returned the 2D statistics map of the same size for each fixed-effect variable. Here to control for multi-comparison problem, the nonparametric cluster-based permutation test cannot be applied because each permutation represents a sample from the null distribution, which is not the case in LMM where it contains a covariance structure induced by multiple levels of relatedness among the individuals. Therefore, we implemented a Bonferroni correction to the statistical threshold to correct for the multiple comparison problem in the resulting statistical maps for each fixed-effect variable. Thus, the resulting statistical map was thresholded with an adjusted alpha level of α = 1.25×10-5 (0.05/4000).

4.4. Results

4.4.1. Experiment 1: Meaningful episodic sequence encoding

4.4.1.1. Free recall for meaningful episodes

In Experiment 1, participants were able to recall on average 2.32 items (SD = 0.456) in each series out of the total possible 5 items included in episodic sequence. Participants tended to recall encoded sequences in the form of “narrative” (e.g., “this is a party, there are balloons, the cake, and after the cake a piñata is broken and a gift comes out, but I don't remember what it was") and we counted the number of picture items included in their recall. The mean percentage of trials across participants that successfully recalled 0,1,2,3,4 and 5 items after the retrieval cue were respectively 24.89% (SD = 9.34%), 8.65% (SD = 3.41%), 14.14% (SD = 4.94%), 24% (SD = 5.68%), 18.91% (SD = 6.49%) and 9.41% (SD = 6.35%). A repeated-measures ANOVA revealed that participants’ memory recall differed as a function of number of items recalled following the retrieval cue ($F_{(5,120)} = 26.227$, $p < 0.001$) (Figure 4.2a). To increase the signal to noise ratio, for later RSA on EEG data, we first adopted a median split approach to separate the trials based on the corresponding task performance. Sequences with 2 or fewer items recalled during the retrieval phase were labeled as Low memory trials and sequences with equal or more than 3 items recalled were labeled as High memory trials. The threshold was selected by its relatively well-balanced separation for number of trials at subject level, as the average percentage of trials after the median split separation was respectively 47.64% (SD = 12.51%) for Low memory condition and 52.36% (SD = 12.51%) for High memory condition (Wilcoxon signed-rank test: $z = 0.821$, $p = 0.412$) (Figure 4.2b).

For each participant, we then calculated recall accuracy for each of the items in the sequence. The results showed a gradual decrease in accuracy for items as a function of their order position in the sequence during encoding: mean = 65.09% and SD = 10.44% for item 2, mean = 56.48% and SD = 10.92% for item 3, mean = 51.76% and SD = 10.43% for item 4, mean = 50.06% and SD = 11.57% for item 5 and mean = 42.91% and SD = 11.35% for item 6 (repeated-measures ANOVA: $F_{(4,96)} = 96.475$, $p < 0.001$) (Figure 4.2c). We also assessed how well item order was preserved during recall by counting, for each trial sequence, the number of items recalled in correct order as a function of the total number of recalled items for that sequence. We found that participants were accurate in recalling in order the items, independent of the total number of items included in their recall ($F_{(3,71)} = 1.464$, $p = 0.231$; mean = 97.17% and SD = 5.40% for 2 items, mean = 95.89% and SD = 5.51% for 3 items, mean = 98.79% and SD = 5.36% for 4 items, mean = 98.01% and SD = 3.26% for 5 items) (Figure 4.2d).

4.4.1.2. Subjective ratings of episodic coherence

The coherence rating provided a subjective measure of the degree of perceived narrative of each sequence. Due to technical issues, data for coherence ratings of 4 participants were not registered and they could not be included in the analysis. On average, for the remaining 21 participants, sequences were rated as 2.6 (SD = 0.41) (on a scale that ranged from 1 to 4), and the mean percentage of trials rated as 1, 2, 3, and 4 were respectively 15.34% (SD = 13.63%), 26.79% (SD = 11.67%), 40.36% (SD = 17.04%) and 17.51% (SD = 14.58%) ($F_{(3,60)} = 9.889$, $p < 0.001$) (Figure 4.2e). After median splitting the trials based on verbal recall performance, trials with High memory showed significantly higher coherence ratings (mean = 2.79, SD = 0.45) compared to trials with Low memory (mean = 2.40, SD = 0.45; paired Student *t*-test: $t_{(20)} = 6.166$, $p < 0.001$, two-tailed) (Figure 4.2f).

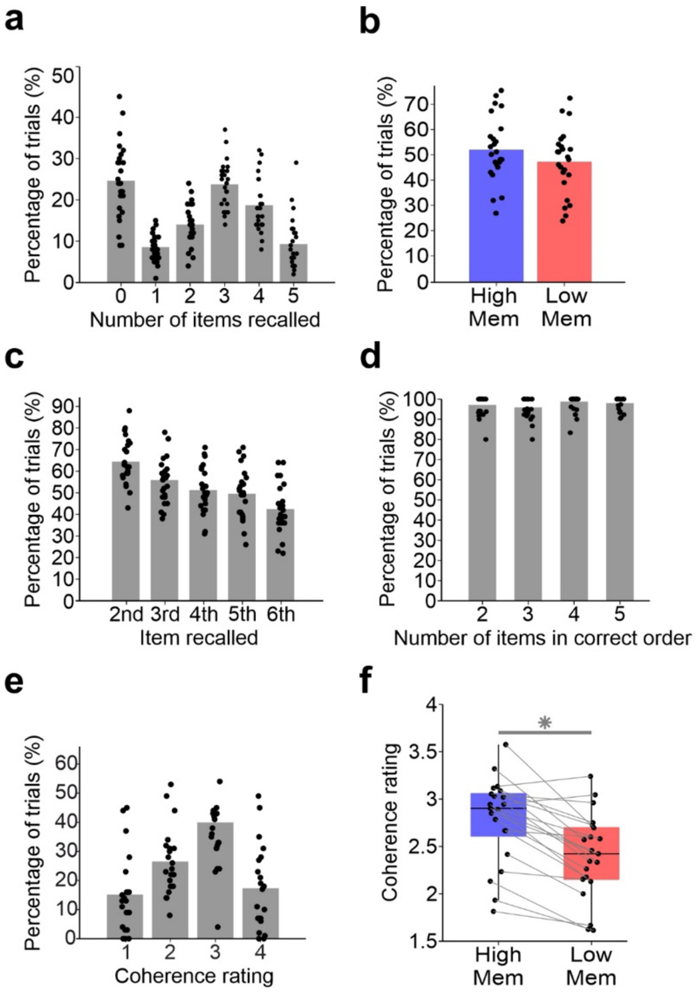

Figure 4.2. Behavioural results for Experiment 1. **(a)**, Percentage of trials at the cued recall task as a function of number of items recalled in each sequence. **(b)**, Percentage of trials included in the High (High Mem) and Low memory (Low Mem) condition in experiment 1 after implementing the median-split approach. Series with at least 3 item pictures correctly retrieved after cue were counted as High memory trials, otherwise the series were counted as Low memory trials. **(c)**, Percentage of trials as a function of the item across sequence order being recalled. **(d)**, Percentage of trials with recall in-order as a function of number of items recalled. A recall trial was considered as in-order when its items were recalled following the same sequential order as the true sequence. **(e)**, Percentage of trials as a function of the participants' degree of subjective coherence rating. **(f)**, Mean coherence rating score for trials included in the High and Low memory condition. separated by median-split verbal recall memory. In (a-c), bars represent the average across participants. Each black dot represents values for an individual participant. For all boxplots in (f), the central mark is the median, the edges of the box are the 25th and 75th percentiles. $*p< 0.05$.

4.4.1.3. RSA between item sequence and episodic offset at encoding

We first asked whether EEG patterns induced at the post-episode offset period correlated to EEG patterns elicited by the just encoded picture items within the episodic sequence, and if so, whether the magnitude of such correlation was associated to memory recall at the test. To address this issue, we implemented a trial-based RSA between EEG data elicited at picture item encoding (1st, 2nd, 3rd ,4th ,5th and 6th) with EEG data at the immediately following episode offset period. For each participant, the resulting trial-based RSA values were averaged separately according to two memory conditions: those associated with the participants ability to recall picture items with High memory (3 or more items) and those with Low memory (less than 3 items). To account for enough number of EEG trials to be included in both conditions and to ensure these trials included all items (1-6) and post-episode offset EEG signal cleaned of artifacts, we set a post-hoc criteria to exclude participants that did not reach a minimum of >15% number of trials in either condition, resulting in 15 participants for the current analysis (Average percentage of trials remained well-balanced between conditions with 35.04% (SD = 8.06%) of total trials included in analysis for Low memory condition and 32.75% (SD = 11.29%) of total trials for High memory condition; Wilcoxon signed-rank test: $z = 0.597$, $p = 0.551$).

In both High and Low memory conditions, the results of this analysis revealed an increase in similarity between EEG patterns induced ~400 ms – 800 ms at the post-episode offset period and EEG signal elicited ~400 ms – 1300 ms at picture item sequence encoding period (Figure 4.3a). However, the nonparametric cluster-based permutation analysis identified one statistically significant cluster where similarity values were higher in the High than in the Low memory trials ($p = 0.001$; mean t-value = 3.242, peak t-value = 5.473) (Figure 4.3b). We next evaluated whether the similarity between item encoding and post-episode offset period was driven by EEG patterns elicited by specific picture items within the just encoded sequence. We extracted the mean similarity values within the identified cluster for each item-offset pair and computed a repeated-measures ANOVA with two factors: trial condition (High vs Low memory) and encoding item (1st, 2nd, 3rd ,4th ,5th and 6th). The results of this analysis showed a significant main effects for both trial condition ($F_{[1,14]} = 15.407$, $p = 0.002$) and encoding item ($F_{[5,70]} = 2.677$, $p = 0.028$), but no significant interaction ($F_{[5,70]} = 0.315$, $p = 0.902$). (Figure 4.3c), indicating that episodic offset increase in similarity was not driven by EEG patterns elicited by specific items from the encoded item sequence. To further evaluate the power of the effect of the difference between High and Low memory conditions, we calculated the Bayes Factor on similarity values of each point on 2D similarity matrix. The results showed strong evidence (Bayes Factor greater than 10) for difference between conditions overlapping the area where significant higher similarity values were found on nonparametric cluster-based permutation analysis (Figure 4.3d).

Having shown that neural similarity increases of the just encoded sequence elements was elicited at episodic offset and that it was functionally associated to later memory recall, we then leveraged this to explore the relationship between the magnitude of memory reinstatement and the numbers of items to be recalled correctly later. At the same time, we asked whether the observed effects could be simply explained by the participants´ subjective feeling of coherence of the episode, as we found that subjective ratings of coherence were higher for High than for Low memory condition. To address these issues, we applied a LMM to our trial-based RSA data (see Methods). Two participants from the previous RSA analysis were not included here due to the missing data for coherence rating, resulting in 13 participants in total for the LMM analysis. Given that previous median-split analysis showed an increase similarity magnitude associated with higher recall performance, we specifically focused on this trend for LMM analysis. The result revealed one time interval that survived the statistical threshold for the fixed effect variable total number of items correctly recalled (one-tailed, mean *t*-value = 5.032, peak *t*-value = 5.621, $p < 1.25\times10^{-5}$, Bonferroni corrected) (Figure 4.3e). The time interval, which covered ~300 ms – 700 ms of post-episode offset and ~300 ms – 900 ms of item encoding, indicated the region where the degree of neural similarity of each item elicited during post-episode offset was significantly positively correlated with total number of items to be recalled in the corresponding sequence. However, no significant point exceeding the statistical threshold was identified on the statistical map accounting for the variable indexing order position in sequence nor for the variable indexing coherence rating.

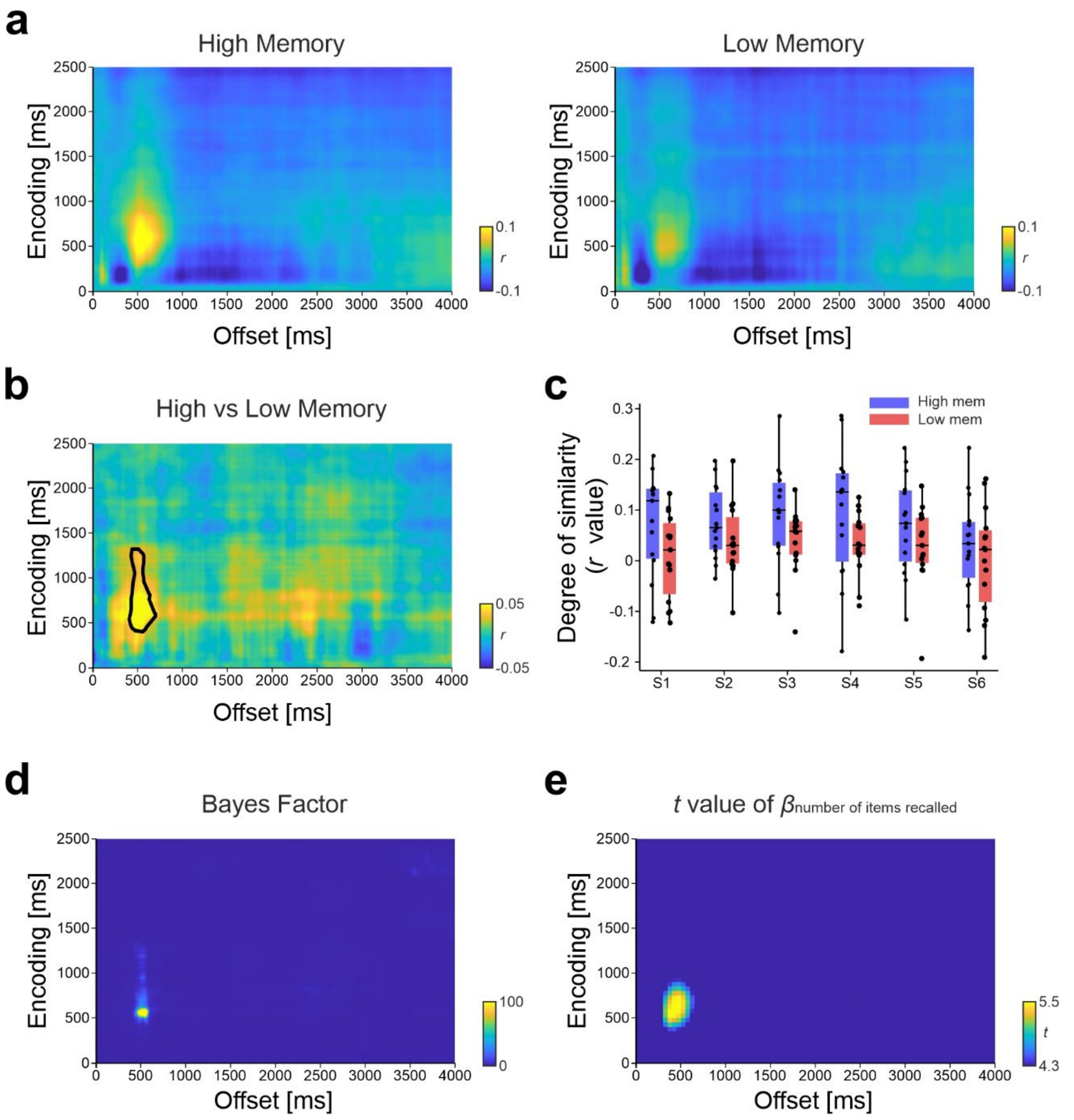

Figure 4.3. Neural similarity results at post-encoding period for Experiment 1. **(a)**, Time-resolved degree of neural similarity between item picture encoding and post-encoding offset for events with High or Low memory at test. **(b)**, Difference between similarity values for the two conditions. Statistically significant ($p < 0.05$, cluster-based permutation test) higher similarity value ~400 ms – 800 ms at the post-episode offset period with ~400 ms – 1300 ms at item picture encoding period was found for events with higher recall performance (indicated by a black thick line). **(c)**, In all boxplots the central mark is the median across participants, the edges of the box are the 25th and 75th percentiles. They depict the degree of similarity within the identified cluster for each item encoding with its corresponding offset period. Each black dot represents values for an individual participant. High memory sequence showed significantly greater similarity across encoding items. **(d)**, Bayes Factor of the difference between similarity values for the two conditions. A Bayes Factor greater than 10 indicates strong evidence for difference between High and Low memory condition. **(e)**, *t*-value map of the variable numbers of item recalled reveals the area that exceeded the significance threshold after Bonferroni correction with adjusted alpha level of $\alpha = 1.25\times10^{-5}$ (one-tailed). No area exceeding the significance threshold was found for *t*-value map of the variable serial position of item in sequence neither for that of the variable coherence rating.

4.4.1.4. RSA between item encoding and sequence item immediate offset

Next, we asked whether the increases in neural similarity between EEG patterns elicited at item sequence encoding and encoding offset were specific to post-episode period, or alternatively, whether they could be also found at offset periods immediately following picture encoding. To address this issue, we implemented the RSA between EEG pattern elicited by each item in the encoding sequence and the EEG signal pattern induced during the immediate post-item offset period. The current analysis was centered in the post-item offset period after the 4th and 5th item as this represented a delay period, as in episodic offset period, that is preceded by the encoding of multiple items from a sequence but differed in that the encoding of episode is not completed yet. At the same time, this research strategy allowed us to implement the same median split analysis used in our previous analysis (i.e., whether or not at least 3 items after the picture cue were correctly remembered), thereby enabling the comparison of the RSA results from the two conditions later on. 8 out of the total 25 participants were excluded for this analysis due to insufficient number of clean EEG trials in either condition (i.e., at least 15% of total number of trials in either encoding item of either condition) and the remaining participants maintained the balanced separation of trials between conditions with 41.06% (SD = 12.89%) of total trials included in analysis for Low memory condition and 37.44% (SD = 9.05%) of total trials for High memory condition (Wilcoxon signed-rank test: $z = 0.355$, $p = 0.722$). The result of this analysis showed no clear increases in neural similarity in either High or Low memory conditions (Figure 4.4a) and that no cluster of similarity values were accounted when the two conditions were compared with a cluster-based permutation test (Figure 4.4b). With the aim to further examine that the increase in neural similarity was specific to the episodic offset period, we directly compared the neural similarity findings at the first 1500 ms of the episodic offset vs at the 1500 ms of the 4th and 5th post-item offset. The results of this analysis confirmed a cluster of significantly higher neural similarity at the episodic offset condition (Figure 4.4c), thereby corroborating the notion that the increase in neural similarity was specific to post-episodic encoding delay period.

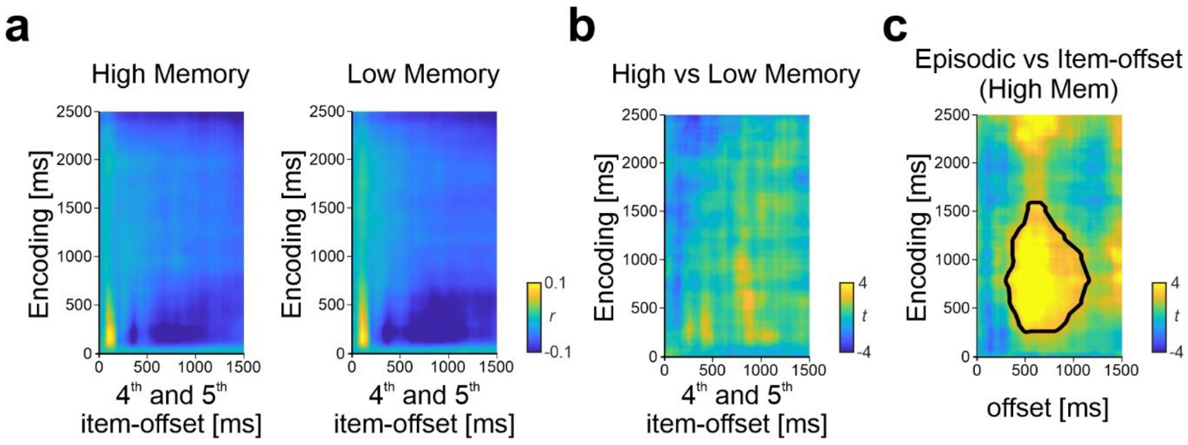

Figure 4.4. (a), Time-resolved degree of similarity between item picture encoding and 4th and 5th post-item following the events that were later recalled with High or Low memory. **(b)**, Difference (expressed in *t* values, uncorrected) between similarity values for the 4th and 5th post-item High and Low memory conditions. No cluster indicated significantly different similarity values between conditions (two-tailed, $p < 0.05$, cluster-based permutation test). **(c)**, Difference (expressed in *t* values, uncorrected) between similarity values for episodic offset and the 4th and 5th post-item offset in the High memory condition. Higher neural similarity was found during the episodic offset compared to item-offset. The significant cluster is indicated by a black thick line. $p < 0.05$, corrected with a cluster-based permutation test.

4.4.2. Experiment 2: Non-meaningful episodic sequence encoding

4.4.2.1. Behavioural results

In general, participants were able to recall on average 0.14 items (SD = 0.176) out of the possible five (picture cue was not included in the counting) in each series. The mean percentage of trials across participants to successfully recall 0,1,2,3,4 and 5 items after the cue were respectively 89.23% (SD = 11.92%), 7.98% (SD = 8.27%), 1.96% (SD = 3.14%), 0.77% (SD = 1.84%), 0.06% (SD = 0.32%) and 0% (SD = 0%) ($F_{[5,135]} = 795.913$, $p < 0.001$) (Figure 4.5a). Even though participants were unable to verbally recall almost any item from an encoded sequence, they showed above chance performance in the order item sequence recognition task.

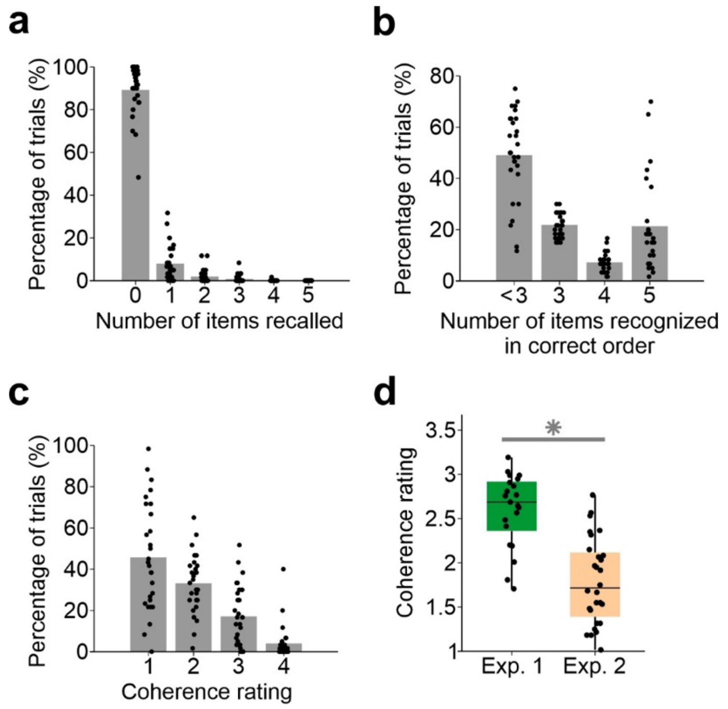

Figure 4.5. Behavioural results for Experiment 2 **(a)**, Percentage of trials separated by number of pictures correctly recalled after the picture. **(b)**, A temporal order sequence recognition memory test was added after the cued recall task. In the recognition memory task, all 6 pictures from the sequence were presented in random positions at the screen, including the cue, and participants had 30 s to order the correct temporal structure of the encoded sequences. The figure showed percentage of trials separated by the sequence recognition score. The score is quantified as the maximum number of pictures correctly ordered consecutively in one trial. A trial with less than 2 correct pictures consecutively ordered following either the cue, the 2nd, the 3rd or the 4th picture (score less than 3) was counted as no sequence recognised. **(c)**, Percentage of trials separated by degree of coherence rating. **(d)**, Mean coherence rating score of sequences in Experiment 1 vs Experiment 2. In (a-c), bars represent the average across participants. Each black dot represents values for an individual participant. For all boxplots in (d), the central mark is the median, the edges of the box are the 25th and 75th percentiles. *$p< 0.05$.

