RHYTHMIC FLUCTUATIONS IN MNEMONIC SIGNATURES DURING ASSOCIATIVE RECALL

BY

CASPER KERRÉN

A thesis submitted to the
University of Birmingham
for the degree of DOCTOR OF PHILOSOPHY

School of Psychology
College of Life and Environmental Sciences
University of Birmingham
July 2020

Abstract

The intricate linking of information processing and neural representations to the underlying hippocampal neural rhythms during episodic memory retrieval are yet to be fully explored in human subjects. In this doctoral thesis, the temporal order of these relationships was investigated, with emphasis on how the processes evolve and take place over time. Empirical evidence and neural network models suggest that hippocampus and the hippocampal theta rhythm play a central role in episodic memory. In the first two studies, different oscillatory dynamics in the hippocampal circuit thought to provide optimal states for encoding and retrieval were investigated. The third study investigated the role of the hippocampal theta oscillation as an adaptive mechanism in regulating competition between similar memories. And lastly, the fourth study investigated sharp-wave ripples in promoting successful episodic memory retrieval. Throughout the four chapters, memory content was decoded using multivariate pattern classification, and the timing of memory reactivation was linked to two prominent oscillatory brain signatures: the hippocampal theta rhythm on the one hand, and hippocampal sharp-wave ripples on the other. In sum, this doctoral thesis provides support for the powerful computations along the hippocampal theta oscillation, and the close interplay between hippocampus and neocortical areas, foremost at time of retrieval.

To Alva

Piece of baby

Acknowledgements

There is nobody I can thank more for the past years in the lab than my supervisor, Maria Wimber. Not only has Maria offered deep knowledge and interesting discussions within and beyond our research field, but she has also been of enormous support and showed a great understanding of my relatively erratic behaviour outside my doctoral thesis. Whenever I have needed someone to talk to, whatever the subject might have been, Maria has always been there. Maria has also shown me that science and scientific life is worth pursuing, encouraged me to believe in my ideas and to think independently, also in times when things have felt tough. Lastly, I appreciate that we have been able to work together on such equal terms and that a visit to the pub together always have felt so comforting and close. I look up to you.

I also want to thank my second supervisor Simon Hanslmayr for always offering important input and feedback. I immensely admire the way in which Simon is enjoying work-life and again that shows me how worth all the effort has been with this thesis.

Another person who cannot go unnoticed in this section is Bernhard Staresina, whom I would like to thank for having made my doctoral years in the lab very educating and interesting, and outside the lab very fun.

Now, my thanks go to my family, Natalie, and Alva, and especially Natalie who has supported me helped me and listened to me babbling about x, y, and z axes, circular variance, hippocampal subfields, and whatnot for numerous hours without ever

complaining. Without you, this thesis would never have been written. Alva, I also thank you for listening and being a baby.

I also want to say thank you to my parents, my sister, and my sister's husband for encouraging and pushing me when I have needed that, and for listening when I have needed that. I would never have finished this without you.

I want to thank every single one of you in the Memory and Attention Group for making science fun, and in particular, Fede, Juan, and Sebastian. You have been of enormous support and source of happiness. Life came in between; one moved to Germany, another to Sweden, a third to America. The fourth got stuck inside due to Covid-19. Regardless of this, we had some good years together.

Publications and Presentations

At the time of this thesis submission, the following publications and conference contributions were derived from this doctoral research.

Publications

Michelmann, S., Treder, M. S., Griffiths, B., **Kerren, C**., Roux, F., Wimber, M., Hanslmayr, S. (2018). Data-driven re-referencing of intracranial EEG based on independent component analysis (ICA). *J Neurosci Methods*, 307, 125-137. doi:10.1016/j.jneumeth.2018.06.021

Kerren, C.*, Linde-Domingo, J.*, Hanslmayr, S., & Wimber, M. (2018). An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval. *Current Biology*, 28(21), 3383-3392 e3386. doi:10.1016/j.cub.2018.08.065

Linde-Domingo, J., Treder, M. S., **Kerrén, C.**, & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, 10(1), 179. doi:10.1038/s41467-018-08080-2

* These authors contributed equally

Conference contributions

Kerren, C., & Wimber, M. (2019). Phase coding of competing memories along the hippocampal theta oscillation in human MEG. Poster at Conference on Cognitive Computational Neuroscience (CCN), Berlin, Germany.

Kerren, C., Linde-Domingo, J., Hanslmayr, Simon. & Wimber, M. (2018). An optimal oscillatory phase for pattern reactivation during memory retrieval. Poster at Learning and Memory 2018, Huntington Beach, USA.

Kerren, C., Linde-Domingo, J., Hanslmayr, Simon. & Wimber, M. (2017). Reactivation of memory patterns in the human EEG rhythmically fluctuates in the theta range. Poster at International Conference for Cognitive Neuroscience (ICON), Amsterdam, The Netherlands.

Table of Contents

Chapter 1 - Introduction 1

Principles of memory 5
1. Episodic memory 5
2. Episodic memory and the brain 7
3. Encoding, consolidation and retrieval computations in episodic memory 9
   3.1. Encoding 9
   3.2. Consolidation 10
   3.3. Retrieval 11
4. Tracking memory-related patterns in the brain 13
5. Neural rhythms 14
   5.1 Theta rhythm 15
6. Computational models of the role of theta 18
   6.1. Encoding-retrieval flip model 18
   6.2. Theta competition model 19
7. Hypotheses 22

Chapter 2 - An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval 24

Abstract 25
1. Introduction 26
2. Methods 28
   2.1. Experimental Model and Subject Details 28
      2.1.1. Participants 28
3. Method details 29
   3.1. Material and setup 29
   3.2. Paradigm 30
4. EEG data analysis 32
   4.1. Preprocessing 32
   4.2. Multivariate pattern analysis 33
   4.3. Power spectrum of the classifier fidelity time series 34
   4.4. Phase-amplitude coupling between EEG data and fidelity values 36
   4.5. Using classifier-locked averages to relate classifier outputs to the phase of the ongoing

EEG-signal 38
4.6. Event-related potential analysis 41
4.7. Source analysis 42
4.8. Distribution of fidelity values across time 43
4.9. Time generalisation 43
4.10. Identifying oscillating frequencies 44
5. Results 46
5.1. Participants retrieve the episodic memories with high accuracy 46
5.2. Power spectrum of classifier shows strongest effects in lower frequencies 47
5.3. Phase-amplitude coupling reveals oscillating patterns at retrieval for 8Hz 50
5.4. Classifier-locked averages reveal a consistent theta phase prior to memory reinstatement 51
5.5. High classifier fidelity is associated with strong theta phase consistency in MTL 54
5.6. Theta phase-locking is unlikely to be produced by early cue-related effects 55
5.7. EEG signals at the exact time points of maximal classifier fidelity show content-dependent differences with a source in anterior temporal lobe 56
5.8. Classifiers that generalise from encoding to retrieval show similar frequency characteristics 58
6. Discussion 60
7. Acknowledgements 66
8. Authors contributions 67
9. Declaration of interests 67
10. Quantification and statistical analysis 67
10.1. Behavioural data 67
10.2. EEG data 68
11. Data and software availability 70
12. Supplemental information 70
12.1. Frequency characteristics of the encoding classifier timecourses 71
12.2. Generalization from encoding to retrieval 72

Chapter 3 - Hippocampal Theta Phase-flip between Encoding and Retrieval 77

Abstract 78
1. Introduction 79

2. Method details 83
2.1. Participants 83
2.2. Experimental procedures 83
2.3. Electrode selection 84
3. Analysis 85
3.1. Preprocessing 85
3.2. Multivariate pattern analysis 86
3.3. Maximum spectral power for hippocampal channels 87
3.4. Phase-amplitude coupling between fidelity values and hippocampal channels 88
4. Results 89
4.1. Behavioural results 89
4.2. Significant decoding performance revealed for both encoding and retrieval. 90
4.3. Maximum power in low frequencies for hippocampal channels 92
4.4. Hippocampal phase modulation of neocortical patterns 92
4.5. Correct rejection mimics encoding patterns 95
5. Discussion 96
6. Acknowledgements 106
7. Authors contributions 106
8. Supplementary material 106

Chapter 4 - Competing Memories are Separated by the Hippocampal Theta Rhythm. 108

Abstract 109
1. Introduction 110
2. Method details 112
2.1. Participants 113
2.2. Material and setup 113
2.2.1. Paradigm 114
3. Analysis 118
3.1. Statistical analysis 118
4. MEG data analysis 119
4.1. Preprocessing 119
4.2. Time-frequency decomposition 121
4.3. Multivariate pattern analysis 122
4.4. Determine peak frequency of fidelity values using MODAL 123
4.5. Source analysis 125
4.6. Phase-amplitude coupling between MEG data

and fidelity values 126
4.7. Extraction of fidelity values 127
4.8. Absolute phase difference between fidelity values for targets and competitors 129
5. Results 129
5.1. Memory performance 129
5.1.1. Lowest memory performance was obtained in the competitive condition 129
5.1.2. Participants picked the competing association more than could be explained by chance 130
5.1.3. Memory performance correlates with number of intrusions and the amount of interference 131
5.2. Reaction time (RT) 132
5.2.1. At retrieval, NC-1 had slowest reaction times 132
5.3. MEG Results 133
5.3.1. Theta power increase for CC compared to NC-2 133
5.3.2. Fidelity values fluctuate at theta frequency 135
5.3.3. Fidelity value extraction 135
5.3.4. Hippocampus theta oscillation clocks the reactivation of target and competitor memories 138
5.3.5. Phase difference shows a functional relationship to behaviour 140
5.3.6. Participant with large decline in competitor reactivations show significant phase difference 142
6. Discussion 142
7. Acknowledgements 151
8. Authors contributions 151

Chapter 5 - Episodic Memory Reinstatement in Neocortex at Time of Hippocampal Sharp-wave Ripples. 152

Abstract 153
1. Introduction 154
2. Method details 156
2.1. Participants 156
2.2. Experimental procedures 157
2.3. Implementation of depth electrodes 158
2.4. Electrode selection 159
2.5. Preprocessing 159
2.6. Ripple detection 160
2.7. Isolation index 161
2.8. Multivariate pattern analysis 161

3. Results 162
3.1. Ripple density 162
3.2. Majority of ripples occurring in isolation 163
3.3. Neocortical reinstatement surrounding ripple events 164
4. Discussion 166
5. Acknowledgements 172
6. Authors contributions 172
7. Supplementary material 173

Chapter 6 – General discussion 174

1. Principal findings 176
1.1. The rhythmic modulation of encoding and retrieval of episodic memories. 176
1.2. Interference along the hippocampal theta rhythm 178
1.3. Neocortical reinstatement surrounding hippocampal SPW-Rs 179
2. Converging and non-converging findings across studies 182
2.1. Phase-coding mechanisms and theta oscillations 182
2.2. The inconsistency of peak frequency 185
2.3. The relationship between theta-locked classification and other commonly used approaches 188
3. A proposed time course of episodic memory retrieval 193
4. Future research 198
4.1. How many competing items can be assigned a slot along the theta oscillation? 199
4.2. Can the post-encoding period act as mini-consolidation? 200
4.3. Are SPW-Rs locked to the phase of the theta oscillation? 202
5. Concluding remarks 205
References 206

List of Figures

i.	Chapter 1. Figure 1	18
ii.	Chapter 1. Figure 2	20
iii.	Chapter 2. Figure 1	46
iv.	Chapter 2. Figure 2	48
v.	Chapter 2. Figure 3	53
vi.	Chapter 2. Figure 4	59
vii.	Chapter 2. Supplementary Figure 1	73
viii.	Chapter 2. Supplementary Figure 2	74
ix.	Chapter 2. Supplementary Figure 3	75
x.	Chapter 2. Supplementary Figure 4	76
xi.	Chapter 3. Figure 1	82
xii.	Chapter 3. Figure 2	91
xiii.	Chapter 3. Figure 3	94
xiv.	Chapter 4. Figure 1	115
xv.	Chapter 4. Figure 2	132
xvi.	Chapter 4. Figure 3	134
xvii.	Chapter 4. Figure 4	136
xviii.	Chapter 4. Figure 5	139
xix.	Chapter 5. Figure 1	156
xx.	Chapter 5. Figure 2	163
xxi.	Chapter 5. Figure 3	166

Abbreviations

Analysis of Variance	ANOVA
Centre for Human Brain Health	CHBH
Complementary Learnings Systems Framework	CLS
Contrastive Hebbian Learning	CHL
Cornu Ammonis	CA
Correct Rejection	CR
Dentate Gyrus	DG
Electroencephalography	EEG
Entorhinal Cortex	EC
Event-related Potential	ERP
Fast Fourier Transformation	FFT
Full-width at Half Maximum	FWHM
Independent Component Analysis	ICA
Intertrial Phase Clustering	ITPC
Intracranial Electroencephalography	iEEG
Irregular-Resampling Auto-Spectral Analysis	IRASA
Kullback-Leibler Distance	KL
Linear Constrained Minimum Variance	LCMV
Linear Discriminant Analysis	LDA
Local Field Potential	LFP
Long-term Depression	LTD
Long-term Potentiation	LTP

Magnetic Resonance Imaging	MRI
Magnetoencephalography	MEG
Mean	M
Medial Temporal Lobe	MTL
Modulation Index	MI
Multivariate Pattern Analysis	MVPA
Phase-locking Index	PLI
Phase-locking Value	PLV
Proactive Interference	PI
Representational Similarity Analysis	RSA
Retrieval-induced Forgetting	RIF
Sharp-wave Ripples	SPW-R
Slow-wave Sleep	SWS
Source Correct	SC
Source Incorrect	SI
Standard Deviation	SD
University of Birmingham Research Ethics Committee	STEM

Chapter 1 - Introduction

Understanding the dynamics, processes and temporal dimensions of human memory retrieval have long been a challenge in the field of cognitive neuroscience. Arguably, who we are as human beings is constructed by what we remember, and indeed what we forget - yet, how and when such memory processes unfold in the brain remains unclear. Early studies on patients diagnosed with amnesia or severe memory deficits during the 1950s highlighted the pivotal role of the hippocampus in associative memory (Scoville & Milner, 1957). Various computational models have later been developed to explain what exact function the hippocampus plays in episodic memory (Alvarez & Squire, 1994; McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; O'Reilly, Bhattacharyya, Howard, & Ketz, 2014; O'Reilly & Norman, 2002; Teyler & DiScenna, 1986; Teyler & Rudy, 2007). For example, the prominent Complementary Learnings systems [CLS] framework (McClelland et al., 1995) attempts to explain why the brain requires on the one hand the hippocampal system and on the other neocortex to complement each other in learning and memory function. Hippocampus is modelled as a fast learner that can capture and store traces of unique, multi-element episodes, while neocortex adjusts its weights much more slowly, in a way that optimally reflects the statistical regularities of the environment. One of the central assumptions of the framework is that at encoding, hippocampus separates incoming patterns to make them as dissimilar as possible so as to avoid *catastrophic interference*, and at retrieval, hippocampus responds to reminders that partially overlap with a stored memory trace by reinstating the original encoding-related pattern in the neocortex (Marr, 1971; Yassa & Stark, 2011). Studies on rodents have yielded support for these computations and furthermore linked them to specific phases of the hippocampal theta rhythm (Amaral & Witter, 1989; Pavlides, Greenstein, Grudman,

& Winson, 1988). However, the intricate linking of information processing and neural representations to the underlying hippocampal neural rhythms are yet to be fully explored in human subjects. Specifically, in this thesis, the focus is on linking the timing of memory reactivation to two prominent oscillatory brain signatures: the hippocampal theta rhythm on the one hand, and hippocampal sharp-wave ripples (SPW-Rs) on the other.

It is from this point that the research questions of this thesis have emerged: it firstly serves to expand the understanding of the human hippocampus and its role in memory retrieval but also aims to bridge important gaps between findings from rodent studies, computational modelling and human brain data.

This thesis is composed of four studies, where brain data in human subjects are analysed at two different anatomical scales: whole-brain recordings using electroencephalography (EEG) (Chapter 2) and magnetoencephalography (MEG) (Chapter 4) on the one hand; and intracranial EEG (iEEG) recordings directly from the hippocampus and surrounding neocortex on the other (Chapters 3 and 5). Together, these separate studies provide a new understanding of the role of the hippocampus in episodic memory, with particular focus on the temporal dynamics of memory retrieval. The first three studies are built on computational models assigning different roles of the hippocampal theta rhythm in episodic memory retrieval. More specifically, these chapters are concerned with the central hypothesis that neural patterns reflecting retrieved mnemonic contents are coupled to the hippocampal theta rhythm, with the aim of investigating how the rhythm separates different, presumably interfering

cognitive operations. The fourth study explicitly tests whether SPW-Rs are the vehicle for such reinstated mnemonic content, and thereby provides a link to the second oscillatory signature, apart from the theta rhythm, that has been described in the relation to the hippocampus.

In the first chapter, EEG data is analysed to test the idea that the phase of the human hippocampal theta rhythm acts as a separator for encoding and retrieving a memory, paralleling findings from rodents. The central prediction from such phase separation, confirmed by the results in the present thesis, is that the neural signatures of memories at the time of encoding and retrieval rhythmically fluctuate, albeit in an anti-phasic (i.e., 180^{o} phase-shifted) manner. The second chapter is a follow-up to this initial study, where iEEG data is analysed to specifically link these fluctuations between encoding and retrieval to the hippocampal theta rhythm, by recording directly from the hippocampus and adjacent neocortical areas. Together, these studies shed light onto the intricate temporal dynamics of memory encoding and recall, and how these processes are coupled to the hippocampal theta rhythm. In the third chapter, MEG data are analysed to test the hypothesis that the hippocampal theta rhythm separates overlapping, competing memories. Lastly, the final chapter presents the results of an intracranial EEG study, testing the idea from rodent research that SPW-Rs trigger the replay of previously stored memories from hippocampus to neocortex. Together, these four independent studies offer novel insights into the temporal dynamics of hippocampal engagement in human episodic memory retrieval, moving beyond the existing body of research on the topic, which has to date mainly relied on results from rodent subjects or computational models.

Principles of memory

Overview

This introduction offers an overview of some basic principles of episodic memory which build the framework upon which the research in this thesis has been based. Theories surrounding the hippocampal role in memory encoding, storage, and retrieval are specifically considered here. For this research, encoding and retrieval are of specific importance with less emphasis placed on consolidation. Following this introduction, theories on the role of neural rhythms are reviewed with particular attention to the theta rhythm. Two computational models are described, both suggesting different roles of the hippocampal theta rhythm: the first highlighting a potential role in encoding and retrieval, and the latter in memory interference resolution. These models lay the foundations of much of the empirical work in the thesis.

1. Episodic memory

Episodic memory is a system that stores multidimensional constructs, containing temporally dated events and the temporal-spatial relationships between them, which allow the subject to mentally travel in time to relive the past, experience the present and predict the future (Howard & Eichenbaum, 2013; Tulving, 1972, 2002; Tulving & Markowitsch, 1998). The ability to encode and retrieve memories is essential for

cognitive functioning and a basis for learning, making these memories a guide for everyday life. Episodic memories differ in their quality from semantic memories. Although both being declarative, the latter refers to general knowledge about the world, such as Malmö being located in the south of Sweden (Collins & Quillian, 1969). However, in contrast to semantic memories, episodic memories are built up from associations between distinct perceptual inputs happening at a specific time and place, which together form personal and unique memories that can be consciously accessed (Cohen & Squire, 1980; Place, Farovik, Brockmann, & Eichenbaum, 2016). A memory of going to lunch with a colleague in Malmö is unique in such a way that the context (e.g., time and place) in which it was encoded is used to individuate that event from all other events (Tulving, 2002). The process of remembering that specific occasion will happen without the event being present in the external world, and it is therefore considered an internal process unique to that specific individual. Note, however, that most retrieval is cue-dependent and can be generated internally or externally. The specificity of the cue is key for retrieval success (Tulving, 1999; Tulving & Thomson, 1973). In cued recall paradigms in the laboratory (and used in Chapters 2 and 4 of this thesis), retrieval is initiated explicitly by a memory cue, that is, a piece of information that was present at encoding, while associated information has to be retrieved entirely from memory. Cued recall differs in this respect from a recognition test (used in Chapters 3 and 5 of this thesis), where the physical information that was present at encoding is again presented at retrieval, and it is the task of the subject to decide whether the material has been seen before or not. Such recognition decisions can be made based on a familiarity signal, or through the recollection of additional contextual information about the original experience, the

latter involving processes more akin to recognition (Norman & O'Reilly, 2003). Today there is an agreement that the processing elicited by the cue at retrieval overlapping with that engaged at encoding is crucial for successful memory retrieval (Rugg, Johnson, Park, & Uncapher, 2008; Tulving, 2002). In the example above, the overlap could be the contextual surroundings of being in Malmö (cued recall), or more unlikely, a video showing you and your friend having lunch (recognition test). Importantly, a memory cue is hardly ever unique to one association and can be tied to several similar memories, for example, another lunch in Malmö. When trying to retrieve that specific lunch with that specific colleague, other lunches are assumed to also be reactivated and trying to compete for retrieval (Anderson, Bjork, & Bjork, 1994). As described further down, the brain presumably has several different processes in place to handle such mnemonic interference.

2. Episodic memory and the brain

Early studies on patients with hippocampal pathology that had led to amnesia and severe memory deficits show that the medial temporal lobe (MTL), and especially the hippocampus, has an important role in episodic memory (Eichenbaum, 2010). However, it was not until the report of 10 patients (among them the famous case of patient Henry Molaison, H.M) by Scoville and Milner (1957) that a universal agreement was reached that put hippocampus at the centre of importance in episodic memory. It is today widely agreed that the MTL, consisting of the hippocampus and surrounding cortical areas are crucial for a successful episodic memory (Frankland, Josselyn, & Kohler, 2019). MTL is not acting in isolation, nor are memories to be

found in one place. Although most theories today hold that neocortex can learn without hippocampus, but much more slowly (i.e., statistical learning) and with less flexibility (McClelland et al., 1995), it is generally agreed that MTL needs to receive input and administer output to other regions of the brain, where special importance is the connection between MTL and widespread neocortical association areas (Eichenbaum, 2010). This widespread network of activity is then stored as long-lasting changes in the brain (Lashley, 1950).