We analysed these data by quantifying the maximum number of pictures correctly ordered consecutively for each given trial sequence. A trial was considered to have been recognised if the participant reported at least 3 items in correct order from the sequence. On average, the percentage of trials to have scored less than 3, equal to 3, 4 and 5 (i.e., all pictures in order following the cue) were respectively 49.11% (SD = 17.31%), 21.85% (SD = 4.72%), 7.26% (SD = 3.55%) and 21.79% (SD = 17.62%) ($F_{(3,81)}$ = 39.777, p < 0.001) (Figure 4.5b). The participants provided correct order recognition for 50.89% (SD = 17.31%) of the trials, which is statistically significantly above chance (chance level = 15%, $t_{(27)}$ = 10.975; p < 0.001, two-tailed). Finally, the average coherence rating for sequences in Experiment 2 was 1.79 (SD = 0.48), and the mean percentage of trials rated as 1, 2, 3, and 4 were respectively 45.71% (SD = 25.96%), 33.15% (SD = 14.20%), 17.14% (SD = 14.68%) and 3.99% (SD = 8.31%)

($F_{(3,81)} = 24.108$, $p < 0.001$) (Figure 4.5c). In general, sequences in Experiment 1 were rated significantly higher than sequences in Experiment 2 ($t_{(47)} = 6.164$, $p < 0.001$, two-tailed) (Figure 4.5d), which suggested that the subjective feeling of coherence matched the general manipulation of the experiment.

4.4.2.2. RSA between item sequence and episodic offset at encoding

As in experiment 1, we computed a trial-based RSA between 2500 ms EEG patterns elicited by each sequence picture item and the corresponding EEG data induced at the 4000 ms episodic offset. The parameters for data smoothing and down sampling was kept the same as Experiment 1. For each participant, the resulting 2D similarity matrix was first averaged within each item-offset pair and then across pairs. 10 out of the total 28 participants were excluded for this analysis due to insufficient number of clean trials for all items (at least 15% of total number of trials in each item-offset pair). Thus, a total of 18 participants were included in the analysis. However, and contrary to as in experiment 1, the result of RSA for Experiment 2 did not show any observable neural similarity increase at the offset period (Figure 4.6a). To assess the extent to which neural similarity patterns seen in experiment 1 differed from those obtained in experiment 2, we separately compared the neural similarity values for High and for Low memory conditions in Experiment 1 with those obtained in Experiment 2. The results of these analyses revealed that neural similarity increase found at early offset period in experiment 1, for both High and Low memory conditions, was statistically different from similarity values at the offset period in experiment 2 (Figure 4.6b). More concretely, this analysis returned one significant cluster ($p = 0.022$ (corrected), mean t-value = 3.354, peak t-value = 5.059), that comprised higher neural similarity values for EEG data within ~400 - 1000 ms time range from post-episode offset period and ~200-1400 ms time range from item picture encoding from High memory trials in experiment 1 over trials from experiment 2. A similar cluster in its timing (~450 - 1000 ms of post-episode offset and ~200 - 1100 ms of item encoding) with significantly higher neural similarity values was found when comparing Low memory trials from experiment 1 with trials from experiment 2 ($p = 0.026$ (corrected), mean t-value = 2.928, peak t-value = 4.237).

Finally, we assessed whether the participants' ability to preserve a temporal order memory, irrespective of their lack of episodic recall could still be associated to increases in neural similarity at the picture sequence offset period. For each individual, we split the encoding trials as a function of whether they were accurate in the temporal order recognition test (defined by trials with 3 items in correct order) and those trials where the participant showed poor temporal order accuracy (i.e., < 3 elements in correct order). In this analysis, 4 more participants were excluded due to insufficient number of clean EEG trials in either condition. The results of this analysis revealed no significant differences at cluster level (*p* > 0.05 two-tailed; permutation test) between the two conditions.

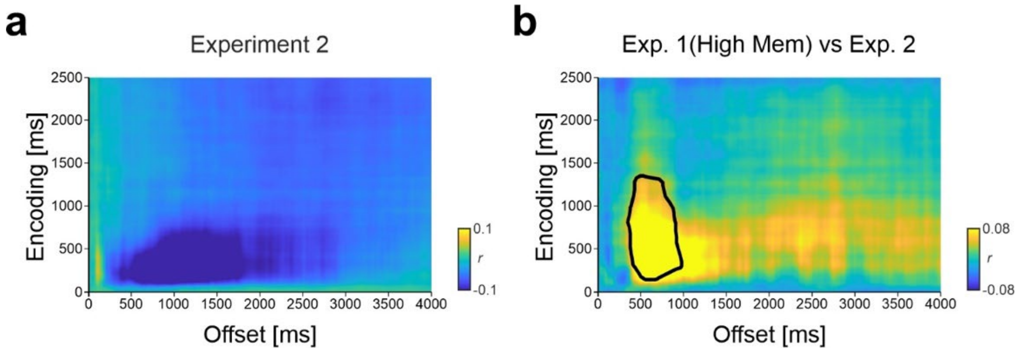

Figure 4.6. Neural similarity in Experiment 2. **(a)**, Neural similarity between item picture encoding at post-encoding period for Experiment 2. **(b)**, Higher neural similarity was found during the offset period in High memory trials in Experiment 1 when compared to the offset period in in Experiment 2 (the significant cluster is indicated by a black thick line. $p < 0.05$, corrected with a cluster-based permutation test).

4.5. Discussion

Here, we investigated in healthy human participants whether memory reinstatement of a just-encoded sequence of episodic items is a mechanism selectively engaged to support episodic memory formation. In two separate experiments, we examined whether memory reinstatement at encoding offset was concomitant to meaningful and/or to non-meaningful sequence of episodic items. In a first experiment, we used representational similarity analysis of scalp EEG recordings during the encoding of sequences of pictures depicting unique episodic-like events and we found that EEG patterns elicited during picture viewing correlated with EEG patterns at the episode offset. The degree of episodic offset-reactivation predicted later memory recall of the encoded picture sequence. In a second experiment, we used a similar analytical approach on EEG recordings while a different set of participants encoded sequences of pictures that were unrelated to each other, thereby preserving similar temporal encoding structure on meaningless episodic sequences. In this experiment, we did not find evidence of post-encoding memory reactivation at the offset. These results suggest that post-encoding memory reinstatement is akin to the rapid formation of unique and meaningful episodes that unfold over time.

Given the unfolding nature of our experience, researchers have started to focus the attention to brain encoding mechanisms that followed online encoding, as this may offer an “optimal” window whereby an unfolding episode can be registered as a bound representation once the ongoing inputs have concluded (Lu et al., 2020). fMRI studies using short videoclips (Zacks et al., 2001; Ben-Yakov et al., 2011; Ben-Yakov et al., 2013), sequence learning tasks alternating picture categories (DuBrow and Davachi, 2014) and, more recently, using long movie clips (Baldassano et al., 2017; Ben-Yakov et al., 2018) offered converging evidence that the brain is sensitive to episodic boundaries during encoding. In line with these findings, we recently showed that event boundaries, operationalized as transition points in the encoding time whereby one episode ends and new one starts, triggered the rapid memory reinstatement of the just encoded event information upon context shifts, and that the degree of memory reinstatement predicted the participants’ ability to preserve temporally adjacent events in a later test (Sols et al., 2017; Silva et al., 2019). The current results extend previous findings in several important ways.

If memory reactivation at episodic offset serves to promote the encoding of unique events into memory, we reasoned, then, it should be observable at the end of an episode and not contemporarily to the beginning of another episodic input. Our results from experiment 1 showed that this is the case, as EEG patterns elicited during sequence encoding correlated to EEG patterns at immediate offset periods that did not contain any stimuli input. Importantly, memory reinstatement was not observed in transition points between pictures of an episode. These findings lend support to the notion that memory reinstatement does not merely reflect an ongoing mechanism that links items associated during encoding. Instead, it suggests that episodic offset memory reactivation is a specific neural signature induced once an individual perceives an unfolding episode concluded.

We also found that post-encoding memory reinstatement followed the encoding of picture sequences that were perceived by individuals as depicting a meaningful episode (i.e., Experiment 1) but not after the encoding of sequences of pictures that were unrelated to each other (i.e., Experiment 2). Why would post-encoding reactivation be important for the formation of memories of meaningful episodes? Previous research has shown that the recall of life-like episodes is organised along representational dimensions beyond their temporal structure, such as the causal (Brownstein & Read, 2007), semantic (van Kesteren et al., 2013; Baldassano et al., 2018) and the relations between the elements embedded in the episodes (Lee and Chen, 2021). This research aligns well to psychological research that emphasized that memory is carved by how people construct a high-order model of the ongoing experience and that the detection of an episodic boundary triggers a set of neural and cognitive processes that would allow the integration of the just-encoded episodic model in memory (Radvansky and Zacks, 2011). Importantly, for this resulting model to be effectively recalled later, it should include several components that vary in the representational hierarchy: from object features to semantics (Radvansky and Zacks, 2011). Recent fMRI research has provided evidence that such representational structure is well reflected along cortical hierarchy during online encoding of realistic stream of stimuli (e.g., Baldassano et al., 2017; Bird, 2020; Reagh and Ranganath, 2021; Lee and Chen, 2021; Heusser et al., 2021) and that such cortical patterns are coupled to hippocampal activity at the detection of high-order event boundaries during encoding to account for later recall (Baldassano et al., 2017). Our current findings contribute to this literature by indicating that memory reactivation is a neural signature by which this high order episodic models can be stored rapidly at episodic encoding offset.

The fact that memory reactivation was found during a delay period immediately following encoding can be seen as a reflection of a mechanism inherently linked to the working memory (WM) maintenance of the encoded sequence of pictures. However, several observations in our results suggest the reinstatement at the episode offset cannot be explained solely by WM processes. First, post-encoding memory reinstatement predicted participants' ability to recollect the episodic picture content in a later test but not their subjective rating of coherence that immediately followed the delay period. Should post-encoding memory reinstatement be a mechanism supporting WM maintenance we would expect it to be at least partially associated with the individual's ability to evaluate the episodic coherence of the encoded picture sequence after the delay period. Second, post-encoding memory reinstatement in our study was circumscribed to the beginning but not throughout the post-encoding delay period. Studies investigating the neural substrates of WM have shown that delay maintenance is associated with a sustained increase in activation of neocortical structures (Fuster and Alexander, 1971). Should post-encoding memory reinstatement be associated with the sustained increase neural activity during the delay we would expect it to be observed over extended portions of the offset delay period and not only at the beginning of it. More recently, it has been argued that such above-threshold delay-period activity may support functions other than information storage per se (D'Esposito and Postle, 2015) and the existence of other neural coding mechanisms such as "activity-silent" states (Stokes, 2015) and dynamic coding schemes (Liu et al., 2020). However, these neural representational formats are still susceptible to be identified with the implementation of multivariate decoding approaches, such as the one implemented in the current design, thereby rendering unlikely they were unobservable throughout the delay period in our study. Third, recent findings from direct recordings at hippocampal and neocortical regions in epileptic human patients showed that the hippocampus marks the conversion from external (perceptual) to internal (mnemonic) representations by signalling cortical reinstatement at ~500 ms after the onset of a retrieval cue (Treder et al., 2021). Our study showing that memory reinstatement detected from scalp EEG signal emerged transient and of a relatively brief duration at ~500 ms at post-episodic encoding period may result from a switch from perception to memory process during encoding itself which would help bind the unfolding information into a memory episodic unit.

A pressing question derived from the findings of our second experiment is why post-encoding memory reactivation was absent when participants encoded sequences of pictures depicting unrelated content. Though the statistical absence of an effect cannot guarantee the absence of the effect, our results lend support to the notion that that post-encoding reactivation strength may not be an automatic mechanism that links items associated during encoding. Instead, it suggests that it contributes to the integration of event components into long-term memory once an individual perceives a meaningful episode is concluded. Previous fMRI research has highlighted that the primary role of hippocampal offset signal in reflecting binding operations of stimuli that just co-occurred within the same spatial-temporal context (Staresina and Davachi, 2009; Ritchey and Cooper, 2020). Similarly, fMRI studies (Baldassano et al., 2017) and electroencephalographic recordings from implanted electrodes in epileptic patients (Michelmann et al., 2021) revealed that the degree to which offset hippocampal activity couples with cortical patterns of activity during a continuous stream of stimuli predicts pattern reinstatement during later recall, thereby indicating that the hippocampus may be responsible for binding cortical representations into a memory trace online during encoding (McClelland et al., 1995; Norman and O'Reilly, 2003; Moscovitch et al., 2005).

Single-cell recordings from the rodent hippocampus during navigational tasks have shown that neural replay can be observed after the first lap on a novel track (Foster and Wilson, 2006). More strikingly, this research has shown that post-encoding replay may preserve the temporal structure of the encoded event sequence in a compressed time-manner (Csicsvari et al., 2007; Diba & Buzsaki 2007; Foster & Wilson 2006; Gupta et al., 2010; Karlsson and Frank, 2009), thereby suggesting that awake neural replay after single-shot learning may reflect the encoding of a model of the experience in long-term memory (Foster, 2017). Our findings based on scalp EEG recordings are blind to whether memory reinstatement at episodic offset period relies on memory replay of a temporally preserved structure of the encoded sequence. Future studies using brain acquisition approaches more sensitive to hippocampal activity, such as MEG (e.g., Liu et al., 2019) or intracortical recordings directly from the human hippocampus (e.g., Vaz et al., 2020) may help disambiguate whether the compressed episodic offset memory reinstatement preserves a temporal structure of an encoded sequence episode.

To conclude, we have shown that episodic offset memory reinstatement is selectively engaged to support successful encoding of sequential picture series with a coherent structure. These results shed light on the neural mechanisms that support the rapid learning of novel episodes that unfold over time in humans and how they serve to selectively transform experiences into long-term memory representations that can be later recalled.

Chapter 5. Study 2

Post-encoding reactivation binds episodic sequences in long-term memory

Please Note: This study has been posted on BioRxiv.
bioRxiv 2022.08.09.503295; doi: https://doi.org/10.1101/2022.08.09.503295

5.1. Summary

In episodic encoding, the unfolding experience is rapidly transformed into a memory representation that binds separate episodic elements into an integrated episodic form so that it can be later recollected. Here, we examined whether this memory transformation occurs rapidly once the encoded episode is completed via a replay-like neural mechanism. We asked participants to encode trial-unique combinations of face-object-scene picture triplets that were subsequently recalled in a test. We used representational similarity analysis of scalp electroencephalography (EEG) recordings during encoding, and we found evidence for memory reactivation of the just encoded episodic elements after encoding, i.e., at the offset following each triplet presentation. Neural patterns associated with face-object-scene categories were obtained from a different set of images and applied to EEG signals to index the magnitude of pattern discriminability across image categories during sequence encoding and at the offset period. We found that the degree of category discriminability decreased as a function of item order in the sequence, indicating the gradual integration of items' categories from the sequence during online encoding. However, we found a smaller degree of category discriminability at offset period presentation (i.e., at 500 ms at offset period) for sequences that would be later recollected compared to those that would be forgotten. Given the temporal overlap of the increase in neural similarity and the decrease in picture category separability measures, and the specificity of the effects in relation to memory performance, we conclude rapid memory reactivation at episodic offset supports an early stage of transformation of the encoded elements within the events into a unique and bound memory trace.

5.2. Introduction

In natural settings, experience unfolds over time, yet when we echo the past, we remember our experience as a collection of distinct and cohesive events. A current hallmark in the research field is to understand how learning systems operate to rapidly transform the ongoing experience into separate and unified memory traces that can be fully recollected later.

Traditionally, human episodic memory research answered this question by focusing on experiments using simple, well-controlled stimuli (e.g., images or words), and analysing the neural underpinnings that predicted their successful retrieval during the online encoding (i.e., when stimuli to be remembered were present) (Paller and Wagner, 2002). This research offered valuable insights into the interactive nature of the hippocampal-neocortical learning system (Eichenbaum, 2000; McClelland et al., 1995; Squire, 1992), their chronometry (Staresina and Wimber, 2019) and the mechanisms that guide the transformation of perceptual inputs into a memory trace in the long term (Lee et al., 2022). Recently, however, researchers have started to study memory in more naturalistic situations, for example, while volunteers watch a movie. These studies complemented results from traditional experiments by showing that episodic encoding occurs selectively over time. For example, results from fMRI (Baldassano et al., 2017; Ben-Yakov et al., 2013; Ben-Yakov and Henson, 2018) or single-neuron activity recordings from epileptic patients (Zheng et al., 2022) studies suggest that episodic encoding occurs preferentially offline, at the ends of events. These results align well with previous findings that showed that successful encoding of an unfolding sequence of items is guided by hippocampal-cortical signalling at event boundaries when there is a contextual shift of an unfolding sequence episode (Dubrow and Davachi, 2014; Clewet et al., 2019). This boundary-locked response predicts subsequent memory performance for the just-completed event (Ben-Yakov and Dudai, 2011; Baldassano et al., 2017), leading researchers to conclude that it is a neural signature of episodic encoding of the just-completed event.

An important question that remains unresolved relates to the representational nature of this episodic immediate offset-locked neural activity supporting the recollection of the just encoded experience. One plausible explanation, inspired by animal literature (Duba and Buzsáki, 2017) and supported by recent scalp electrophysiological findings in humans (Sols et al., 2018; Silva et al., 2019; Wu et al., 2021), would be that the completion of an encoded sequential event triggered the reactivation of the just-encoded episodic elements. The possibility of episodic offset-locked neural reactivation taking place is mechanistically attractive because of several reasons. It would bring the possibility to rapidly implement identified neural processes of memory formation and consolidation (Dudai et al., 2015) and adjust them to the dynamical timing of the episodes’ start and end in naturalistic experiences. It would support the notion that encoding-retrieval processes can rapidly intertwin during the awake experience (Hasselmo, 2003), thereby indicating that retrieval-like operations are engaged throughout the encoding experience enabling the integration of novel and past memory traces (Griffith and Fuentemilla, 2019). It would enable the rapid storage of the just-encoded elements accurately and cohesively by minimising the possible interference driven by the continual encoding of the ongoing experience. Alternatively, post-encoding neural reactivation may promote the rapid transformation of the just encoded event into an integrated form, whereby the distinct elements within the event are bound together in long-term memory as a unified memory trace akin to a rapid memory consolidation process (e.g., Liu et al., 2021). In the current study, we attempted to arbitrate around this latter issue by examining whether episodic immediate offset-locked neural signature promotes the reactivation of an accurate representation of the specific elements of a just-encoded episode or, alternatively, whether it aids an early stage of memory transformation of the encoded event into a bound representation to achieve long term memory recollection.

We recorded scalp EEG signal while participants encoded trial-unique combinations of face-object-scene picture triplet sequences to be subsequently recalled in a test. We used time-resolved representational similarity analysis in combination with a novel pattern classification analysis to quantify the degree of fidelity of an EEG signal to a general picture category representation (faces vs objects vs scenes). We applied these analyses during the encoding and episodic offset period. We found an increase in neural similarity between neural signals elicited at ~ 400 ms during sequence encoding and at episodic offset. We also found that the degree of trial-based neural similarity at episodic offset was greater for the recollected triplet episodes than those forgotten. Interestingly, neural signals elicited during picture encoding and at the early stages of the offset period, at ~ 400 ms, resembled general category picture representations. However, the degree of category representation at the offset, but not during picture encoding, was lower for later recollected than forgotten triplet episodes. The increase in trial-based similarity measures and the picture category pattern separation at the offset period for recollected episodes suggest that rapid offset-locked signals trigger a reactivation of a bound representation of the elements of a just-encoded experience.

5.3. Materials and Methods

5.3.1. Participants

Thirty-two native Spanish speakers were recruited for the current experiment and compensated by 10 €/hour for their participation. All participants had a normal or corrected-to-normal vision and reported no medical, neurological or psychiatric disorders history. Two participants were excluded from the study due to technical problems during the EEG recordings. Data from 30 participants (17 females; age range 18 - 32 years, M = 23.77, SD = 4.38) were analysed. Informed consent was obtained from all participants in accordance with procedures approved by the Ethics Committee of the University of Barcelona.

5.3.2. Stimuli

The experimental design included 312 images (350×350 pixels each): 104 images of famous faces (52 male and 52 female), 104 images of famous places, and 104 object images. Famous face and scene images were selected from a larger sample of the image database consisting of 284 and 184 pictures of each category, respectively. The selection was carried out by a separate sample of 10 Spanish university students (5 females; age range 21-39 years) who rated their familiarity with each image on a scale from 1 to 4 (1: Not recognised; 2: Familiar; 3: Recognised but don´t know the name; 4: Know the name). The final set of 104 face and place images were those that received the highest mean score by 10 external raters (mean score equal to or higher than 3.44 for male, 2.89 for female and 2 for places). The 104 objects were selected from available object picture databases and covered 6 categories (clothing, food, tools, transport, work and leisure). For each participant, 60 images (20 object images, 20 face images of famous people and 20 images of famous places) were randomly selected for the Localizer phase. Among the 312 images, 60 (20 object images, 20 face images of famous people and 20 images of famous places) were randomly selected for the localisation block. For the main task, 36 images (12 object images, 12 face images and 12 place images) were used for example trials and the rest 216 images (72 object images, 72 face images and 72 place images) were used for the encoding trials, this separation was kept the same across participants.

5.3.3. Experimental design

The experiment consisted of the localiser phase and the task phase. In the localisation phase, 60 images (20 faces, 20 scenes and 20 objects) were presented in random order to participants. Each trial started with a 1000 ms fixation cross, followed by a 2500 ms image presentation. Then, a text displayed on the screen indicated the need of the participants to inform the category of the just presented image (Figure 5.1). Participants had a maximum of 10 seconds to respond. The next trial started immediately once a response was given or passed the maximum time limit. There was a brief break between every 20 trials where participants could briefly rest and decide to continue whenever they felt ready.

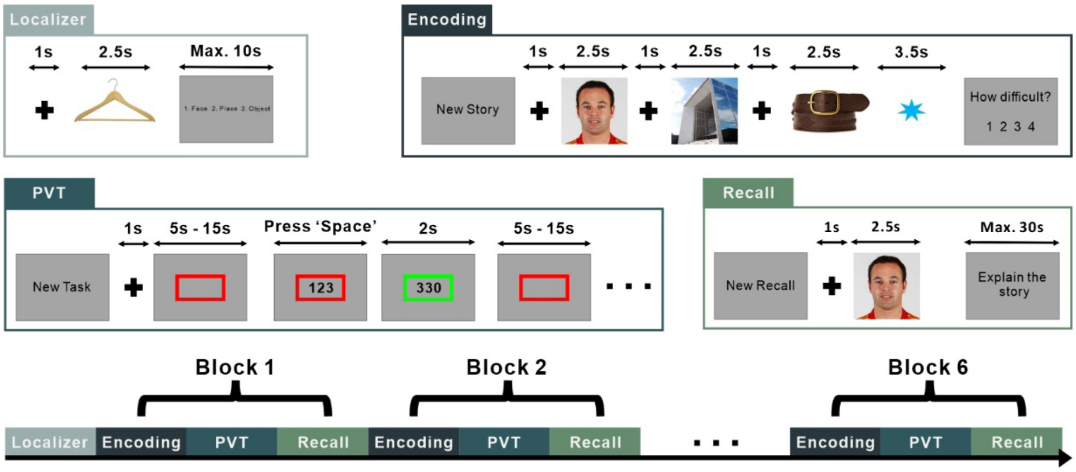

Figure 5.1. Experimental design. For the localisation task, 60 different images from 3 categories (face, place, and object, 20 each) were presented 2.5s on the screen. Participants were asked to indicate the image category within a maximum of 10s after its presentation. Each block started the with encoding phase, where 12 trials of triplets, each consisting of image of an object, a famous face, and a famous place, with fixed category order, were presented to participants. Participants were encouraged to construct stories using the three elements for later memory test. Each image was presented on the screen for 2.5s following 1s of fixation cross. A blue asterisk appeared at the end of each triplet for 3.5s, then participants needed to rate the sense of difficulty for story construction. Two example encoding trials were presented at the beginning of each encoding phase. For the Psychomotor vigilance task, participants were instructed to pay attention to the centre of the screen waiting to react on the onset of a text timer by pressing the 'Space' button. After a random interval between 5s to 15s, text timer started counting in the middle of the red square indicating the real passing time in milliseconds. Once the button was pressed, the timer stopped with the presentation of the final reaction time for 2s. Then the text was cleared, and the new trial started. There were in total 12 trials with no pause in the middle. Retrieval phase began after the PVT task. There were 12 recall trials, each of which used the first image of the previously presented triplets to cue the free recall of the other two images. Participants had maximumly 30s to verbally recall after the 2.5s of cue image presentation. One block was completed after the

retrieval phase, and the next block started following a brief break. The experiment consisted of 6 blocks in total.

The task phase started after the localiser phase. The task phase included 6 blocks, each of them including an encoding task, a Psychomotor vigilance task (PVT) and a retrieval test. Each block was independent of each other, so that picture images presented in one block were never shown in any other of the blocks, but the task instructions and their order of alternation were the same in each block. In the encoding task, participants were instructed to encode 12 series of three images, namely an object (O), a famous face (F), and a famous place (P). Participants were encouraged to construct stories using triplet elements in the form of a narrative (e.g., Iniesta went to Paris and purchased an expensive belt), and they were informed the triplet information would be tested later. In total, 72 triplets were randomly generated for each participant from 216 images (72 object images, 72 face images and 72 place images), and each image was used only once in the experiment. In each block, the presentation order of the image categories in a series was fixed (e.g., always ordered as Face-Place-Object in one block). There are in total 6 possible presentation orders, each of which was used in one of the 6 blocks with no repetition and randomly generated for each participant. At the beginning of each block, two example trials were presented, indicating the order of presentation of the image categories. Participants were instructed to use the two example trials to rehearse the upcoming series encoding in the block and that the example trials would not be tested later. Each encoding trial began with the presentation of the text 'New Story' for 3000 ms, which marked the start of a new triplet series. Triplet images were then presented sequentially on a white screen for 2500 ms each after a 1000 ms black fixation cross. Immediately after the presentation of the last image in each triplet, a blue asterisk appeared on the screen, indicating a post-episode offset period of 3500 ms, during which participants were instructed to avoid rehearsing the just-encoded triplet series. The asterisk remained visible on the screen during the offset period. Participants were then asked to provide a degree of subjective feeling of the difficulty of constructing a coherent episode with the just presented triplet of images by a button press on a scale from 1 ('very easy') to 4 ('very difficult'). The next trial began immediately after a response was given, or no response was given after a time limit of 10 seconds. A small break of ~10 sec was provided after completing 6 trials.

A block of the PVT task followed the encoding phase. In each PVT block, participants were instructed to pay attention to the screen's centre and press the space button as quickly as possible once the timer started counting. The task commenced with the text presentation 'New Task' for 3000 ms. Then an empty red square was displayed at the centre of the screen following a 1000 ms fixation cross. After a random interval between 5 sec to 15 sec, the timer started counting in the middle of the square indicating the real passing time in milliseconds. The timer counted maximumly until 3500 ms if no response was given. Once the participants pressed the button during the counting period, the timer stopped with the presentation of the final reaction time in the centre of the screen for 2000 ms. In cases where no response to the timer was given within the time limit, the presentation would be the final counting time of the timer (i.e., 3500 ms). The new PVT trial started immediately after the reaction time presentation. In total, 12 repetitions of response were required with no interruption in the middle. A block of a PVT task lasted around 3 minutes.

The PVT task was followed by a cued-recall task. During this task, participants were presented with the first image of all the encoded triplets in the current block in random order. They were required to verbally recall the story episode containing the other two images associated with the cue image. Each trial began with the text 'New Recall' for 3000 ms, followed by the cue image on the screen for 2500 ms and a 1000 ms fixation cross. Then, the text 'Explain the story' was displayed on the screen, which indicated to the participants they could start the verbal recall. The verbal recall had a maximum duration of 30 s, during which the text instruction remained visible on the screen all the time. Participants could skip to the next trial when finished with their recall or if they were unable to recall any associated image by pressing the space bar. A brief break of ~20 s separated the start of the next block.

5.3.4. EEG recording and preprocessing

During the experiment, EEG was recorded with a 64-channel system at a sampling rate of 512 Hz, using a eego™ amplifier and Ag/AgCl electrodes mounted in an electrocap (ANT neuro) located at 59 standard positions ((Fp1/2, AF3/4, Fz, F7/8, F5/6, F3/4, F1/2, FCz, FT7/8, FC5/6, FC3/4, FC1/2, Cz, T7/8, C5/C6, C3/4, C1/2, CPz, TP7/8, CP5/6, CP3/4, CP1/2, Pz, P7/8, P5/6, P3/4, P2/1, POz, PO7/8, PO5/6, PO3/4, Oz, O1/2) and at the left and right mastoids. Horizontal and vertical eye movements were monitored with electrodes placed at the right temple and at the infraorbital ridge of the right eye. Electrode impedances were kept below 10 kΩ. EEG was re-referenced offline to the linked mastoids. Bad channels were interpolated, and a band-pass filter (0.5 Hz - 30 Hz) was implemented offline. Blinks and eye movement artifacts were removed with independent component analysis (ICA) before the analysis.

5.3.5. Behavioural data analysis

During the retrieval phase of the experiment, participants were instructed to verbally recall the constructed story episode associated with the picture cue. Verbal recall of each trial was recorded through an audio recorder, and the audio files were later analysed. A successful recall of the image was considered as either correctly mentioning the name or describing it in precise detail. Memory for each triplet was quantified by the number of images (excluding the cue) correctly recalled.

5.3.6. EEG data analysis

For each participant, we extracted epochs of EEG activity surrounding pictures presented in the localiser and the encoding tasks. These EEG trial epochs had a duration of 2500 ms (1280 data points given the 512 Hz EEG recording sampling rate), and they were baseline corrected to the pre-stimulus interval (-100 to 0 ms). We also extracted EEG epochs of 2100 ms (1024 data points) from the offset period following the encoding of each triplet series. EEG signal to the offset period was baseline corrected to the -100 to 0 ms averaged EEG activity. EEG trial epochs that exceeded ± 100 μV were discarded for further analysis. EEG trials were then Gaussian smoothed by averaging data via a moving window of 100 ms (excluding the baseline period) and then downsampled by a factor of 5.

5.3.7. Representational Similarity Analysis (RSA)

RSA was performed timepoint-to-timepoint and included spatial features (i.e., scalp voltages from all the 28 electrodes) (Silva et al., 2019; Wu et al., 2021). The similarity analysis was calculated using Pearson correlation coefficients, which are insensitive to the absolute amplitude and variance of the EEG response.

We conducted a trial-based RSA between the EEG signal elicited by each encoding item (1st, 2nd, and 3rd, regardless of the image category) and the EEG signal elicited at the offset period following the encoding of the triplet series. After smoothing and down-sampling, EEG epoch data elicited by each picture in the triplet included 205 sample points (given the 512 Hz EEG recording sampling rate) covering the 2000 ms of picture presentation and EEG data from post-triplet offset contained 359 time points, equivalent to 3500 ms. Point-to-point correlation values were then calculated, resulting in a 2D similarity matrix with the size of 205×359, where the x-axis represented the episodic offset time points and the y-axis represented the picture encoding time points. The output 2D matrix depicted the overall degree of similarity between EEG patterns elicited by each encoding image and the subsequent post-episodic offset interval.

To account for RSA differences between conditions, we employed a nonparametric statistical method (Maris & Oostenveld, 2007), which identifies clusters of significant points on the resulting 2D similarity matrix and corrects for multiple comparison based on cluster-level randomisation testing. Statistics were computed on values between conditions for each time point, and adjacent points in the 2D matrix that passed the significance threshold ($p < 0.05$, two-tailed) were selected and grouped together as a cluster. The cluster-level statistics took the sum of the statistics of all time points within each identified cluster. This procedure was then repeated 1000 times with randomly shuffled labels across conditions. Cluster-level statistics with the highest absolute value for each permutation was registered to construct a distribution under the null hypothesis. The nonparametric statistical test was calculated by the proportion of permuted test statistics that exceeded the true observed cluster-level statistics.

5.3.8. Linear Discriminant Analysis (LDA)

To identify the multivariate pattern of brain activity for image processing of different categories, a Linear Discriminant Analysis (LDA) was trained and tested on the EEG sensor patterns of localiser trials (pre-processed signal amplitude from 59 channels). The classifier was trained independently per participant and at each time point during localiser image presentation, then tested with a leave-one-out cross-validation procedure. Given that three categories were included in the current experiment (face, place, and object), at the training stage, the classifier was trained repetitively three times, including each possible pair out of the three classes. For each of the two classes, the classifier found the decision boundary that best separated the pattern activity. We then asked the classifier to estimate the unlabelled pattern of brain activity for each of the three decision boundaries (one for each pair of classes). The output of the classifier for each two trained classes at a given time point was the distance value to the decision boundary, which represents how probable the pattern of brain activity belonged to one of the two included classes, with the sign indicating the class while the magnitude reflecting how confidence was the classifier. The distance value for each pair of classes was then sigmoid transformed to get the probability of either class that unlabelled pattern activity belonged to (e.g., a distance value of 0 will return 50% for either class). After normalising and averaging values across the three possible pairings, the class with the highest probability was marked as the final label for the testing data. To access the general separability between the three classes in a compound measure, we defined a separability index as the sum of the absolute of the three distance values to each of the decision boundaries, with the assumption being that the greater the separability index, the higher the probability that the given activity pattern belonged to a specific class rather than assimilating to all three classes with equal distinctiveness (i.e., closer to zero).

This training-test procedure was repeated until every single localiser trial had been classified. The predicted labels for all trials at every given time point were then compared to the true classes to assess the accuracy of the classifier across all localiser time.

To evaluate how face, object and scene category representations accounted for EEG patterns elicited during picture encoding and at the offset period, we first identified the time point where the cross-validation of the classifier reached the peak accuracy. Then using patterns of activity surrounding 10 time points around the peak (- 50 ms to 50 ms with the peak time point in the middle), we trained the classifier per participant with all localiser trials and predicted all sample points separately for encoding and offset trials. The results were then averaged across localiser time points, resulting in a 1D separability index line for each trial where each sample point represented the encoding/offset time points.

5.3.9. Linear-mixed effect model

To further explore how the separability of pattern activity between picture categories changed along the encoding sequence and whether it is predictive for behavioural memory on a trial basis, we implemented a Linear Mixed Effect Model (LMM) on the resulting general distance value of each encoding image classified by patterns trained on trials from the localiser task. We further smoothed the resulting 1D distance value for each predicting encoding trial by averaging over a moving window of 100 ms, then introduced in our LMM value on each time point as the independent variable and both image order in triplet series (1^{st}, 2^{nd}, and 3^{rd}) and recall memory (successfully recalled 0, 1, or 2 images following the cue) as well as the interaction of the two as fixed effect variables. Subject was introduced in the model as the grouping variable, with random intercept and a fixed slope for each fixed effect variable. The statistical significance was then evaluated using Bonferroni correction for each fixed effect variable at each timepoint thresholded with an adjusted alpha level of $\alpha = 2.44\times10^{-4}$ (0.05/205). The procedure was repeated to compare high and low memory conditions during offset. The output 1D distance line across offset was averaged for each condition across subjects. The *t*-statistics was then computed at each time point, and the significance was evaluated at the cluster level after cluster-based permutation.

5.4. Results

5.4.1. Localisation task

For the localisation task, 26 out of 30 participants reached 100% accuracy in identifying the image category, and the mean accuracy across the 30 participants was 99.72% (SD = 0.77%).

5.4.2. Recall of picture triplets

Participants were able to recall on average 1.12 pictures (SD = 0.37) following the cue, with the mean percentage of trials recalling 0, 1, and 2 items being respectively 35.61% (SD = 16.41%), 16.30% (SD = 6.52%) and 48.09% (SD = 20.68%) (Figure 5.2a). We also found that the average number of images recalled upon the picture cue did not vary between blocks, indicating that encoding and retrieval accuracy did not vary throughout the task (repeated-measures ANOVA with block number as the main factor: $F_{(5,140)} = 0.415$, $p = 0.838$). However, there was a significant difference of recall performance depending on the category order of the triplet series. More concretely, we found that encoding blocks that included triplets with face as the first picture (i.e., Face-Place-Object or Face-Object-Place) were less accurately recalled (Face-Place-Object: mean = 0.914, SD = 0.098; Face-Object-Place: mean = 0.905, SD = 0.082) than triplets from blocks where place (Pace-Face-Object: mean = 1.253, SD = 0.081; Pace-Object-Face: mean = 1.224, SD = 0.081) or object (Object-Face-Place: mean = 1.213, SD = 0.081; Object-Place-Face: mean = 1.230, SD = 0.084) was presented first (repeated measures ANOVA: $F_{(5,140)} = 9.798$, $p < 0.001$).

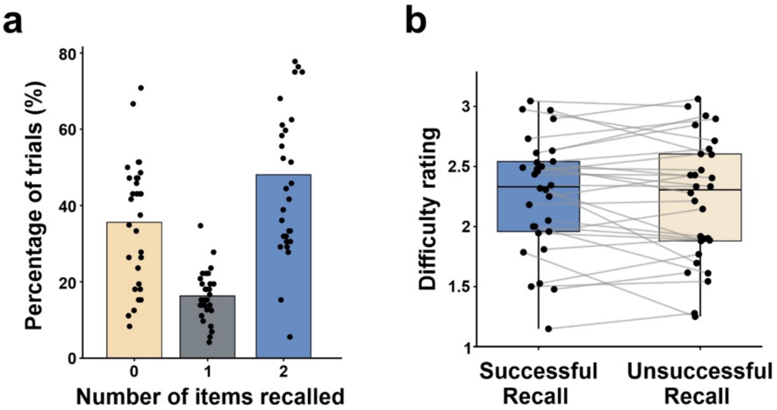

Figure 5.2. Behavioural results. **(a)** Percentage of trials as a function of numbers of images correctly retrieved during free recall. **(b)** Subjective rating of the difficulty of triplet encoding separated by whether or not the triplet was later successfully recalled (with both images associated with the cue being correctly recalled). Each dot on both plots represents the value for an individual in the corresponding condition. Each grey line on the boxplot connects the value of an individual in two conditions.

For RSA analysis, we adopted a median-split approach (Wu et al., 2021) to separate the trials based on whether the entire triplet images were correctly retrieved. Triplets with 2 images recalled after the cue were labelled as successful recall, and triplets with either 1 image or no image recalled were labelled as unsuccessful recall. The average percentage of trials were respectively 48.09% (SD = 20.68%) for successful recall condition and 51.91% (SD = 20.68%) for unsuccessful recall condition (Wilcoxon signed-rank test: $z = -0.43$, $p = 0.67$). RSA results for successful and unsuccessful recall trials were then compared using a point-to-point paired t-test. The statistical difference between the two conditions was then assessed with a cluster-based permutation approach.

5.4.3. Participants' ratings of encoding difficulty

On average, triplets were rated as 2.24 (SD = 0.47) (on a scale that ranged from 1: no difficulty to 4: very difficult), and the mean percentage of triplets rated as 1, 2, 3, and 4 were respectively 27.91% (SD = 21.73%), 33.79% (SD = 13.80%), 24.65% (SD = 13.13%) and 13.65% (SD = 12.03%). Based on the median-split criteria, difficulty ratings for trials with successful recall (mean = 2.26 and SD = 0.48), and for trials with unsuccessful recall (mean = 2.22 and SD = 0.51) did not differ statistically between each other (paired Student *t*-test: $t_{(29)} = 1.01$, $p = 0.32$, two-tailed) (Figure 5.2b).

5.4.4. RSA between item sequence and episodic offset at encoding

We first examined the existence of encoding-offset neural similarity differences between trials that were successfully or unsuccessfully recalled. This analysis revealed that EEG patterns elicited during the encoding of picture triplets that were later recalled showed, compared to unsuccessfully recalled trials, a higher degree of neural similarity during the episodic offset period (Figure 5.3a and 5.3b). This result was

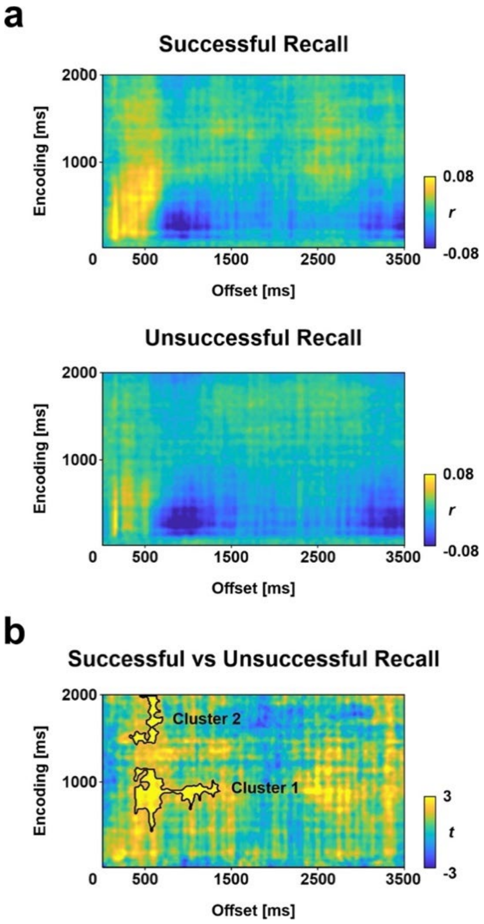

Figure 5.3. RSA for image encoding and its corresponding post-triplet offset. **(a)**: Time-resolved degree of neural similarity between image encoding and post-triplet offset for trials with successful subsequent recall (upper) and unsuccessful recall (lower). **(b)**: Difference between similarity values for the two conditions. Statistically significant (p < 0.05, cluster-based permutation test) higher similarity value was found for trials successful recall centred in two areas (indicated by black contour lines. $P_{\text{cluster1}} < 0.001$; $P_{\text{cluster2}} = 0.002$).

corroborated statistically with the cluster-based permutation test, which showed two clusters of increased neural similarity starting at ~400 ms at offset period (Cluster 1: $p < 0.001$, mean t-value = 2.98, peak t-value = 4.74; Cluster 2: $p = 0.002$, mean t-value = 3.10, peak t-value = 4.92) (Figure 5.3b).

5.4.5. Classification accuracy and separability of picture category

We adopted the LDA approach to classify and predict the image category being processed based on the elicited EEG pattern in the localiser task. The classifier was trained independently per participant and at each time point during picture encoding, then tested with a leave-one-out cross-validation procedure. Two output values were extracted for each time of training/testing, namely the category of tested data predicted by the model with the highest probability among three alternatives (i.e., accuracy) and a general distance value (*D* value) of tested data to the classification plane among categories (i.e., separability) (Figure 5.4a).

The results of this analysis showed that picture category could be reliably predicted rapidly at picture onset (i.e., ~130 ms), showing a peak classification accuracy at ~180 ms (t_{29} = 8.41, $p_{\text{corr}} < 0.001$) (Figure 5.4b).

Expectedly, the pattern separability analysis showed similar temporal dynamics as the accuracy ones. More specifically, pattern separability became significant as the distance value increased compared to surrogate trials, with the difference emerging from ~170 ms. *D* value reached the local maximum at respectively ~180 ms (D = 6.57, t_{29} = 5.032, p_{corr} = 0.005) and at ~380 ms (D = 6.60, t_{29} = 7.25, $p_{\text{corr}} < 0.001$) (Figure 5.4c).

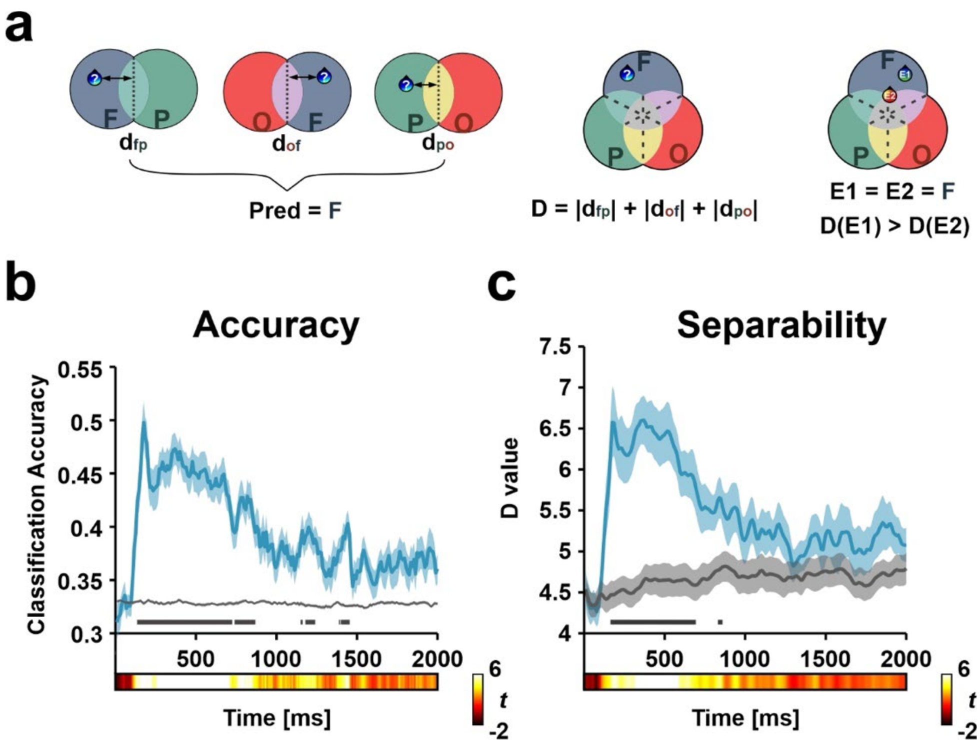

Figure 5.4. Two Cross-validated LDA classifier output measures using localisation trials. **(a)**: Abstract illustration of the classifier output calculation. The distance value for each pair of classes was sigmoid transformed to get either class's probability. The class with the highest probability after normalising and averaging values across three pairs was marked as the final label for the testing data (left). The general distance value (*D* value) was defined as the sum of the absolute of the three distance values to each of the decision boundaries (middle). Pattern examples 1 & 2 can be both classified accurately as 'Face' images. However, Pattern example 2 showed a more similar pattern also to the 'Place' and 'Object' category, which a smaller *D* value can indicate. **(b)**: Classifier accuracy estimated using the leave-one-out method. Image categories can be reliably decoded compared to surrogate trials starting around ~130 ms after image onset, with peak value reaching ~180 ms. **(c)**: Pattern separability among image categories quantified by the general distance value, which evolved similarly across time compared to the accuracy measure. *D* value reached the peak ~180 ms and ~380 ms. In plots (***b***) and (***c***), the shaded area indicated SEM across participants, and statistical significance compared to surrogate trials was Bonferroni-corrected and marked in dark grey line.

5.4.6. Gradual integration of picture category information during sequence encoding

We examined whether the sequential encoding of pictures from different categories in the encoding task would involve a gradual integration of the just encoded images from the sequence and whether this process predicted memory recall. To address this issue, we extracted the -50 to 50 ms EEG pattern surrounding the peak (i.e., 180 ms; Figure 5.4b) LDA accuracy during the encoding of images in the localiser task. Then, we used these EEG patterns as the training data in a new LDA and tested on EEG patterns elicited at each time point from each picture from the sequence on the encoding task.

We then averaged across all training time points at trial-level and included the resulting distance value at each time point of encoding into LMM as the dependent variable. For each trial, the number of items recalled, the encoding order of the image in the triplets (i.e., 1st, 2nd and 3rd), and the interaction of the two were included in the model as fixed-effect variables. Subject was introduced into the model as the grouping variable, with random intercept and a fixed slope for each fixed-effect variable.

This analysis showed that the *D* value correlated negatively with the order of picture in the sequence (Figure 5.5a) and that such effect emerged ~460 ms after picture onset and persisted until ~860 ms. However, we found that *D* value did not correlate with later picture recollection at test nor the interaction of picture order and memory. This suggested that the picture category integrative process takes place during sequence encoding and that this had no impact on the later ability of the participants to retrieve the sequence episode. To control for the possibility that the observed effect was not merely due to a decrease in the specific category classification accuracy as a function of the order of the picture in the sequence, we extracted the mean accuracy across image order within the time window where the significant decrease in pattern separability was identified (Figure 5.5b). A repeated-measure ANOVA showed significantly above-chance accuracy value ($F_{(1,29)} = 111.57$, $p < 0.001$) with no main effect for image order ($F_{(2,58)} = 0.43$, $p = 0.65$).