With the capability of the brain to store and recall an almost infinite amount of information comes, as mentioned before, the inevitability of storing overlapping memories. A high degree of overlap between memories can result in catastrophic interference (McClelland et al., 1995; McCloskey & Cohen, 1989), meaning that new information will overwrite previously learned material. The present thesis will mainly refer to two frameworks or theories, both suggesting that the brain utilises two interacting systems to avoid catastrophic interference: The Complementary Learnings System framework (McClelland et al., 1995), and the hippocampal memory indexing theory (O'Reilly et al., 2014; Teyler & DiScenna, 1986; Teyler & Rudy, 2007). Both are inspired by David Marr's early work (Marr, 1971), in which he proposed that the difference in histology between the hippocampus and neocortex allows for complementary computations. Hippocampus, as one part of this system, connects to each sensory modality and can do simple one-shot associative memorizing, whereas neocortex, the second part of the system can based on statistical regularities, reclassify the input and build networks between associated and interconnected hubs representing

the content of a memory (Marr, 1971; McClelland et al., 1995; Teyler & DiScenna, 1986; Teyler & Rudy, 2007).

3. Encoding, consolidation and retrieval computations in episodic memory

3.1. Encoding

Perception of a visual stimulus starts in the retina and reaches the visual cortex in the occipital lobe through the thalamus. From here, signals are transferred to association areas of the cortex. More specifically, the ventral pathway is concerned with *what* is being perceived, whereas the dorsal pathway codes for *where* it is being perceived (Goodale & Milner, 1992). Both streams end up in entorhinal cortex (EC), with an intermediary station in the perirhinal and parahippocampal cortex, respectively. Although memories are widely distributed in the brain, an injury to MTL can be detrimental specifically for forming new episodic memories. This is a result of the projection of activity from the perirhinal and parahippocampal cortex to the hippocampal formation, where activity from these two streams is assumed to be combined and transformed into an index representing the cortical activity elicited during encoding of the particular event (Teyler, DiScenna, 1986). The processing of items is assumed to involve the lateral EC projecting to Cornu Ammonis 1 (CA1) of the hippocampus and back (Hargreaves, Rao, Lee, & Knierim, 2005), whereas the processing of where and when involves the medial EC projecting to CA1 and back. As mentioned earlier, the neural patterns of a specific lunch will overlap with stored

neural patterns of other similar lunches. Trying to remember one of these episodes will inevitably reactivate other, overlapping episodes. To avoid potential interference at retrieval, EC additionally sends input to the dentate gyrus (DG) at encoding. In DG, patterns are thought to be orthogonalised (pattern separated), that is, they are transformed into being as different from each other as possible. Orthogonalised signals are then sent to CA3, where the sparsely coded input from DG is combined with signals from EC (Marr, 1971; McNaughton & Morris, 1987). CA3 then back-projects to itself using Schaffer collaterals, and to CA1 using the commissural pathway. Presumably, the memory system is dependent on and will not work sufficiently without pattern separation, as neural representations of old memories similar to what is to-be-encoded will otherwise be overwritten (Yassa & Stark, 2011). The goal of the encoding process is to integrate this new information into long-term memory (Hebb, 1949; Roy et al., 2019).

3.2. Consolidation

Post encoding, the labile memories are being stabilised through consolidation (Dudai, 2004). The mere fact that the hippocampus holds sparsely coded representations makes it a bad contestant for the storage of the entire memory trace. Instead, the hippocampus' role in most models is to teach and migrate the information to, neocortex (Buzsaki, 1996; McClelland et al., 1995). During sleep, the encoding of new material cannot disturb the consolidation process and makes this the optimal time for redistributing and strengthening new memories without the problem of interference (Diekelmann & Born, 2010). Specifically, during slow-wave sleep (SWS), slow

rhythms in neocortex bias hippocampus to reactivate the encoded events. At the same time, slow rhythms trigger the release of sleep spindles, originating in the thalamus (Sirota, Csicsvari, Buhl, & Buzsaki, 2003). Spindles in turn bundle hippocampal SPW-Rs, which originate in CA3 (sharp waves) and CA1 (ripples), and are thought to represent compressed information of previous events (Lee & Wilson, 2002). The synchrony between spindles and ripples is assumed to promote the transfer of compressed versions of previously experienced events from the hippocampus to the neocortex for long-term storage (Diekelmann & Born, 2010; Peyrache, Khamassi, Benchenane, Wiener, & Battaglia, 2009; Staresina et al., 2015). Over time, consolidation changes the memory representation and the content of new memories will be stabilised and integrated into already existing long-term memory networks (system consolidation) in neocortical areas (Diekelmann & Born, 2010). Recently, evidence is emerging that memories are stabilised not only during sleep, but also during other offline periods after learning, and during active retrieval (Antony, Ferreira, Norman, & Wimber, 2017). It is today an open question of whether awake stabilisation involves the same neural processes as sleep-dependent consolidation.

3.3. Retrieval

Retrieval starts when an internally or externally generated cue activates parts of the original input in the neocortex, or as Semon wrote: “the influences which awaken the mnemic trace or engram out of its latent state into one of manifested activity” (Semon, 1921, p. 12). Given that the input pattern to the neocortex is similar enough to what has been laid down at encoding, CA3 is assumed to complete the missing neural

patterns representing the original experience. As described before, CA3 is also implicated in pattern separation, a process incompatible with pattern completion (Bakker, Kirwan, Miller, & Stark, 2008). To ensure that incorrect retrieval is not sent back to EC, and to know when CA3 is supposed to pattern separate or pattern complete, CA1 has been suggested to match input from CA3 with that of EC to subsequently shift the system from one process to the other (Eichenbaum & Buckingham, 1990; Levy, 1989). In later paragraphs, a computational model that has capitalised on the shift of processes will be explained (Figure 1). If the pattern in CA3 calls for the system to enter a retrieval mode, it triggers activity in the recurrent connections within CA3 and strengthened connections between CA3, CA1, and subiculum. Subiculum subsequently combines the signal from CA1 with that of EC and projects to EC, which in turn projects to neocortical areas (Chrobak, Lorincz, & Buzsaki, 2000; Staresina & Wimber, 2019).

Based on theoretical work and studies in rodents, it is believed that a loop between the hippocampus and cortical areas is governed by the means of SPW-Rs (Buzsaki, 1996, 2015; Swanson, Levenstein, McClain, Tingley, & Buzsáki, 2020). During replay and retrieval of previous experiences, cortical areas bias hippocampal processing and thereby the information sent by hippocampal SPW-Rs back to neocortical areas for reinstatement. Intriguingly, new evidence from studies in human subjects shows that during awake memory recall, reinstatement in (the lateral temporal) cortex is related to an increased density of ripples in MTL, and to be strongest when ripples in temporal cortex follow ripples in MTL as expected during an outwards information flow from the hippocampus to the neocortex (Vaz, Inati, Brunel, & Zaghloul, 2019). These

findings show a new role of ripples outside that of offline consolidation and replay, an area that is further explored in Study 4.

After retrieval signals have left the hippocampus, regions that were active during the initial experience of that event will now be targeted, and as time unfolds and the interaction between hippocampus and neocortex continues, the subject experiences a full conscious memory (Danker & Anderson, 2010; Staresina & Wimber, 2019; Vaz et al., 2019). The intricate temporal dynamics of the communication between the hippocampus and the neocortex, and of the resulting memory reinstatement, are only starting to be investigated, and are the central focus of the present thesis.

4. Tracking memory-related patterns in the brain

Novel analysis techniques, such as multivariate pattern analysis (MVPA), have advanced the ability to capture wide-spread memory representations in the brain as they are reactivated during retrieval (Norman, Polyn, Detre, & Haxby, 2006). The current thesis heavily relies on indices of memory reinstatement derived from such multivariate pattern classification algorithms. For the majority of experiments in the present thesis, these algorithms are used in the context of associative-learning paradigms where participants learn to associate two unrelated bits of information (e.g. a word and an image), and are later asked to recall one (e.g. the image) when cued with the other (e.g., the word). Mathematical models can then be built based on known (training/encoding) data, where for example neural patterns of when a subject learns a word-scene association is contrasted with when they learn a word-object association.

When at retrieval, the word prompts the subject to retrieve the association made at encoding, the trained models are used to make predictions about unknown (test) data. The central rationale is that if the classifier can pick up a signal during cued recall that reliably differentiates the retrieved content (e.g. type of image, animate vs inanimate object), despite the image not being physically present, then the fidelity of such a classifier can be used as an index for the successful reactivation of memory.

The classifiers used in the present thesis solely rely on spatial patterns for each sample point to estimate how much evidence there is for the pattern belonging to one or the other class. A fidelity value (i.e., an index of memory reactivation) can thus be obtained for each single time point after cue onset, uncovering the temporal dynamics of memory retrieval. Previous findings using time-resolved classification have offered a new understanding of the temporal dynamics of object recognition during visual perception (Cichy, Pantazis, & Oliva, 2014), and episodic memory encoding (perception) and retrieval (Linde-Domingo, Treder, Kerrén, & Wimber, 2019).

5. Neural rhythms

So far a summary of how the brain constantly encodes, consolidates, and retrieves personal memories - memories that are in constant transformation - has been provided. The neural processes behind these computations are many, and most of them seem to depend on the MTL and foremost hippocampus. Without this region, it is difficult if not impossible to store new memories, making the individual lost in the past. What is

missing so far is an overarching process that can orchestrate these computations in time, and can link together the neural assemblies coding for different parts of a memory. Described in the next part of this thesis is one prominent candidate for such an overarching process, namely neural rhythms.

Summed activity from a large number of neurons creates waves of fluctuating activity in the local field potential (LFP) (Buzsaki, 2002). If neurons fire in synchrony, the sum of the activity can be detected and recorded on different anatomical scales using recording methods such as EEG, MEG, and iEEG. Neurons that fire together temporarily create ensembles thought to represent a specific aspect of memory (Hebb, 1949). Across the brain, it is thought that several of these ensembles of neurons together represent a memory in full, and the necessary synchronous firing within and across ensembles is aided by neural oscillations. More specifically, transferring of information from one group of neurons to another, and strengthening of synapses between them, demand that these two groups are excitable at the same time, something that can only be achieved through phase synchronisation, most likely in the lower frequency bands (Fell & Axmacher, 2011; Fries, 2005).

5.1. Theta rhythm

Theta oscillations are one of the most prominent oscillations in the (rat) brain. They are most likely generated in the medial septum (Vertes & Kocsis, 1997), with the neurons in medial septum being phase-locked to neurons in the hippocampus (Hangya, Borhegyi, Szilagyi, Freund, & Varga, 2009); however see Lubenov and Siapas (2009)

for an alternative view. The clear functional relationship to behaviour has been shown in numerous studies. In one of the first studies, it was shown that the theta amplitude was larger when rats were moving than when standing still (Vanderwolf, 1969). Similarly, whisking and sniffing in rats and cats have been shown to follow the frequency of theta (Berg & Kleinfeld, 2003; Macrides, Eichenbaum, & Forbes, 1982). However, one of the most intriguing findings both for the understanding of the brain in general, but in particular for the studies in the present thesis, is that of the relationship between theta oscillations and place cells. Originally, this phase-coding mechanism in the hippocampus was discovered in 1993 by John O'Keefe and Michael Recce (John O'Keefe & Michael L. Recce, 1993), when they showed that firing of cells that code for a specific location in an environment (i.e., place cells) was timed to the phase of the slow (theta) oscillation in the surrounding cells. A related phenomenon is phase precession, in which cells are firing earlier in relation to the underlying oscillation when the rat approaches the place field. This means that as the rat moves closer to the place field, the cell fires further and further away from the theta peak (John O'Keefe & Michael L. Recce, 1993), indicating that the phase of the hippocampal theta oscillation is carrying important information about an animal's location in space (Burgess & O'Keefe, 1996). This seemingly adaptive coding along the oscillation has been further shown to support the timing for changing the synaptic strength between neurons (Pavlides et al., 1988), which has been modelled in several theta-based computational models (Hasselmo, Bodelon, & Wyble, 2002; Kunec, Hasselmo, & Kopell, 2005; Manns, Zilli, Ong, Hasselmo, & Eichenbaum, 2007) (Figure 1). Although early work was exclusively done on rodents, the theta code seems to be more similar between rodents and humans than has previously been implicated

(Bohbot, Copara, Gotman, & Ekstrom, 2017). Outside spatial navigation, human theta oscillations have been implicated in working memory and episodic memory encoding and retrieval (Buzsaki & Moser, 2013; Kahana, 2012). Earlier work focused on the amplitude of the hippocampal theta oscillation (Buzsaki, 2002), but more recent work suggests that theta phase might play a similar role as in the rodent hippocampus. For example, it was recently shown that neocortical gamma bursts are phase-locked to hippocampal theta during a sequential episodic encoding task (Heusser, Poeppel, Ezzyat, & Davachi, 2016), when subjects needed to separate six items for future retrieval. Similar phase separation has been shown in a goal-directed spatial memory task (Kunz et al., 2019). Furthermore, the phase of the theta oscillation also shows a functional role in neocortical areas. In a working-memory task, posterior fast oscillations were nested into different phases of the frontal theta depending on cognitive demands (Berger et al., 2019). Based on these empirical findings and the neural network models previously mentioned, three out of four projects in this thesis capitalised on the important characteristics of the hippocampal theta oscillation in human episodic memory, with a specific focus on the role of theta phase during episodic memory retrieval. In Chapters 2, 3, and 4, it is assumed that the oscillation is clocking memory reactivation. Specifically, in Chapters 2 and 3 it is investigated whether the different presumed hippocampal dynamics for encoding and retrieval result in decodability of memories in different phases of the oscillation. In Chapter 4, a specific model suggesting that overlapping, competing memories are separated along the oscillation is tested.

In the following section two computational models that have capitalised on the relevance of the hippocampal theta rhythm for synaptic modifications are described.

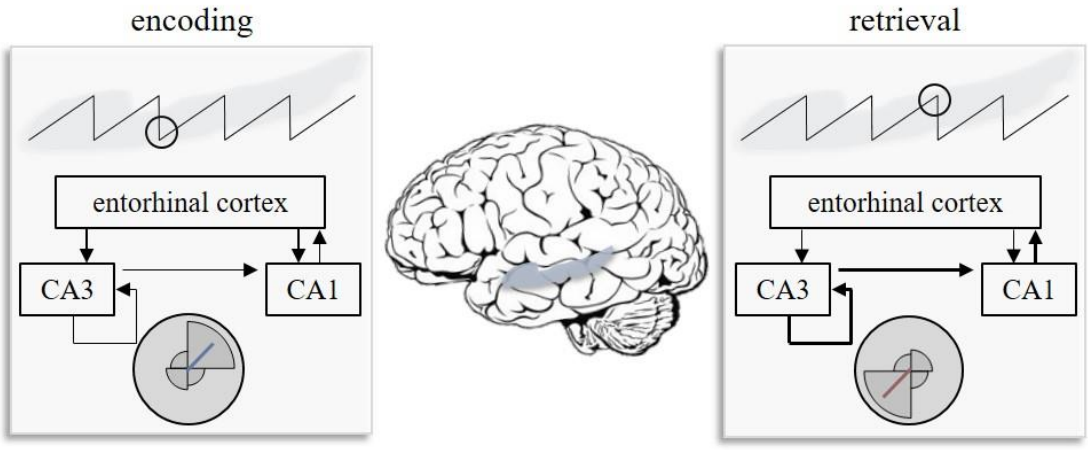

Figure 1. **Encoding and retrieval dynamics in hippocampus.** At encoding, the hippocampal formation receives strong input from EC, at the same time as there is weak activity in CA3 and between CA3 and CA1. At retrieval, the input from EC is weak, at the same time as the activity in CA3 and between CA3 and CA1 is strong. These different dynamics of the hippocampal system set the stage for different processes at encoding and retrieval.

6. Computational models of the role of theta

6.1. Encoding-retrieval flip model

Building on the association between the theta rhythm and phasic changes in the magnitude of synaptic currents in different layers of hippocampus, theta-based models suggest encoding an event is associated with one particular phase of the rhythm, whereas retrieving the event is associated with the opposite phase of the rhythm (Hasselmo, 2005; Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007)

(Figure 1). The phase that has been related to the encoding of a new event is characterized by strong synaptic input from EC to the hippocampal regions, while at the same time there is only weak activity in all recurrent collaterals in CA3, and along the output pathways from CA3 to CA1, and from the hippocampus to EC. This connectivity pattern thus optimally sets the stage for encoding an event, a process assumed to be associated with long-term potentiation (LTP). During the opposite phase of the theta rhythm related to retrieval, input from EC into the hippocampus is weak, while the intraregional signalling and output pathways from the hippocampus to the neocortex are strongly active. The models assume that synapses are therefore more likely to undergo de-potentiation or long-term depression (LTD) rather than LTP, which prevents encoding of potentially incorrect retrieval activity (Hasselmo et al., 2002). The separation of encoding and retrieval occurs recurrently and is crucial for separating material that could potentially interfere with each other (Hasselmo, 2005). Although evidence for this model has been shown in rodents (Colgin et al., 2009; Douchamps, Jeewajee, Blundell, Burgess, & Lever, 2013; Griffin, Eichenbaum, & Hasselmo, 2007), evidence in human subjects is lacking. The aim of the two first studies in Chapter 2 and 3 was, therefore, to test the assumption in human subjects that across the different phases of the hippocampal theta rhythm, the memory system fluctuates between representing incoming information relevant for the encoding of new events, and reactivated internal information representing past events.

6.2. Theta competition model

The memory system encodes, stores, and modifies memories constantly. While awake, encoding and retrieval are necessary processes to interact with the world. Because we encounter the same contexts, people, and themes repeatedly, many of our memories are very similar. Retrieval cues are thus highly likely to simultaneously reactivate several competing memories. Co-activation can give rise to interference between overlapping memories, which has long been theorised to be one of the main causes of forgetting (Baddeley & Logie, 1999; Underwood & Postman, 1960).

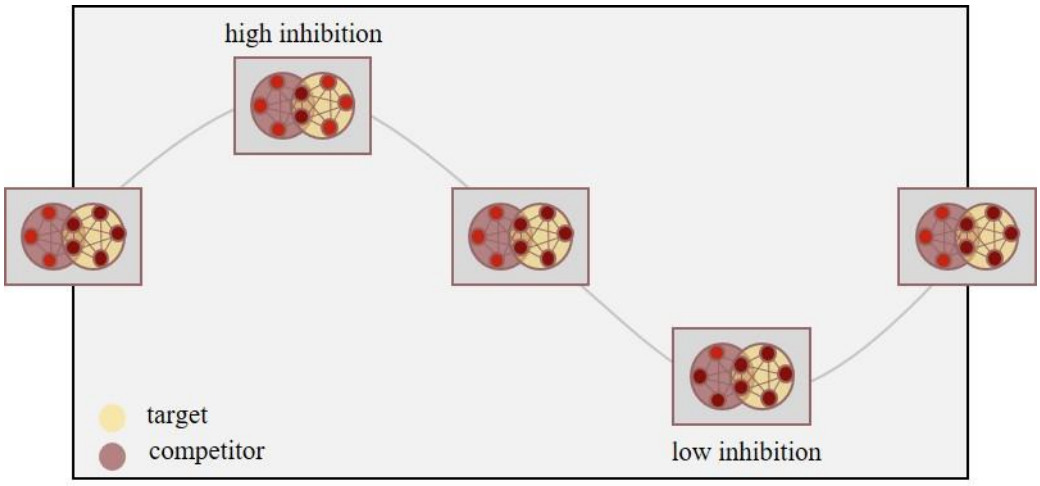

Figure 2. **Theta competition model.** Target and competitor memories are being strengthened and weakened at different phases of the oscillation (indicated by the grey line). When inhibition is going from normal to high and back to normal, target memories are strengthened through LTP. When inhibition is going from normal to low and back to normal, competitor memories are punished. This pruning of memories continues for a few cycles and is assumed to solve competition between similar memories by reducing neural overlap.

A promising computational model is using the theta rhythm to model how the brain separates overlapping memories during retrieval (Norman, Newman, Detre, & Polyn, 2006; Norman, Newman, & Detre, 2007). The model capitalises on the strong link between inhibitory interneurons and the theta rhythm (Buzsaki, 2002), to show how target memories are strengthened by LTP at one phase, whereas competing memories

are weakened through LTD at another phase. More specifically, by using a contrastive Hebbian learning (CHL) weight change equation (Ackley, Hinton, & Sejnowski, 1985), a more desirable state ("pure" target memory state) is contrasted to a less desirable state (co-activation of target and competing memory). The desirable state is strengthened at the peak of the theta rhythm when inhibition is high, resulting in weak features of the target memory falling under the inhibition threshold. Adjusting network weights allows for strong features of the target memory to spread their activation to the weak features and thereby strengthen the target memory (by LTP) for the next-coming cycle. In the less desirable state inhibition is low, which allows strong features of the competing memory to reach above the threshold, in addition to the target features. The transition from this state to a higher inhibition level allows for the strong competitor features to be identified and down-regulated by LTD (Norman et al., 2007).

The model was successful in showing evidence for how the theta rhythm (or, in fact, any slow oscillation) can function as a modulator for the strengthening and weakening of overlapping memories. However, empirical evidence from human subjects supporting this model is still missing, and the question of whether the theta rhythm has this functional role therefore remains unanswered. The third project, in Chapter 4 of this thesis, was concerned with testing the assumptions of the theta competition model.

7. Hypotheses

In the present thesis, the goal is to gain an understanding of the temporal aspects of human episodic memory retrieval by investigating the relationship between the neural reinstatement of memories, and neural rhythms of the hippocampus. There are three main hypotheses.

1. The first hypothesis states that the memory system fluctuates between representing incoming information relevant for the encoding of new events, and reactivated internal information representing past events (as described in section 6.1). Evidence for this idea has been shown in rodents, but never in human subjects. Two separate experiments will address this question, both doing so by decoding the content of mental representations in EEG patterns and linking these time-resolved indices of memory reinstatement to the phase of the theta rhythm. In the first study, scalp EEG in healthy participants will be used to track the dynamic representations emerging during encoding and retrieval. In the second study, the same rationale is applied to intracranial EEG obtained from 11 patients suffering from pharmaco-resistant epilepsy, allowing the extraction of the theta rhythm directly from the hippocampus.