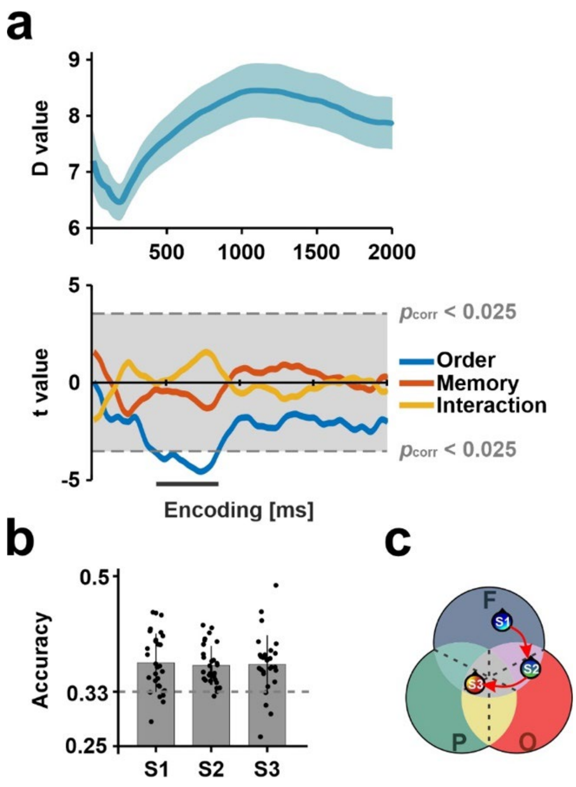

Figure 5.5. Pattern separability during image encoding predicted by LDA classifier trained on localisation trials. **(a)**: Classifier predicted *D* value during image encoding averaged across participants (upper, shaded area indicated SEM across participants) with statistical significance Bonferroni-corrected for the main effect of image order, subsequent memory and their interaction (lower). Shaded grey area and light grey dash line marked the significance threshold boundaries (two-tailed) adjusted by Bonferroni correction. Time window where the main effect passed the threshold was marked in dark grey line below. **(b)**: Classifier accuracy averaged across the time window where a significant effect for image order was found. Category of image was classified equally accurate across image order (p = 0.65) yet significantly above chance (grey dash line) ($p <$ 0.001). Each black dot represents values for an individual participant. The central mark is the median, and the edges of the box are the 25th and 75th percentiles. **(c)**: Abstract illustration of the speculated integration process. While the classifier continued to predict the image category accurately, there was a trend for a 'integrated' pattern indicated by a gradually decreased pattern separability.

5.4.7. Picture sequence integration and memory at episodic offset period

We next examined whether an integrated form of the just encoded sequence could predict memory for the episode right after their online encoding, that is, once the encoding ended, at the offset period. If this was the case, we should observe that *D* value was reduced at the offset period for successful compared to unsuccessful recalled picture sequences.

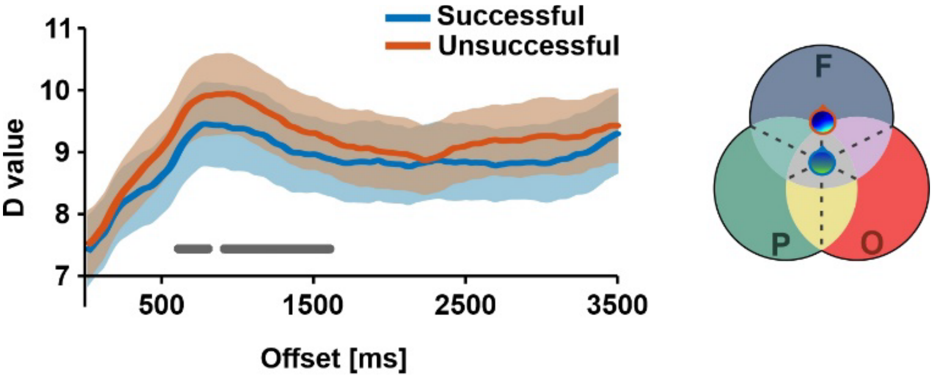

Figure 5.6. Pattern separability during post-triplet offset period predicted by LDA classifier trained on localisation trials. Lower distance value for trials that were later successfully emerged from ~630 ms at triplet offset. Shaded area indicated SEM across participants. Dark grey line marked statistical significance adjusted by cluster-based permutation. An abstract illustration of separability pattern for successfully and unsuccessfully remembered trials were included on the right. Successfully remembered trials showed a more reduced D value as a sign of successful integration among the three categorical representations.

To address this issue, we again extracted the EEG pattern elicited during ~140 ms - 230 ms by picture presentation in the localiser task. We applied it to each time point of the offset period from the encoding task. The resulting D value from the model was then averaged across all training timepoint. We then separated D values for successful and unsuccessful memory trial conditions and averaged them for each participant. Statistical comparisons between conditions were assessed and then assessed with a cluster-based permutation approach.

Confirming our hypothesis, the results of this analysis showed significant lower D value for successful compared to unsuccessful recalled trials at ~630 ms to ~820 ms ($p < 0.001$, mean t-value = -2.41, peak t-value = -2.42) and at ~940 ms to ~1640 ms ($p < 0.001$, mean t-value = -2.59, peak t-value = -2.78) at offset period (Figure 5.6). Importantly, this time window coincided with the increased neural reactivation for successfully recalled triplets identified previously in the RSA, suggesting an overlapping functional role between post-encoding reactivation and integration.

5.5. Discussion

The current study asked whether the rapid neural reinstatement at the end of a sequence episode involves an accurate replay of the just encoded elements within the episode or, alternatively, it reflects an early stage of transformation of the encoded information into a different form. Our findings revealed greater neural similarity for EEG patterns elicited during image sequence encoding at the episodic offset period for those episodes that were later recollected compared to those that were forgotten. In addition, we used a novel analytical approach that quantified the degree of fidelity of an EEG signal to a general picture category representation (faces vs objects vs scenes), and we found that the degree of picture category pattern separability at the offset period, but not during picture encoding, was lower for later recollected than forgotten triplet episodes. The increase in trial-based similarity measures and the picture category pattern separation at the offset period for recollected episodes suggest that rapid offset-locked signals trigger a reactivation of a bound representation of the elements of a just-encoded experience.

Consistent with previous findings (Cichy et al., 2014; Wimmer et al., 2020; Jafarpour et al., 2014), we successfully identified the neural patterns associated with the encoding of picture categories at an early time window from image presentation onset). However, instead of registering only the output from the classifier, generally defined as the predicted class with maximal likelihood among possible alternatives, we used it to develop an index that quantified the classifier's ability to distinguish among all the possible classes in a given time point, the separability or the *D* index. In other words, the *D* index expresses the degree to which a tested neural pattern assimilated or deviated from all the possible trained categories. By extracting the *D* index during each of the pictures from the encoding triplet, we found a gradual decrease of pattern separability from neural patterns elicited at early temporal stages from the pictures sequence, being higher in the first picture and lower in the 3rd picture of the sequence. It is important to note that during this identified time window, the accuracy of the classifier to the correct category remained above chance and similar throughout each of the pictures from the triplet sequence. Thus, the observed gradual decrease in pattern separability cannot simply be attributed to a weak classification performance but instead indicated a gradual reduction in the specificity of activity patterns to a particular picture category (scene, object or face). One possible explanation of this finding may be attributed to an attenuated neural activity by prior expectation, given that participants could anticipate the category of the upcoming image since the order of presentation was fixed within each block. Though prior studies revealed that anticipation might reduce response in neurons tuned for expected stimulus (Kok et al., 2012; Kumar et al., 2017), multivariate approaches have instead shown a 'sharpening' effect for perceptual representations in cortical regions due to a more selective population response (De Lange et al., 2018), resulting in a more accurate pattern classification (Kok et al., 2012). Our findings that the decrease in picture category pattern separability is taking place at around 500 ms from picture onset, however, may not be explained by 'sharpening' effects because they are thought to occur earlier in the temporal course of processing (i.e., < 400 ms from stimuli onset). Instead, we argue that the gradual reduction in pattern separability following the sequential presentation of images reflected a continuously additive category-specific processing, which promoted the encoding of multiple categorical information in parallel, supported by various overlapping cortical regions. In fact, different yet overlapping cortical regions (e.g., various regions on the lateral surface of occipitotemporal cortex) are selectively sensitive to stimuli from different categories when presented in isolation, including face, objects and scenes (Silson et al. 2016). In naturalistic scenarios, the processing of multiple categorical information embedded in the encoding experience takes place simultaneously, and the neural signature of such processes can be decoded in different cortical regions (Cooper and Ritchey, 2020). In the context of our study, the ongoing need to associate each appearing picture with the previously encoded pictures from the sequence may have promoted integrative processes online during the encoding of the picture.

An interesting finding from our study is that a decrease in the degree of picture category pattern separability at the episodic offset, when the episode was completed, but not during online encoding, was in fact predictive of later episodic recollection at test. In addition, we found that later successfully recollected episodes showed greater trial-based neural similarity at an overlapping temporal window at the offset period. All in all, these findings suggest that successfully encoded episodic events triggered a rapid neural reactivation that promoted an early transformation of the encoded elements within the events into a unique and bound memory trace. Previous fMRI literature using different input types such as picture sequences (DuBrow and Davachi, 2014), short video clips (Zacks et al., 2001; Ben-Yakov et al., 2011; Ben-Yakov et al., 2013), and movie clips (Baldassano et al., 2017; Ben-Yakov et al., 2018) highlighted the sensitivity of the hippocampal-neocortical system to detect episodic offsets, suggesting that the end of a long-timescale event triggers memory encoding processes that occur after the event has ended. fMRI representational analysis revealed that offset-locked pattern of neural response might reflect an updated input to long-term memory of multimodal, abstract representations of the features of the just encoded event (Baldassano et al., 2017). Our findings also align well with a recent study that combined direct electrophysiological recordings from human hippocampus and deep neural network analysis that showed that early representation transformation of visual picture information in the first second after stimulus offset was associated with better long-term memory (Liu et al., 2021).

Rodent literature has shown that memory reply right at the end of an encoded event, when a trial is completed, is a core neural mechanism to aid the formation of long-term memories (Foster and Wilson, 2006). However, neural replay in rodents is not limited to playing a veridical recording of experience. Instead, it can make inferences - piecing together multiple past events as if building and sampling from a model of the world (Foster, 2017; Gupta et al., 2010; Karlsson and Frank, 2009). Similarly, human research has proposed that past information could be integrated with incoming sensory information (Eichenbaum, 2000), thereby assisting the registering of an interpretable and meaningful memory trace of incoming information (Stawarczyk et al., 2021; Chen et al., 2016; Keidel et al., 2018; Cohn-Sheehy et al., 2021). The notion that memory transformation takes place at early stages of encoding fits the observation that remembering is a reconstructive process that prioritises more meaningful components of an event over other, more shallow aspects (Schacter, 2012; Schacter et al., 2011). Our findings contribute to this literature by proposing one of the earliest stages whereby this memory transformation takes place is at the end of an encoded event through a replay-like neural mechanism.

In conclusion, we found a gradual integration process of perceptual representations as encoding experience unfolded and the neural mechanisms elicited at episode offset period to promote the transformation of the elements of a just-encoded event to a bound memory trace in the long term.

Chapter 6. Study 3

Contextual incongruency triggers memory reinstatement and the disruption of neural stability

Please Note: This study has been posted on BioRxiv.
bioRxiv 2022.07.26.501077; doi: https://doi.org/10.1101/2022.07.26.501077

6.1. Summary

Schemas, or internal representation models of the environment, are thought to be central in organising our everyday life behaviour by providing stability and predictiveness to the structure of the world. However, when an element from unfolding event mismatches the schema-derived expectations, the coherent narrative is interrupted, and an update to the current event model representation is required. Here, we asked whether the perceived incongruence of an item from an unfolding event and its impact on memory relied on the disruption of neural stability patterns preceded by the neural reactivation of the memory representations of the just encoded event. Our study includes data from 3 different experiments whereby participants encoded images of target objects preceded by trial-unique sequences of events depicting day-to-day routine. We found that neural stability patterns gradually increased throughout the ongoing exposure to a schema-consistent episodic narrative, and that the brain stability pattern was interrupted when the encoding of an object of the event was lowly congruent within the ongoing schema representation. We found that the decrease in neural stability for low congruent items was best seen at ~1000 ms from object encoding onset when compared to high congruent items and that this effect was preceded by an enhanced N400 ERP and an increased degree of neural reactivation of the just encoded episode for low congruent items observed between ~200 to 1000 ms from picture onset. Current results offer new insights into the neural mechanisms and their temporal orchestration that are engaged during online encoding of schema-consistent episodic narratives and the detection of incongruencies.

6.2. Introduction

Experience is guided by internal representation models of the environment, or knowledge schemas, with an impact on perception and memory (Gilboa and Marlatte, 2017). Schemas are thought to be central in organising our everyday life behaviour by providing stability and predictiveness to the structure of the world (Gershman et al., 2014). Thus, despite the ever-changing sequence of inputs of our experience, schemas bring relatedness and comprehension of unfolding events by anticipating stereotyped or congruent-like elements to encounter next. A computational advantage to the memory systems is, therefore, that schema-consistent items can be added to an existing schema without requiring alterations or extensions to it (McClelland et al., 2020). Accordingly, when elements of the unfolding experience are congruent with expected representations from a current activated schema, they are rapidly integrated into the memory model of the event (McClelland et al., 2020; Tse et al., 2007, 2011). However, when our predictions are incorrect, we must update our internal models of the world to support adaptive behaviour. Yet, the neural mechanisms that support memory integration and updating of an unfolding event remain unclear.

If an internal memory representation is stable over time, then some properties of its underlying neural implementations may also exhibit invariance during encoding. Indeed, neuroimaging studies in humans observed stable brain patterns of activity during the encoding of continuous streams of audio-visual inputs and that shifts in neural stability are coincident with the detection of unexpected elements in the unfolding stream (i.e., event boundaries) (Baldassano et al., 2017). Similarly, Sinclair et al. (2021) recently showed that hippocampal activation patterns stabilised during the encoding of a narrative episode, akin to sustained representations accumulated during an unfolding schema-congruent event. Intriguingly, this study also revealed that when the narrative was suddenly interrupted, the ongoing stability of the neural activity became disrupted, reflecting the need to update the sustained representation of the event model.

However, the notion that the detection of incongruencies of an unfolding event engenders a disruption of the ongoing representation challenges a set of findings that found no memory disturbance or even improvement for surprising events (e.g., Greve et al., 2017; Greve et al., 2019; Quent et al., 2021; Frank et al., 2018, 2020; Rouhani et al., 2018; Chen et al., 2015; Pine et al., 2018). This literature relies on the idea that mnemonic prediction error enhances hippocampal biases toward encoding (Bein et al., 2020) and that this shift in encoding strategy reflects the need to evaluate and, if necessary, update the representational content of the ongoing experience with the current incongruent event.

How does the brain accommodate these two seemingly opposite lines of research evidence, namely, that mnemonic prediction errors disrupt ongoing neural representations of unfolding event and, at the same time, promotes the update of the ongoing memory model during encoding? Here, we asked whether this process is supported by distinct brain mechanisms that occur rapidly (in the order of milliseconds) but sequentially orchestrated over time. More specifically, drawing on past theoretical (McClelland et al., 2020) and empirical research (Sols et al., 2017; Silva et al., 2019; Wu et al., 2021), we hypothesise that the subjective degree of item's congruence with an unfolding experience is determined by an evaluation process guided by a rapid reactivation of the encoded event. The concomitant representation of the new element and the reactivated memory of the just encoded event would promote the effective and rapid assessment of the extent to which the novel element match or mismatches expectancies driven by the unfolding event. As a results, the brain would be able to either assimilate the new item with the ongoing memory representation by preserving a stable state of neural pattern of activity or, alternatively, disrupt it to promote their update.

To test this hypothesis, we recorded scalp electrophysiological (EEG) activity while healthy participants encoded images of target objects preceded by trial-unique sequences of four pictures of events depicting an everyday life routine (Figure 6.1a). The sequence of pictures preceding the target object image was thought to mimic a realistic unfolding episodic event with the aim to provide a gradual schema consistent narrative that determined whether specific target objects matched or mismatched expected occurrences within that context. Importantly, participants were instructed to rate the perceived congruence of the item in relation to the previously encoded event sequence episode, thereby allowing us to assess the degree of perceived congruence of the target object for every single trial at individual level. To examine how object congruence shaped memory for the target object, a surprise recognition memory test was administered to the participants after the encoding phase. In Experiment 1, we first asked participants to indicate whether a label word referred to object pictures encoded in the previous phase, and if so, to recognise which from two very similar pictures was the exact one presented during the encoding phase (Figure 6.1b). This later test allowed assessing the extent to which encoding congruence detailed memory representation (e.g., Bein et al., 2020). We ran two additional follow-up behavioural experiments on a separate sample of participants to further scrutinise congruence-shaped long-term memory. The two experiments consisted of a similar structure and materials used in Experiment 1 but differed in the format of the recognition phase. In

Experiment 2, we presented pictures that corresponded to the ones shown during the encoding phase and pictures that depicted similar objects but differed in small visual details (lures), thereby allowing us to assess whether possible differences in memory as a function of encoded congruence was independent of the format of the test (Figure 6.1c). In Experiment 3, we asked participants to indicate whether object pictures were encoded together with a selected picture of the episodic sequence (Figure 6.1d). This test allowed examining whether the perceived congruence of the object within an episodic narrative influenced how the two become associated in long-term memory.

6.3. Material and Methods

6.3.1. Participants

Participants were healthy college students from the University of Barcelona who had a normal or corrected-to-normal vision and reported no history of medical, neurological or psychiatric disorders. Thirty-three participants (26 females, M = 20.94 years, SD = 3.24 years) were recruited and were paid 10€/h for their participation in Experiment 1. Four participants were excluded due to loss of EEG data for technical reasons. Thirteen (11 female, M = 22.17 years, SD = 2.33) and eighteen (16 female, M = 23.05 years, SD = 6.55) participants were recruited and paid 5€/hour for their participation in the follow up Experiment 2 and 3, respectively. All participants signed informed consent, approved by the University of Barcelona Ethics Committee.

6.3.2. Stimuli

Experimental stimuli consisted of 160 photographs of household objects and 80 episodic sequences, each formed by 4 photographs. There were 80 different household objects included, each had two slightly different versions, for a total of 160 photographs. Episodic sequences consisted of 4 snapshots in temporal order depicting a person moving around and interacting with the surroundings in different house rooms during a short interval of time. Each sequential episode was designed to match with one of the 80 household objects. The object images were taken from the Stark lab set of stimuli, freely available at (http://faculty.sites.uci.edu/starklab/mnemonic-similarity-task-mst/) The pictures of the episodic sequences did not actually contain the matching household object. Instead, the sequences were designed in a way that the matching object could fit in or make sense with the given sequence. In other words, the matching object could be expected to be encountered in the situation depicted in the episodic sequence. Each episodic sequence was designed to be congruent with its specific corresponding (congruent) object.

6.3.3. Experimental design

The experiment design of the three studies consisted of an encoding and a test phase. The encoding phase was the same in all three studies. Participants were presented in trial with an episodic sequence followed by a picture object. Participants were asked to rate the degree of congruence on a scale from 1 (i.e., does not fit in) to 4 (i.e., fits in very well) of the target picture object in relation to the context formed by the succession of the 4 preceding episodic sequence images (Figure 6.1a). During the encoding phase included a total of 80 trials, each consisting of an episodic sequence followed by an object. Two versions of the encoding phase were constructed so that 40 of the episodic sequence – picture objects could be perceived as high congruent by the participant. This yielded a total of an a priori possible 40 high and 40 low congruent sequence – object pairing. The order of the trials at encoding was randomised for each participant.

Each trial started with the appearance of a fixation cross on the screen for a random duration of 2000 to 4000 ms. Afterwards, an episodic sequence consisting of four photographs was presented. Each of the four photographs was presented on a white background for 2000 ms, one at a time in temporal order, separated by the presentation of a fixation cross for 500 ms. After the episodic sequence was presented, a fixation cross appeared on the screen for 2000 to 3000 ms, separating the episodic sequence from the presentation of the following object. The picture of the object was presented on a white background for 500 ms, followed by the appearance of a fixation cross for 2 seconds. Finally, a screen was presented with the word ‘Congruence?’ and the digits ‘1-2-3-4’ below, upon which participants had to indicate, within a maximum of 4 seconds, the degree of congruence between the object and the just encoded episodic sequence by pressing 1,2,3 or 4 in the keyboard. Participants were previously instructed to respond thoughtfully as fast and accurate as possible. As soon as they responded, a fixation cross was presented for 500 ms, and the next trial began. The encoding phase lasted around 30 min. After the 80 episodic sequences and objects were presented, the encoding phase was finished.

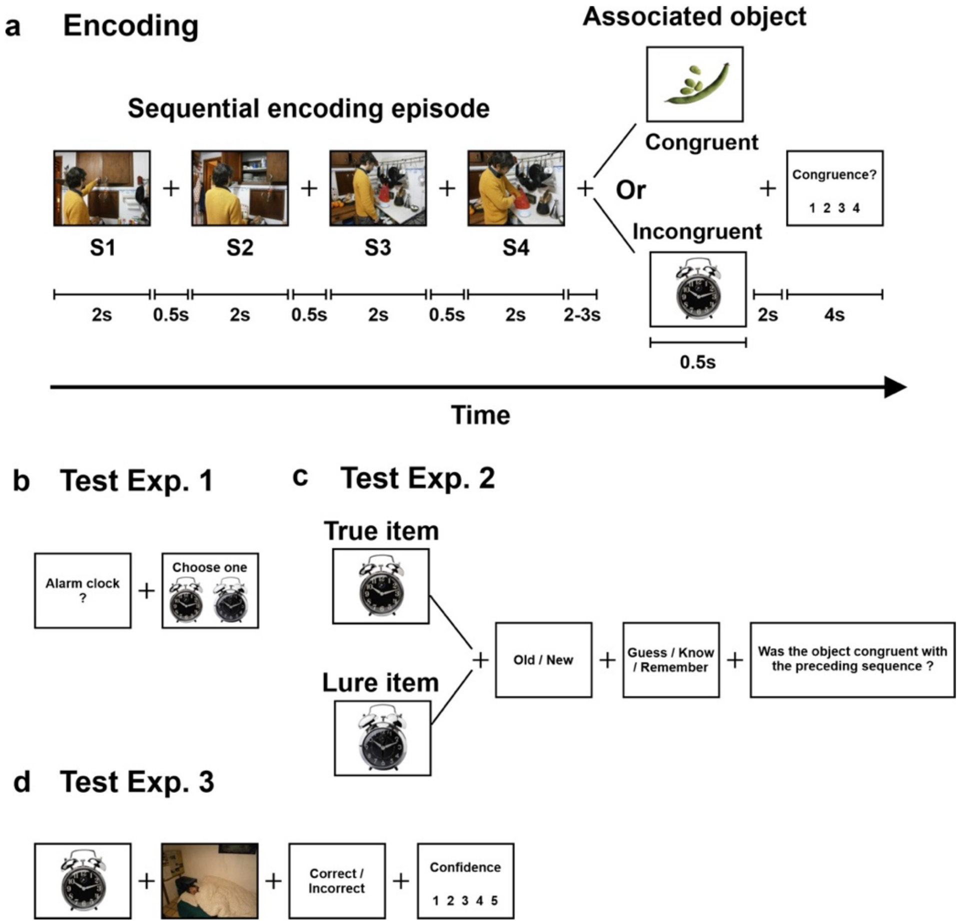

Figure 6.1. Encoding and recognition memory task design for Experiments 1, 2 and 3. **(a)** During the encoding phase for all experiments, participants encoded episodic sequences composed of 4 photographs depicting a routine domestic episode. These were followed by highly congruent or low congruent/incongruent object pictures. Participants indicated the degree of congruence between the episode and the object. **(b)** In Experiment 1, memory for object pictures was tested by the object word label followed by a true and a lure item of the same object. Participants had to indicate the correct picture presented during the encoding phase. **(c)** In Experiment 2, memory for the encoded objects was assessed by requesting participants to discriminate whether true or lure items corresponded to those presented during the encoding phase. This was followed by a 'guess/know/remember' task and ended with asking the participants to indicate whether the picture object was encoded within a congruent episodic event. **(d)** In Experiment 3, each object picture presented during the encoding phase was displayed together with one image from an episodic sequence. Participants were required to indicate whether the object and episodic image picture corresponded with the episodic + object picture presented in the same trial during the encoding phase.

The encoding phase was followed by a ~ 10 min interference task consisting of choosing the correct answer to simple additions and subtractions that appeared on the screen. Participants were told to respond as soon as possible, although no time limit was imposed. The distraction task ensured the participants would not rehearse the pictures they had previously seen.

6.3.4. Recognition memory test

After a break of ~10 minutes, a recognition memory test was presented unexpectedly to the participants in the three studies.