2. The second hypothesis is built on the principle of how phase coding enables separation of overlapping memories, as described in section 6.2. MEG is used to track the reactivation of overlapping but competing memories, with the hypothesis being that the reactivation of neural representations belonging to target memories are phase-

shifted compared to the reactivation of neural representations belonging to competing memories.

3. The third hypothesis is based on recent findings in human subjects showing an important role for hippocampal SPW-Rs in awake episodic memory retrieval. In Chapter 4, the relationship between neocortical neural reinstatement and SPW-Rs is therefore investigated, with specific emphasis in understanding the temporal lag between these two areas, and its functional relevance for memory success.

Chapter 2 - An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval

At the time of thesis submission, this chapter has been published in Current Biology: CB 28 (21), 3383–3392. https://doi.org/10.1016/j.cub.2018.08.065 e6, but has been modified for reasons of consistency of the various chapters.

Abstract

Computational models and in vivo studies in rodents suggest that the hippocampal system oscillates between states optimal for encoding and states optimal for retrieval. We here show that in humans, neural signatures of memory reactivation are modulated by the phase of a theta oscillation. EEG was recorded while participants were cued to recall previously learned word-object associations, and time-resolved pattern classifiers were trained to detect neural reactivation of the target objects. Classifier fidelity rhythmically fluctuated at 7-8Hz, and was modulated by theta phase across the entire recall period. The phase of optimal classification was shifted approximately 180° between encoding and retrieval. Inspired by animal work, we then computed "classifier-locked averages" to analyse how ongoing theta oscillations behaved around the time points at which the classifier indicated memory retrieval. We found strong theta (7-8Hz) phase consistency approximately 300ms before the time points of maximal neural memory reactivation. Our findings provide important evidence that the neural signatures of memory retrieval fluctuate and are time-locked to the phase of an ongoing theta oscillation.

Keywords

Oscillations, episodic memory, hippocampus, theta oscillations, phase coding

1. Introduction

Our episodic memory defines us by storing a record of our past experiences and allowing us to consciously access these records. It is widely agreed that the hippocampus and neocortical areas work in conjunction during the formation and later retrieval of a memory (McClelland et al., 1995; Rolls, 1996; Teyler & DiScenna, 1986; Tulving & Markowitsch, 1998). At encoding, the hippocampus is thought to continuously store a sparse and non-overlapping index that points to ongoing activity patterns in cortical space. This hippocampal index can later be reactivated by a reminder, and lead to the reconstruction of a previously stored memory pattern in neocortex (Alvarez & Squire, 1994; Marr, 1971; McClelland et al., 1995; Norman & O'Reilly, 2003; Teyler & DiScenna, 1986; Teyler & Rudy, 2007). Many recent studies have tested these computational assumptions by tracking the reinstatement of memory-related brain activity patterns during retrieval. The basic premise that content-specific neural patterns are reactivated during retrieval has been confirmed using fMRI (for reviews, see (Danker & Anderson, 2010; Rissman & Wagner, 2012)) and more recently also EEG and MEG (Jafarpour, Fuentemilla, Horner, Penny, & Duzel, 2014; Johnson, Price, & Leiker, 2015; Kurth-Nelson, Barnes, Sejdinovic, Dolan, & Dayan, 2015; Michelmann, Bowman, & Hanslmayr, 2016; Staudigl, Vollmar, Noachtar, & Hanslmayr, 2015; Waldhauser, Braun, & Hanslmayr, 2016; Wimber, Maass, Staudigl, Richardson-Klavehn, & Hanslmayr, 2012). However, no study has so far investigated the temporal fluctuations of memory-related patterns in human long-term memory, and whether they are systematically linked to brain oscillations.

A major computational challenge for our memory system is to effectively separate the information arriving from external sensory sources from the information generated in internal circuits. In other words, if the brain constantly pattern completes, how does it make sure that the neural coding of this internally (and possibly incorrectly) generated information does not interfere with the coding of new, incoming information? One promising explanation suggests that this is accomplished by means of neural oscillations. In particular, it has been argued that the phase of the hippocampal theta oscillation supports the chunking of mnemonic information such that the neural assemblies involved in encoding and retrieval are temporally segregated (Hasselmo et al., 2002; Nyhus & Curran, 2010). In a seminal paper, Pavlides et al. (1988) showed that stimulating a hippocampal assembly at one phase of the theta rhythm induced long-term potentiation (LTP), whereas stimulating at the opposite phase induced long-term depression (LTD). This finding has since been replicated many times in rodents (Huerta & Lisman, 1993; Hyman, Wyble, Goyal, Rossi, & Hasselmo, 2003), and implemented in computational models of episodic memory and the hippocampus (Buzsaki, 2002; Hasselmo et al., 2002; Hasselmo & Eichenbaum, 2005; Kunec et al., 2005; Parish, Hanslmayr, & Bowman, 2017). These models share the assumption that successful retrieval is most likely at one specific phase of the hippocampal theta rhythm, opposing the optimal encoding phase (Hasselmo, 2005; Hasselmo et al., 2002). Memory retrieval should be a continuously oscillating process that is locked to the hippocampal theta phase.

Direct evidence for theta phase modulation in human long-term memory still remains elusive. FMRI studies by nature are blind to the sub-second temporal dynamics that

could mediate memory reinstatement, and electrophysiological studies have so far not investigated rhythmic fluctuations in memory reactivation. To our knowledge, only one previous study exists that has shown evidence for periodic reactivation, and this was during a working memory task (Fuentemilla, Penny, Cashdollar, Bunzeck, & Duzel, 2010). In human long-term memory it is therefore unknown whether neural signatures of memory reactivation are locked to a theta rhythm. The present study was aimed at directly testing this hypothesis. EEG data was recorded while participants encoded novel word-object associations, and were later cued with the words to retrieve the objects. EEG-based pattern classifiers were trained to detect memory-related neural patterns during recall with high temporal precision. We demonstrate that within each retrieval period, classifier fidelity fluctuates at 7-8Hz within each retrieval period, and that this index of memory reactivation is locked to a particular phase of the same theta rhythm.

2. Methods

2.1. Experimental model and subject details

All experimental procedures in the present study were approved by and conducted in accordance with the University of Birmingham Research Ethics Committee (STEM). Written informed consent was obtained from participants before they took part in the experiment.

2.1.1. Participants

Twenty-four healthy participants (19 female) aged 18-32 years (mean = 22.1, SD = 4.7 years) received credits or monetary payment for participation. Participants had normal or corrected-to-normal vision and reported no history of neurological disorders.

3. Method details

3.1. Material and setup

The material consisted of 64 images depicting animate objects (equal number of mammals, birds, insects, and marine animals) and 64 images depicting inanimate objects (equal number of electronic devices, clothes, fruits, and vegetables), taken from BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and from online royalty-free databases, and was used due to previous success at distinguishing these categories using multi-variate pattern analysis (T. Carlson, Tovar, Alink, & Kriegeskorte, 2013). All images were scaled to 500 x 500 pixels. A black-and-white drawing version of each image was manually created using GNU imaging manipulation software (www.gimp.org). The photographs vs. drawings served as an additional perceptual category (not of interest for the purpose of our current analyses). In addition to the material used for the experiment, 16 images were used for demonstrative purpose. Images from both semantic classes were randomly split into 16 sets, so that each set consisted of 8 images, 4 animate and 4 inanimate. Each set

constituted one learning block. In addition, a list of 128 action verbs was generated for the experiment, serving as cue words in the cued recall task.

The experiment was set up via custom written MATLAB 2016a (©The Mathworks, Munich, Germany) code using functions from the Psychophysics Toolbox Version 3 (Brainard, 1997). The presentation was done on a 15-inch computer screen with Windows 64 bit.

3.2. Paradigm

Participants received instructions about the task and first performed two practice blocks. All participants then performed 16 experimental blocks (8 trials per block), each consisting of an associative learning phase, a distractor task, and a retrieval test (Figure 1). A learning trial consisted of a jittered fixation cross (between 500 and 1500ms), a unique action verb (1500ms), a fixation cross (between 500 and 1500ms), followed by a picture of an object that was presented in the centre of the screen for a minimum of 2 and a maximum of 10 seconds. The task was to come up with a vivid mental image that involved the object and the action verb presented in the current trial. As soon as they had a clear association in mind, participants pressed the up-arrow key on the keyboard, which led to the onset of the next trial. Participants were aware of the later memory test, and knew that they had to pay attention to perceptual and meaningful aspects to perform the memory test.

A distractor task followed each learning phase. Here participants had to respond if a given random number (between 1 and 99) presented on the screen was odd or even. They were instructed to accomplish as many trials as they could in 45 seconds, and received feedback about their accuracy at the end of each distractor block.

After the distractor task, participants' memory for the 8 verb-object associations learned in the immediately preceding learning phase was tested in random order. Each trial consisted of a jittered fixation cross (500-1500ms), followed by one of the action verbs as a reminder cue for the association. Participants were asked to bring back to mind the object that had been associated with this word as vividly as possible. To capture the particular moment when participants consciously recalled a specific object, they were asked to press the up-arrow key as soon as they had a complete image of the associated memory in mind; or the down-arrow if they were unable to remember the association. The reminder was presented on the screen for a minimum of 2 seconds and until a response was made. Immediately following the button press, a blank square with the same size as the original images was displayed, and participants were asked to hold the retrieved object in mind for 3000ms. After a short fixation interval (1500ms), two questions were displayed sequentially, asking participants whether the associated object was a photograph or line-drawing (perceptual question), or an animate or inanimate object (semantic question). The order of questions was pseudo-random across trials such that the semantic question was asked first on half of the trials, and second on the other half.

Each semantic category was presented equally often in each type of perceptual level per participant. The action verbs were randomly assigned to the word-object pairs, and the distribution of object categories for perceptual and semantic features was equally distributed across the first and second half of the experiment.

4. EEG data analysis

The electroencephalogram (EEG) was recorded using a BioSemi Active-Two Recording System (BioSemi, Amsterdam, the Netherlands) with a 128-channel electrode cap, sampled at 1024Hz.

4.1. Preprocessing

Preprocessing was done twice using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom written MATLAB code: First before implementing multivariate pattern analysis, and again after re-epoching the data based on the maxima of the classifier output. The data was baseline corrected based on the whole trial before implementing the independent component analysis (ICA), and down-sampled to 256 Hz for the second preprocessing step, but kept at 1024Hz for the first. The down-sampling was done in order to decrease computational time for the classifier-locked average analyses, where the time-frequency transformation diminishes temporal resolution anyway.

Data were divided into trials from 700ms pre-stimulus to 2000ms post-stimulus onset (before implementing multivariate pattern analysis [MVPA]), or 2500ms before the classification maxima to 2500ms after the classification maxima (epochs created based on points of maximum fidelity). A high-pass filter of 0.1Hz, a low-pass filter of 195 Hz, and a band-stop filter (48 to 52Hz; 99 to 101Hz, and 149 to 151Hz), were applied to the data. At the edges of each trial, 500ms was then cut out to remove edge artifacts from filtering the epoched data. Trials were visually inspected before an ICA was computed to remove components related to eye-blink artifacts and muscle tension. After components were removed, all trials were again visually inspected, and trials still containing artifacts were manually removed. On average 112 out of 128 trials were kept (min = 100, max = 124, SD = 7). Bad channels were interpolated using the triangulation method. Data were then re-referenced to average.

4.2. Multivariate pattern analysis

In order to attenuate unwanted noise, a Gaussian window with a full-width at half maximum (FWHM) in the time-domain of 40ms was applied to the signal before classification. A Linear Discriminant Analysis (LDA) was then trained and tested on the EEG sensor patterns (pre-processed signal amplitude on each of the 128 channels), independently per participant and at each time point during retrieval from 200ms pre-cue up to 1500ms post-cue. The classifier was trained to detect systematic differences between trials where participants were recalling an animate or inanimate object. A leave-one-out cross-validation procedure was used to train and test the classifier. The LDA reduces the data from 128 channels into a single decoding time course per trial,

and we used these single-trial, time-resolved output of the classifier as an index of memory reinstatement. During training, the classifier found the decision boundary that could best separate the patterns of activity from the two classes (animate or inanimate) in a high-dimensional space. We then asked the classifier to estimate whether the unlabelled pattern of brain activity was more similar to one or the other class. This training-test procedure was repeated until every single retrieval trial had been classified. To avoid overfitting, the covariance matrix was regularized using shrinkage regularization (Blankertz, Lemm, Treder, Haufe, & Muller, 2011). The output of the classifier on a single-trial level indicates the distance to the decision boundary in a high-dimensional space, at a given time point. This parametric value is called a fidelity value or distance (d-)value, and can intuitively be regarded as reflecting how confidently the classifier predicted that the pattern of brain activity belonged to one or the other of the two classes, with the assumption being that the farther away from the boundary the more confident the classifier was (T. A. Carlson, Ritchie, Kriegeskorte, Durvasula, & Ma, 2014). Note that all the central LDA analyses in this study were based on retrieval data. To relate retrieval phase to encoding, the same LDA approach was also applied to the encoding data. Moreover, additional results from classifiers trained on encoding and tested during retrieval are reported in the Supplemental Materials.

4.3. Power spectrum of the classifier fidelity time series

The first analysis investigated the frequency characteristics of the classifier timeseries using Fast Fourier Transformation (FFT). This and all subsequent phase locking

analyses were limited to the classifier outputs from 200ms until 1200ms after onset of the reminder. We choose this time-window of interest because based on the existing literature, memory reinstatement is highly unlikely to occur within the first 200ms post-cue, and in order to reduce influences of early, stimulus-evoked ERP components. For each participant, the trials were averaged and tapered with a Hann window before conducting the FFT. To better visualize the power spectrum, a least-squares linear regression was used to subtract the 1/f background signal (Miller, Sorensen, Ojemann, & den Nijs, 2009; Voytek et al., 2015). The signal was log-transformed in the time and frequency domain and fitted with a regression line. The regression line was then subtracted from the power spectrum, and only the data that differed from the subtracted regression line were retained.

A baseline for the LDA outputs was created using a classifier with randomly shuffled labels. The labels of the two classes that the classifier later used for training and testing were shuffled pseudo randomly (to keep the same number of photographs and line drawings in each class), and fed into the LDA 25 times for each participant, such that the newly created groups had approximately the same number of trials from both classes. The parameters for running the classifier were the same as previously described for the real labels. In line with the procedure outlined in (Stelzer, Chen, & Turner, 2013), and identical as for the real data, for each participant we drew (with replacement) 100 random accuracy maps (i.e., either a baseline that was created using shuffled labels, or the real classification of the data), which were then averaged within participants. These accuracy maps were tapered with a Hann window, frequency transformed, and averaged into a group accuracy map. The background 1/f signal was

subtracted using a least-squares linear regression, as described above. This procedure was repeated 1000 times, and resulted in an empirical chance distribution, which allowed us to investigate whether the results from the real-labels classification had low probability of being obtained due to chance ($p<.05$) (i.e., exceeding the 95th percentile).

4.4. Phase-amplitude coupling between EEG data and fidelity values

To investigate the relationship between the continuous classifier outputs and the EEG data, the Modulation Index (MI) was computed in accordance with (Tort, Komorowski, Eichenbaum, & Kopell, 2010). Following the same procedure as outlined under Source Analysis below, we projected the data from scalp level to source level, where each filter was computed using baseline corrected pre-processed data (-.2 – 0 sec), and frequencies below 15Hz (i.e., -200 before to 1500 ms after cue onset). Epochs were then reconstructed for 2015 virtual electrodes, rather than the original 128 electrodes. The phase of the EEG signal was estimated by convolving the data with a complex Morlet wavelet of 6 cycles. Each complex value data point was then point-wise divided by its magnitude (absolute value or complex modulus), which gave us a 4D-matrix of phase values, containing trials*channels*frequencies*time. We then binned the phase values at a given electrode (e.g. a virtual hippocampal electrode), and at a given frequency of interest (e.g. 8Hz), into 10 adjacent bins, ranging from $-\pi$ to π. The z-scored amplitudes (d-values) of the classifier output from corresponding time points were then sorted into their corresponding phase bins, and the mean amplitude of each phase bin was calculated. Following this sorting procedure at a given

frequency, the modulation index was calculated. The MI was computed by comparing the distribution of classifier fidelity values across the 10 phase bins against a uniform distribution (using the mean across bins to construct the uniform distribution). The Kullback-Leibler (KL) distance was then calculated using the equation in (Tort et al., 2010)

A statistical control analysis was then performed to infer whether the MI was significantly different from a distribution that could be obtained by chance. The baseline was computed by running the same analysis as described above, but by cutting the classifier outputs into two segments at a random time point, and inserting the second data segment at the beginning of the trial. This procedure is recommended in (Cohen, 2014), because it keeps the temporal structure of the classifier outputs largely intact while randomizing their relationship to the EEG phase at any given time point. The newly created random classifier outputs were then paired with the real EEG phase time series from their corresponding trial, and were binned in the same way as the real data. This procedure was repeated 500 times, and the MI was calculated for each iteration. The 95th percentile across iterations was determined, and the real modulation index for each subject was compared against this subject's 95th percentile using a paired samples t-test. Note that this is a very conservative analysis, resulting only in statistically significant phase modulation, if across participants real phase modulation values significantly exceed the 95th percentile of the time-permuted baseline.

Based on our initial FFT findings, all phase modulation analyses were focused on the oscillatory phase at 8Hz (Figure 2). The phase modulation index was calculated as

described above for each virtual channel in source space, and a mask including left and right hippocampus (from AAL atlas as implemented in FieldTrip, see Figure 2E) was then applied to specifically extract the modulation index from our main region of interest. This was done separately for the phase modulation during retrieval, and the phase modulation during encoding. To directly compare the preferred phase during encoding and retrieval, the bin containing the highest classifier amplitudes was identified in each participant, separately for encoding and retrieval. A Rayleigh test (implemented using circ_rtest in the Circular Statistics Toolbox for Matlab) was then used to statistically test the extent to which the distribution of phase angles at encoding and retrieval differed from each other.

4.5. Using classifier-locked averages to relate classifier outputs to the phase of the ongoing EEG-signal

The third, classifier-locked average analysis was aimed at characterizing the EEG phase of the time points where the classifier showed the highest fidelity. To this end, three criteria were established in order to identify times of maximum fidelity. In order to be considered a maximum, a fidelity value was required to have an amplitude that exceeded the 95th percentile of a baseline constructed from the random-label classifications. For each participant, we drew (with replacement) the fidelity timeseries from random trials 1000 times to obtain the baseline distribution. In addition, a maximum included in the final analysis was also required to remain above the 95th percentile threshold for more than 30ms, and to occur later than 200ms after reminder onset, for the same reasons as mentioned above. The average number of classification

maxima extracted per trial was 2.27 (SD = 0.26). The onsets of the classifier maxima in each trial were then marked, and the corresponding time stamps were located in the raw, continuous EEG recordings. New epochs were created that were centred on each classifier maximum and contained 2.5 secs before and after the maximum, which were then cut during preprocessing to 2 secs before and after the maximum. These new epochs were used for all subsequent phase-locking analyses.

A phase-locking analysis was conducted on the new epochs to test whether classifier maxima were related to a consistent phase of a theta oscillation. For every frequency between 1 and 20 Hz, we estimated phase by convolving the data with a complex Morlet wavelet of 6 cycles. Resulting complex values were then point-wise divided by their magnitude (absolute value or complex modulus), and the mean phase was computed over all trials within each participant. The magnitude of this resulting complex value is a single value (the phase-angle time series) for each time-frequency-channel point averaged over all the trials. The value reflects the consistency of frequency-specific phase across trials and has a minimum of 0 and a maximum of 1, also called phase-locking value (PLV), phase-locking index (PLI) or Intertrial Phase Clustering (ITPC) (Cohen, 2014).

A baseline was calculated for each trial and each participant by shifting single-trial EEG epochs randomly between 0ms and 150ms (roughly one theta-cycle) forward or backward in time, relative to the centre (i.e., the classifier maxima). By doing so, the temporal structure of the analysed signal was kept intact, but the signal was shifted relative to the classifier maxima. The phase-locking index was calculated as described

above for the “real”, non-shuffled data. Shuffling was done 25 times per participant and thereafter averaged together.

First, paired samples t-tests were computed between the real data and the time-shuffled baseline to investigate the difference in phase-consistency when using all maxima. To account for the multiple comparisons problem, the t-statistics for each time point (-500ms to 500ms), frequency band (6 to 14 Hz), and electrode were subjected to nonparametric cluster-based permutation testing, as implemented in the FieldTrip software. The threshold for the statistical testing was set to an alpha level of 0.025. The minimum number of neighbouring channels that were considered a cluster was set to two. T-values above the threshold of 0.1 were then summed up, and compared against a distribution where condition labels were randomly assigned 5000 times with the Monte-Carlo method, following the standard method implemented in FieldTrip.

Phase consistency is strongly biased by number of trials. For our first analysis comparing all maxima against the time-shuffled baseline, the real data and shuffled baseline contained an equal number of trials. We also ensured that all subsequent comparisons were made between conditions with exactly equal trial numbers, within each participant, including an analysis contrasting classifier maxima of high fidelity and maxima of lower fidelity, and two analyses excluding early maxima (see following two paragraphs). For the analysis contrasting conditions with high and low fidelity values, we additionally controlled the average time of the high and low classifier maxima. This was done by creating 8 time bins of equal size between 200ms and 1500ms post-cue. Fidelity values in each time bin were median split into high and low fidelity values, resulting in two matrices representing high and low fidelity trials,

equally distributed across time. To calculate the phase consistency, we then followed the same procedure as described above for all maxima, except that instead of using the shuffled baseline the two groups of trials were directly compared using a non-parametric cluster-based permutation test.

To investigate the degree to which our phase-locking effects were mainly produced by classifier maxima close to the reminder word, which would be strongly influenced by the early stimulus-elicited event-related potential (ERP), we conducted two additional analyses excluding early classifier maxima that occurred in the first 400ms and the first 600ms post-cue, respectively, from further analysis. Otherwise, these analyses followed the same method as described for all maxima, with the same time-shuffled baseline. Similarly, an analysis using only the highest classifier maximum per trial used the same procedures and baseline described in this section for all maxima.

4.6. Event-related potential analysis

Event-related analyses were mainly conducted as sanity checks, on the one hand to investigate average signal differences between the retrieval of animate and inanimate objects locked to cue onset; and on the other hand to evaluate the average signal differences and their topography/source around the time points at which the classifier showed maximal confidence that the correct category was reinstated. For the classifier-centred analysis, we only used the 20% classifier maxima with the highest fidelity values in each of the to-be-compared classes (i.e., animate and inanimate retrieval trials), in order to enhance signal-to-noise ratio. This latter analysis included on

average 48 (SD = 7.10) trials per participant. Cluster-based statistics for ERPs were conducted in the same way as for phase, except that we here focused on a narrower time window from 200ms pre- until 200ms post-maximum.

4.7. Source analysis

A linear constrained minimum variance (lcmv) beamforming approach (Gross et al., 2001) was used to reconstruct EEG epochs in source space. The source-level results were used to obtain an approximation of the hippocampal theta phase for the phase modulation analysis, and to reconstruct classifier-locked averages (i.e., phase consistency and ERP effects) in source space (Oostenveld et al., 2011). Since individual MRI scans were not available, a standard MRI model was used to construct the boundary element model. The boundary element model was used in combination with individual electrode positions obtained from a Polhemus system (Colchester, Vermont, USA) to reconstruct the activity on a source level. To project the phase consistency effect from scalp level to source level, each filter was computed using frequencies below 15Hz and the entire time-window from the preprocessed data (i.e., 1500ms before to 1500ms after classifier maxima), and the original epochs were then reconstructed on 2015 virtual electrodes. Thereafter the phase-locking analysis followed the same procedure as done on scalp level. For calculating the filters for the ERP effect, we used all frequencies below 20Hz, and a time-window of 300ms pre-maxima to 300ms post-maxima. The ERP was then calculated in the same way as on a scalp level. Note that the full-brain source reconstructions or the classifier-locked effects are only used to illustrate the most likely sources of the effects observed on

scalp level (see above). We do not report additional statistics at source level, since these would be circular relative of the already known effects on scalp level. Labels of MNI coordinates were assigned based on the Lancaster et al. (2000) Talairach atlas.

4.8. Distribution of fidelity values across time

To statistically test whether the distribution of fidelity values was different from a uniform distribution across the entire retrieval time window, we manually created a uniform distribution, by producing linearly spaced values between the minimum and maximum of the real values. We then binned the data into 20 linearly spaced bins and calculated the chi square statistic using the crosstab function as implemented in MATLAB, which tests whether the proportion of items in one cell is equal to the product of the proportion in that row (Figure S2A).

4.9. Time generalisation

To characterise the temporal dynamics of the classifiers, we calculated the full time generalization matrices from encoding and retrieval. These matrices show where in time classification accuracy was maximal, to which degree a classifier trained at one time point generalises to a different time point, indicating temporal stability of the underlying neural code (King & Dehaene, 2014). All analyses were performed using LDA as implemented in the MVPA-Light toolbox, running on MATLAB (https://github.com/treder/MVPA-Light). Two different analyses were run: training at each time point at encoding and testing at each time point at encoding (Figure S1A);

training at each time point at retrieval and testing at each time point at retrieval (Figure S1B). When analysing encoding-to-encoding generalization, data were baseline corrected (-200 to 0ms), and then z-scored per trial before running the classification. We used a k-fold cross-validation approach with 5 folds, which was repeated twice with randomly assigned folds. When training and testing at each time point at retrieval, we did not baseline correct before the classification. However, baseline correction was applied after the classification in both analyses.

4.10. Identifying oscillating frequencies

An alternative method for detecting oscillations in time series was used in addition to our FFT approach in order to corroborate our claim that classifier outputs oscillate. This method finds time points of oscillations in the data by investigating the change in phase per unit time. We followed the method detailed in (M.X. Cohen, 2014), with a modification for dynamic filter edges only using minimum and maximum of frequencies exceeding the 1/f distribution, made in line with (Watrous, Miller, Qasim, Fried, & Jacobs, 2017). Briefly, we started with raw time series data, which in our case was the z-scored fidelity values averaged within participants. Instead of creating a plateau-shaped band-pass filter based on an a priori defined frequency range, the filter was constructed based on the lowest and highest frequencies exceeding the fitted line in log-log space using robustfit in MATLAB (Lega, Jacobs, & Kahana, 2012). The analytic signal was obtained by applying the Hilbert transform to the data, from where we extracted the phase angle time series. To obtain the frequency and phase at each sample, frequency sliding was applied to the data as follows: (sampling

frequency*diff(unwrap(signal))/(2*π)). After this step, in order to attenuate "phase slips", we applied median filters with different length in the time domain (50ms to 400ms), wherefrom we took the median, in accordance with (Cohen, 2014). Frequencies that did not exceed the 1/f-fitted line were then excluded, which gave us a vector for each participant containing the frequencies and time points where an oscillation was present. We then calculated the average probability across time (200 to 1200ms post-cue, as in all other analyses using classifier output) for observing an oscillation in a given frequency between 1 and 15 Hz.

To infer whether the result that we obtained was significantly different from chance, we randomly picked one averaged random label classifier per participant. The same procedure as has been described above was applied. An average of this value was then calculated, and stored. This was done 1000 times, and resulted in an estimated chance distribution. The 95th percentile was then calculated for each frequency, and compared that to the real data (Figure 2D).

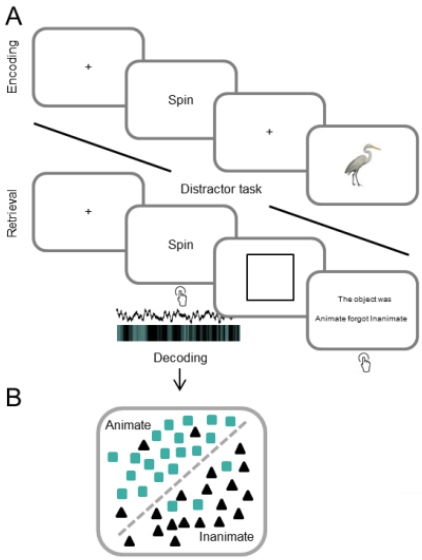

***Figure 1*. Trial structure and Multivariate Pattern Analysis. (A)** At encoding, participants associated action verbs with images depicting either an animate or inanimate object. After a short distractor task, participants were tested on the previously learned associations. The action verb was shown as a cue, asking participants to retrieve the associated object, and to indicate with a button press when the object came back to mind. They then had to respond to the question whether it was an animate or inanimate object. (**B)** For each time point and each trial from cue onset at retrieval, a linear discriminant analysis (LDA) classifier was trained and tested on detecting evidence for retrieval of the correct object category. The output of the classifier was a parametric value for each time point, reflecting the fidelity of the classifier to differentiate between the two object classes.

5. Results

5.1. Participants retrieve the episodic memories with high accuracy

The paradigm was a simple word-object associative memory task designed to yield a high number of correct trials (Figure 1A). Participants studied associations between action verbs and objects in random pairings, and were later cued with the word to retrieve the object. Two measures of memory accuracy confirmed that participants performed the task well. The first was a subjective measure where participants

indicated, via a button press after cue onset, whether and when they recalled the associated object. Participants on average indicated that they remembered the object on 94.21% (SD = 5.75%) of the trials. A second, more objective measure was accuracy in response to a question about the object's semantic category (animate vs inanimate), which appeared at the end of each retrieval trial, and which participants answered correctly on 88.20% (SD = 6.57%) of the trials. These two measures were highly correlated (rSpearman = 0.60, $p < .05$). Average accuracy for perceptual detail (photograph vs line drawing) was 85.31% (SD = 6.45%).

Reaction times for the first button press when retrieving animate (Mean = 3.03 secs, SD = .95 secs, min = 1.28 secs, max = 6.01 secs) and inanimate (Mean = 2.96 secs, SD = .77 secs, min = 1.47 secs, max = 4.24 secs) objects did not differ significantly, t(1,23) = .57, $p = .58$. The time window used for classification (-200ms to 1500ms around the cue) thus only minimally overlapped with the button press window.

5.2. Power spectrum of classifier shows strongest effects in lower frequencies

Our primary goal was to test whether the neural signatures of memory retrieval wax and wane in a theta oscillatory rhythm. Our neural index of memory retrieval was obtained from a linear discriminant analysis (LDA) trained to detect evidence for the reactivation of the correct object category (animate vs. inanimate) during retrieval (Figure 1B, see methods for details). The LDA was trained and tested independently per participant at each retrieval time point starting with the onset of the word cue, using

a leave-one-out procedure. The input into the LDA was a feature vector containing the signal amplitudes from all 128 EEG channels at a given time point. The major output of interest was the fidelity (distance, or d-) values available for each trial and time point. These values represent the distance from the hyperplane that optimally separates the two classes of retrieved objects (animate vs inanimate), and their timecourses served as our time-resolved, parametric index of memory reactivation. For the purpose of this study, the LDA was trained and tested during cued recall in order to isolate a purely retrieval-based signature or memory retrieval, which could then (below) be compared with a purely encoding-based index of memory classification. Additional analyses using classifiers trained on encoding and tested at retrieval are reported in the supplementary materials (Figure S1 and S4).

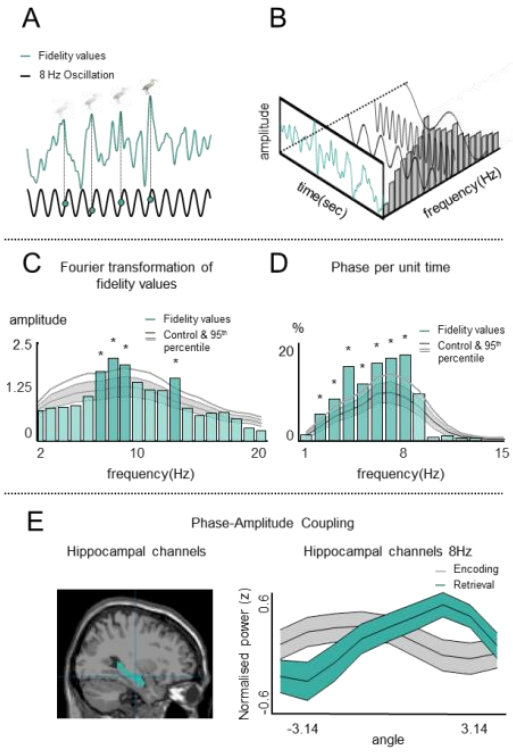

***Figure 2*. Analysis rationale and results of the time-frequency analyses relating classifier fidelity to theta oscillations and phase-modulation. (A)** Example of a single-trial output from the LDA,

reflecting the fidelity of the classifier in detecting the retrieved object's correct category at each time point during a retrieval trial. The black line represents a theta oscillation to illustrate our assumption that neural indices of memory reinstatement (i.e., the d-value time series) rhythmically fluctuate, and that the time points of maximal classifier fidelity should be consistently related to a particular phase of the underlying oscillation. **(B)** D-values were subjected to a Fourier transformation which reveals the power in each frequency band. **(C)** The resulting power spectrum shows significant deviations from an empirical null distribution at 7 to 9 Hz and 13Hz. The baseline power spectrum was obtained from a combination of random label classifiers and bootstrapping, and is shown in grey (mean and SD). Values of the real classifier outputs exceeding the 95th percentile of the baseline distribution are marked as significant. **(D)** Frequency decomposition of the classifier time series using an alternative approach to detect frequencies [38], again showing above baseline power at slow frequencies including 7-8Hz. Figure showing mean ± SEM for baseline (grey lines) and 95th percentile (thick grey line). **(E)** Phase-amplitude coupling between EEG phase and classifier fidelity at source level revealed a significant modulation index averaged over hippocampal virtual channels (mask shown on the left) for 8Hz. Figure showing mean ± SD. See also Figure S4.

We first asked whether evidence could be found for an oscillation in these time-resolved indices of memory reactivation (Figure 2A-B). Fidelity timecourses from the recall task were averaged across trials per participant and subjected to a Fourier Transformation. If memory reactivation fluctuates in a theta rhythm, the resulting power spectra will show a selective increase in a band-limited lower (theta) frequency band. We compared the power spectra obtained from the real classifier outputs with a bootstrapped baseline (Oostenveld et al., 2011), the latter using the d-value outputs from classifiers that were trained and tested on the same EEG trials but with randomly shuffled category labels (see Method section). This procedure controls for spurious power peaks that are driven by the frequency characteristics of the raw data (e.g. a dominant oscillation in the single trials). Significant power differences between the real and shuffled data were found in frequency bins at 7-9Hz and 13Hz, all exceeding the 95th percentile of the empirical null distribution (Figure 2C). Power at 7-9Hz was significantly higher, $t(23) = 1.9425$, $p = .03$, when including only correctly retrieved trials that when including all trials, suggesting a relationship of the classifier fluctuation to memory success . An alternative method with more stringent criteria to

determine the presence of oscillations (Watrous et al., 2017) confirmed that oscillatory power in the classifier time series was increased above baseline in the 7-9Hz frequency range (Figure 2D). Moreover, a similar power spectrum was found when the classifier was trained on encoding and tested on retrieval (Figure S4.) The frequency characteristics of the classifier fidelity time courses thus suggest a rhythmic fluctuation in memory reactivation that was most consistent in the 7-9Hz frequency range.

5.3. Phase-amplitude coupling reveals oscillating patterns at retrieval for 8Hz

Our next two analyses were aimed at specifically testing for coupling between neural reactivation (i.e., classifier timeseries) and the phase of hippocampal theta-band oscillations. For this purpose, the raw EEG trials were projected into source space using an LCMV beamforming algorithm (Gross et al., 2001; Van Veen, van Drongelen, Yuchtman, & Suzuki, 1997), and a hippocampal mask was used to extract the 8Hz phase of the hippocampal virtual channels for each trial and time point. We computed a phase modulation index (MI) (Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997) reflecting the strength of coupling between the hippocampal 8Hz phase and the amplitude of the classifier output. Classifier fidelity as a function of hippocampal theta phase is plotted in Figure 2E (green line). This analysis revealed a significant modulation index (M = .0071, SD = .0042; baseline: M = .0056, SD = .0006), $t(23) = 1.8191$, $p < .05$, one-sided t-test, indicating that fidelity of the retrieval classifier was modulated by the phase of the hippocampal 8Hz oscillation (Figure 2E).

We next directly compared the theta phase at which classifier fidelity was maximal during encoding and retrieval. All basic analysis steps were repeated for the encoding EEG data, where an LDA discriminating animate from inanimate objects was trained and tested at each time point from 200ms before until 1500ms after object onset. The full time generalization matrices showing classifier performance for encoding and retrieval can be found in Figure S1. The 8Hz phase at encoding was then extracted from hippocampal virtual channels to calculate the phase modulation index. Classifier fidelity as a function of hippocampal theta phase during encoding is shown in Figure 2E (grey line). A significant phase modulation was found also for encoding (M = .0068, SD = .0029; baseline: M = .0052, SD = .0007), t(23) = 2.7494, $p < .05$, one-sided t-test). In order to directly compare the encoding and retrieval phases, we identified the phase at which encoding or retrieval classification was optimal in each subject. A Rayleigh circular statistic comparing the absolute phase angles at which encoding and retrieval classification was maximal revealed that these angles significantly differed from each other, z(23) = 5.5342, $p = .001$. Similar statistics were obtained by fitting a sine wave to the data and identifying and extracting the phase at which classification was optimal. Together, the results of the phase modulation analyses show that retrieval fluctuates as a function of hippocampal theta (8Hz) phase, and that the optimal retrieval phase is on average 188 degrees phase shifted compared with the optimal phase during encoding.

5.4. Classifier-locked averages reveal a consistent theta phase prior to memory reinstatement

Having established that the neural retrieval patterns oscillate and are coupled to an 8Hz oscillation, we next investigated the temporal relationship between theta phase and memory reinstatement. The analysis was inspired by the use of spike-triggered averages in animal intracranial work (Douchamps et al., 2013; Rieke et al., 1997). We here adopted a similar approach computing classifier-locked averages around the time points of maximal memory reactivation (see Methods for details). On each single trial, those time points of maximal classifier fidelity that exceeded the 95th percentile of a bootstrapped baseline were marked as new events of interest, the corresponding time stamps were located in the raw EEG epochs, and the ongoing EEG signal surrounding these maxima was then analysed for phase consistency across all electrodes (Figure 3A). We used a non-parametric cluster-based permutation test to compare the real data with a temporally shuffled baseline that keeps the EEG trial structure intact but produces a random temporal alignment between the classifier maxima and the ongoing phase (see Method section). Comparing the "real" times of maximum classifier fidelity with the temporally shuffled baseline revealed a cluster of significant ($p_{corr} < .05$) phase consistency from 500ms to 50ms before the classifier maxima, centred at 7Hz (Figure 3B). Note that in this analysis, several classifier peaks per trial can exceed the 95th percentile criterium, and many of the classifier-locked EEG epochs will thus overlap, resulting in temporal smearing of the phase-locked activity. When running the same analysis extracting only one maximum per trial (Figure S2C), we found a similar cluster of phase locking but with a more narrow temporal extent from 500ms to 150ms pre-maxima, suggesting that the strongest phase-consistency effect was present roughly two theta cycles (corresponding to 2*143ms = 286ms) before mnemonic information could most confidently be decoded. This finding supports our

primary hypothesis that memory reinstatement shows a consistent oscillatory timing across trials and participants, in the same 7-9Hz frequency band at which the classifier fluctuates (Figure 2C).

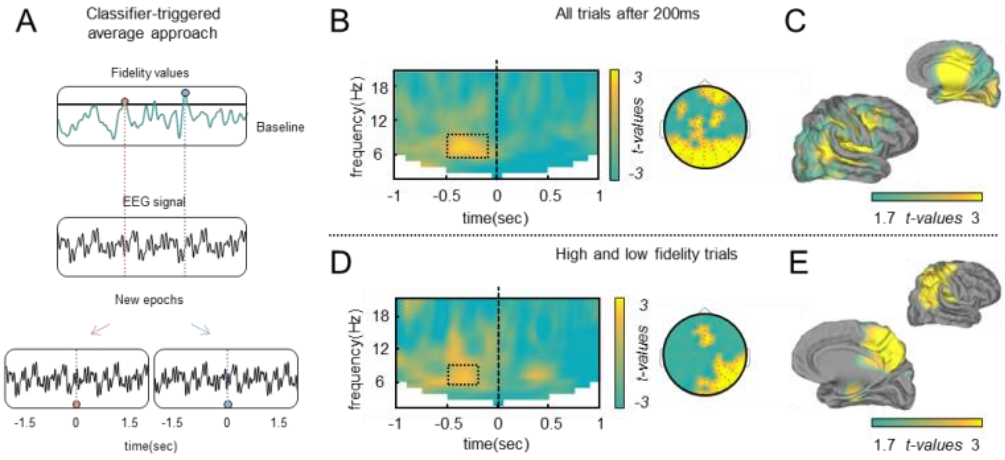

***Figure 3*. Rationale for classifier-locked average analysis to test for a functional relationship between classification maxima and neural memory reinstatement**. **(A)** Classifier d-values exceeding the 95th percentile of the chance distribution were marked, corresponding time stamps were found in the ongoing EEG data, and the EEG was then re-epoched relative to the classifier maxima. This procedure resulted in new epochs with the classifier maxima at time zero. **(B)** Results of the classifier-locked average analysis relating classifier maxima to ongoing EEG phase. A non-parametric cluster-based permutation test revealed a significant cluster of phase consistency centred at 7-8Hz, spanning from 500ms to 50ms before the maxima. **(C)** At source level, the maximal phase consistency was observed in occipital and right temporal lobe. (**D**) Contrasting maxima of high fidelity and maxima of lower fidelity, a significant cluster was again found at 7-8Hz, from 500ms to 200ms before the maxima. **(E)** At source level, the maximal phase consistency effect was located in parietal and temporal lobes, including MTL, when contrasting high and low fidelity trials. Time-frequency plots highlight the significant cluster in time and frequency. Topographical and source level plots show values above the critical t-threshold (t-value of 1.7, 23 degrees of freedom, one-sided test) for significance. See also Figure S2.

It might seem counterintuitive that the strongest phase consistency was observed prior to the time points of maximum classification fidelity, rather than at the maxima themselves. However, this temporal relationship is to be expected if the phase-locked signal originates from a different, upstream region in the processing hierarchy

compared to the signal that the classifier's decision is based on. Our findings are consistent with a model where the re-instantiation of a memory trace is triggered at a consistent phase of a hippocampal/MTL theta oscillation, followed by memory reinstatement in a broader range of neocortical regions representing the stored memory (Buzsaki, 2002; McClelland et al., 1995; O'Reilly et al., 2014; O'Reilly & Norman, 2002). The aim of the next analysis was to identify the brain regions involved in producing the observed clusters of theta phase consistency, with the hypothesis that the effect should be present in MTL areas (McClelland et al., 1995; Teyler & DiScenna, 1986).

Trial time-courses were projected into source space using a beamforming algorithm (Gross et al., 2001; Oostenveld et al., 2011), and we then looked for the sources showing the strongest phase consistency. Contrasting all classifier maxima with the shuffled baseline (identical to the scalp level analysis), we found an activation cluster spanning large regions of occipital, temporal and frontal cortex, primarily in the right hemisphere (maximum at MNI coordinates xyz = 10 -10 10, Thalamus, Figure 3C). While these sources included medial temporal lobe areas, they do not suggest a specific role of the hippocampus in producing the theta phase-locked signal preceding the classifier maxima.

5.5. High classifier fidelity is associated with strong theta phase consistency in MTL

We next wanted to test whether the theta phase consistency systematically varied with the strength of neural reinstatement. We hypothesized that phase consistency would be highest when the classifier correctly detected neural reactivation with high fidelity, and lower when the classifier was correct, but less confident.

Comparing classifier maxima of higher and lower fidelity revealed a significant (*pcorr* < .05) cluster at 7Hz preceding the maxima by 500ms to 200ms (Figure 3D). This cluster highly overlapped in timing, frequency and topography with our previous classifier-triggered average analyses. When conducting the same analysis in source space, we found sources that spanned the parietal and the right medial temporal lobes (maximum MNI coordinates xyz = 50 -30 30, inferior parietal lobule; Figure 3E), strongly reminiscent of the core recollection or memory success network typically found in fMRI studies (Rugg & Vilberg, 2013). Our data thus suggest that the neural signatures of memory retrieval are linked to a specific phase of a theta oscillation, and this phase relationship becomes stronger with more confident neural reactivation. The source level analysis additionally confirms our a priori assumption that the phase-locked signal that precedes memory reactivation involves the MTL and other core recollection areas.