In Experiment 1, the test included 160 object words and 160 object picture pairs of each word. Eighty words and object pictures corresponded to previously presented objects in the encoding phase (‘Old’), whereas the other 80 were non-related (‘New’) objects (Figure 6.1b). Each picture pair depicted the same object but with small changes in specific features between each other (e.g., orientation, colour, etc.). Each trial began with a fixation cross lasting from 3 to 5 seconds at random. Subsequently, each word was presented for a maximum of 6 seconds with a question mark below ‘?’. Participants were instructed to press ‘1’ on the keyboard if they considered the word referred to an object presented during encoding (Old) or ‘2’ if not (New). If the participant responded ‘Old’ to a word, then two pictures of the object word were presented on the screen for a maximum of 6 seconds (Figure 6.1b). Pictures were presented on the computer screen, with one item to the left and one to the right of fixation. The left/right assignment was randomly chosen on each trial. The two pictures from each pair were almost identical, but only one corresponded to the exact one presented in the encoding phase (true), whereas the other one served as a lure. For example, if the participant had seen the photograph of a hammer and later, during the test, correctly identified the word ‘hammer’ as one of the objects she saw, then two photographs of similar-looking hammers appeared. The participant was instructed to identify the picture object exactly like the one presented in the encoding phase by pressing 1 if it was the left photograph or 2 if it was the right one. Picture pair presented when a participant misclassified as ‘Old’ a new word was almost identical, though none of the two pictures had been seen in the encoding phase. The order of the presentation of word + picture pairs in the test was randomised before each participant started the test. The recognition memory test lasted ~20 min.

In Experiment 2, memory for the encoded objects was assessed by requesting participants to recognise them throughout a set of pictures presented randomly in the test phase. The test included 40 target objects previously presented at encoding (True items) and 40 objects highly similar to the target objects presented at encoding but with some changes in their specific features (e.g., orientation, colour, etc.) (Lure items) (Figure 6.1c). In total, the recognition memory test was comprised of 80 items. 40 of the picture objects (20 True and 20 Lure) related to pictures encoded in high congruent trials, and the other 40 picture objects (20 True and 20 Lure) related to pictures encoded in low congruent trials at the encoding phase. Two versions of the test phase counterbalancing True/Lure and high/low congruency conditions were prepared and assigned randomly to the participant's sample. Each trial began with a fixation cross lasting from 3 to 5 seconds at random. Then, one object picture remained on the screen for a maximum of 8 seconds with the question 'Did this object appear before?', and participants had to indicate on a keyboard whether the same item was presented during encoding ('1' – Old and '2' – New). Participants were told only items the same as items presented during encoding were correct answers. 'Old' responses were followed first by a 'guess/know/remember' judgment of the picture and later by a question referring to the semantic context: 'Was this object encountered in a high/low congruent context?'. Participants had a maximum of 8 seconds to respond to each question.

In the test phase of Experiment 3, the 80 object pictures presented during the encoding phase were included in the test. In each trial, each object picture was presented together with one image from each of the episodic sequences presented at the encoding phase (Figure 6.1d). Participants were requested to indicate whether the object picture and episodic image picture matched the episodic sequence and object picture presented during the encoding phase. Half of the object+episode picture pairs presented in the test matched the encoding ones, whereas the other was randomly paired with each other. The total set of 80 picture pairs was constructed so that 40 old ones included 20 object+episode pairs encoded in the high congruency condition and 20 in the low congruency condition. The same distribution pattern was used to construct the set of object+episode new picture pairings (i.e., those that do not match the trials presented at encoding). Two different versions of 40 old and 40 new sets of picture pairings were created by controlling that in one version, a picture object was paired with a matched image from the encoded episodic sequence (Old) and to an unmatched image (New) in the other. The two versions were assigned randomly to the participant's sample. Each trial began with a fixation cross lasting from 3 to 5 seconds at random. Then, the object and the episodic sequence picture were presented on the screen. Participants were instructed to answer whether both photographs had been presented together in the same trial during the encoding phase (by pressing '1' on the keyboard) or not (by pressing '2'). Participants were asked to rate their confidence in their previous response from 1 ('no confidence') to 5 ('absolute confidence') with the same numbers on the keyboard. Object and episodic pictures were presented on the computer screen, with one item to the left and one to the right of fixation. The left/right assignment was randomly chosen on each trial.

6.3.5. Behavioural data analyses

Paired Student t-test was used to compare participants' performance (measured in percentage) between conditions. Repeated measures ANOVA was used to statistically assess differences between participants' performance when they included more than two variables. Statistical significance threshold was set at $p < 0.05$.

6.3.6. EEG recording and preprocessing

In study 1, EEG was recorded using a 32-channel system at a sampling rate of 500 Hz, with an online band-pass filter set at 0.01-100 Hz, using a BrainAmp amplifier and tin electrodes mounted in an electrocap (Electro-Cap International) located at 29 standard positions (Fp1/2, Fz, F7/8, F3/4, Fcz, Fc1/2, Fc5/6, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T5/6, Po1/2, Oz) and at the left and right mastoids. An electrode placed at the lateral outer canthus of the right eye served as an online reference. EEG was re-referenced offline to the average of all channels. Vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. Electrode impedances were kept below 5 kΩ. EEG was low-pass filtered offline at 30 Hz. We applied the Parks-McClellan Notch filter using the toolbox ERPLAB (http://erpinfo.org/erplab).

For each participant, we extracted the EEG epochs for each encoding image. Epochs had a duration of 2000 ms for images from the episodic sequence and 2500 ms for images of the picture object and were baseline corrected to the pre-stimulus interval (-100 to 0 ms). Epochs with maximum absolute amplitude over 100 µV were discarded for further analysis. For later analysis, all the epochs were smoothed by averaging data via a moving window of 100 ms (excluding the baseline period) and then downsampled by a factor of 5.

6.3.7. Neural stability analysis

To account for whether the ongoing encoding of an episodic sequence of pictures elicited a gradual increase in stable brain activity patterns, we implemented a temporally resolved similarity analysis using Pearson correlation coefficients, which are insensitive to the absolute amplitude and variance of the EEG response. The correlation analysis on EEG data was made at the individual level and to each time point separately and included spatial (i.e., scalp voltages from all the 29 electrodes) features of the resulting EEG single trials. To examine how a schema consistent sequence's unfolding modulated the stabilisation of activity patterns, we correlated the EEG patterns of activity elicited by the 1st and the 2nd pictures and compared them to the correlation obtained by the 3rd and the 4th picture of the episodic sequence. This analysis was computed at the trial level by randomly creating pairs of correlation analysis from pictures from different episodic sequences. For each participant, we created 200 permutations of possible unique pairings, and the final cross-correlation output resulted from averaging the point-to-point Fisher's *z* scores correlation values across the 200 permutations.

A similar analysis was performed on EEG patterns elicited by picture objects. For each participant, we randomly created 200 sets of unique picture pairs that the participant rated as High congruent (i.e., rated as 1 or 2 during the encoding phase) or Low congruent (i.e., rated as 3 or 4 during the encoding phase).

We implemented a cluster-based permutation test to account for neural stability differences between picture order within the episodic sequence and between High and Low congruence conditions (Maris and Oostenveld, 2007). It identifies clusters of significant points in the resulting 1D matrix in a data-driven manner and addresses the multiple-comparison problem by using a nonparametric statistical method based on cluster-level randomisation testing to control the family-wise error rate. Statistics were computed for every time point, and the time points whose statistical values were larger than a threshold ($p < 0.05$, two-tail) were selected and clustered into connected sets based on adjacency points in the 1D matrix. The observed cluster-level statistics were calculated by taking the sum of the statistical values within a cluster. Then, condition labels were permuted 1000 times to simulate the null hypothesis, and the maximum cluster statistic was chosen to construct a distribution of the cluster-level statistics under the null hypothesis. The nonparametric statistical test was obtained by calculating the proportion of randomised test statistics that exceeded the observed cluster-level statistics.

6.3.8. Representational Similarity Analysis (RSA)

RSA was performed timepoint-to-timepoint at trial level and upon spatial features (i.e., scalp voltages from all the 29 electrodes) (Silva et al., 2017; Wu et al., 2021). RSA was conducted between the EEG signal of each encoding image (i.e., the image at 1st, 2nd, 3rd, 4th position in a sequence) with the EEG signal of the corresponding offset period (i.e., the matching object with the sequence and the followed fixation cross). Point-to-point Pearson correlation values were then calculated, resulting in a 2D similarity matrix with the size of 200×250 (each time point represents 10 ms, given the 500 Hz EEG recording sampling rate and a down-sampling factor of 5). The *x*-axis of the matrix represented the object and offset time points, and the *y*-axis represented the time points of sequence picture encoding. The output 2D matrix represents the overall degree of neural pattern similarity between EEG elicited for each pair of encoding image and its corresponding sequence offset.

6.3.9. Linear-mixed effect model

To further explore on the trial base how the neural pattern was associated with the subjective feeling of the congruence and its impact on subsequent memory of the object image, we applied a Linear Mixed Effect Model (LMM) on ERPs elicited during the offset period as well as the pattern similarity between encoding and offset.

For offset ERPs, we first group 29 electrodes into 6 scalp regions (Figure 6.4a). To obtain more stable spatial patterns, border electrodes between regions were included in each neighbouring region (Lu et al., 2015; Sols et al., 2017). As a result, the 6 regions were defined as the following: region 1 (Fp1, Fz, F3, F7, Fcz, Fc1, Fc5); region 2 (Fp2, Fz, F4, Fcz, Fc2, Fc6); region 3 (Fcz, Fc1, Fc5, Cz, C3, T3, Cp1, Cp5); region 4 (Fcz, Fc2, Fc6, Cz, C4, T4, Cp2, Cp6); region 5 (Cp1, Cp5, Pz, P3, T5, Po1, Oz); region 6 (Cp2, Cp6, Pz, P4, T6, Po2, Oz). The LMM was constructed separately for each region. ERPs voltages at each offset time point were averaged across electrodes within the region and then introduced in our LMM as the dependent variable. The rating of the object's congruence to the sequences (on a scale of 1 to 4) and the memory for the object (whether the word and the image were correctly recognised during the test) were included in the model as fixed effect variables. Then, subject was introduced into the model as the grouping variable, with random intercept and a fixed slope for each fixed effect variable. The statistical significance was adjusted by Bonferroni correction for each fixed effect variable taking into account both the number of regions and the number of time points, resulting in a corrected alpha level of $\alpha = 3.33\times10^{-5}$ (0.05/(6*250)).

We also applied LMM on output from RSA. We identified on the resulting 2D similarity matrix the time point of encoding and offset where the pattern similarity reached the peak value. We then averaged at a single-trial level the similarity value across ±50 ms (11 data points) around the peak encoding time point for each offset time point, resulting in a 1D similarity value with the length of 250 time points, covering the whole offset period. For each time point, we constructed the LMM with similarity value as the dependent variable. Then 3 fixed effect variables were introduced into the model: the image position in the sequence (1st, 2nd, 3rd, and 4th), congruence rating and memory for the object. Subject was included in the model as the grouping variable, with random intercept and a fixed slope for each fixed effect variable. The statistical significance for each fixed effect variable was Bonferroni corrected with a thresholded alpha level of $\alpha = 2\times10^{-4}$ (0.05/250).

6.4. Results

6.4.1. Behavioural results from Experiment 1

In Experiment 1, the proportion of trials rated as highly congruent ('1' or '2') was 45.17% (SD = 8.59%) and 15.43% (SD = 6.42%), respectively, and lowly congruent ('3' or '4'), respectively, was 13.45% (SD = 6.36%) and 25.39% (SD = 7.47%) with the episode.

In general, participants were accurate in recognising words referring to picture objects learnt during the encoding phase (Mean = 71.09%, SD = 12.80%; above chance: $t_{(28)} = 8.41$, $p < 0.001$). They showed greater accuracy for words related to picture objects encoded with high (Mean = 76.07%, SD = 13.27%) than low (Mean = 66.11%, SD = 13.89%) congruency to the episodic context during the encoding phase (paired *t*-test: $t_{(28)} = 5.89$, $p < 0.001$) (Figure 6.2a). For those correctly recognised words, in the next picture recognition test, participants were accurate in correctly identifying the exact object picture presented in the encoding phase (Mean = 73.13%, SD = 8.02%; above chance: $t_{(28)} = 15.53$, $p < 0.001$). They performed similarly for pictures rated as high (Mean = 74.61%, SD = 11.63%) and low (Mean = 72.61%, SD = 8.61%) congruency to the episodic context (paired t-test: $t_{(28)} = 0.39$, $p = 0.878$) (Figure 6.2b).

To specify the subsequent memory strength for objects, for all later analyses, we classified the memory performance of a trial based on whether the object image was correctly recognised during the test. Therefore, successful recognition of both gist word and object image was considered as a 'remembered item' condition, while either failing in recognition of gist word or image was considered as a 'forgotten item' condition. This separation also rendered a balanced percentage of trials across conditions, with 53.53% (SD = 11.43%) for 'remembered' condition and 46.47% (SD = 11.43%) for 'forgotten' condition (Wilcoxon signed-rank test: $z = 1.74$, $p = 0.08$).

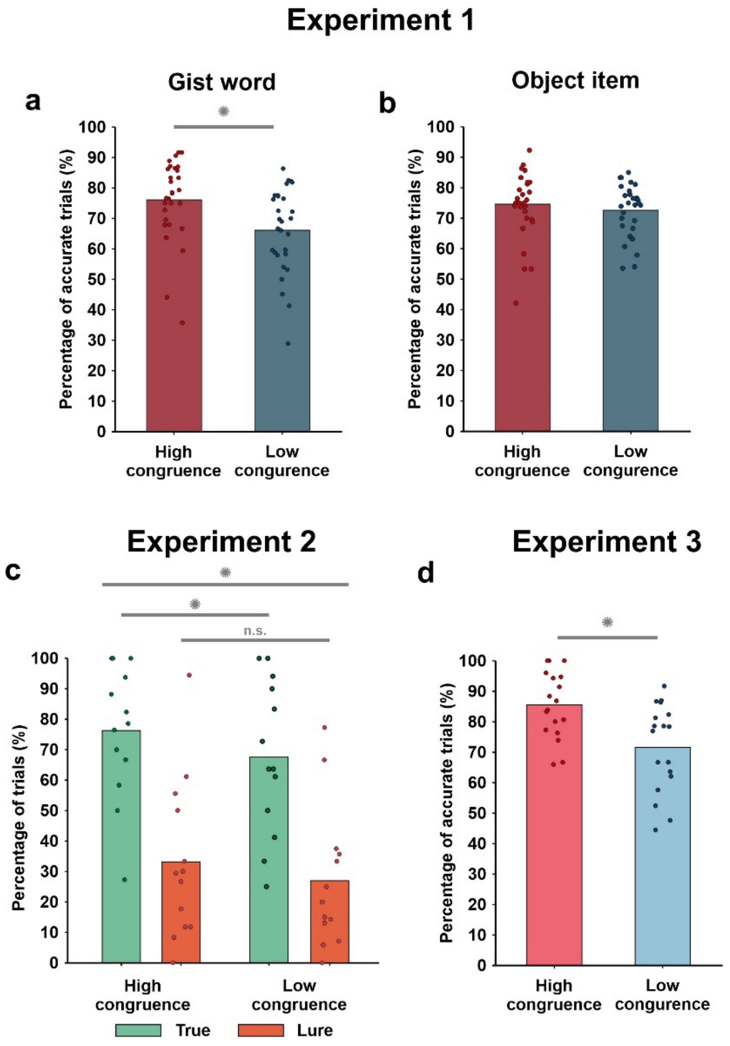

Figure 6.2. Behavioural results of Experiment 1, 2 and 3. **(a)** Participants' test accuracy on recognising the gist word separated by whether the picture object was perceived by the participant as highly or lowly congruent with the episodic sequence during encoding in Experiment 1. **(b)** Participants' test accuracy on identifying the exact object picture presented in the encoding phase after correctly recognising the 'Old' gist word. The accuracy was separated by whether the picture object was perceived by the participant as high or low congruent with the episodic sequence during encoding in Experiment 1. **(c)** Participants' memory accuracy (in percentage) at the recognition test of Experiment 2 for True (hit) and Lure (correct rejection) as a function of whether the picture object was perceived by the participant as high or low congruent with the episodic sequence during encoding. **(d)** Participants' memory accuracy (in percentage) in the recognition memory test of Experiment 3. In all plots, dots represent values for an individual subject. * indicates $p < 0.05$. 'n.s.' indicates $p > 0.05$.

6.4.2. Behavioural results from the Experiments 2 and 3

The results of Experiment 1 revealed that memories of objects that were perceived highly congruent with the episodic context were later better remembered. However, the results of this study relayed on a test that changed the perceptual format at test, that is, objects were encoded as a picture and tested as a word at test, which lacks the visual details of the picture object and engenders semantic processing. A concern of this change could be that memory congruency benefits were to some extent explained by differential processes taking place during retrieval. Information that is highly congruent with prior knowledge is often found to be better remembered than low congruent information, putatively because of an increase in semantic associations and relational integration (Staresina et al., 2009; Atienza et al., 2011; van Kesteren et al., 2014; Bein et al., 2015; also see Craik and Tulving, 1975). Experiment 2 addressed this concern by testing object memories in the same visual format depicted at the encoding phase.

In Experiment 2, the proportion of trials rated as high ('4') and low congruent ('1') were 50.10% (SD = 10.58%) and 42.21% (SD = 4.98%), respectively, and the average proportion of intermediate levels of congruency was very low ('2': Mean = 2.69%, SD = 3.42%; and '3': Mean = 5%, SD = 5.68%). A comparison of the proportion of trials rated as high congruent ('4') and low congruent ('1') revealed they were not significantly different ($t_{(12)} = 2.03$, $p = 0.06$). Given the low proportion of trials rated by the participants with intermediate level of congruence (i.e., < 10% on average), and to ensure a proper orthogonalization of possible memory effects at the test driven by encoding congruency, we included in the subsequent analyses trials rated as high ('4') or low congruent ('1') by the participants.

In general, participants were highly accurate in recognising True Old object pictures that were encoded on a highly congruent (Mean = 0.99, SD = 0.01) and on a lowly congruent episodic context (Mean = 1, SD = 0). They also showed to be prone to misclassify Lure items as Old in the two encoding conditions (High congruent: Mean: 0.48, SD = 0.23; Low congruent: Mean = 0.47, SD = 0.22). To investigate whether retrieval accuracy differed as a function of encoding congruency, we applied a repeated measures ANOVA, including picture type (True vs. Lure), congruence (High vs Low congruent) and subjective retrieval quality (remember vs. know/guess), as a within subject factors. The results confirmed that participants were more accurate in correctly recognising True items than rejecting Lure items as Old (main effect of picture type: $F_{(1,12)} = 81.76$, $p < 0.001$) but, as expected, participants' differed in their accuracy for True and Lure items as a function of whether the retrieval was catalogued as ‘remember’ or ‘guess/know’ (main effect of remember: $F_{(1,12)} = 7.88$, $p = 0.016$ and an interaction picture type x remember: ($F_{(1,12)} = 16.97$, $p = 0.001$). However, we did not find a main effect of congruence ($F_{(1,12)} = 0.35$, $p = 0.85$) or a congruence x item type interaction ($F_{(1,12)} = 0.16$, $p = 0.67$), but found a significant retrieval quality x congruence interaction ($F_{(1,12)} = 6.95$, $p = 0.02$), which suggested that congruence at encoding may affect retrieval accuracy only for when participants were capable to retrieve object information as a function of the ability to retrieve their context (Figure 6.2c). To unpack this finding, we ran separate repeated measures ANOVA, including item type and congruence as within subject factors, for participants’ accuracy as a function subjective retrieval quality. The results indicated that participants were more accurate in correctly identifying True items than in misclassifying as Old the Lure items in the remember condition (main effect of item type: $F_{(1,12)} = 103.22$, $p < 0.001$) and that participants’ accuracy differed for high and low congruent encoded items (main effect of congruence: $F_{(1,12)} = 9.52$, $p = 0.009$; congruence x item type interaction: $F_{(1,12)} = 0.44$, $p = 0.52$). A post-hoc paired student t-test indicated higher accuracy for high than low congruent True items ($t_{(12)} = 2.65$, $p = 0.02$) and a trend towards statistical significance for Lure items ($t_{(12)} = 2.07$, $p = 0.06$). The ANOVA on response accuracy for ‘guess/know’ condition showed a similar trend directions in the effects, though none of the effects reached statistical significance (main effect of item type: $F_{(1,12)} = 4.17$, $p = 0.06$; main effect of congruency: $F_{(1,12)} = 3.59$, $p = 0.08$; item type x congruency interaction: $F_{(1,12)} = 0.78$, $p = 0.39$). Collectively, the results of Experiment 2 replicated the effects observed in Experiment 1, by showing that retrieval accuracy was higher for items that were preceded by a high than a low congruent episodic context with the object. However, they also reveal that items encoded in a high congruent context are more prone to errors at retrieval, suggesting the possibility that the benefit of congruence may also come with a decreased detailed item representation in long-term memory.

We next sought to examine whether the retrieval benefits of encoding congruence observed in Experiments 1 and 2 were also accompanied by an increased degree of binding of the object to the preceding episodic context. Thus, in Experiment 3, participants were required to judge whether pairs of pictures, including one of the encoded objects and one picture from the sequence episodes, were seen together or not during encoding. As in Experiment 1 and 2, the proportion of trials rated as high

(‘4’) and low congruent (‘1’) were 35.39% (SD = 12.63%) and 37.67% (SD = 10.06%), respectively, and the average proportion of intermediate levels of congruency was much lower (‘2’: Mean = 12.92%, SD = 9.97%; and ‘3’: Mean = 10.81%, SD = 5.68%). A comparison of the proportion of trials rated as high congruent (‘1’) and low congruent (‘4’) revealed they were not statistically different ($t_{(17)}$ = -0.85, p = 0.41). Therefore, as in Experiment 2, we analysed memory accuracy for items classified as ‘1’ or ‘4’ at encoding. The results revealed that participants were accurate in correctly identifying the encoded correspondence between objects and the episodic context (Mean = 0.79, SD = 0.12), but that the accuracy was higher (paired student *t*-test, $t_{(17)}$ = 5.59, $p <$ 0.001) and more confident ($t_{(17)}$ = 6.78, $p <$ 0.001) when the association between the object and the context at encoding was high rather than low congruent (Figure 6.2d).

6.4.3. Neural stability and episodic congruence

We first examined whether EEG states of stability gradually increased throughout the encoding of a schema consistent images of the narrative sequence. Our hypothesis was that schema consistency for picture sequences would promote neural stability, and that this would be observed as a gradual increase in across item EEG pattern similarity (Figure 6.3a) of the EEG patterns elicited throughout the picture sequence. To address this issue, we calculated a time-resolved neural similarity analysis between EEG patterns elicited by initial (i.e., 1st and 2nd) and final (3rd and 4th) picture items across the sequences. Confirming our hypothesis, we found that the 3rd and the 4th picture elicited more stable patterns of EEG activity from 1030 ms to 2500 ms from picture onset (Cluster statistics: p = 0.009, corrected for multiple comparisons, sum T-values = 259.78) (Figure 6.3b).