5.6. Theta phase-locking is unlikely to be produced by early cue-related effects

While consistent with our hypotheses, this pattern of results could in theory also be explained by an early ERP elicited by the reminder word, since ERPs are generally

associated with strong phase locking in slow frequencies (Gruber, Klimesch, Sauseng, & Doppelmayr, 2005). Such an explanation would assume that our classifier maxima tend to occur at a consistent time point within each retrieval trial with a delay to the reminder-elicited ERP of approximately 300ms. Several observations speak against this alternative. First, the classifier maxima were relatively evenly distributed across the entire retrieval period and did not tend to cluster around early time points. A slight increase in density was found in the typical recollection time window (Yonelinas, 2002) from 400-800ms post-cue, but the overall distribution of the maxima did not significantly differ from uniform ($\chi 2$ = (1, N = 6007) = 7376600, p = .375) (Figure S2A). Second, we repeated the classifier-triggered average analysis excluding all classifier maxima that occurred earlier than 400ms or 600ms post-cue, respectively, excluding the time delays that would be most strongly affected by early ERPs. Both analyses revealed a significant phase-locking effect (*pcorr* < .05) in a very similar time window and frequency band as in the original analysis (Figure S2E and F). This result indicates that the theta phase-locked process preceding memory reinstatement can occur at various delays in a recall trial.

5.7. EEG signals at the exact time points of maximal classifier fidelity show content-dependent differences with a source in anterior temporal lobe

In order to correctly classify a trial as belonging to one category or another, linear classifiers including LDA require a consistent EEG signal difference across trials. If these signal differences additionally have consistent timing and topography across

participants, we should on average be able to observe a robust signal difference between animate and inanimate objects at time points of confident classification. We therefore conducted two confirmatory ERP analyses comparing the average waveforms for animate and inanimate objects during retrieval. The first of these analyses contrasted animate and inanimate trials time-locked to the onset of the word cue (Figure S3A-B). This analysis shows that the strongest average signal differences were present over frontal channels, although this cluster did not survive correction for multiple comparisons using cluster-based permutation statistics (*pcorr* = .64). The lack of significance could be due to variance in retrieval latency across trials, varying topographies across participants, or in fact due to an oscillating process that makes it difficult to observe a coherent cluster in time. Interestingly, when conducting an FFT on the average ERP differentiating animate and inanimate object retrievals in each participant, these signal differences showed power increases above baseline at 6-9Hz (Figure S3C), in the same range revealed by our frequency transformation of the classifier fidelity values. This finding confirms that the 8Hz oscillation is inherent in the signal difference that the LDA relies on.

The second ERP analysis again contrasted animate and inanimate trials, but this time locked to the time points of maximal classifier fidelity (as used in previous analyses). A cluster-based permutation test revealed a significant cluster (*pcorr* < .05) over frontal electrodes, spanning from 90ms before to 120ms after the classifier maxima (Figure 4A). Reconstructed at source level (Figure 4B), this effect showed a maximum in left anterior temporal lobe (maximum MNI coordinates xyz = -30 10 -40, superior temporal gyrus; and -40 0 -40, inferior temporal gyrus). The results confirm that the

single-trial classifier maxima indeed reflect a meaningful difference in the neural patterns elicited by retrieving different types of objects, rather than reflecting random fluctuations in classifier performance. The most likely source of the effect was found in anterior temporal lobe, an area strongly linked to semantic memory processing (Duvernoy, Cattin, & Risold, 2007; Patterson, Nestor, & Rogers, 2007), where previous studies found tight links between classifier fidelity and the speed at which participants behaviourally categorize objects as animate or inanimate (Oostenveld et al., 2011). Together, the two ERP analyses validate our LDA approach and provide converging evidence that retrieval-related differences between animate and inanimate objects fluctuate in the theta range and are most pronounced over neocortical regions involved in high-level semantic processing (Tyler et al., 2013).

5.8. Classifiers that generalise from encoding to retrieval show similar frequency characteristics

The results reported so far focus on an index of memory reactivation derived from classifiers trained and tested on the retrieval data. Below, we report additional analyses conducted on classifiers that were trained on the encoding data, and then tested either on the encoding or on the retrieval data. Encoding-to-retrieval classification has been commonly used in previous studies (Waldhauser et al., 2016). We conducted the additional analyses to confirm that such classifiers can also successfully detect memory reactivation, and that their frequency characteristics are similar to our main, purely retrieval-based metric. The results are summarized in Figure S4.

Encoding analyses were conducted on epochs time-locked to the onset of the animate and inanimate objects (-200ms to 1500ms). As a first step, an LDA was trained on encoding and also tested during encoding (Figure S4A). In line with the existing literature on object perception (Brodeur et al., 2010), animate vs inanimate category membership could be best decoded in a time window around 300ms after object onset, with an accuracy peak at 305ms. The classifier fidelity timecourses were then averaged within participants and subjected to a Fourier Transformation, following the same procedure as for the retrieval data. The resulting spectra (Figure S4B) showed the strongest power in lower frequencies with peaks at 3, 5, and 6 Hz exceeding the 95th percentile of the random label chance distribution.

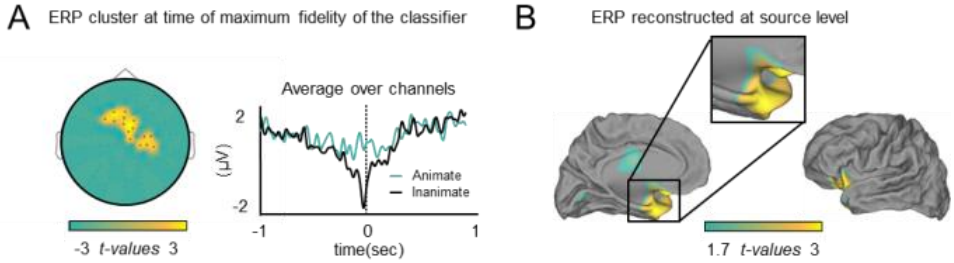

***Figure 4*. Event-related potentials centred around classifier maxima, on scalp and source level.** **(A)** ERPs locked to the time points of maximum classifier fidelity. A non-parametric cluster-based permutation test revealed a significant ($p < .05$, cluster-corrected) difference in the average signal produced by animate and inanimate recall trials, confirming that a robust difference between retrieved object classes was present at the time points of maximum classifier fidelity. The ERP plot shows the average of the significant channels for descriptive purposes. **(B)** The classifier-locked ERP reconstructed at source level shows a maximum in anterior temporal lobe, regions assumed to be involved in high-level semantic processing. Source level plot show values above the critical t-threshold (t-value of 1.7, 23 degrees of freedom, one-sided test). See also Figure S3.

During the time window where the LDA performed best, we also found a univariate ERP cluster ($pcorr < .05$) from 240-340ms with a frontal topography that significantly differentiated animate from inanimate objects during encoding (Figure S4C). Note that this cluster had a frontal topography similar to the main cluster differentiating animate

from inanimate objects during retrieval (as shown in Figure S3A), providing a first indication that content-specific processes engaged during encoding might be re-engaged during retrieval.

Based on this observation, we next tested explicitly whether classifiers trained to distinguish animate from inanimate objects during encoding could successfully discriminate those categories during retrieval. For this analysis, the classifier was trained on each time point within the 240-340ms encoding interval identified above, and tested at each time point at retrieval (see Figure S4D). This approach revealed the highest decoding accuracy in a retrieval time window from approximately 800-1500ms, a window typically associated with successful recollection (Yonelinas, 2002). We then assessed the frequency characteristics of the encoding-retrieval classifiers using the same FFT method as before, but this time applied to the classifiers trained on the activity patterns between 240-340ms during encoding, and tested at each time point during retrieval (see Figure S4E). The resulting power spectra showed the maximum peak at 9Hz (5 and 9 Hz exceeding the 95th percentile), with a similar distribution but at a slightly higher frequency peak compared with results obtained when training and testing at retrieval (see main Figure 2C).

6. Discussion

Memory retrieval, or at least the neural reactivation process underlying it, is often thought of as a static process that happens in an all-or-none fashion once a reminder has reactivated a past experience. However, evidence from rodents suggests that

pattern completion fluctuates on a sub-second time scale, and that these fluctuations are determined by a hippocampal theta oscillation that shifts the network between states optimal for encoding, and states optimal for retrieval (Hasselmo et al., 2002; Kunec et al., 2005). We here sought to investigate these oscillating retrieval dynamics in humans in a cued recall task. Several findings from the present experiment indicate that the neural signatures of memory reactivation in fact do fluctuate within a single recall trial in the human brain, and are tightly linked to a specific phase of a theta oscillation.

Our main metric of interest was a parametric, time-resolved index of memory reactivation for each trial that we obtained from a multivariate classifier trained to detect the semantic category of the recalled object. First, we found that this index in itself fluctuates at 7-8Hz. This oscillating pattern was evident in the average classifier fidelity time courses from each participant (Figure 2C), relative to a baseline which used the output from random label classifiers. The effect can thus not readily be explained by the frequency structure of the data that served as input to the classifier (e.g., a dominant 7-8Hz rhythm inherent in the EEG epochs). The 7-9Hz fluctuation was stronger for successfully remembered than for all associations including misses, and it was also present in the average ERP waveforms differentiating the retrieval of animate and inanimate objects (Figure S3C). These findings suggest a fluctuation in the signals differentiating the two classes of retrieved mnemonic representations, consistent with a rhythmic memory reactivation process.

Second, we investigated whether the classifier-based indices of memory reactivation systematically varied as a function of theta phase. We found that classifier fidelity was significantly modulated by the phase of an 8Hz oscillation extracted from virtual hippocampal channels (Figure 2E). The phase of peak classification fidelity during recall was 188 degrees shifted compared to the phase of peak fidelity during encoding. These results support two of the central claims of the Hasselmo model: that neural signatures of memory reactivation are tightly coupled to a particular phase of a hippocampal 8Hz oscillation; and that the optimal phase for memory retrieval is flipped relative to the optimal encoding phase along this same theta oscillation (Hasselmo et al., 2002).

Third, to scrutinize the temporal relationship between memory retrieval and theta phase, we tested whether the time points where our classifier indicated maximal neural memory reinstatement were time-locked to a consistent phase in the same frequency range, as would be the case if retrieval was initiated at a particular theta phase. A classifier-locked EEG analysis, inspired by animal work, revealed significant phase alignment at 7-8Hz, preceding the time points of maximal memory reactivation by approximately 200-300ms (Figure 3B-C). This cluster remained robust irrespective of whether we included only one classifier maximum or several maxima per trial (Figure S2C), when including correct trials only (Figure S2D), and when excluding early maxima close to the onset of the word cue (Figure S2E-F). Together, these findings suggest a close functional relationship between the phase of an ongoing theta oscillation, and neural memory reinstatement as measured by EEG classifiers, in line

with the computational models that motivated our hypotheses (Hasselmo et al., 2002; Ketz, Morkonda, & O'Reilly, 2013; Kunec et al., 2005).

The functional coupling between memory reinstatement and oscillatory phase is further corroborated by an analysis that contrasted phase consistency between classifier maxima of high and low fidelity, used as a proxy for strong vs weak memory reactivation (Figure 3D-E). Phase consistency in the 7-8Hz frequency and -500 to -200ms time range was higher for high-fidelity trials. The sources producing the difference between high and low fidelity maxima spanned medial and lateral parietal regions, and medial temporal lobe areas including the hippocampus. These regions are typically engaged during successful recollection (Rugg & Vilberg, 2013) and show strong functional connectivity with the hippocampus (Wang et al., 2014). While we cannot establish the hippocampus as a unique source of the theta phase-locking effect, our results are at minimum consistent with a hippocampal theta oscillation that extends into the functionally connected core recollection network. A link to medial temporal is also corroborated by the first analysis showing modulation of memory reactivation by the hippocampal 8Hz phase (Figure 2E). Together with the phase-locking results, our findings thus support theories suggesting that episodic memory retrieval relies on periodic cycles of communication between storage/retrieval systems in medial temporal lobe and neocortical areas that represent the various components of an episode (McClelland et al., 1995; Teyler & DiScenna, 1986).

The exact time course of the interaction between hippocampus and neocortex during retrieval is still not fully understood. Electrophysiological studies using time-resolved

multivariate methods have detected memory reactivation in the typical recollection time window (Jafarpour et al., 2014; Johnson et al., 2015; Michelmann et al., 2016). Consistent with this timing, our classifier maxima had a tendency to cluster in the recollection window around 400-800ms post-cue (Figure S2A). Our main interest in this study, however, was whether neural reactivation was linked to a consistent oscillatory phase in the theta band irrespective of when exactly it is triggered within a trial. Our findings provide strong evidence for such phasic modulation within a recall trial, in line with models suggesting that memory retrieval is initiated at an optimal phase of a hippocampal theta oscillation (Hasselmo et al., 2002).

At the exact time of the classifier maxima, we observed a significant difference in the ERPs distinguishing between the different types of retrieved memories (i.e., animate vs inanimate, Figure 4). The main source of this difference was localized to the anterior temporal lobe, consistent with this region's role in representing abstract object information (Patterson et al., 2007). Note that it is not surprising that we observed such an ERP effect, since the classifier requires a reliable signal difference in order to detect differences in reactivated content. The source of this signal is interesting, however, indicating that the classifier's decisions are based on information that originates from neocortical sources that are likely to represent the reactivated memory's content, and have little overlap with the sources of the theta phase-locked signal. Overall, our findings suggest that a few hundred milliseconds before the brain reinstates a memory in neocortex, an oscillating process in the MTL initiates retrieval, leading to a memory signal that oscillates and is modulated by the hippocampal theta phase (Teyler & DiScenna, 1986; Teyler & Rudy, 2007).

To our knowledge, our study is the first that directly links memory reinstatement to theta phase in human long-term memory. Previous studies have investigated the role of theta phase in working memory, and have provided first evidence for a phase shift between encoding and retrieval (Rizzuto, Madsen, Bromfield, Schulze-Bonhage, & Kahana, 2006). They also suggest that theta phase plays a role in orchestrating gamma (30-80Hz) oscillations during periods of working memory maintenance (Fell & Axmacher, 2011; Jensen, 2006). High frequency activity in the gamma range is thought to represent the firing of cell assemblies that code for the content of mental representations, and lower frequencies presumably provide the time windows for the firing of these assemblies (Fell & Axmacher, 2011; Fuentemilla et al., 2010; Jensen, 2006; Nyhus & Curran, 2010; Tallon-Baudry & Bertrand, 1999). Following this logic, Fuentemilla et al. (2010) used a delayed match-to-sample working memory task to investigate how gamma patterns representing the encoded material re-emerged during maintenance. Reactivation took place several times over a 5-sec delay, and these reactivations were phase-locked to a theta oscillation. Rodent work also suggests a link between gamma oscillations and theta phase. Different hippocampal subfields produce faster or slower gamma oscillations depending on whether the animal is encoding novel information or retrieving familiar information, and these two gamma rhythms are coupled to distinct phases of the hippocampal theta rhythm (Colgin et al., 2009). Our results provide the first evidence for a similar relationship in human long-term memory, using a classifier-based metric rather than gamma oscillations as a proxy for memory reinstatement and its relationship to the ongoing EEG.

We hope that our method will prove useful as a general approach for probing the relationship between information coding and the phase of slow oscillations. Phase coding has been suggested as an important mechanism outside the memory domain, including attentional selection (Jensen, Bonnefond, & VanRullen, 2012) and spatial navigation (John O'Keefe & Michael L. Recce, 1993). Within memory, our approach could be used to directly test whether distinct parts of a sequence of events are represented at different phases along a theta oscillation (Heusser et al., 2016), or whether memories are reactivated at specific phases of slow oscillations during sleep (Hanert, Weber, Pedersen, Born, & Bartsch, 2017; Staresina et al., 2015). Computational models (Norman, Newman, et al., 2006) also postulate that phase coding is crucial for resolving mnemonic competition when several memories are simultaneously reactivated by a reminder. Building on our method and findings, follow-up studies can directly test phase coding as a mechanism of organizing memories (e.g. according to their relevance) during encoding, during offline periods following encoding, and when reactivating memories during retrieval.

In sum, the present experiment shows that memories – or their neural signatures – wax and wane on a millisecond time scale within a trial, and that their neural reactivation follows the phase of a 7-8Hz theta rhythm. These findings provide the first direct support for theta phase encoding-retrieval models in the human brain, and thus bridge an important gap between computational, rodent and human work.

7. Acknowledgements

This work was supported by a fellowship from Stiftelsen Olle Engkvist Byggmästare awarded to M.W. and C.K., and a Starting Grant from the European Research Council awarded to M.W. (ERC-2016-STG-715714). We also thank James Lloyd-Cox for help with data acquisition, and Matthias Treder, Benjamin Griffiths and Sebastian Michelmann for their helpful conceptual input during data analysis.

8. Authors contributions

Conceptualization, C.K., J.L.-D. and M.W.; Methodology, C.K., J.L.-D., S.H., and M.W.; Investigation J.L.-D.; Formal Analysis, C.K. and J.L.-D.; Writing – Original Draft, C.K. and M.W.; Writing – Review & Editing, C.K., J.L.-D., S.H., and M.W.; Visualization, C.K.; Funding acquisition, C.K. and M.W.

9. Declaration of interests

The authors declare no competing financial interests.

10. Quantification and statistical analysis

10.1. Behavioural data

N = 24 for all behavioural analyses.

Correlation between the two measures of remembering were highly correlated, using the Spearman's rank correlation coefficient implemented in the MATLAB function corr, and can be seen on page 4 (rSpearman = 0.60, p = .002).

Reaction times for the first button press when retrieving animate (Mean = 3.03 secs, SD = .95 secs, min = 1.28 secs, max = 6.01 secs) and inanimate (Mean = 2.96 secs, SD = .77 secs, min = 1.47 secs, max = 4.24 secs) objects did not differ significantly, t(23) = .57, p = .58. The time-window used for classification (-200ms to 1500ms around the word cue) thus only minimally overlapped with the time window where participants made a button press, and can be seen on page 4.

10.2. EEG data

N = 24 for all EEG analyses.

Power spectrum of classifier output was calculated by using the Fieldtrip function ft_freqanalysis, implemented in MATLAB. The baseline was calculated as described on page 27. Every frequency that exceeded the 95th percentile was considered significant. This was done on all 24 participants, and the results can be seen in Figure 2C.

Phase-amplitude coupling between EEG data and classifier output was calculated as described on page 6. The real data and the time-shuffled baseline were subjected to a

paired-samples t-test, for hippocampal virtual channels for retrieval, t(23) = 1.8191, $p < .05$, one-sided t-test (Figure 2E), and encoding, t(23) = 2.7494, $p < .05$, one-sided t-test (Figure 2E).

All phase-consistency analyses were calculated using the following procedure. The different conditions were inserted in the Fieldtrip function ft_freqstatistics on a scalp level, and ft_sourcestatistics on a source level, implemented in MATLAB, which performs a non-parametric cluster-based permutation testing.

The p-values for the different analyses were:

For all peaks at scalp level: $p = .0002$, Figure 3B.

For all peaks at source level: $p = .0002$, Figure 3C.

High vs low fidelity trials at scalp level: $p = .003$, Figure 3D.

High vs low fidelity trials at source level: $p = .009$, Figure 3E.

ERP at scalp level: $p = .04$, Figure 4A.

ERP at source level: $p = .003$, Figure 4B.

One Peak: $p = .001$, Figure S2C.

Only correct trials: $p = .002$ and .005, Figure S2D.

Excluding 400ms: $p = .001$ and .006, Figure S2E.

Excluding 600ms: $p = .049$, Figure S2F.

Testing for a uniform distribution, the MATLAB function crosstab was used. The function provides a chi-square test, to obtain significant difference between two distributions. The results revealed no significant difference, $\chi 2$ (19) = 20.00; $p = 0.333$), and can be seen on page 11, Figure S2A.

To identify oscillating frequencies, we implemented the procedure described on page 34. The results were compared to a constructed baseline, and only frequencies exceeding the 95th percentile of the baseline were considered significant (Figure 2D). To test for difference between the power spectra for all trials and only correct trials, the two matrices were subject to a one-sided paired samples t-test, where we expected higher power for only correct trials in the 7-9Hz frequency range of interest, $t(23) = 1.9425$, $p = .03$ (Figure S2B).

11. Data and Software Availability

Custom MATLAB code as well as data additional to the already published on http://dx.doi.org/10.17632/h4vcpxt4sr.1 will be made available upon request (fulfilled by Lead Contact, C.Kerren@pgr.bham.ac.uk). Since consent for sharing data at the level of the individual participant was not received originally, we can only make summary data available online or upon request.

12. Supplemental information

The results reported in the main text focus on an index of memory reactivation derived from classifiers trained and tested on the retrieval data. This purely retrieval-based index is used throughout the main results to demonstrate its relationship to theta phase. Below, we report additional analyses conducted on classifiers that were trained on the encoding data, and then tested either on the encoding or on the retrieval data.

Encoding-to-retrieval classification has been commonly used in previous studies (Waldhauser et al., 2016). We conducted the additional analyses to confirm that such classifiers can also successfully detect memory reactivation in our paradigm, and that their frequency characteristics are similar to our main, purely retrieval-based metric. The results are summarized in Figure S4.

12.1. Frequency characteristics of the encoding classifier timecourses

Encoding analyses were conducted on epochs time-locked to the onset of the animate and inanimate objects (-200ms to 1500ms). As a first step, an LDA was trained on encoding and then also tested during encoding (Figure S4A). In line with the existing literature on object perception, animate vs inanimate category membership could be best decoded in a time window around 300 ms after object onset, with an accuracy peak at 305 ms. The classifier fidelity timecourses were then averaged within participants and subjected to a Fourier Transformation, following the same procedure as for the retrieval data reported in the main text. The resulting spectra (Figure S4B) showed the strongest power in lower frequencies with peak at 3, 5, and 6 Hz exceeding the 95th percentile of the random label chance distribution.

During the time window where the LDA performed best, we also found a univariate ERP cluster ($p_{corr} < .05$) from 240-340ms with a frontal topography that significantly differentiated animate from inanimate objects during encoding (Figure S4C). Note that this cluster had a frontal topography similar to the main cluster differentiating animate from inanimate objects during retrieval (as shown in Figure S3A), providing a first

indication that content-specific processes engaged during encoding might be re-engaged during retrieval.

12.2. Generalization from encoding to retrieval

Based on this observation, we next tested explicitly whether classifiers trained to distinguish animate from inanimate objects during encoding could successfully discriminate those categories during retrieval. For this analysis, the LDA-based classifier was trained on each time point within the 240-340ms encoding interval identified above, and tested at each time point at retrieval (see Figure S4D). This approach revealed the highest decoding accuracy in a retrieval time window from approximately 800-1500ms, a window typically associated with successful recollection (Yonelinas, 2002). We then assessed the frequency characteristics of the encoding-retrieval classifiers using the same FFT method as before, but this time applied to the classifiers trained on the activity patterns between 240-340ms during encoding, and tested at each time point during retrieval (see Figure S4E). The resulting power spectra showed the maximum peak at 9Hz (5 and 9 Hz exceeding the 95th percentile), with a very similar distribution but at a slightly higher frequency peak compared with results we obtained when training and testing at retrieval (see main Figure 2C).

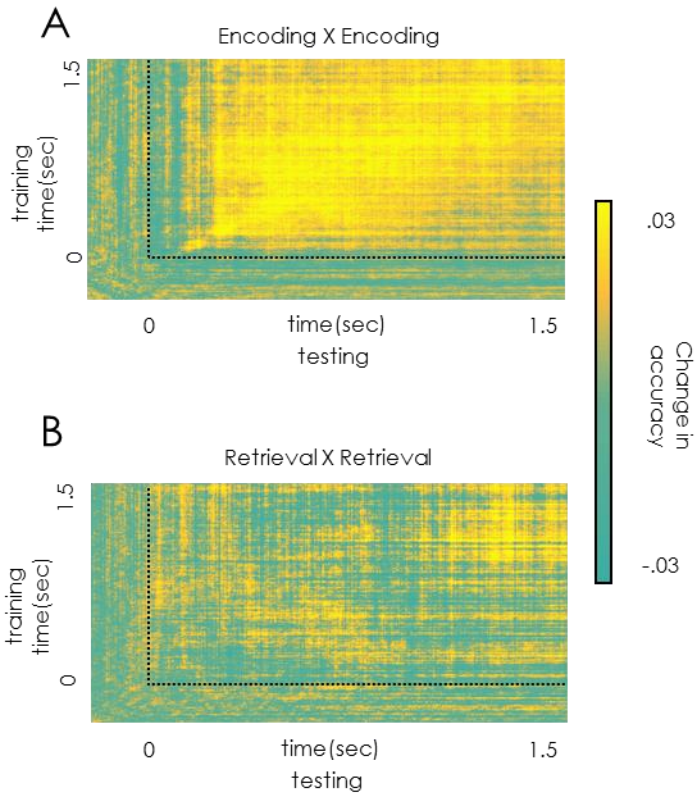

Figure S1. **Time generalisation matrices for encoding and retrieval, with time zero indicating the onset of the object during encoding, and the onset of the reminder word during retrieval.**
(A) Training and testing at encoding showed sustained high classifier accuracy from approximately 500-600ms to the end of the time window. **(B)** Training and testing at retrieval shows that accuracy is generally above baseline after cue onset, and indicates that participants reinstated the memory at different time points, and possibly several times. Unlike at encoding, the retrieval pattern suggests that there is not a sustained state across the entire time period, consistent with periodic reactivation. Each of the matrices in panels A-B is based on an LDA classification using a 5-fold cross-validation.

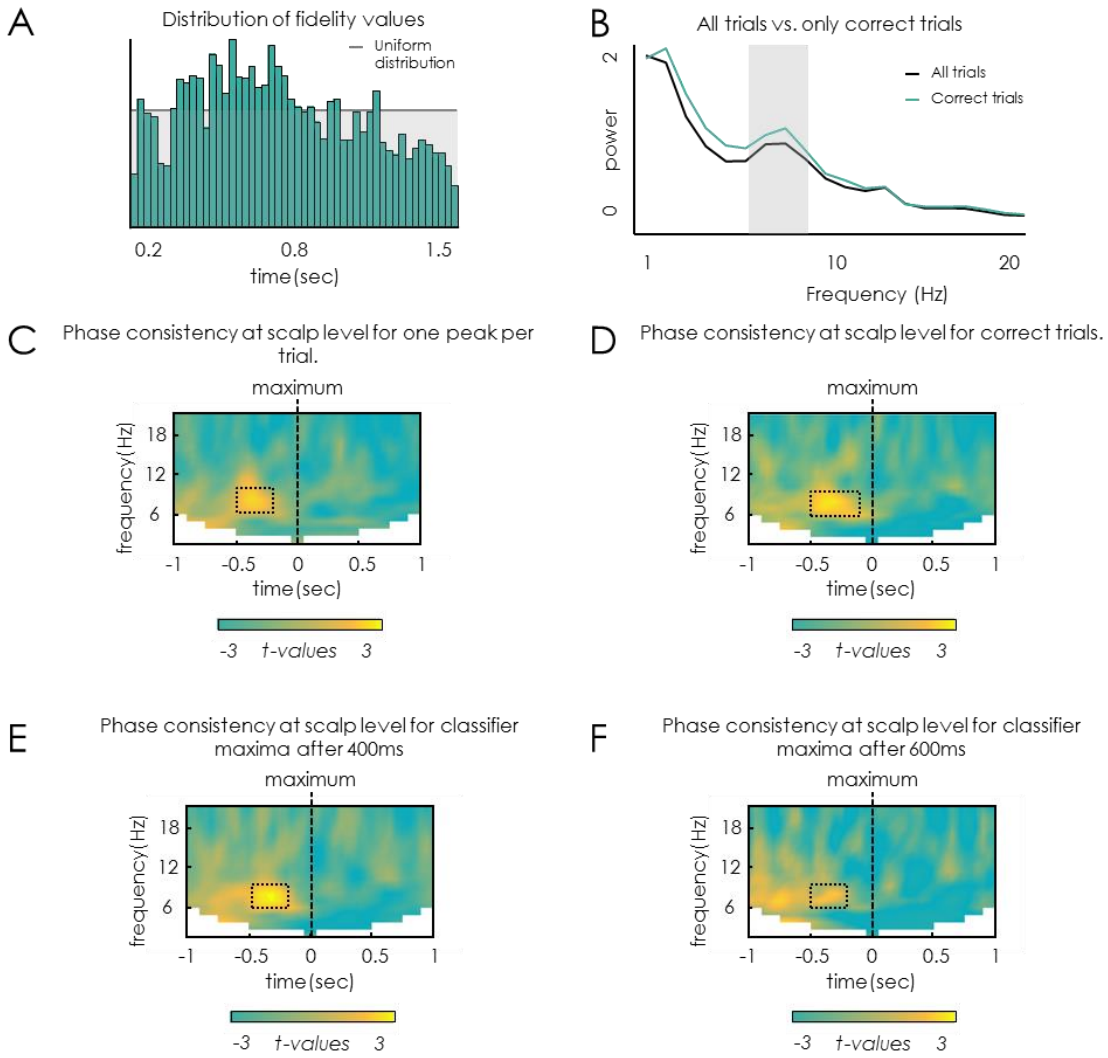

Figure S2. **Distribution of classifier maxima across participants and time, behavioural relationship to power spectra, and phase consistency for various control analyses.**
(A) The distribution of classifier maxima, accumulated across participants, showed no significant deviation from a uniform distribution, indicating that the maxima were evenly distributed across the entire retrieval period, with a noticeably increased density around 400-800ms. This is in line with previous studies showing strongest memory reinstatement in the recollection period. **(B)** To evaluate the relationship between the power spectra and memory performance, we compared the power spectra for all trials and correct trials only for 7-9Hz, which revealed a significantly stronger effect for correct trials compared to all trials. Note that a direct comparison between correct and incorrect trials was not possible due to a low number of incorrect trials in the cued recall task. **(C)** Classifier-locked averages showing phase consistency when using only the highest maximum per trial, and thus excluding all overlapping epochs. As expected, the phase consistency is less temporally smeared, with a cluster from -500ms to -150ms pre-maximum. **(D)** Same analysis as shown in main Figure 3, but limited to correct trials, showing a cluster of significant phase consistency 500-150ms before the classifier maxima. **(E)** Removing the first 400 ms of classifier maxima did not change the phase consistency effect, neither did removing the first 600 ms of classifier maxima **(F)**. When using only very late maxima, an even earlier cluster of 7-8Hz phase consistency becomes evident, with the later cluster at -500ms to -250 ms remaining significant. This results likely reflects several cycles of a 7-8Hz oscillation.

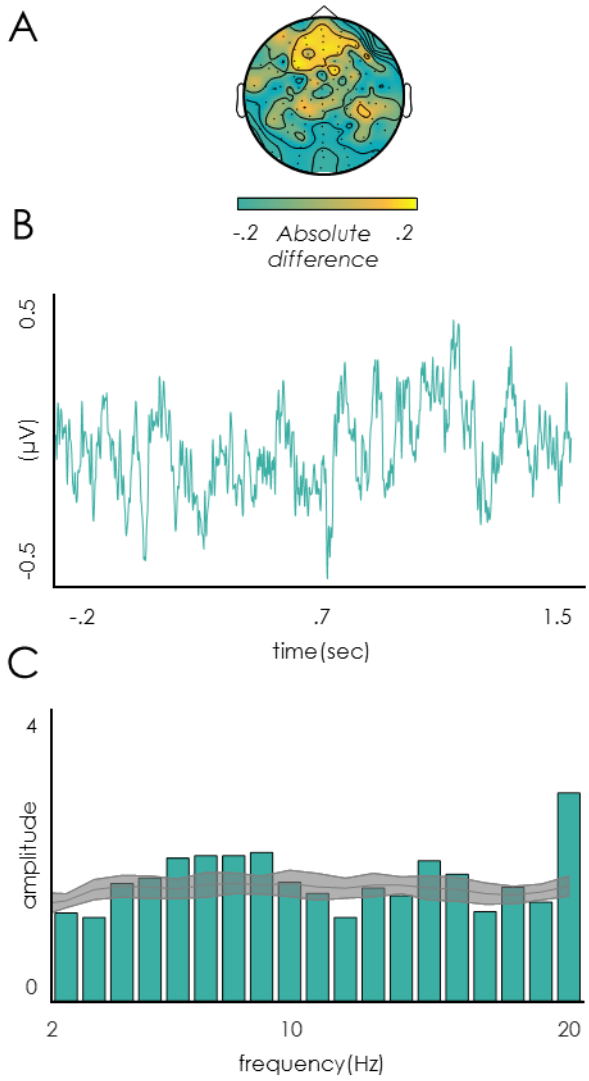

Figure S3. **Average difference between animate and inanimate object retrievals shown in the time and frequency domain.**
(A) Topographies of the absolute EEG difference between the recall of animate and inanimate objects, showing a frontal maximum during retrieval (600-1200ms). **(B)** Average difference signal between animate and inanimate objects during retrieval, interestingly showing a visible rhythmicity. **(C)** Applying the Fourier-transform, we can see above baseline power increases in spectral frequencies between 6-9Hz, the same frequencies that also show power increases in the classifier time series.

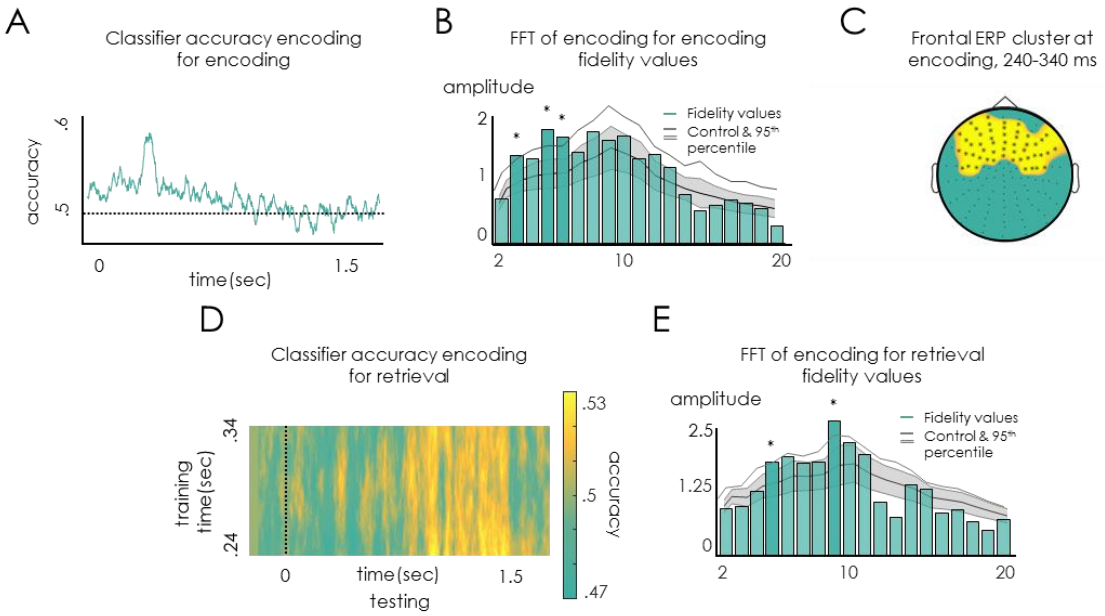

Figure S4. **Encoding for encoding and encoding for retrieval analyses.**
(A) Training and testing at encoding revealed a peak of classifier performance at ~300ms, a time window commonly seen when investigating encoding activity for semantic memory. **(B)** The averaged fidelity values were subjected to a Fourier Transformation, and showed a peak in the lower frequencies. **(C)** At encoding, a frontal cluster survived a non-parametric cluster-based permutation test, indicating an overlap with retrieval activity seen in Figure S3A. **(D)** Using the time points from 240-340ms during encoding, where animate vs inanimate differences showed an overlapping topography compared with retrieval (see Fig. S4A), a fluctuating pattern is also visible in the time generalisation matrix of a classifier trained on encoding and tested during retrieval. **(E)** The power spectra of this encoding-retrieval classifier revealed a peak at 9Hz.

Chapter 3 - Hippocampal Theta Phase-flip between Encoding and Retrieval.

At the time of thesis submission, this chapter represents preliminary analyses that are being prepared for publication (Casper Kerrén, Bernhard Staresina, Maria Wimber).

Abstract

Encoding and retrieving an event has shown to demand different computations in hippocampal circuits. Encoding is presumably characterised by strong input from EC into subfields CA3 and dentate gyrus, whereas retrieval is thought to rely on strong intra-regional signalling in area CA3, and output from CA3 via CA1 to EC. Empirically, evidence for such differential connectivity has been shown in rodents, but only one recent study using scalp EEG has been able to investigate these dynamics in human subjects (see Chapter 2). In the present study, intracranial EEG data from 11 patients were recorded whilst they were participating in a word-object associative recognition task, intending to replicate the previous findings showing that the decodability of memories oscillates and is phase shifted between encoding and retrieval, and importantly, to directly link these fluctuations to the hippocampal theta rhythm. To this end, an identical approach as in the previous study was used, decoding from multivariate patterns over neocortical electrodes, and linking the fluctuations in decoding fidelity to the theta phase obtained from hippocampal electrodes. Once again, evidence is shown for encoding and retrieval being oscillating processes that are phase-shifted along the hippocampal rhythm. These results suggest that there is an agreement between recording data on different anatomical scales, but more importantly, that the neural patterns reflecting the relevant mnemonic information during encoding and retrieval are time-locked to different phases of the hippocampal oscillation.

1. Introduction

Two fundamental mechanisms of episodic memory are the ability to encode and recall event-specific information (Tulving, 1972). At encoding, the perceptual input from external sensory sources arrives in the hippocampus. Here, it is assumed that an index of this event is laid down, with pointers to neocortical areas where the actual content of that memory is being represented. At retrieval, this index can be reactivated by appropriate reminders, signalling assemblies in the neocortex to reinstate the event. One major challenge for the memory system is to know when to encode a new event, and when to recall a previously stored event. A promising framework for how the brain avoids interference between encoding and retrieval computations is by temporally separating these two mechanisms in time along the hippocampal theta rhythm (Hasselmo, 2005; Hasselmo, Bodelon, & Wyble, 2002; Kunec, Hasselmo, & Kopell, 2005; Manns, Zilli, Ong, Hasselmo, & Eichenbaum, 2007).

Previous studies in rats have shown that the hippocampal theta rhythm supports the chunking of information employing phase-dependent long-term potentiation (LTP) and long-term depression (LTD) (Nyhus & Curran, 2010). One study found that a neural assembly stimulated at one phase of the hippocampal theta rhythm was strengthened through LTP, whereas stimulating at the opposite phase weakened the assembly through LTD (Pavlides, Greenstein, Grudman, & Winson, 1988). In humans, coding different information along the phase of the theta rhythm has been proven important in sequence encoding in episodic memory (Heusser, Poeppel, Ezzyat, & Davachi, 2016) and in working memory (Fuentemilla, Penny, Cashdollar, Bunzeck, &

Duzel, 2010). Furthermore, the hippocampal theta rhythm also seems to separate competing memories in goal-directed navigation (Kunz et al., 2019), and the different computations during encoding and retrieval in episodic memory (Kerren, Linde-Domingo, Hanslmayr, & Wimber, 2018). On a cellular level, different phases of the theta rhythm are related to changes in synaptic currents in the hippocampal system. Encoding new information creates strong input from the entorhinal cortex (EC) to CA1, at the same time as the input from CA3 is weak. This creates perfect dynamics for encoding new material through LTP (Hasselmo et al., 2002). On the contrary, EC input into the hippocampus during retrieval is weak, but the output from CA3 back to EC is strong, which presumably leads to the activity being propagated to synapses that were modified at encoding, leading to the retrieval of these stored associations. LTP is low at this point, which might be adaptive in the sense that incorrectly retrieved information is prevented from being encoded (Hasselmo et al., 2002).

Evidence for this model has been shown in several studies in rodents (e.g., Douchamps, Jeewajee, Blundell, Burgess, & Lever, 2013), however, only one study has shown direct support in the human brain (Kerren et al., 2018). In this study (Chapter 2), scalp electroencephalography (EEG) was recorded whilst participants completed an associative memory task. At encoding, they associated an action verb with an image. At retrieval, participants were asked to recall the image when prompted with the word cue. A multivariate pattern classifier was used to discriminate the spatial patterns corresponding to the different image classes at encoding and retrieval (animate and inanimate), and classification fidelity on a single trial level was used as an index of perception and memory retrieval, respectively. A coupling between the

amplitude of this memory index and the phase of the theta rhythm was conducted, showing that during the encoding of episodic memory, information was preferentially coded at a theta phase that was approximately 180 degrees phase-shifted relative to the preferred phase during the retrieval of the same episodic memory. Critically, the theta phase was, in this former study, extracted from virtual sensors in the hippocampus, and thus relied on an accurate source reconstruction of hippocampal activity.

In the present experiment, the goal was to replicate these findings, but this time by analysing data from 11 patients suffering from pharmaco-resistant epilepsy. Depth electrodes were implanted bilaterally, and only electrodes from the healthy hemisphere were analysed. Neocortical contacts were used to train the pattern classifier and detect mnemonic information during encoding and retrieval, while contacts in the hippocampus proper were used to define the theta rhythm. This small-scale anatomical approach allowed us to test whether it is truly the hippocampal theta phase that clocks neocortical reinstatement, as concluded from the scalp EEG data in Chapter 2. Data were recorded whilst participants performed an associative memory paradigm. The task was to encode a word together with either a scene (indoor or outdoor) or a colour (blue or red). At retrieval, participants were shown both previously seen (old) and unseen (new) words together with scene and colour choice options and were asked whether they recognised a word as old or new. If responding old, they were further asked to identify the specific image associate presented together with a word during encoding from the choice options given. Classifiers were trained on neocortical activity using the same retrieval time window as in the previous study (200ms to 1200ms post-cue), and coupled the output with the phase of the data from one anterior

and one posterior hippocampal electrode contact per participant. The findings from the previous study were replicated, demonstrating agreement between source localising scalp EEG data and intracranial data. Importantly, it is in the present study again shown that the brain separates encoding and retrieval along the hippocampal theta rhythm to minimise interference between these computationally incompatible processes.

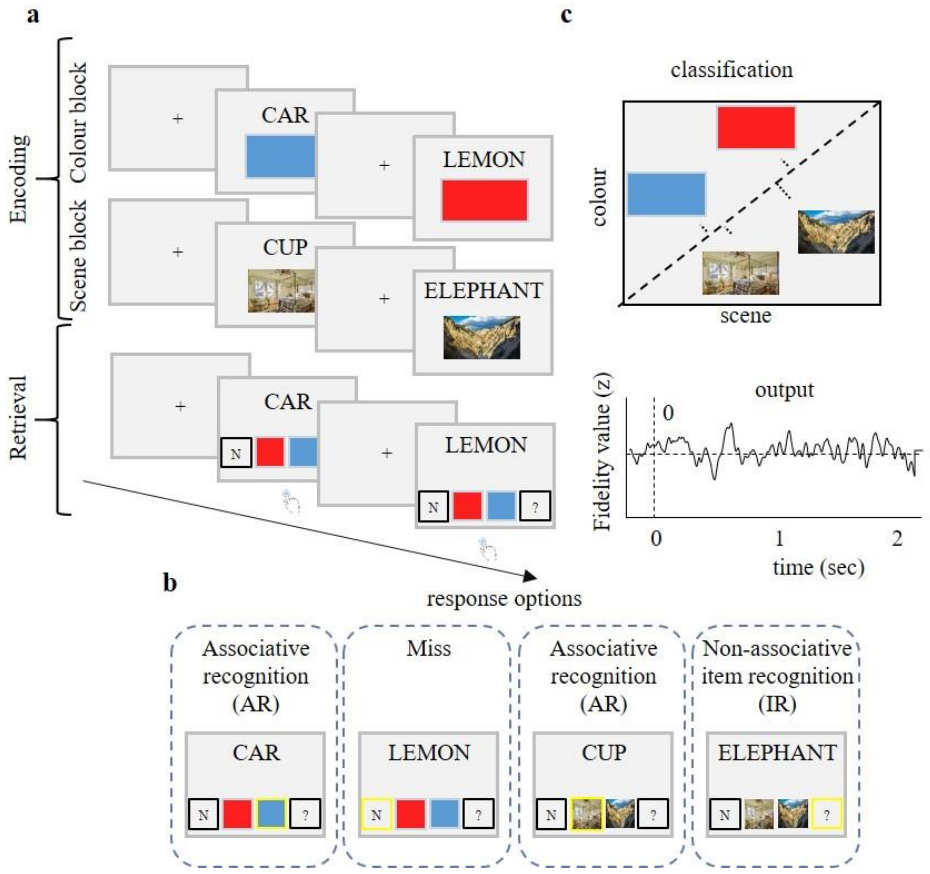

***Figure 1*. Paradigm and rationale for decoding analysis. A)** At encoding, subjects learned words and images. The images were of colours (red and blue) or scenes (indoor and outdoor). **B).** At retrieval, all words from encoding were shown, intermixed with new words, and a number of response options. The task for the subject was to judge whether the word had been seen before or not, and if seen, to select the original colour or scene associate that was paired with the word during encoding. **C)** Colours versus scenes were used as the two classes that we fed into the classifier. The output on a single trial level was the fidelity value, indicating the distance to the hyperplane best separating the two classes at any given time point. This metric was used as an index of perception at encoding, and as an index of memory reinstatement at retrieval.