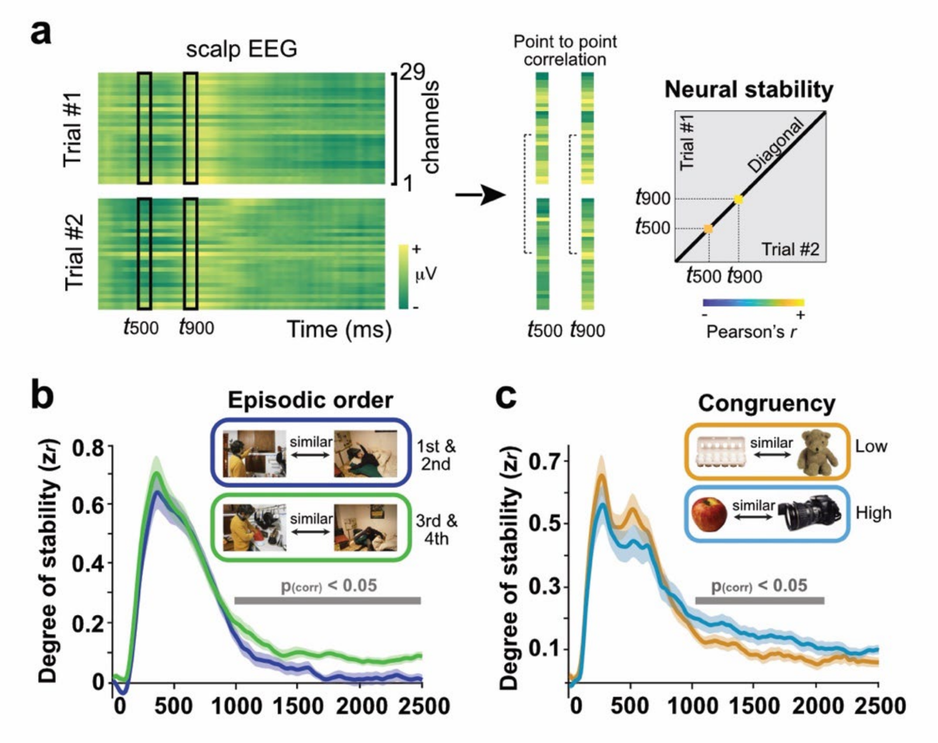

Figure 6.3. Neural stability and congruence. **(a)** Schematic representation of the analysis. A temporal cross-stimuli correlation matrix is generated from the EEG data for each participant. **(b)** Point-to-point participants' average degree of neural stability between 1st and 2nd and 3rd and 4th picture across different episodic sequences. **(c)** Point-to-point participants' average degree of neural stability across object pictures rated lowly and highly congruent with the preceding episodic event at encoding. The shaded area represents standard error (SEM) across subjects. The thick grey line depicts the timing of the significant cluster between conditions ($p < 0.05$, cluster-based permutation test).

Next, we evaluated whether EEG states of stability were associated with the encoding of object pictures that were high or low congruent to the preceding episode. We hypothesised that congruency induced higher states of stability and that this elevated state of neural stability would be reflected as an increase in neural similarity upon object encoding, rendering them more similar than objects perceived as lowly congruent with the preceding episodic context. To address this issue, we compared the cross-temporal correlation analysis to EEG activity patterns from different pictures within each condition. We reasoned that if congruency induced patterns of neural stability, this should be reflected as a general persistent pattern of activity during encoding, independently of the depicted object. Please, note that the assignment of episode-object associations was counterbalanced between participants, and therefore, an object encoded highly congruently with an episode in one participant was encoded lowly congruently by another participant. The results of this analysis showed that high congruent items elicited more stable patterns of EEG activity than low congruent items from 1036 ms to 2008 ms from picture onset (Cluster statistics: *p* = 0.02, corrected for multiple comparisons, sum *T*-values = 452.90) (Figure 6.3c). These results indicate that context-dependent neural state of stability is modulated by episodic congruency at encoding.

6.4.4. N400 signals the detection of incongruent objects within an episodic context

We hypothesised that the disruption of neural stability elicited by the detection of low congruent objects from the unfolding schema consistent episodic narrative would be preceded by a prediction error signal in the brain. We aimed to identify such an error signal in the EEG by means of a transient increase in the N400 ERP component, which has been widely related to incongruence detection in the literature (Kutas and Federmeier, 2011). To assess this issue in our data, we grouped the 29 electrodes into 6 regions and averaged the epochs across electrodes within each region. Then, we introduced the voltage value into the LMM as the dependent variable and included participants' ability to correctly recognise the picture image at test congruence rating provided at encoding as the main fixed-effect variables. Subject was included as the grouping variable. This analysis was conducted for each time point and each scalp region separately.

The results of the LMM analysis revealed significant effects at specific scalp regions and temporal points for both memory and congruence. For memory, later forgotten objects elicited significantly more negative ERPs amplitude at 520 ms to 870 ms from picture onset distributed over the frontal scalp region. For congruence, we found more substantial negative amplitude with objects rated less congruent with the preceding episodic context. This effect resulted significant from 410 ms to 730 ms from picture onset and was distributed over frontocentral scalp regions (Figure 6.4a).

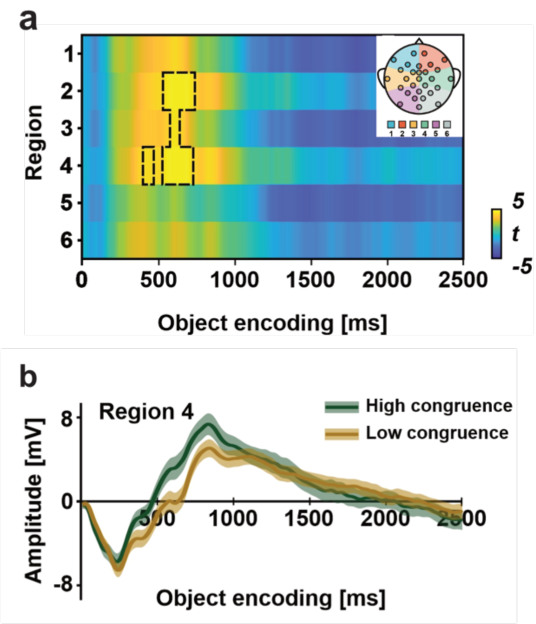

Figure 6.4. LMM on ERPs elicited by object pictures. 29 electrodes were grouped into 6 scalp regions, and boundary electrodes were included in either neighbouring region to obtain more stable spatial patterns. **(a)** *t*-value resulting from the LMM analysis at each region and time point as a function of congruence. **(b)** Participants' averaged ERPs from representative region 4 (note that for visual illustration, trials with congruence ratings of 1 and 2 were grouped and averaged as Low congruence and trials with congruence ratings of 3 and 4 were grouped as High congruence). The shaded area represents standard error (SEM) across subjects. A black dash line on the statistical map marks the area where the *t*-statistics exceed the significance threshold ($p < 0.05$) with the alpha level adjusted with Bonferroni correction.

6.4.5. Neural reinstatement induced by object incongruence

We next sought to test that prediction error elicited by low congruent items would be accompanied by more robust reactivation of the just encoded episodic elements. To address this issue, we implemented a trial-based and temporally-resolved neural similarity analysis between EEG data elicited at picture object and EEG data elicited by each image of the preceding episodic context. The results of this approach revealed an increase in neural similarity between EEG patterns from ~100 ms – 700 ms at object picture onset and EEG patterns of activity between ~100 ms – 750 ms from the onset of pictures within the episodic sequence (Figure 6.5a).

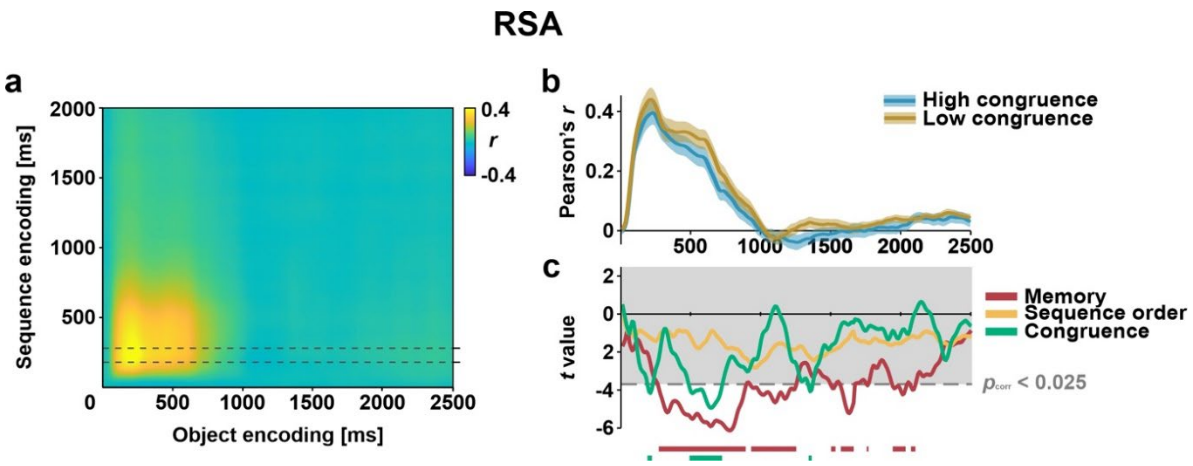

Figure 6.5. Neural similarity between episode and object. **(a)** Time-resolved degree of neural similarity between EEG patterns elicited by picture images from the episodic sequence and EEG patterns elicited during the encoding of the associated picture object. Grey dash line showed the ±50 ms time window where similarity reached the peak (190 ms - 290 ms at encoding time). **(b)** Neural similarity values for the High and Low congruence conditions averaged across participants. Shaded area represents SEM of the participants' sample. **(c)** *t*-statistics from the output of LMM on similarity value along picture object encoding period. Grey dash line marked the one-sided significant threshold after Bonferroni correction. Grey shade showed the area with $p_{\text{corrected}} > 0.05$ (two-tailed).

To assess whether the degree of neural reactivation observed during the encoding of picture objects was modulated by their perceived congruency with the encoding episodic context and their later memory accessibility at test, we ran an LMM analysis including neural similarity, participant congruence rating and recognition accuracy at test. We also included a variable accounting for the picture order of pictures in the episodic sequence to control for the possibility that any effect could be accounted for by the temporal proximity of the encoding pictures of the preceding episode with the picture object. To do so, we first identified on the similarity matrix the exact time point where the similarity value reached the peak across participants (i.e., at 240 ms from episodic sequence picture onset and at 220 ms from object picture onset). We then ran a single trial LMM analysis on the averaged similarity value ±50 ms around the peak (Figure 6.5b). We found that the degree of neural reactivation correlated negatively with the participants' ability to correctly recognise an item at test. The predictive negative relationship between neural reactivation and memory started to be significant at 270 ms from picture object onset and remained significant throughout almost the entire epoch until 2100 ms from the object picture onset (Figure 6.5c). In addition, we also found a negative correlation between neural similarity and congruence rating. However, the significant effects were more transient but comparably distributed along with the object picture encoding epoch. Notably, the first significant timepoint was very early, at 190 ms from picture onset, which preceded the relationship effects between neural similarity and memory (Figure 6.5c). A more persistent negative correlation was also found later, between 490 ms – 720 ms and at 1340 ms from picture onset. Finally, no significant point exceeded the statistical significance threshold for the variable picture order position, indicating that the significant relationship between neural similarity and memory and congruence was not driven by specific neural similarity measures between picture objects and pictures of the episode.

6.5. Discussion

In this study, we tested whether the perceived incongruence of an item from an unfolding event and its impact on memory relied on the disruption of neural stability patterns preceded by the neural reactivation of the memory representations of the just encoded event. Our findings, derived from combining behavioural data from 3 different experiments and the implementation of multivariate pattern analysis on EEG signal during encoding of one of them, confirmed our hypothesis by showing that neural stability patterns gradually increase throughout the ongoing exposure to a schema-consistent episodic narrative, and that the brain stability pattern is interrupted when the encoding of an object of the event is lowly congruent within the ongoing schema representation. We found that the decrease in neural stability for low congruent items was best seen at ~1000 ms from object encoding onset when compared to high congruent items and that this effect was preceded by an enhanced N400 ERP and an increased degree of neural reactivation of the just encoded episode for low congruent items observed between ~200 to 1000 ms from picture onset. Current results offered new insights into the neural mechanisms and their temporal orchestration that are rapidly engaged during online encoding of schema-consistent episodic narratives and the detection of incongruencies.

Central in our findings is that the degree of neural reactivation of the encoded episode by the final target object correlated negatively with the perceived congruence and the participant's ability to later recognise the picture in a memory test. The notion that memory reactivation benefits memory formation is well established in previous research. Most of it showed that the reactivation strength drives long-term memory formation by mimicking neural replay phenomena thought to promote rapid consolidation processes seen in rodent studies (i.e., Carr et al., 2011). Other studies have revealed that when novel encoding inputs reactivate previously encoded information that overlaps in content, the long-term memory representations of the two event contents become integrated, promoting generalisation and adaptive behaviour (Shohamy and Wagner, 2008; Schlichting and Preston, 2015). Though the notion that memory reactivation may potentially benefit memory formation, another set of findings described the opposite effect. These studies found that memory reactivation of overlapping memories may yield competition between the two, resulting in interference with a negative impact on memory (Kuhl et al., 2011). Our results that object incongruence elicited greater memory reactivation of the preceding event and that the degree of memory reactivation correlated negatively with memory accuracy for the object picture may align well with the latter view.

An additional possible explanation for the proactive interference effect found between memory reactivation and incongruence detection is that surprise itself creates an event boundary (Antony et al., 2021), sectioning off the preceding and the current elements as distinct events in memory. In fact, theoretical models propose that mnemonic prediction errors would promote the encoding of distinct memory traces (McClelland et al., 1995; Love et al., 2004; Gershman et al., 2014; Frank et al., 2020). That is, events that violate our expectations should be allocated a unique memory representation distinct from other existing memories. This may facilitate memory for the unexpected event while mitigating interference with existing memories that may still be relevant. Our findings that high and low congruent items were retrieved with similar accuracy when memory was tested with a detailed visual representation in Experiment 2 and in correctly rejecting lure items in Experiment 1 would fit to this idea, and it would also converge with recent behavioural findings that mnemonic prediction errors do not increase gist-based mistakes of identifying similar lures as old (Bein et al., 2021).

Our findings that surprising or unexpected elements of the unfolding experience elicited the rapid reinstatement of the just encoded picture sequence are in line with previous findings that showed that sudden shifts in an ongoing episodic context (i.e., event boundaries) induce the rapid reactivation of preceding episodic information (Sols et al., 2017; Silva et al., 2019; Wu et al., 2021). Event boundaries are thought to represent moments in time whereby a continuous stream of incoming information is segmented into different memory units (Zack et al., 2011). In this model, the process of event segmentation starts when your current understanding of the world is destabilised by a new observation that does not fit our current expectations. Viewed from this perspective, high levels of surprise (defined here as a high degree of inconsistency of the object picture to the schema or internal model representation activated during the preceding episodic sequence) refers to a substantial change in our understanding of the current inputs from experience. This engenders additional resources to re-evaluate the current internal model in the face of the new observation, which may benefit from the greater reactivation of the memory representations to resolve it.

In summary, the current study offers three novel findings. It shows that the detection of low congruent elements of an episodic experience elicited a rapid memory reactivation of the just encoded event episodic information, that this is concomitant to a mnemonic prediction error signal during encoding and that the result of this computation leverages the disruption of stable patterns of neural activity elicited during the schema consistent episodic event. These findings inform about the rapid but sequential structure of the distinct neural mechanisms supporting the detection of incongruencies during encoding and their consequences on memory. We speculate that these same processes may take place in realistic scenarios of our everyday experience.

Chapter 7. General discussion

7.1. Summary of the study results

The main objective of the three studies described in this thesis is to explore humans' immediate post-encoding memory reactivation mechanism and its functional role in episodic memory formation. The results of the three studies helped advance the understanding of this neural signature as an early stage of memory formation once an episode is perceived to end. In the followings, the main results of each study will be briefly outlined.

7.1.1. Study 1: Post-encoding reactivation promotes one-shot learning of episodes in humans

In Study 1, participants encoded sequences of pictures depicting unique and meaningful episodic-like events. We applied representational similarity analysis on scalp electroencephalography recordings at the encoding stage and found evidence of immediate post-encoding memory reactivation. The degree of memory reactivation predicted later recall for that episode. Importantly, memory reactivation was not observed between successive elements within an episode, indicating memory reactivation was specifically induced once participants perceived the episode to be complete. Finally, the memory reactivation mechanism was not present at the offset of sequences that were not perceived as meaningful episodes. Altogether, these results suggest that the post-encoding memory reactivation mechanism is akin to the rapid formation of unique and meaningful episodes that unfold over time.

7.1.2. Study 2: Post-encoding reactivation binds episodic sequences in long-term memory

In Study 2, we asked participants to encode trial-unique combinations of face-object-scene picture triplets that were subsequently recalled in a test. We first replicated the findings in Study 1. Memory reactivation of the just encoded episodic elements after encoding was found at the early stage of the offset period following sequence encoding (i.e., at 500 ms at offset period onset), and the magnitude of memory reactivation at the offset predicted whether the picture triplets would be successfully recalled as a whole. Using a pattern classifier approach, we found that the degree of category feature discriminability gradually decreased as a function of image order in sequences, indicating the gradual integration of categorical representation during encoding. However, the degree of category representation at the offset, but not during picture encoding, was lower for later recollected than forgotten triplet episodes. The increase in trial-based similarity measures and the decreased degree of category-based neural representations at the offset period for recollected episodes suggest that rapid offset-locked signals trigger a reactivation of a detailed representation of the elements of a just-encoded experience to be bound as a cohesive episodic memory at the long term.

7.1.3. Study 3: Contextual incongruency triggers memory reactivation and the disruption of neural stability

In Study 3, participants encoded images of target objects preceded by trial-unique sequences of events depicting day-to-day routine. In the first experiment, we found behavioural evidence supporting that object item congruent to the preceding episodic context was remembered more accurately. In the two follow-up behavioural experiments, we further confirmed that the increase in accuracy in recognition memory for the congruent object was also accompanied by a decreased detailed item representation, yet a stronger association between the object and context. At the neural level, we found that the processing of congruent objects was underlined by an increased state of neural stability. And importantly, compared to the incongruent items, such an increase was preceded by a less pronounced N400 ERP and a reduced neural reactivation of the preceding episode. Collectively, these results offered new insights into the neural mechanisms and their temporal orchestration that need to be rapidly engaged during online encoding of schema-consistent episodic narratives and the detection of incongruencies.

7.2. Does post-encoding neural reactivation reflect working memory maintenance?

Having shown the correlational relationship between post-encoding neural reactivation and later memory performance, a remaining question is whether the rapid post-encoding memory reactivation represents a carry-over effect from previous episode encoding or it contains a unique role in the memory formation. This question was first discussed in Study 1, where two main arguments were listed to favour the latter claim. First, the degree of memory reactivation predicted only the long-term memory strength of the episode but did not affect the coherence rating that immediately followed the offset period. Second, memory reactivation was present only at the early stage of the offset period within a delimited time window, while delay maintenance is associated with a sustained increase in activation of neocortical structures (Fuster and Alexander, 1971). Here, this debate could be extended with further insights provided by the results in Study 2 and 3.

With different experiment paradigms, Study 2 and 3 replicated the existence of rapid neural reactivation centred only at the early stage of the post-encoding period. However, a previous study with similar multivariate decoding approaches showed that working memory allowed transient information to guide upcoming behaviour by periodically reactivating the just-encoded representation throughout the entire maintenance interval (Fuentemilla et al., 2010). Such temporal nature of the putative mechanism supporting working memory was not observed in our results. Also, in our three studies, the degree of post-encoding neural reactivation was all linked directly to posterior memory performance, being either predictive of the memory strength for the preceding episode (Study 1 and 2) or the successfulness of integration of object item (Study 3) in the long term. This long-term impact of neural reactivation suggests a function beyond simple information maintenance.

On the other side, it has been recently suggested that persistent neural activity might neither be necessary to maintain representations held in working memory (D'Esposito and Postle, 2015). This claim is also compatible with our results that a distinctive neural pattern was not observed during the time window where working memory mechanisms could take place. For example, in Study 1, picture encoding neural pattern was not shown to re-emerge during transition points between pictures of an episode. Also, in Study 3, the neural reactivation was not present following the image sequence presentations where there were 2 – 3s of fixation cross presentation before the onset of the object image.

Putting all together, although our findings may not directly preclude the co-existence of a working memory mechanism during the offset period, it is plausible to claim that the post-encoding neural reactivation mechanism observed in our results functions beyond a sole maintenance process for preceding information, further it reflects a crucial step to organise the just-encoded experience for memory formation.

7.3. Is the post-encoding neural reactivation an event boundary effect?

Ever since the proposal of the Event Segmentation Model (Zacks et al., 2007), accumulating neuroimaging studies have provided consistent evidence for the existence of the neural mechanism triggered by event boundary and its impact on information processing (Baldassano et al., 2017; Brunec et al., 2018) as well as memory formation (DuBrow and Davachi, 2014; Liu et al., 2022). It has been shown to be a relevant framework to address how continuous stimuli can be chunked into clusters of information and transformed into event representation in memory. Considering the experimental paradigms adopted in our three studies, there are some aspects to be discussed: Is the end of sequence encoding in our studies (e.g., the offset fixation cross in Study 1 and 2; the final object presentation in Study 3) different from a classic event boundary? Is the post-encoding neural reactivation observed in our findings the exact mechanism as the neural activity triggered by event boundary reported in previous studies? (e.g., Sols et al., 2017). And finally, how do the findings linked to the 'end of event sequences' help advance our understanding of the human neural reactivation mechanism?

Previous studies exploring the event boundary effect mainly manipulated the encoding paradigm under the continuous unfolding temporal context (e.g., sequential presentation of images), and an event boundary can be induced by an abrupt change of perceptual stimulus (Sols et al. 2017); by switching the stimulus category and task type (DuBrow and Davachi, 2016); or relying on the event transitions points embedded in more naturalistic stimulus such as movies (Baldassano et al., 2017; Ben-Yakov et al., 2018). In Study 1 and 2, we adopted a similar sequence encoding paradigm. However, we did not specifically insert any transition point to break the continuity of event sequences. Instead, each encoding trial contained a clear mark of starting and end points previously known to the participants. In this case, unlike an event boundary that marks a change in event context for the upcoming stimuli, the offset sign in our studies rather represents the endpoint of a one-shot encoding experience, analogous to the end of a trajectory in rodent navigation task (Diba and Buzsáki, 2007).

Recent studies have indeed discovered that an induced event boundary could efficiently trigger the neural reactivation immediately after its detection (Sols et al., 2017; Silva et al., 2019). According to the EST, this neural signature time-locked to the event boundary reflects an event model update process where a mismatch, mainly at the perceptual level, to the current state urges the system to re-evaluate and, at the same time, reset the context representation and lead the event model to a new state (Zacks et al., 2007, 2011). Recent empirical evidence is also compatible with such a view by illustrating the memory-related function of neural reactivation for the proceeding but not the upcoming event (Silva et al., 2019). However, the two putative functional roles of this neural signature, namely 'to chunk' the ongoing stimulus and 'to integrate' the previously extracted event into memory, might be overlayed in the same transient temporal window, making it difficult to single out and to stress the particular function underlying how the system achieves the 'wrap up' process to generate a concrete representation for the proceeding event. However, this is not the case in our studies, where the fixed single-trial structure allows us to pinpoint the exact mechanism of neural reactivation that particularly serves the memory formation process.

Another critical aspect of the current findings is that the neural reactivation observed following the end of an event sequence was not triggered by abrupt changes in external information per se. Indeed, it has been recently suggested that an abrupt change in perceptual input is not necessary nor sufficient to determine how event representation is extracted and generated from experience (Shin and DuBrow, 2021). This view extended the possibility that the neural concomitant to the event boundaries previously reported during passive viewing task (Baldassano et al., 2017; Ben-Yakov and Henson, 2018; Silva et al., 2019) could be elicited independent of perceptual change and even when transitions are predictable. Our findings are compatible with such a view. In Study 1 and 2, neural reactivation was found to be endogenously initiated once and only at sequence offset, where the perceptual changes were predictable, and the end of an event was expected. Our findings thus support the notion that post-encoding neural reactivation reflects a natural process particularly engaged in memory formation rather than passively triggered by ongoing information processing. To take a further step, we directly manipulated the end of sequences in Study 3. The results consistently showed that initiation of the 'wrap up' process awaited the completion of encoding sequences. However, how the series ended directly impacted how the brain 'closed up' the event, thus favouring the claim of neural reactivation as a flexible mechanism that can be manipulated by a top-down control over the expectancy of the event schema.

It should be noted that current findings do not preclude the possibility that the neural reactivation triggered by the event boundary and the end of event encoding represents the same mechanism. For example, one can argue that under our sequential encoding paradigm of the three studies, the end of a sequence is essentially an event boundary in the sense that it marks the transition point from task encoding to the off-task period. However, we believe that the specific exploration centred on the end of event encoding, as implemented by our paradigms, helped highlight the particular functional role of neural reactivation in rapidly organising and strengthening the just-encoded episode and demonstrated the features of such mechanism to be flexible and active.