2. Method details

2.1. Participants

A total of 11 participants participated, six were female, and the mean age was 34 years (range 23–44). Due to a non-reliable determination of anterior and posterior electrode contacts, one additional patient was excluded from the current data analysis, as compared to Chapter 5. Informed consent was obtained from all participants, and the study was approved by the ethics committee of the Medical Faculty of the University of Bonn. All participants were patients suffering from pharmaco-resistant epilepsy, who underwent presurgical monitoring via depth-electrodes (see below for electrode placements) to determine the epileptic focus.

2.2. Experimental procedures

Participants were sitting upright in a sound-attenuated room in front of a laptop computer at an approximately 50-cm distance and partook in an associative learning paradigm (Figure 1), where each block consisted of an encoding phase, a distracter phase, and a retrieval phase. During encoding, a German noun was paired with either a colour or a scene, depending on which run they were in. Colour and scene runs alternated, wherein colour runs, the noun was either paired with a red or blue square and in scene runs with either an indoor or an outdoor image. The encoding task was to

associate the word and the picture stimulus (colour or scene) by imagining, as vividly as possible, the entity described by the noun together with the colour (e.g., a red lemon) or scene (e.g. an elephant in the mountains), and to rate the plausibility of that self-created image. Participants were given a maximum of 3 seconds to make their plausibility judgment. Each trial was preceded by a jittered intertrial interval (700–1300 ms, mean = 1000 ms) during which a fixation cross was shown in the centre of the screen. Participants' button press terminated the trial. During retrieval, participants were presented with 50 previously seen words intermixed with 25 novel words, together with four choice options. The task was to indicate, with a single button press, whether the word was new ('N' response), whether it was old but the target association could not be retrieved ('?' response), or whether the word was old and the target association was also remembered (in which case the correct colour or scene response was to be selected). Responses were given in a self-paced manner but had an upper time limit of 5 seconds. Again, each trial was terminated with the button press and was preceded/followed by a jittered intertrial interval (700– 1300 ms, mean = 1000 ms) showing a fixation cross. Each run lasted approximately 9 minutes. Eight participants completed all six runs and three participants completed five runs. Two participants completed the six runs in two different sessions.

2.3. Electrode selection

Intracranial EEG depth electrodes were implanted during the presurgical evaluation, either laterally or along the longitudinal axis of the hippocampus. Seven of the eleven participants were implanted according to the longitudinal axis implementation scheme,

whereas four participants were implanted laterally via the temporal lobe (Figure 2A). Participants also had additional electrodes placed on the scalp at position Cz, C3, C4, and Oz, according to the 10-20 system. These electrodes were later removed from further analyses. Participants received anticonvulsive medication (plasma levels within the therapeutic range). The drug regimen at the time of recording for each participant is listed in Supplementary Table 1. In consideration of recent literature suggesting different computations along the longitudinal axis of the hippocampus (Poppenk, Evensmoen, Moscovitch, & Nadel, 2013), one anterior and one posterior hippocampal depth electrode was selected per participant based on anatomical and functional criteria. The anatomical evaluation was done by two researchers independently, whereas the functional evaluation was done by the means of pair-wise channel coherence (4-8Hz) to reveal functional transitions between regions along the axes of the hippocampus (Mormann et al., 2008; Staresina, Fell, Do Lam, Axmacher, & Henson, 2012). If multiple channels were anatomically located in the anterior or posterior hippocampus and were clustered based on the pairwise channel coherences, the channel with the highest signal quality (based on event-related potentials [ERPs]) was selected, following the same approach as used in (Staresina, Fell, Do Lam, Axmacher, & Henson, 2012).

3. EEG Analyses

3.1. Preprocessing

Data processing was performed with FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and standard MATLAB functions. Depth electroencephalograms were referenced to linked mastoids and recorded with a sampling rate of 1 kHz (bandpass filter: 0.01Hz to 300Hz). For both encoding and retrieval trials, data were divided from 1000ms pre-stimulus to 3000ms post-stimulus onset. We made sure that the two selected channels in the hippocampus were not physically defect, but other than that no further artifact rejection was conducted.

3.2. Multivariate pattern analysis

Preprocessed trial time courses from all but the hippocampal channels were used for classification. A Gaussian sliding window of full-width at half maximum of 40ms in the time-domain was used to attenuate unwanted noise. A baseline correction was applied by subtracting from all channel time courses the mean voltages over a time window from 400ms to 200ms pre-cue. Two separate LDA analyses were conducted on the encoding and retrieval trials (following Kerren et al., 2018), where the classifier was in each case trained and tested on each time point from 200ms pre-cue to 3000ms post-cue on its ability to discriminate the two different classes of images (colour and scene). A leave-one-out approach was used, such that cross-validation was performed by leaving one trial out when training the data, and testing the trained classifier on the remaining trial (Grootswagers, Wardle, & Carlson, 2017). This procedure was repeated until all trials had been tested. For a first overview of LDA performance, accuracy was averaged across all tested trials, resulting in the probability that the classifier had taken the correct decision based on the true label (Figure 2B and C). On

a single-trial level, the distance from the optimal decision-boundary that could best separate the brain patterns belonging to the two classes relevant for classification was extracted. The amplitude of these fidelity values at any given time point was used for the phase-amplitude coupling, with the computing of the hippocampal phase described in a later paragraph.

3.3. Maximum spectral power for hippocampal channels

The Irregular-Resampling Auto-Spectral Analysis (IRASA) method has shown to be robust in finding and separating the oscillatory signal both in ECoG and MEG data and was used to quantify the oscillatory signal of hippocampal channels (Wen & Liu, 2016). More specifically, the brain produces task-related rhythmic (oscillatory) components, but also arrhythmic scale-free (fractal) components, and these are probably produced through different brain mechanisms (Buzsaki & Draguhn, 2004) and should, therefore, be separated. The rhythmic (oscillatory) components are regular across time, whereas the arrhythmic (fractal) components are irregular (Wen & Liu, 2016). In short, IRASA resamples a time-series signal and computes a geometric mean of every pair (oscillatory and fractal) of the resampled signal. The median of the geometric mean is taken, which is then used to extract the fractal power spectrum. The difference between the original power spectrum and the fractal power spectrum is the estimate of the power spectrum of the oscillatory component of the signal (Wen & Liu, 2016), and the results of applying this analysis to all hippocampal channel time series can be seen in Figure 2E.

3.4. Phase-amplitude coupling between fidelity values and hippocampal channels

To obtain an estimate of the phase of the hippocampal channels, data were convolved with a complex Morlet wavelet of 6 cycles. To normalise the data, each entry was divided with its magnitude. Before coupling the phase of the hippocampal channels to the output of the classifier, we determined the time points in the raw data that contained oscillations, using the MODAL algorithm (Watrous, Miller, Qasim, Fried, & Jacobs, 2017), with the goal to only include time points when there was an oscillation in the theta range (<9Hz) (see: Kunz et al., 2019 for same procedure). The logical matrix was point-wise multiplied with the complex number matrix obtained from the Fourier transformation, and the angle was extracted using MATLAB's function *angle*. The data were then split into ten adjacent phase bins ranging from minus pi to pi (-3.14 to 3.14). Next, the output from our LDA-based classifier, trained and tested on non-hippocampal channels, was related to the theta phase determined from the hippocampal channels as follows. The z-scored amplitude of the fidelity values corresponding to a given phase bin, derived from classification anywhere between 200ms to 1200ms post cue onset at encoding and retrieval was computed, closely following the procedure in Kerren et al. (2018). The binned classification values were averaged across trials, leaving one value per bin per participant. The modulation index (MI) was calculated by comparing the distribution of fidelity values across the ten bins to a uniform distribution, the latter obtained by simply calculating the mean fidelity across all bins (Tort et al., 2010). To infer whether the distribution was different from a uniform distribution, the Kullback-Leibler (KL) distance was calculated. The values

from this calculation range between 0 and 1, with 1 being a Dirac-like distribution (1 for one bin and 0 for the rest) and 0 is a uniform distribution. To infer whether the MI was significantly different from what can be expected by chance, we randomised the trials in the classification output before coupling them with the hippocampal phase and sorted them into their corresponding phase bins and calculating the MI. This was done 5000 times, and the 95th percentile of these MI values was taken, against which the value from the real MI was compared.

4. Results

4.1. Behavioural results

Behavioural results showed that participants on average had 51.9% hits (range = 26.9% to 78.1%, standard deviation, SD = 16.5%), calculated as source correct/(source correct + source incorrect + source don't know). The large range of hit rates between participants indicated that the task was harder for some than for others, as expected in a patient sample. However, all participants were well within the boundaries for the exclusion criterion interval (±2.5SD cut-off from the mean). Out of the unstudied words, participants correctly rejected (correct rejection/(correct rejection + false alarm) 78.9% (SD=16.4%), indicating that they were on average very accurate at detecting new words as being new. All source correct trials were used for subsequent analyses, both for classification and calculating MI. For the contrast between source correct and correct rejection, again all trials from these conditions were used.

4.2. Significant decoding performance revealed for both encoding and retrieval.

As an initial control for subsequent analyses, it was of interest to ensure that the decoding of colour versus scenes created reliable results during encoding and retrieval. To this end, an LDA classifier was trained and tested at each time point, using a leave-one-out cross-validation approach. Average classification performance significantly (t-test) exceeded chance-level (50%) early on for both encoding and retrieval (Figure 2B and C). For encoding, a distinct peak was visible approximately 200ms after the onset of cue and image. At retrieval, again an early peak was visible, most likely reflecting the visual processing of the selection screen, but this time accompanied by a later peak at around 1000ms post-onset, the latter more likely to reflect genuine memory reactivation (Staresina & Wimber, 2019).

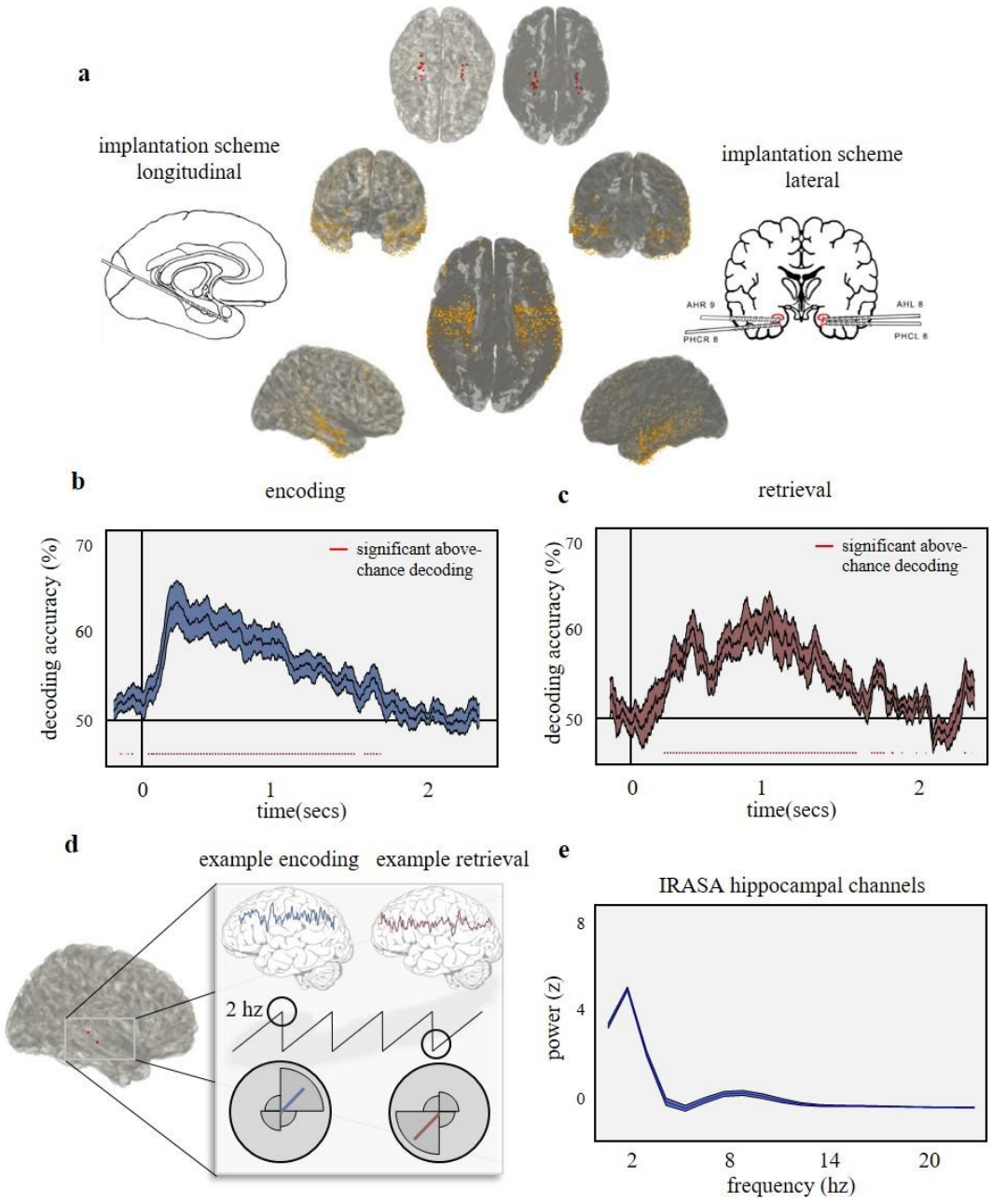

***Figure 2*. Decoding, electrode coverage, peak frequency and rationale for phase-amplitude coupling. A)** The electrode coverage for extra-hippocampal electrodes for all patients can be seen in yellow, and in red for hippocampal channels. On the left and right edges, the two different implantation schemes are shown. **B)** Decoding colour versus scene at encoding resulted in reliable above-chance performance from cue onset. **C)** At retrieval, the performance reached significance early on, and lasted until around 1500ms. **D)** Hypothesis that encoding and retrieval are maximally decodable at different phases of the hippocampal theta oscillation. **E)** The frequency with the strongest oscillatory component in hippocampal channels at retrieval was 2Hz, with a secondary peak at 8Hz.

4.3. Maximum power in low frequencies for hippocampal channels

Having established that the decoding accuracy was reliable and significant across participants, the next step was to investigate the phase-amplitude coupling between the fidelity values of the decoding and the hippocampal phase (Figure 2D). For phase-amplitude coupling, the frequency with the strongest spectral power between 2 to 8 Hz was used as the phase-modulator. To this end, one anterior and one posterior channel was selected for each of the 11 participants, following Staresina et al. (2012). The oscillatory signal for anterior and posterior hippocampal channels was calculated separately for encoding and retrieval from 200 to 1200ms post screen onset, using the IRASA method (Wen & Liu, 2016) (Figure 2E). Averaged over channels and participants, 2Hz was the strongest frequency and was therefore used for all subsequent phase analyses. This pattern was unchanged for encoding anterior, encoding posterior, retrieval anterior, and retrieval posterior.

4.4. Hippocampal phase modulation of neocortical patterns

Next, the amplitude of the classification output on a single trial basis was coupled to the phase of the hippocampal rhythm separately for the anterior and posterior channel in the time window of interest (see Method section, 3.4). The Kullback-Leibler distance was calculated between the real data and a randomly shuffled trial data. Significance was determined in two complementary ways (see Method section, 3.4).

First, we tested for a significant modulation of classifier fidelity by the phase of a hippocampal 2Hz oscillation, identified as the strongest oscillatory component in the hippocampus by the IRASA algorithm. A significant MI (exceeding the 95th percentile of the chance distribution, p-value .05) was obtained for both encoding and retrieval, and with respect to anterior and posterior hippocampal channels (encoding anterior real data mean = .013, SD = .0045, estimated chance distribution mean = .0086, SD = .0013; encoding posterior real data mean = .011, SD = .0031, estimated chance distribution mean = .009, SD = .0015; retrieval anterior real data mean = .012, SD = .0042, estimated chance distribution mean = .0062, SD = .0009; retrieval posterior real data mean = .011, SD = .0048, estimated chance distribution mean = .0064, SD = .001). This evidence support the hypothesis that decodability of the perceptual input at encoding, as well as the recalled content at retrieval, are significantly biased towards a specific phase of the hippocampal theta rhythm at a 2Hz frequency (Figure 3A and B). For comparison, Figure 3A and B show the MI, relative to the chance distribution, for each frequency between 2 and 8 Hz. Modulation indices exceeded the 95th percentile of the chance distribution only for lower frequencies (2-3Hz for encoding, 2-4Hz for retrieval). Since our previous results from the IRASA analysis identified 2 Hz as the strongest oscillatory component in the hippocampus, we chose to continue all further analysis focused at this 2Hz frequency.

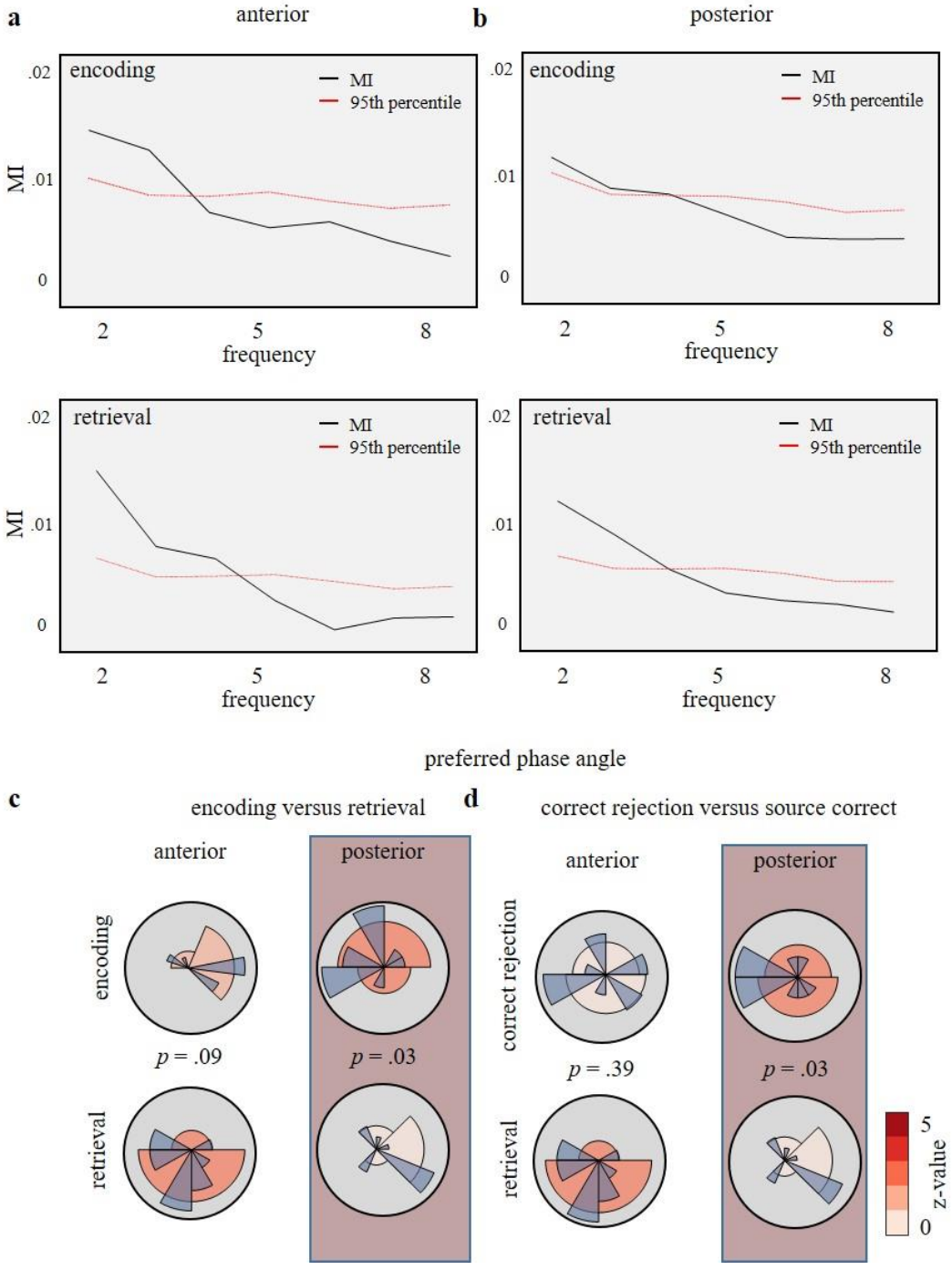
a
anterior
b
posterior
encoding
retrieval
MI
95th percentile
.02
.01
0
MI
2
5
8
frequency
preferred phase angle
c
encoding versus retrieval
d
correct rejection versus source correct
anterior
posterior
encoding
retrieval
p = .09
p = .03
correct rejection
p = .39
p = .03
5
0
z-value

***Figure 3*. Modulation index and absolute phase difference encoding and retrieval. A)** Phase modulation of classification fidelity by an oscillation extracted from anterior electrodes. The MI at 2 and 3Hz was significantly above the estimated chance distribution at encoding (top). At retrieval (bottom), modulation at 2-4Hz reached significance. **B)** Posterior channels showed the strongest modulation at 2Hz, with 2 and 3Hz being significant at encoding, and 2 to 4Hz being significant at retrieval. In A and B, the black line depicts phase modulation of the real data, and the red line rdepicts the 95th percentile of the estimated chance distribution. **C)** The preferred (i.e., median) phase angle where classification performance was maximal for encoding and retrieval. This preferred angle was phase-shifted by approximately 180 degrees between encoding and retrieval. **D)** Correct rejections showed a similar phase distribution as encoding, and was significantly different from retrieval at posterior channels. P-values between encoding and retrieval show the result of test of equal medians.