7.4. How does the post-encoding neural reactivation in humans bridge with the replay mechanism found in rodents?

Although exploration of neural reactivation was initially inspired by the replay mechanism found in rodents, up to date, there still exists interpretational gaps between the two presumably homologous mechanisms. The primary concern regards the different forms of reactivation/replay observed in humans and rodents. In the rodent hippocampus, neural replay has been found during the off-task period as the organised sequential reactivation of the place cells (Burgess and O'Keefe, 1996) that encode the initial location during the navigation task (Foster and Wilson, 2006; Diba and Buzsáki, 2007). However, in humans, the detection of neural reactivation has mostly relied on multivariate analytical approaches to neuroimaging data, and neural reactivation is generally reflected by the re-emergence of pattern activity from large-scale brain regions (Xue, 2018). Being aware of the differences in experimental and analytical approaches, studies have focused on the shared properties between human neural reactivation and rodent replay mechanism. Several attractive features in rodent replay make it plausible to be treated as the putative mechanism supporting memory consolidation (Tambini and Davachi, 2019). We believe our results fit with these features to a certain extent, thus contributing to advancing the understanding of the neural reactivation mechanism in humans and its role underlying the memory formation process.

The first iconic feature observed in rodent neural replay consists in its temporally compressed format that occurs during brief (100–200 ms) sharp-wave ripple (SWR) events (Lee and Wilson, 2002; Foster and Wilson, 2006; Diba and Buzsáki, 2007; Dupret et al., 2010). Primary fMRI approaches have been able to detect neural reactivation for specific event patterns in the human hippocampus (Tambini and Davachi, 2013; Grube et al., 2016; Hermans et al., 2017), yet owing to the limited temporal precision, evidence for its temporal nature is still elusive. Leveraging the high temporal resolution of EEG, our results from three different studies are compatible with such feature in a way that neural reactivation has been observed centred in the narrow time window (e.g., ~400 ms - 800 ms) at the early stage of the post-encoding period, and at the same time, what has been reactivated covered the neural pattern elicited across the extended encoding period (e.g., ~400 ms - 1300 ms). Although it remains open for discussion whether cortical patterns reflect the same mechanism observed in the rodent hippocampus, our paradigm brings up the possibility of exploring human neural reactivation with fast timescale information.

The second important feature of rodent replay lies in its high-fidelity decoding of prior experience, which is grounded on the precise regional specificity of data sampling and domain-specific experimental paradigm (i.e., constrained spatial navigation task). However, these prerequisites are hard to be met in human studies, given the high complexity of episodic memory and the constraints on measuring techniques. Multivariate approaches have the advantage of deciphering the nature of representation at different memory stages yet lack the power in content-specificity to illustrate what is precisely being reactivated. Although these obstacles remain present in our experiments, our paradigms help narrow down the alternative interpretations, providing the results compatible with the experience-specific nature of neural replay found in rodents. In the three studies, the existence and the memory-related function of neural reactivation have been detected after a one-shot episodic encoding experience. Such necessity for trial-specific neural reactivation in memory formation was especially highlighted in rodent studies where replay has been shown to contribute to one-trial learning (Diba and Buzsáki, 2007; Carr et al., 2011). Following this trend, in Study 2, we further found that the increase in neural pattern similarity to the encoding experience at the event offset period temporally overlapped with a reduced decodability of categorical-feature representations, indicating that the content of neural reactivation might be more trial-unique or experience-specific with less containing generic categorical feature representations.

Finally, rodent replay has been linked to robust hippocampus-cortical interaction, which has been considered the neural basis for reorganising memory representation across hippocampal–cortical networks (Ólafsdóttir et al., 2016; Rothschild et al., 2016). The neural reactivation captured at the cortical level in our data is congruent with recent findings in rodents where neural reply at the cortex was observed during sleep (Ji and Wilson, 2007) and awake period (O'Neill et al., 2017). However, owing to the limited spatial resolution of EEG, it is not feasible to direct link the current evidence to the previously reported post-encoding neural activity in the human hippocampus (e.g., Ben-Yakov et al., 2011), thus failing to reveal the specific cortico-hippocampal dynamics of the reactivation mechanism. This limitation will be further discussed in Section 6 of this chapter.

Taking it all together, our paradigms and analytical approaches revealed the temporal and mechanistic properties of the post-encoding neural reactivation mechanism similar to the neural replay mechanism found in rodents, thus bridging the neural reactivation/replay findings in the two species.

7.5. Limitations and future directions

7.5.1. Sequentiality of the post-encoding reactivation mechanism

In rodents, the directionality of the sequential replay, being either forward or backwards, bears different functional implications. While forward replay enables to pre-play the future trajectory to guide memory-related decision making, backward replay is often observed immediately after the very first one-trial navigation experience, suggesting a process to promote task learning (Carr et al., 2011).

In Study 1, 2 and 3, we speculate that within the crucial temporal window where the reactivation was found, the representations being reactivated should also contain information regarding the sequentiality. And inspired by rodent literature (Diba and Buzsáki, 2007; Carr et al., 2011), we expect such reactivation should be displayed in reverse order, as it fits with the putative role of the rapid reactivation mechanism in strengthening the one-shot learning experience. However, representational patterns captured in scalp EEG in Study 1, 2 and 3 were limited to reflect a global neural state engaged during encoding and offset, with their similarity bridging putative off-task period to the reinstatement of the encoding neural state. In other words, the neural patterns per se have limitations in capturing the specificity of content representation, thus lacking the precise decodability to disentangle the temporal dynamics of the fluctuation of accurate feature representations in a compressed time window (e.g., the early stage of the post-encoding period).

Future studies should adopt a paradigm with finer regularities over the stimulus features and apply statistical approaches that combine nonlinear classification with linear temporal modelling (e.g., Temporally delayed linear modelling (Liu et al., 2021)) in search for possible sequentiality during post-encoding reactivation.

7.5.2. Spatial contribution to the post-encoding neural reactivation

Also, the latest evidence supports that the hippocampus plays an important role in making a ‘switch’ from external information encoding to an internal mnemonic process (Treder et al., 2021). Such process is physiologically marked by a sustained increase in hippocampal high gamma power (55 to 110 Hz) that emerged 500 ms after cue onset during the memory retrieval. Consistent evidence was reported in our three studies where similar temporal feature in the post-encoding period was found in cortical regions, suggesting that the reactivation of the encoding pattern may result from a switch from perception to memory process so that a stable memory trace can be successfully generated. Indeed, it has been long proposed that the hippocampus engages in binding separate just-encoded episodic elements into an integrated memory trace. Such a process is hypothesised to be achieved by establishing a specific neural pathway linking cortical representations with neuronal ensembles in the hippocampus that store their spatiotemporal patterns (Teyler and DiScenna, 1986; Teyler and Rudy, 2007). However, due to the lack of spatial resolution of scalp EEG recordings, our results were limited in providing direct evidence of the region-specific contribution of the post-encoding reactivation mechanism. Thus, the crucial puzzle to complete the whole picture of the cortical-hippocampus dynamics supporting such transformation remains unknown.

For future studies, it will be essential to evaluate how different brain regions are orchestrated functionally and temporally to conduct such a rapid and efficient reactivation process. We propose extending this research line by collecting data with the invasive brain mapping technique intracranial electroencephalography (iEEG). In collaboration with Hospital Clínic Barcelona, we have already implemented a modified version of the paradigm of Study 2 to start testing on drug-resistant epileptic patients with iEEG implantation. Data from 5 patients have been successfully recollected. Data with more patients and future analysis are estimated to unveil the particular transient and region-specific features of the post-encoding neural reactivation mechanism with fine-grain temporal and spatial resolution.

7.5.3. Endogenously elicited neural reactivation mechanism

In both Study 1 and Study 3, we found evidence supporting the post-encoding reactivation mechanism to be a flexible process sensitive to the general context of the task. Its elicitation, apart from being memory relevant, was shown to be conditional to the global presentation paradigm of the episode. For example, we observed its presence only when an episode is perceived to be complete but not before (Study 1), and its magnitude can be manipulated by how convincible the item following the event sequences perceived to be an end of the episode (Study 3).

As previously discussed, experiences could be grouped together following the inference-based framework so that boundaries can occur without perceptual changes or when transitions between events can be predicted (Shin and DuBrow, 2020). This brings up the possibility that not only the ‘wrap up’ process can be endogenously initiated once the event encoding reaches the end, but when an end is conceived should also be internally guided. However, in our three studies, owing to the lack of variability in the length of event sequences and the fixed trial structure, participants did not have control over the temporal property of event sequences, thus impeding a direct measure of such possibility. It will be interesting to specifically explore the interaction between a subjective decision of an episode to finish and neural concomitant underlying the memory formation process once it does. Future experiments can address this assumption by adopting a varied length of encoding sequence series or a self-guided learning paradigm.

Chapter 8. Conclusions

- Neural reactivation is rapidly elicited during the post-encoding period following the one-shot experience of the event episode. Such reactivation can be quantified as an increased neural similarity of the previous encoding experience in a limited time window. The degree of neural reactivation is predictive memory recollection for the just-encoded event.
- Memory reactivation is induced at the end of the continuous ongoing episodic memory encoding, once participants perceive the episode to be complete but not between successive elements within an episode. Also, it fails to be elicited at the offset of sequences that were not perceived as meaningful episodes. Together, it suggests post-encoding memory reactivation as a natural process selectively engaged in supporting memory formation for meaningful episodes.
- Successful event encoding requires a gradual integration of disparate episodic elements that can be bound as a cohesive memory trace in the long term. Such integration can be promoted by offset-locked neural activity that triggers the reactivation of a detailed representation of the experience-unique episodic elements of the preceding encoding experience.
- The perceived incongruence of the item within an unfolding event can disrupt the neural stability established throughout the ongoing exposure to a schema-consistent episodic narrative. Such mismatch detection triggers a rapid neural reactivation to wrap up the just-encoded episodic events and urge the system to update the current state representation.
- The temporal and representational properties of the post-encoding neural

reactivation mechanism are similar to those of the neural replay found in rodents, serving to bridge the experimental gaps between the two species and to advance the understanding of why, when, and how such neural signature in humans may be associated to memory formation.

References

Andersen, P, Bliss TVP, Skrede KK (1971) Lamellar organization of hippocampal exitatory pathways. Exp Brain Res 13:222-238. https://doi.org/10.1007/BF00234087

Andersen, P., Morris, R., Amaral, D., Bliss, T., & O'Keefe, J. (Eds.). (2006). The hippocampus book. Oxford university press. https://doi.org/10.1093/acprof:oso/9780195100273.001.0001

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2021). Behavioral, physi- ological, and neural signatures of surprise during naturalistic sports viewing. Neuron, 109, 1-14. https://doi.org/10.1016/j.neuron.2020.10.029

Aggleton JP, Brown MW (1999) Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. Behav Brain Sci 22: 425-489 https://doi.org/10.1017/S0140525X99002034

Atienza, M., Crespo-Garcia, M., and Cantero, J.L. (2011). Semantic congruence enhances memory of episodic associations: role of theta oscillations. J. Cogn. Neurosci. 23, 75-90. https://doi.org/10.1162/jocn.2009.21358

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

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.
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. Journal of Neuroscience, 38(45), 9689-9699.
https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Bart Krekelberg (2021). bayesFactor (https://github.com/klabhub/bayesFactor), GitHub. Retrieved December 11, 2021.
https://doi.org/10.5281/zenodo.5707551

Behrens, T. E. J., Woolrich, M. W., Walton, M. E. and Rushworth, M. F. S. (2007) Learning the value of information in an uncertain world. Nature Neuroscience, 10,1214-1221.
https://doi.org/10.1038/nn1954

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., and Maril, A. (2015). Delineating the Effect of Semantic Congruency on Episodic Memory: The Role of Integration and Relatedness. PLOS ONE 10, e0115624.
https://doi.org/10.1371/journal.pone.0115624

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

Ben-Yakov, A., Eshel, N., & Dudai, Y. (2013). Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. Journal of Experimental Psychology: General, 142(4), 1255.
https://doi.org/10.1037/a0033558

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

Bergmann, T. O., Mölle, M., Diedrichs, J., Born, J., & Siebner, H. R. (2012). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. Neuroimage, 59(3), 2733-2742.
https://doi.org/10.1016/j.neuroimage.2011.10.036

Bird, C. M. (2020). How do we remember events?. Current Opinion in Behavioral Sciences, 32, 120-125.
https://doi.org/10.1016/j.cobeha.2020.01.020

Braver, T.S. et al. (2001) Context processing in older adults: evidence for a theory relating cognitive control to neurobiology in healthy aging. J. Exp. Psychol. Gen. 130, 746-763
https://doi.org/10.1037/0096-3445.130.4.746

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

Brownstein, A. L., & Read, S. J. (2007). Situation models and memory: The effects of temporal and causal information on recall sequence. Memory, 15(7), 730-745.
https://doi.org/10.1080/09658210701539596

Brunec, I. K., Moscovitch, M., & Barense, M. D. (2018). Boundaries shape cognitive representations of spaces and events. Trends in cognitive sciences, 22(7), 637-650.
https://doi.org/10.1016/j.tics.2018.03.013

Burgess, N. & O'Keefe, J. Neuronal computations underlying the firing of place cells and their role in navigation. Hippocampus 6, 749-762 (1996).
https://doi.org/10.1002/(SICI)1098-1063(1996)6:6<749::AID-HIPO16>3.0.CO;2-0

Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nature neuroscience, 14(2), 147.
https://doi.org/10.1038/nn.2732

Chen, J., Cook, P. A., & Wagner, A. D. (2015). Prediction strength modulates responses in human area CA1 to sequence violations. Journal of Neurophysiology, 114(2), 1227-1238.
https://doi.org/10.1152/jn.00149.2015

Chen, J., Honey, C. J., Simony, E., Arcaro, M. J., Norman, K. A., & Hasson, U. (2016). Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. Cerebral cortex, 26(8), 3428-3441.
https://doi.org/10.1093/cercor/bhv155

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

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. Nature neuroscience, 17(3), 455-462.
https://doi.org/10.1038/nn.3635

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. Hippocampus, 29(3), 162-183
https://doi.org/10.1002/hipo.23074

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. Science, 210(4466), 207-210.
https://doi.org/10.1126/science.7414331

Cohen, N. J., & Eichenbaum, H. (1995). Memory, amnesia, and the hippocampal system. MIT press.

Cohn-Sheehy, B. I., Delarazan, A. I., Reagh, Z. M., Crivelli-Decker, J. E., Kim, K., Barnett, A. J., ... & Ranganath, C. (2021). The hippocampus constructs narrative memories across distant events. Current Biology, 31(22), 4935-4945.
https://doi.org/10.1016/j.cub.2021.09.013

Conway, M. A. (2009). Episodic memories. Neuropsychologia, 47(11), 2305-2313. https://doi.org/10.1016/j.neuropsychologia.2009.02.003

Cooper, R. A., & Ritchey, M. (2020). Progression from feature-specific brain activity to hippocampal binding during episodic encoding. Journal of Neuroscience, 40(8), 1701-1709. https://doi.org/10.1523/JNEUROSCI.1971-19.2019

Corkin, S. (2013). Permanent present tense: The man with no memory, and what he taught the world. Penguin UK.

Cowley, B. R., Snyder, A. C., Acar, K., Williamson, R. C., Byron, M. Y., & Smith, M. A. (2020). Slow drift of neural activity as a signature of impulsivity in macaque visual and prefrontal cortex. Neuron, 108(3), 551-567. https://doi.org/10.1016/j.neuron.2020.07.021

Craik, F.I., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. J. Exp. Psychol. Gen. 104, 268. https://doi.org/10.1037/0096-3445.104.3.268

Csicsvari, J., O'Neill, J., Allen, K., & Senior, T. (2007). Place-selective firing contributes to the reverse-order reactivation of CA1 pyramidal cells during sharp waves in open-field exploration. European Journal of Neuroscience, 26(3), 704-716. https://doi.org/10.1111/j.1460-9568.2007.05684.x

Davachi, L. (2006). Item, context and relational episodic encoding in humans. Current opinion in neurobiology, 16(6), 693-700. https://doi.org/10.1016/j.conb.2006.10.012

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception?. Trends in cognitive sciences, 22(9), 764-779. https://doi.org/10.1016/j.tics.2018.06.002

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends in cognitive sciences, 11(9), 379-386. https://doi.org/10.1016/j.tics.2007.08.001

Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. Nature neuroscience, 10(10), 1241-1242. https://doi.org/10.1038/nn1961

Diekelmann, S., & Born, J. (2010). The memory function of sleep. Nature Reviews Neuroscience, 11(2), 114-126. https://doi.org/10.1038/nrn2762

Duarte, A., Ranganath, C., & Knight, R. T. (2005). Effects of unilateral prefrontal lesions on familiarity, recollection, and source memory. Journal of Neuroscience, 25(36), 8333-8337. https://doi.org/10.1523/JNEUROSCI.1392-05.2005

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

DuBrow, S. and Davachi, L. (2014) Temporal memory is shaped by encoding stability and intervening item reactivation. The Journal of Neuroscience, 34, 13998-14005. https://doi.org/10.1523/JNEUROSCI.2535-14.2014

DuBrow, S. and Davachi, L. (2016) Temporal binding within and across events. Neurobiology of Learning and Memory, 134, 107-114. https://doi.org/10.1016/j.nlm.2016.07.011

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift?. Current opinion in behavioral sciences, 17, 141-146. https://doi.org/10.1016/j.cobeha.2017.08.003

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram?. Annu. Rev. Psychol., 55, 51-86. https://doi.org/10.1146/annurev.psych.55.090902.142050

Dudai, Y. (2012). The restless engram: consolidations never end. Annual review of neuroscience, 35, 227-247.
https://doi.org/10.1146/annurev-neuro-062111-150500

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. Neuron, 88(1), 20-32.
https://doi.org/10.1016/j.neuron.2015.09.004

Duncan, C. P. (1949). The retroactive effect of electroshock on learning. Journal of comparative and physiological psychology, 42(1), 32.
https://doi.org/10.1037/h0058173

Dupret, D., O'Neill, J., Pleydell-Bouverie, B., & Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. Nature neuroscience, 13(8), 995-1002.
https://doi.org/10.1038/nn.2599

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. Annual review of psychology, 66, 115-142.
https://doi.org/10.1146/annurev-psych-010814-015031

Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. science, 318(5853), 1147-1150.
https://doi.org/10.1126/science.1148979

Ebbinghaus, H. (1885). Über das gedächtnis: untersuchungen zur experimentellen psychologie. Duncker & Humblot.

Eichenbaum, H. (2000) A cortical-hippocampal system for declarative memory. Nature Reviews Neuroscience 1(1): 41–50.
https://doi.org/10.1038/35036213

Eichenbaum, H. (2004) Hippocampus: cognitive processes and neural representations that underlie declarative memory. Neuron 44, 109-120
https://doi.org/10.1016/j.neuron.2004.08.028

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. Annu. Rev. Neurosci., 30, 123-152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

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

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex (New York, NY: 1991), 1(1), 1-47. https://doi.org/10.1093/cercor/1.1.1

Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature, 440(7084), 680-683. https://doi.org/10.1038/nature04587

Foster, D. J. (2017). Replay comes of age. Annual review of neuroscience, 40, 581-602. https://doi.org/10.1146/annurev-neuro-072116-031538

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. Cognitive, Affective, & Behavioral Neuroscience, 1(2), 137-160. https://doi.org/10.3758/CABN.1.2.137

Frank, D., Montaldi, D., Wittmann, B., & Talmi, D. (2018). Beneficial and detrimental effects of schema incongruence on memory for contextual events. Learning & Memory, 25(8), 352-360. https://doi.org/10.1101/lm.047738.118

Frank, D., Montemurro, M. A., & Montaldi, D. (2020). Pattern Separation Underpins Expectation-Modulated Memory. The Journal of Neuroscience, 40(17), 3455-3464. https://doi.org/10.1523/JNEUROSCI.2047-19.2020

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

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Düzel, E. (2010). Theta-coupled periodic replay in working memory. Current Biology, 20(7), 606-612. https://doi.org/10.1016/j.cub.2010.01.057

Fuhs, M. C., & Touretzky, D. S. (2007). Context learning in the rodent hippocampus. Neural Computation, 19, 3173-3215 https://doi.org/10.1162/neco.2007.19.12.3173

Fujisawa, S., Matsuki, N., and Ikegaya, Y. (2006). Single neurons can induce phase transitions of cortical recurrent networks with multiple internal states. Cereb. Cortex 16, 639-654. doi: 10.1093/cercor/bhj010 2236(87)90011-7 https://doi.org/10.1093/cercor/bhj010

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. Science, 173(3997), 652-654. https://doi.org/10.1126/science.173.3997.652

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. Current opinion in neurobiology, 20(2), 251-256. https://doi.org/10.1016/j.conb.2010.02.008

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. Psychological review, 117(1), 197. https://doi.org/10.1037/a0017808

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. Learning & behavior, 40(3), 255-268. https://doi.org/10.3758/s13420-012-0080-8

Gershman, S. J., Radulescu, A., Norman, K. A., & Niv, Y. (2014). Statistical computations underlying the dynamics of memory updating. PLoS computational biology, 10(11), e1003939. https://doi.org/10.1371/journal.pcbi.1003939

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. Trends in cognitive sciences, 21(8), 618-631. https://doi.org/10.1016/j.tics.2017.04.013

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning?. Journal of memory and language, 94, 149-165.
https://doi.org/10.1016/j.jml.2016.11.001

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. Journal of Experimental Psychology: General, 148(2), 325-341.
https://doi.org/10.1037/xge0000498

Gupta, A. S., Van Der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. Neuron, 65(5), 695-705.
https://doi.org/10.1016/j.neuron.2010.01.034

Hasselmo, M. E., & Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. Neural Networks, 18, 1172- 1190.
https://doi.org/10.1016/j.neunet.2005.08.007

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. science, 303(5664), 1634-1640.
https://doi.org/10.1126/science.1089506

Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron, 57(3), 452-462.
https://doi.org/10.1016/j.neuron.2007.12.009

Hebb, D. (1949). The Organization of Behavior (New York: Wiley & Sons).

Hermans, E. J., Kanen, J. W., Tambini, A., Fernández, G., Davachi, L., & Phelps, E. A. (2017). Persistence of Amygdala-Hippocampal Connectivity and Multi-Voxel Correlation Structures During Awake Rest After Fear Learning Predicts Long-Term Expression of Fear. Cerebral cortex (New York, N.Y. : 1991), 27(5), 3028-3041.
https://doi.org/10.1093/cercor/bhw145

Heusser, A. C., Ezzyat, Y., Shiff, I. and Davachi, L. (2018) Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. Journal of Experimental Psychology: Learning, Memory, and Cogni- tion, 44, 1075-1090.
https://doi.org/10.1037/xlm0000503

Heusser, A. C., Fitzpatrick, P. C., & Manning, J. R. (2021). Geometric models reveal behavioural and neural signatures of transforming experiences into memories. Nature Human Behaviour, 1-15.
https://doi.org/10.1038/s41562-021-01051-6

Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: A theory. Behavioral Biology, 12, 421-444
https://doi.org/10.1016/S0091-6773(74)92231-7

Hodges JR (1994) Retrograde amnesia. in Handbook of memory disorders, eds Baddeley A, Wilson BA, Watts F (Wiley, New York), pp 81-107.

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K. and Burgess, N. (2016) The role of spatial boundaries in shaping long-term event representations. Cognition, 154, 151-164.
https://doi.org/10.1016/j.cognition.2016.05.013

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. Journal of mathematical psychology, 46(3), 269-299.
https://doi.org/10.1006/jmps.2001.1388

Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (2005). The temporal context model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. Psychological Review, 112, 75-116.
https://doi.org/10.1037/0033-295X.112.1.75

Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D., & Seamans, J. K. (2012). Contextual encoding by ensembles of medial prefrontal cortex neurons. Proceedings of the National Academy of Sciences of the United States of America,

109, 5086-5091
https://doi.org/10.1073/pnas.1114415109

Ishizuka N, Weber J, Amaral DG (1990) Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. J Comp Neurol 295:580-623.
https://doi.org/10.1002/cne.902950407

Jafarpour, A., Fuentemilla, L., Horner, A. J., Penny, W., & Duzel, E. (2014). Replay of very early encoding representations during recollection. Journal of Neuroscience, 34(1), 242-248.
https://doi.org/10.1523/JNEUROSCI.1865-13.2014

Jarrard, L. (1993). On the role of the hippocampus in learning and memory in the rat. Behavioral and Neural Biology, 60, 9-26.
https://doi.org/10.1016/0163-1047(93)90664-4

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. The American Journal of Psychology, 35(4), 605-612.
https://doi.org/10.2307/1414040

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. Journal of Neuroscience, 30(46), 15558-15565.
https://doi.org/10.1523/JNEUROSCI.1337-10.2010

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. Nature neuroscience, 10(1), 100-107.
https://doi.org/10.1038/nn1825

Jonas P, Major G, Sakmann B (1993) Quantal components of unitary EPSCs at the mossy fibre synapse on CA3 pyramidal cells of rat hippocampus. J Physiol 472: 615-663.
https://doi.org/10.1113/jphysiol.1993.sp019965

Kahana, M. J. (1996). Associative retrieval processes in free recall. Memory & cognition, 24(1), 103-109.
https://doi.org/10.3758/BF03197276

Karlsson, M. P., & Frank, L. M. (2008). Network dynamics underlying the formation of sparse, informative representations in the hippocampus. Journal of Neuroscience, 28(52), 14271-14281.
https://doi.org/10.1523/JNEUROSCI.4261-08.2008

Karlsson, M. P., & Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. Nature neuroscience, 12(7), 913-918.
https://doi.org/10.1038/nn.2344

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the american statistical association, 90(430), 773-795.
https://doi.org/10.1080/01621459.1995.10476572

Keidel, J. L., Oedekoven, C. S., Tut, A. C., & Bird, C. M. (2018). Multiscale integration of contextual information during a naturalistic task. Cerebral Cortex, 28(10), 3531-3539.
https://doi.org/10.1093/cercor/bhx218

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. Nature neuroscience, 22(10), 1598-1610.
https://doi.org/10.1038/s41593-019-0467-3

Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. Neuron, 75(2), 265-270.
https://doi.org/10.1016/j.neuron.2012.04.034

Kudrimoti, H. S., Barnes, C. A., & McNaughton, B. L. (1999). Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. Journal of Neuroscience, 19(10), 4090-4101.
https://doi.org/10.1523/JNEUROSCI.19-10-04090.1999

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. Proceedings of the National

Academy of Sciences, 108(14), 5903-5908.
https://doi.org/10.1073/pnas.1016939108

Kumar, S. et al. (2017) Encoding of predictable and unpredict- able stimuli by inferior temporal cortical neurons. J. Cogn. Neu- rosci. 29, 1445-1454
https://doi.org/10.1162/jocn_a_01135

Kumaran, D. and McClelland, J.L. (2012) Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. Psychol. Rev. 119, 573-616
https://doi.org/10.1037/a0028681

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. Trends in cognitive sciences, 20(7), 512-534.
https://doi.org/10.1016/j.tics.2016.05.004

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. Trends in cognitive sciences, 12(2), 72-79.
https://doi.org/10.1016/j.tics.2007.11.004

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). Annual review of psychology, 62, 621.
https://doi.org/10.1146/annurev.psych.093008.131123

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L., & Pennartz, C. M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. PLoS biology, 7(8), e1000173.
https://doi.org/10.1371/journal.pbio.1000173

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 years of consolidation-remembering Müller and Pilzecker. Learning & memory, 6(2), 77-87.
https://doi.org/10.1101/lm.6.2.77

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. Neuron, 36(6), 1183-1194. https://doi.org/10.1016/S0896-6273(02)01096-6

Lee, H., & Chen, J. (2021). Narratives as Networks: Predicting Memory from the Structure of Naturalistic Events. bioRxiv 2021.04.24.441287; doi: https://doi.org/10.1101/2021.04.24.441287

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. Trends in cognitive sciences, 15(8), 343-351. https://doi.org/10.1016/j.tics.2011.06.004

Li X-G, Somogyi P, Ylinen A, Buzsaki G (1994) The hippocampal CA3 network: an in vivo intracellular labeling study. J Comp Neurol 339:181-208 https://doi.org/10.1002/cne.903390204

Lisman, J. E. (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. Neuron, 22(2), 233-242. https://doi.org/10.1016/S0896-6273(00)81085-5

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. (2019). Human replay spontaneously reorganizes experience. Cell, 178(3), 640-652. https://doi.org/10.1016/j.cell.2019.06.012

Liu, Y., Dolan, R. J., Higgins, C., Penagos, H., Woolrich, M. W., Ólafsdóttir, H. F., ... & Behrens, T. E. (2021). Temporally delayed linear modelling (TDLM) measures replay in both animals and humans. Elife, 10, e66917. https://doi.org/10.7554/eLife.66917

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

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. Psychological review, 111(2), 309-332. https://doi.org/10.1037/0033-295X.111.2.309

Lu, Y., Wang, C., Chen, C., and Xue, G. (2015). Spatiotemporal neural pattern similarity supports episodic memory. Curr. Biol. 25, 780-785. https://doi.org/10.1016/j.cub.2015.01.055

Lu, Q., Hasson, U., & Norman, K. A. (2020). When to retrieve and encode episodic memories: a neural network model of hippocampal-cortical interaction. bioRxiv 2020.12.15.422882; doi: https://doi.org/10.1101/2020.12.15.422882 https://doi.org/10.1101/2020.12.15.422882

Manning, J. R., Norman, K. A., & Kahana, M. J. (2015). The role of context in episodic memory. Appears in Gazzaniga M, Ed. The Cognitive Neurosciences.

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. Neuron, 56(3), 530-540. https://doi.org/10.1016/j.neuron.2007.08.017

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177-190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review, 102(3), 419. https://doi.org/10.1037/0033-295X.102.3.419

McClelland, J. L., McNaughton, B. L., & Lampinen, A. K. (2020). Integration of new information in memory: new insights from a complementary learning systems perspective. Philosophical Transactions of the Royal Society B, 375(1799), 20190637. https://doi.org/10.1098/rstb.2019.0637

McNaughton, B. L., and Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends

Neurosci. 10, 408-415.
https://doi.org/10.1016/0166-2236(87)90011-7

Michelmann, S., Price, A. R., Aubrey, B., Strauss, C. K., Doyle, W. K., Friedman, D., ... & Norman, K. A. (2021). Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. Nature communications, 12(1), 1-15.
https://doi.org/10.1038/s41467-021-25376-y

Miles, R., & Wong, R. K. (1983). Single neurones can initiate synchronized population discharge in the hippocampus. Nature, 306(5941), 371-373.
https://doi.org/10.1038/306371a0

Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. Neuron, 20(3), 445-468.
https://doi.org/10.1016/S0896-6273(00)80987-3

Morris RG (2003) Long-term potentiation and memory. Philos Trans R Soc Lond B Biol Sci 358: 643-647
https://doi.org/10.1098/rstb.2002.1230

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. Journal of cognitive neuroscience, 4(3), 257-267.
https://doi.org/10.1162/jocn.1992.4.3.257

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., ... & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. Journal of anatomy, 207(1), 35-66.
https://doi.org/10.1111/j.1469-7580.2005.00421.x

Murdock Jr, B. B. (1962). The serial position effect of free recall. Journal of experimental psychology, 64(5), 482.
https://doi.org/10.1037/h0045106

Nadel, L. (1995). The role of the hippocampus in declarative memory: A comment on Zola-Morgan, Squire, and Ramus (1994). Hippocampus, 5, 232-239. https://doi.org/10.1002/hipo.450050309

Nadel, L., & MacDonald, L. (1980). Hippocampus: Cognitive map or working memory? Behavioral and Neural Biology, 29, 405-409. https://doi.org/10.1016/S0163-1047(80)90430-6

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. Current opinion in neurobiology, 7(2), 217-227. https://doi.org/10.1016/S0959-4388(97)80010-4

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. Hippocampus, 10(4), 352-368. https://doi.org/10.1002/1098-1063(2000)10:4<352::AID-HIPO2>3.0.CO;2-D

Nadel, L., Winocur, G., Ryan, L., & Moscovitch, M. (2007). Systems consolidation and hippocampus: two views. Debates in Neuroscience, 1(2), 55-66. https://doi.org/10.1007/s11559-007-9003-9

Nakazawa K, Quirk MC, Chitwood RA, Watanabe M, Yeckel MF, et al. (2002) Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science 297: 211-218. https://doi.org/10.1126/science.1071795

Nakazawa K, Sun LD, Quirk MC, Rondi-Reig L, Wilson MA, et al. (2003) Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. Neuron 38: 305-315. https://doi.org/10.1016/S0896-6273(03)00165-X

Nassar, M. R., Rumsey, K. M.,Wilson, R. C., Parikh, K., Heasly, B. and Gold, J. I. (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. Nature Neuroscience, 15, 1040-1046. https://doi.org/10.1038/nn.3130

Nassar, M. R.,Wilson, R. C., Heasly, B. and Gold, J. I. (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. The Journal of Neuroscience, 30, 12366-12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Nichols EA, KaoYC, Verfaellie M, Gabrieli JD (2006) Working memory and long-term memory for faces: evidence from fMRI and global amnesia for involvement of the medial temporal lobes. Hippocampus 16:604-616. https://doi.org/10.1002/hipo.20190

Nolde, S. F., Johnson, M. K., & Raye, C. L. (1998). The role of prefrontal cortex during tests of episodic memory. Trends in cognitive sciences, 2(10), 399-406. https://doi.org/10.1016/S1364-6613(98)01233-9

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychological review, 110(4), 611. https://doi.org/10.1037/0033-295X.110.4.611

Olsen RK, Nichols EA, Chen J, Hunt JF, Glover GH, Gabrieli JD, Wagner AD (2009) Performance-related sustained and anticipatory activity in hu- man medial temporal lobe during delayed match-to-sample. J Neurosci 29:11880-11890 https://doi.org/10.1523/JNEUROSCI.2245-09.2009

O'Keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. New York, NY: Oxford University Press.

O'Neill, J., Pleydell-Bouverie, B., Dupret, D., & Csicsvari, J. (2010). Play it again: reactivation of waking experience and memory. Trends in neurosciences, 33(5), 220-229. https://doi.org/10.1016/j.tins.2010.01.006

O'Neill, J., Boccara, C. N., Stella, F., Schönenberger, P., & Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. Science, 355(6321), 184-188. https://doi.org/10.1126/science.aag2787

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. Trends in cognitive sciences, 6(12), 505-510. https://doi.org/10.1016/S1364-6613(02)02005-3

O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. Trends in neurosciences, 33(8), 355-361. https://doi.org/10.1016/j.tins.2010.05.002

O'Reilly, J. X. (2013) Making predictions in a changing world-inference, uncertainty, and learning. Frontiers in Neuroscience, 7, 105. https://doi.org/10.3389/fnins.2013.00105

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. Cognitive science, 38(6), 1229-1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

Ólafsdóttir, H.F. et al. (2016) Coordinated grid and place cell re- play during rest. Nat. Neurosci. 19, 792-794 https://doi.org/10.1038/nn.4291

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. Trends in cognitive sciences, 6(2), 93-102. https://doi.org/10.1016/S1364-6613(00)01845-3

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. Journal of neuroscience, 9(8), 2907-2918. https://doi.org/10.1523/JNEUROSCI.09-08-02907.1989

Peigneux P, et al. (2006) Offline persistence of memory-related cerebral activity during active wakefulness. PLoS Biol 4(4):e100. https://doi.org/10.1371/journal.pbio.0040100

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., & Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep.

Nature neuroscience, 12(7), 919-926.
https://doi.org/10.1038/nn.2337

Piantoni, G., Van Der Werf, Y. D., Jensen, O., & Van Someren, E. J. (2015). Memory traces of long-range coordinated oscillations in the sleeping human brain. Human brain mapping, 36(1), 67-84.
https://doi.org/10.1002/hbm.22613

Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. Nature Communications, 9(1), 1-14.
https://doi.org/10.1038/s41467-018-03992-5

Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. Trends in cognitive sciences, 12(1), 24-30.
https://doi.org/10.1016/j.tics.2007.10.010

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. Psychological review, 116(1), 129.
https://doi.org/10.1037/a0014420

Postle, B.R. (2006) Working memory as an emergent property of the mind and brain. Neuroscience 139, 23-38
https://doi.org/10.1016/j.neuroscience.2005.06.005

Qin, Y. L., McNaughton, B. L., Skaggs, W. E., & Barnes, C. A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 352(1360), 1525-1533.
https://doi.org/10.1098/rstb.1997.0139

Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. Neurobiology of Learning and Memory, 179, Article 107382.
https://doi.org/10.1016/j.nlm.2021.107382

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. Mem Cognit, 34(5), 1150 - 1156
https://doi.org/10.3758/BF03193261

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. WIREs Cognitive Science, 2 (6), 608-620.
https://doi.org/10.1002/wcs.133

Radvansky, G. A. (2012). Across the event horizon. Current Directions in Psychological Science, 21(4), 269-272.
https://doi.org/10.1177/0963721412451274

Ranganath, C., & D'Esposito, M. (2001). Medial temporal lobe activity associated with active maintenance of novel information. Neuron, 31(5), 865-873.
https://doi.org/10.1016/S0896-6273(01)00411-1

Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. Hippocampus, 20(11), 1263-1290.
https://doi.org/10.1002/hipo.20852

Rajji T, Chapman D, Eichenbaum H, Greene R (2006) The role of CA3 hippocampal NMDA receptors in paired associate learning. J Neurosci 26: 908-915.
https://doi.org/10.1523/JNEUROSCI.4194-05.2006

Reagh, Z. M., & Ranganath, C. (2021). A cortico-hippocampal scaffold for representing and recalling lifelike events. bioRxiv 2021.04.16.439894; doi:
https://doi.org/10.1101/2021.04.16.439894

Redish AD (1999) Beyond the cognitive map: from place cells to episodic memory. Cambridge, MA: MIT Press
https://doi.org/10.7551/mitpress/1571.001.0001

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. Cognitive science, 31(4), 613-643.
https://doi.org/10.1080/15326900701399913

Ritchey, M., & Cooper, R. A. (2020). Deconstructing the posterior medial episodic network. Trends in cognitive sciences.
https://doi.org/10.1016/j.tics.2020.03.006

Rothschild, G., Eban, E., & Frank, L. M. (2017). A cortical–hippocampal–cortical loop of information processing during memory consolidation. Nature neuroscience, 20(2), 251-259.
https://doi.org/10.1038/nn.4457

Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. Journal of Experimental Psychology: Learning Memory and Cognition, 44(9), 1430-1443.
https://doi.org/10.1037/xlm0000518

Rudy, J. W., & O'Reilly, R. C. (1999). Contextual fear conditioning, conjunctive representations, pattern completion, and the hippocampus. Behavioral Neuroscience, 113, 867-880.
https://doi.org/10.1037/0735-7044.113.5.867

Saez, A., Rigotti, M., Ostojic, S., Fusi, S., & Salzman, C. D. (2015). Abstract context representations in primate amygdala and prefrontal cortex. Neuron, 87(4), 869-881.
https://doi.org/10.1016/j.neuron.2015.07.024

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

Schacter, D. L. (2012). Constructive memory: past and future. Dialogues in Clinical Neuroscience, 14:1, 7-18.
https://doi.org/10.31887/DCNS.2012.14.1/dschacter

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. Nature neuroscience, 16(4), 486-492.
https://doi.org/10.1038/nn.3331

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. Current opinion in behavioral sciences, 1, 1-8. https://doi.org/10.1016/j.cobeha.2014.07.005

Schon K, Hasselmo ME, Lopresti ML, Tricarico MD, Stern CE (2004) Persistence of parahippocampal representation in the absence of stimulus input enhances long-term encoding: a functional magnetic resonance imaging study of subsequent memory after a delayed match-to-sample task. J Neurosci 24:11088-11097. https://doi.org/10.1523/JNEUROSCI.3807-04.2004

Schönauer, M., Alizadeh, S., Jamalabadi, H., Abraham, A., Pawlizki, A., & Gais, S. (2017). Decoding material-specific memory reprocessing during sleep in humans. Nature communications, 8(1), 1-9. https://doi.org/10.1038/ncomms15404

Schreiner, T., Petzka, M., Staudigl, T., & Staresina, B. P. (2021). Endogenous memory reactivation during sleep in humans is clocked by slow oscillation-spindle complexes. Nature communications, 12(1), 1-10. https://doi.org/10.1038/s41467-021-23520-2

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of neurology, neurosurgery, and psychiatry, 20(1), 11. https://doi.org/10.1136/jnnp.20.1.11

Semon, R. (1923). Mnemic psychology (b. duffy, trans.). Concord, ma: george allen & unwin.(original work published 1909).

Shin, Y. S., & DuBrow, S. (2021). Structuring memory through inference-based event segmentation. Topics in Cognitive Science, 13(1), 106-127. https://doi.org/10.1111/tops.12505

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron, 60(2), 378-389. https://doi.org/10.1016/j.neuron.2008.09.023

Siefke, B. M., Smith, T. A. and Sederberg, P. B. (2019) A context-change account of temporal distinctiveness. Memory & Cognition, 79, 1-15. https://doi.org/10.3758/s13421-019-00925-5

Silson, E. H., Groen, I. I., Kravitz, D. J., & Baker, C. I. (2016). Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. Journal of vision, 16(6), 14-14. https://doi.org/10.1167/16.6.14

Sinclair, A. H., Manalili, G. M., Brunec, I. K., Adcock, R. A., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories. Proceedings of the National Academy of Sciences, 118(51). https://doi.org/10.1073/pnas.2117625118

Silva, M., Baldassano, C., & Fuentemilla, L. (2019). Rapid memory reactivation at movie event boundaries promotes episodic encoding. Journal of Neuroscience, 39(43), 8538-8548. https://doi.org/10.1523/JNEUROSCI.0360-19.2019

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. Psychonomic bulletin & review, 8(2), 203-220. https://doi.org/10.3758/BF03196157

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. Current Biology, 27(22), 3499-3504. https://doi.org/10.1016/j.cub.2017.09.057

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. Journal of memory and language, 53(1), 125-140. https://doi.org/10.1016/j.jml.2005.02.009

Squire LR. (1992) Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychological Review 99(2): 195–231.

Squire, L. R., Slater, P. C., & Chace, P. M. (1975). Retrograde amnesia: temporal gradient in very long term memory following electroconvulsive therapy. Science, 187(4171), 77-79.
https://doi.org/10.1126/science.1109228

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. Current opinion in neurobiology, 5(2), 169-177.
https://doi.org/10.1016/0959-4388(95)80023-9

Staresina, B. P., & Davachi, L. (2008). Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. Journal of cognitive neuroscience, 20(8), 1478-1489.
https://doi.org/10.1162/jocn.2008.20104

Staresina, B. P., & Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. Neuron, 63(2), 267-276.
https://doi.org/10.1016/j.neuron.2009.06.024

Staresina, B. P., Alink, A., Kriegeskorte, N., & Henson, R. N. (2013). Awake reactivation predicts memory in humans. Proceedings of the National Academy of Sciences, 110(52), 21159-21164.
https://doi.org/10.1073/pnas.1311989110

Staresina, B. P., & Wimber, M. (2019). A neural chronometry of memory recall. Trends in cognitive sciences, 23(12), 1071-1085.
https://doi.org/10.1016/j.tics.2019.09.011

Stawarczyk, D., Bezdek, M. A., & Zacks, J. M. (2021). Event representations and predictive processing: The role of the midline default network core. Topics in Cognitive Science, 13(1), 164-186.
https://doi.org/10.1111/tops.12450

Sterpenich, V., van Schie, M. K., Catsiyannis, M., Ramyead, A., Perrig, S., Yang, H. D., ... & Schwartz, S. (2021). Reward biases spontaneous neural reactivation during sleep. Nature Communications, 12(1), 1-11.
https://doi.org/10.1038/s41467-021-24357-5

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. Trends in cognitive sciences, 19(7), 394-405. https://doi.org/10.1016/j.tics.2015.05.004

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. J Exp Psychol Gen, 138(2), 236 - 257
https://doi.org/10.1037/a0015631

Tambini A, Ketz N, Davachi L (2010) Enhanced brain correlations during rest are related to memory for recent experiences. Neuron 65(2):280-290. https://doi.org/10.1016/j.neuron.2010.01.001

Tambini, A., & Davachi, L. (2013). Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. Proceedings of the National Academy of Sciences of the United States of America, 110(48), 19591-19596. https://doi.org/10.1073/pnas.1308499110

Tambini, A., & Davachi, L. (2019). Awake reactivation of prior experiences consolidates memories and biases cognition. Trends in cognitive sciences, 23(10), 876-890.
https://doi.org/10.1016/j.tics.2019.07.008

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. Behavioral neuroscience, 100(2), 147.
https://doi.org/10.1037/0735-7044.100.2.147

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

Tonegawa, S., Liu, X., Ramirez, S., & Redondo, R. (2015). Memory engram cells have come of age. Neuron, 87(5), 918-931.
https://doi.org/10.1016/j.neuron.2015.08.002

Treder, M. S., Charest, I., Michelmann, S., Martín-Buro, M. C., Roux, F., Carceller-Benito, F., Ugalde-Canitrot, A., Rollings, D. T., Sawlani, V., Chelvarajah, R., Wimber,

M., Hanslmayr, S., & Staresina, B. P. (2021). The hippocampus as the switchboard between perception and memory. Proceedings of the National Academy of Sciences of the United States of America, 118(50), e2114171118.
https://doi.org/10.1073/pnas.2114171118

Tse, D. et al. (2007) Schemas and memory consolidation. Science 316, 76-82
https://doi.org/10.1126/science.1135935

Tse, D. et al. (2011) Schema-dependent gene activation and memory encoding in neocortex. Science 333, 891-895
https://doi.org/10.1126/science.1205274

Tulving, E. (1972). 12. Episodic and semantic memory. Organization of memory/Eds E. Tulving, W. Donaldson, NY: Academic Press, 381-403.

Tulving, E. (1983). Elements of episodic memory.

van Kesteren, M. T., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. Neuropsychologia, 51(12), 2352-2359.
https://doi.org/10.1016/j.neuropsychologia.2013.05.027

van Kesteren, M.T.R., Rijpkema, M., Ruiter, D.J., Morris, R.G.M., and Fernández, G. (2014). Building on prior knowledge: schema-dependent encoding processes relate to academic performance. J. Cogn. Neurosci. 26, 2250-2261.
https://doi.org/10.1162/jocn_a_00630

Vaz, A. P., Wittig, J. H., Inati, S. K., & Zaghloul, K. A. (2020). Replay of cortical spiking sequences during human memory retrieval. Science, 367(6482), 1131-1134.
https://doi.org/10.1126/science.aba0672

Wilber, A. A., Skelin, I., Wu, W., & McNaughton, B. L. (2017). Laminar organization of encoding and memory reactivation in the parietal cortex. Neuron, 95(6), 1406-1419.
https://doi.org/10.1016/j.neuron.2017.08.033

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. Science, 265(5172), 676-679. https://doi.org/10.1126/science.8036517

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. Neuron, 81(2), 267-279. https://doi.org/10.1016/j.neuron.2013.11.005

Wimmer, G. E., Liu, Y., Vehar, N., Behrens, T. E., & Dolan, R. J. (2020). Episodic memory retrieval success is associated with rapid replay of episode content. Nature Neuroscience, 23(8), 1025-1033. https://doi.org/10.1038/s41593-020-0649-z

Wu, X., Viñals, X., Ben-Yakov, A., Staresina, B. P., & Fuentemilla, L. (2021). Post-encoding reactivation promotes one-shot learning of episodes in humans. bioRxiv. https://doi.org/10.1101/2021.04.13.439658

Xue, G. (2018). The neural representations underlying human episodic memory. Trends in Cognitive Sciences, 22(6), 544-561. https://doi.org/10.1016/j.tics.2018.03.004

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. https://doi.org/10.1037/0033-2909.127.1.3

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., ... & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. Nature neuroscience, 4(6), 651-655. https://doi.org/10.1038/88486

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. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. Journal of Cognitive Neuroscience, 23(12), 4057-4066. https://doi.org/10.1162/jocn_a_00078

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: building memories to navigate future decisions. Frontiers in human neuroscience, 6, 70. https://doi.org/10.3389/fnhum.2012.00070

Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. Nature communications, 9(1), 1-11. https://doi.org/10.1038/s41467-018-06553-y

Zheng, J., Schjetnan, A. G., Yebra, M., Gomes, B. A., Mosher, C. P., Kalia, S. K., ... & Rutishauser, U. (2022). Neurons detect cognitive boundaries to structure episodic memories in humans. Nature neuroscience, 25(3), 358-368. https://doi.org/10.1038/s41593-022-01020-w

Zola-Morgan, S. M., & Squire, L. R. (1990). The primate hippocampal formation: evidence for a time-limited role in memory storage. Science, 250(4978), 288-290. https://doi.org/10.1126/science.2218534