To find the hippocampal theta phase associated with maximum classification fidelity, the phase with the highest mean fidelity value for 2Hz was extracted individually per participant. The first analysis contrasted encoding trials with retrieval trials, using a non-parametric multi-sample test for equal medians. This test is similar to a Kruskal-Wallis test for linear data (Fisher, 1995) and was implemented using the *circ_cmtest* in MATLAB. For posterior hippocampal channels, this test revealed a significant difference of the median phase between encoding (mean = 157 degrees) and retrieval (mean = -36 degrees), z(10) = 4.7, p = .03 (Figure 3C). For anterior channels, this difference was trending but not significant (anterior encoding mean = -4 degrees, anterior retrieval mean = -121 degrees), z(10) = 2.93, p = .09.

4.5. Correct rejections mimic the phase patterns during encoding

Next, a contrast was conducted between trials on which participants correctly rejected new words as being new and trials on which they correctly remembered old words. The rationale for this analysis was that new, unfamiliar stimuli should elicit a novelty

signal that puts the hippocampal system in an encoding mode (Douchamps et al., 2013; Duncan, Sadanand, & Davachi, 2012; Lever et al., 2010). Based on this assumption, it was hypothesized that the phase of maximum image decodability on correctly rejected trials should show a similar circular distribution as on encoding trials and that the phase of maximal decodability should furthermore be significantly different from that during correct source retrieval. A non-parametric multi-sample test for equal medians (again using *circ_cmtest*) revealed a significant difference between distributions with respect to the phase of posterior hippocampal contacts (CR: mean = -173 degrees, correct source retrieval: mean = -36 degrees), z(10) = 4.7, p = .03) (Figure 3D). No significant difference between distributions was found with respect to the phase of anterior hippocampal electrode contacts (CR: mean = -147 degrees, correct source retrieval: mean = -121), z(10) = .73, p = .39. To investigate whether CR was similar to encoding the MATLAB function *circ_vtest* was used, testing whether the distribution of CR trials was uniformly distributed around the circle, or if it had a mean angle of the same as the encoding trials (157 degrees). The result revealed that CR trials were significantly clustered at the same phase as the encoding trials, z(10) = 4.86, p = .02. These results suggest that CR trials and encoding trials share a similar mean angle and that the phase difference between correct rejection trials and retrieval trials mimics that of encoding trials and retrieval trials.

5. Discussion

Forming new memories is assumed to require different circuit dynamics within the hippocampal system compared to retrieving a memory from the past (Hasselmo, 2005;

Hasselmo et al., 2002; Manns et al., 2007). The present study set out to replicate and expand a previous finding using scalp EEG in humans (Kerren et al., 2018), where a significant difference was found between the optimal phase for encoding and retrieval. In the present study, data were analysed from intracranial electrodes in 11 patients suffering from pharmaco-resistant epilepsy. Patients took part in an associative memory paradigm, where they were asked to associate a concrete noun with either a colour or a scene. At retrieval, old and new words were shown, and participants had to respond whether they had seen the word before or not, and if judged old, what image had been associated with the word during encoding. First, evidence that the classifier fidelity values were significantly modulated by the hippocampal 2Hz oscillation was provided. Importantly, both encoding and retrieval, anterior and posterior, were significantly modulated by the hippocampal 2Hz oscillation. These results, in principle, replicate the results of the previous study (Chapter 2), but also go beyond the results in Chapter 2 by providing a direct link to hippocampal theta. Second, when comparing the coupling to the hippocampal phase at which an image was maximally decodable at encoding and retrieval (Figure 3C), a significant difference in the preferred phase specifically for posterior hippocampal channels, again at the dominant hippocampal frequency of 2Hz, was shown. Thus, by analysing data directly from the hippocampal anterior and posterior regions for encoding and retrieval, it is demonstrated that the hippocampal theta oscillation modulates the neocortical traces of encoding and retrieval and assigns them different phases along the oscillation, as to keep them separated.

The hippocampal theta oscillation has been both modelled (Hasselmo, 2005; Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007) and empirically shown (Douchamps et al., 2013; Lever et al., 2010) to play a prominent role in shifting the encoding and retrieval dynamics in rodents. Along the hippocampal theta oscillation, different phases are related to changes in synaptic currents in the hippocampal system. It is believed that when a participant is required to encode a novel event, the hippocampal system is driven by strong input from EC at the same time as the intraregional activity in CA3 is weak, which dampens the retrieval-associated activity. At retrieval, the system is more dependent on the pattern completion process in CA3 wherefore the input from EC is weak, which leads to the activity being spread to synapses that were modified at encoding, thereby retrieving the stored associations (Hasselmo, 2005; Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007). This flip of focus puts the system into appropriate states of synaptic plasticity. In human subjects, the evidence for this flip is scarce. In one behavioural study, Duncan et al. (2012) showed that participants were biased towards identifying subtle changes to previously seen object images on trials that directly followed the encoding of new material when the brain presumably is in a novelty-triggered encoding mode. On the other hand, retrieving old objects on a given trial biased participants to integrate material (i.e., more likely to judge item as old and ignore the subtle changes) on the subsequent trial, similar to that of pattern completion. The authors argued that the processes engaged at encoding and retrieval had a lingering impact on a time scale of several seconds on subsequent mnemonic processing through a bias towards pattern separation (encoding) and pattern completion (retrieval) (Duncan et al., 2012). This is consistent with rodent work (Douchamps et al., 2013) showing that acetylcholine can

cause such slower shifts in circuit dynamics (Hasselmo & Fehlau, 2001). However, the model (Hasselmo et al., 2002) also predicts such fluctuations on a sub-second time scale along the theta rhythm. The first study to provide evidence for sub-second computation along the hippocampal theta oscillation and how it might modulate this flip was conducted using scalp EEG (Kerren et al., 2018). Although this study found evidence for a modulatory effect of the theta oscillation, no study has previously been able to simultaneously measure activity directly from the hippocampus and neocortical regions to directly assess the modulatory effect the hippocampal theta oscillation has on neocortical brain activity elicited at encoding and retrieval of an associative memory task.

On a single-trial level, the distance to the hyperplane best separating the two classes functioned as an index of both perception and memory reinstatement. Thus, on average it was assumed that fidelity values at theta phases when the system was in an optimal state for encoding would be significantly higher than during the opposite phase. Likewise, at retrieval, fidelity values would, at time points when the system was, in terms of theta phase, in an optimal state for retrieving previously encoded memories, be significantly higher than other phases. Furthermore, the maximal fidelity values at encoding would differ in phase along the hippocampal theta oscillation as compared to retrieval fidelity values, in line with the theta-based models tested in the present study (Hasselmo, 2005; Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007) (Figure 2D). Fidelity values on a single trial level were coupled to the phase of the hippocampal 2Hz oscillation, for anterior and posterior electrodes separately. The modulation index (MI) was calculated for both encoding and retrieval decoding time

series. The first evidence that the hippocampal theta rhythm modulates encoding and retrieval was shown in a significant MI for both processes (Figure 3A and B). This means that higher amplitudes were on average more biased to one phase than to others along this slow theta rhythm.

With the possibility of recording data from the longitudinal axis of the hippocampus, we could obtain separate measures for how the slow oscillation at anterior and posterior sites in the hippocampus modulated neocortical encoding and retrieval patterns. We found evidence for a modulatory effect of both encoding and retrieval separately, and a significant difference between the two processes. First, a significant difference of approximately 180 degrees between encoding and retrieval was found at posterior electrodes (Figure 3C). The posterior hippocampus has previously been implicated in pattern separation and in representing both more local spatial information and memory for more detailed contextual information (Poppenk, Evensmoen, Moscovitch, & Nadel, 2013; Poppenk & Moscovitch, 2011), consequently making posterior hippocampus better suited for noticing changes in mnemonic processing dynamics. Secondly, the strongest individual effects (see the red colour scale for corresponding z-values in Figure 3C) for encoding and retrieval were in the posterior and anterior hippocampus, respectively, with anterior electrode contacts showing significant phase clustering of decoding values at retrieval. This is again in line with previous studies showing anterior hippocampus to be more related to pattern completion and representing gist-like features of memories, whereas posterior hippocampus is related to pattern separation and representing more detailed mnemonic information (Poppenk et al., 2013; Schlichting, Mumford, & Preston, 2015).

Consequently, the findings are consistent with the idea that encoding dynamics require pattern separation to reduce the overlap between representations (Axmacher et al., 2010; Yassa & Stark, 2011), a process assumed to take place in the hippocampal subfield of the dentate gyrus (DG). At retrieval, on the other hand, the system relies on a pattern completion process taking place in subregion CA3, which increases the overlap between incoming cue representation and the stored memory, which supports the reactivation of the target memory.

Previous studies have shown stimulus-specific reactivation of previously encoded material during a working memory maintenance task is tightly coupled to the hippocampal theta oscillation (Fuentemilla et al., 2010). Similarly, the theta oscillation has been hypothesised to phase separate distinct working memory contents along the oscillation (Lisman & Idiart, 1995), and this assumption has recently been supported by evidence from intracranial recordings in human subjects (Bahramisharif, Jensen, Jacobs, & Lisman, 2018). Rizzuto et al. (2003) investigated phase reset in a working memory task and found a phase shift between encoding and maintenance in the first few cycles after cue onset. The work in the present study goes beyond this finding showing that not only does a stimulus reset the phase of the rhythm, but the hippocampus also continues to keep the phase consistent for encoding and retrieval at least the first second. Although not tested for in this study, it might be that the separation of encoding and retrieval is only present during the initial processing of the material, in line with a study showing that the hippocampal theta oscillations being relatively short bursts of events only lasting a few cycles in human subjects (Bush et al., 2017). Further evidence that the process might be short.-lived was shown in a

previous study by Staresina et al. (2012), in which they investigated the retrieval processing dynamics of the hippocampus using intracranial EEG recordings. In this study, the hippocampus showed a retrieval effect (source recognition relative to item recognition) with an onset of approximately 250ms after the recognition cue. At approximately 1 second, hippocampal signals shifted to show a novelty effect (correct rejection relative to item recognition). This means that after approximately one second, successful memory retrieval is followed by a re-encoding of the retrieved information. These results indicate that the hippocampal system might be biased towards certain mnemonic processing for just the amount of time needed for the task to be achieved (however, see Duncan et al., (2012), for evidence for longer-lasting biases).

An associative recognition paradigm was used in the present study as opposed to the cued recall paradigm used in Chapter 2. Participants made old/new decisions at retrieval, allowing for two different analyses. In addition to the contrast between encoding and retrieval trials (targeted at detecting mode effects), a contrast between correctly rejected (CR) new trials and correct old recognition trials was also conducted. To reject a trial as new during retrieval, arguably the participant needs to first encode it, as has been hypothesised elsewhere (Stark & Okado, 2003). Therefore, it was theorised that the contrast between correct old recognition trials and CR trials would show the same pattern as the contrast between encoding and retrieval. Additionally, as both of these processes took place at retrieval, we expected the difference to be found only in anterior hippocampus, due to its association with foremost retrieval dynamics (Poppenk et al., 2013). However, as has been shown before, the separation process is more associated with the posterior hippocampus, and it is therefore equally likely that

the difference would be found over posterior electrodes. A significant absolute phase difference between CR and correct old recognition trials was found on a group level for posterior electrodes (Figure 3D). Furthermore, when directly comparing the distribution of CR to the mean phase of encoding trials, a significant result was obtained, meaning that CR trials were clustered around the same phase as encoding trials. Previous studies have suggested that there is a lingering effect of the different mnemonic processes, such that when having enacted in an encoding or retrieval process you are biased to stay in this state (Duncan et al., 2012; Hasselmo & Fehlau, 2001) for some time. The results in the present study expand the understanding of the hippocampal system by also showing that the hippocampal system can switch between encoding and retrieval computations on a fast (sub-second) time scale within a trial.

Since encoding processes have shown to be more related to posterior regions of hippocampus and retrieval to anterior regions of the hippocampus (Poppenk et al., 2013; Schlichting et al., 2015), a better contrast of phase angles might have been the one between posterior encoding and anterior retrieval. There is currently a debate whether it is possible to contrast phases along the axis of hippocampus, with one side arguing that theta oscillations function as a clock that synchronises the entirety of hippocampus with zero phase lag (Buzsaki, 2002), whereas the other side has shown that theta oscillations are traveling along the long (septotemporal) axis of the hippocampus (Lubenov & Siapas, 2009). The latter result would suggest that each region is clocked by theta oscillations in this particular region, thereby making it impossible to contrast phase differences between regions. In a previous study in rodents showing evidence for encoding-retrieval dynamics, brain activity was

recorded from the region CA1 of the hippocampus (Douchamps et al., 2013), due to CA1's ability to act as a comparator by receiving input from both EC and CA3 at different phases of the oscillation (Vinogradova, 2001). They showed that at encoding, input from EC to CA1 preferably takes place at the theta peak, whereas retrieval input from CA3 to CA1 takes place at the trough. These results would thus suggest that it is beneficial to record encoding and retrieval activity from the same site.

Given the relatively big discrepancy between the peak frequency with which the fidelity values were coupled in Chapters 2 and 3 it might be worth noting that there is a controversy as to which frequency band functionally corresponds to the rodent theta band (4-8Hz) in the human brain. Some authors argue that hippocampal theta in humans is much slower (Jacobs, 2014), while others have found hippocampal memory effects is in the 6-9Hz range (Fuentemilla et al., 2010; Rutishauser, Ross, Mamelak, & Schuman, 2010), and still others have suggested a slow and a fast theta rhythm plays a role for episodic memory retrieval (Lega et al., 2012; Pastötter & Bäuml, 2014). Thus, both 2Hz and 8Hz are within the theta frequency band. However, given the temporo-occipital sources of the effects in foremost Chapter 2, it could be argued that the phase-locked activity in this study is more likely to reflect a neocortical process in the alpha band and that is why such a fast oscillation is picked up. Alpha oscillations have typically been related to information gating in sensory cortices (Bonnefond & Jensen, 2012; Klimesch, Sauseng, & Hanslmayr, 2007), and are thought to modulate the signal-to-noise ratio in cortical regions, allowing only the most relevant neurons to fire at phases of high inhibition. Concerning the findings in Chapter 2, it is thus possible that once a memory trace has been reactivated, neural memory reinstatement

is present in neocortex continuously but the classifier is most successful at detecting the signal during those phases of the alpha rhythm where the signal-to-noise ratio is high. While such a mechanism could produce a 7-8Hz peak in the power spectrum of the classifier output, as found in the present study, it would also predict that the time points of maximal classifier performance will coincide with the time points of maximal phase locking, because the signal can most easily be detected at a specific alpha phase. This alternative explanation is difficult to reconcile with the finding that phase consistency preceded neural reactivation by approximately 300ms.

Another possibility is that hippocampal theta detected on a scalp level in Chapter 2 is faster due to its synchronisation with a larger brain network, spanning from mid-frontal/septal to parietal areas (Canolty et al., 2006; Friese et al., 2013), which is easier to detect, or more dominant, when recording on a scalp level as compared to directly from hippocampus proper, consequently resulting in two different peak frequencies for the two studies.

Brain dynamics associated with encoding an event are clearly different from those of retrieving something that has already been experienced before. In this Chapter, the focus was on the role of the hippocampal theta oscillation in modulating and organising the dynamics optimal for encoding and retrieval states. The study first shows that the neural patterns representing to-be-encoded and to-be-retrieved event information are modulated by the phase of a slow hippocampal oscillation. Secondly, brain patterns for encoding events are significantly phase-separated from brain patterns when retrieving an event, in particular with respect to theta phase in the

posterior hippocampus. This study replicates previous findings, and further expands them by showing recordings directly from the hippocampus in patients suffering from pharmaco-resistant epilepsy.

6. Acknowledgements

This work was supported by a fellowship from Stiftelsen Olle Engkvist Byggmästare awarded to M.W. and C.K., and a Starting Grant from the European Research Council awarded to M.W. (ERC-2016-STG-715714).

7. Authors contributions

Conceptualization, C.K., B.S and M.W.; Methodology, C.K., B.S and M.W.; Investigation C.K..; Formal Analysis, C.K. .; Writing – Original Draft, C.K. and M.W.; Writing – Review & Editing, C.K. and M.W.; Visualization, C.K.; Funding acquisition, C.K. and M.W.

8. Supplementary material

Table 1: Participant's drug regimen at the time of recordings

Participant	Anticonvulsant	Antidepressant
p01	Clobazam, Valproat	-
p02	Lamotrigin, Levetiracetam	-
p03	Lacosamide, Levetiracetam	-
p04	Lamotrigin, Levetiracetam	-
p05	Lamotrigin, Oxcarbazepin	-
p06	Lamotrigin, Oxcarbazepin	Citalopram

p07	Lamotrigin	-
p08	Levetiracetam	-
p09	Lamotrigin	Sertralin
p10	Levetiracetam, Oxcarbazepin	-
p11	Oxcarbazepin	-

Chapter 4 - Competing Memories are Separated by the Hippocampal Theta Rhythm.

At the time of thesis submission, this chapter represents preliminary analyses that are being prepared for publication (Casper Kerrén, Benjamin Griffiths, Maria Wimber).

Abstract

Mnemonic overlap is a major challenge for our memory system. To resolve this conflict, a computational model suggests that the hippocampal slow oscillation can up- and down-regulate target and competitor memories, respectively, through contrastive Hebbian learning. Empirical evidence is here shown for this computation. MEG was recorded while participants recalled previously learned word-object associations. Proactive interference was created by associating some words with more than one associate, and asking participants to selectively recall the most recent associate. Similar to the two previous studies in this thesis, time-resolved pattern classifiers were trained to detect neural reactivation of the memories. At the time of retrieval, evidence for co-activation of target and competitor memories was found. In line with our central hypothesis, these reactivations were locked to different phases of the hippocampal theta oscillation. By contrasting participants who experienced high and low levels of interference, participants with low levels had a significantly larger phase difference between target and competitor memories. These results provide important evidence that the neural signatures for target and competitor memories are time-locked to different phases of the theta oscillation, and that this phase separation plays a functional role in aiding the selective recall of episodic memories under conditions of high mnemonic competition.

1. Introduction

Every day we form episodic memories (i.e., declarative memories acquired in a specific context at a given time) (Tulving & Markowitsch, 1998) of highly similar situations; for example, having dinner with your family. When trying to recall one specific dinner happening hours, days, or weeks ago, other dinners will inevitably be co-activated. The mnemonic overlap is a major challenge for our memory, and it is still unclear how the brain manages such interference. In this study, the idea that the human brain uses a phase code to adaptively separate overlapping memories along the hippocampal theta oscillation is tested.

Although interference has long been proposed as one of the main causes of forgetting (Baddeley & Logie, 1999; Underwood & Postman, 1960), few mechanistic models exist that can explain how the brain resolves this interference. Most neuroscientific literature and computational models have focused on how the brain separates overlapping memories at encoding (see: Yassa & Stark (2011)). Here, it is believed that the hippocampus orthogonalises patterns of stimuli, to minimize overlap in their neural representation (Marr, 1971; McClelland et al., 1995; Norman & O'Reilly, 2003; O'Reilly et al., 2014; Teyler & DiScenna, 1986; Teyler & Rudy, 2007). However, the hippocampus cannot completely separate patterns and residual overlap between similar memories remains, which can lead to interference (Norman & O'Reilly, 2003), suggesting a need to further separate competing memories.

Repeated learning has been shown to differentiate patterns of activity belonging to competing memories (Favila, Chanales, & Kuhl, 2016; Schlichting, Mumford, & Preston, 2015). The hippocampus is also assumed to play a major role in differentiating overlapping memories during retrieval, where repeated reactivation via active recall prunes mnemonic representations (Antony et al., 2017; Hulbert & Norman, 2015; Ritvo, Turk-Browne, & Norman, 2019). This pruning of memories for successful retrieval is assumed to be dependent on the down-regulation of competing memories, and such down-regulation can make these memories temporarily less accessible for future recall (Anderson et al., 1994). Although the idea that competitors are punished has received extensive support (Anderson, 2003), only one known computational model has directly implemented such down-regulation of competitors (Norman, Newman, et al. (2006). This model (see General Introduction, Figure 2) offers a mechanism by which interference is reduced during retrieval (Norman, Newman, et al., 2006), based on a learning algorithm that takes advantage of the strong link between inhibitory interneurons and slow oscillations (Buzsaki, 2002). More specifically, when inhibition is high, only strong features of the target memory will be active. During the transition to a lower inhibition state, activation is spread from strong to weaker features of the target memory, and those weaker features are strengthened through LTP. Conversely, at the opposite phase, when inhibition goes from normal to low and back to normal, overly strong features of the competing memory become active and are in turn down-regulated through LTD. This mechanism is repeated across several cycles of an oscillation, which ultimately leaves the initially overlapping memories in a state where they share fewer active nodes and are thus less likely to interfere. The model can explain the basic finding of retrieval-induced forgetting (RIF)

and under what conditions it is most likely to occur (Anderson et al., 1994; Norman, Newman, et al., 2006; Norman et al., 2007). Specifically, the model predicts that strongly active target memories tend to get strengthened, while moderately active competitors are most likely to get weakened, a central assumption that is at the core of the "non-monotonic plasticity hypothesis" (Ritvo et al., 2019) and has been corroborated by empirical evidence (Keresztes & Racsmany, 2013; Newman & Norman, 2010; Poppenk & Norman, 2014). However, the central tenet that the reactivation of target and competing memories is shifted in theta phase has never been tested.

In Chapter 2 a method for quantifying rhythmic reactivation was developed and showed that in a scenario where only one single memory is linked to a reminder, the neural pattern of memory retrieval fluctuates at a theta frequency. Similar theta-fluctuating reactivations have been shown in working memory (Bahramisharif, Jensen, Jacobs, & Lisman, 2018; Fuentemilla et al., 2010) and during memory-guided navigation (Kunz et al., 2019), in line with rodent work emphasizing the importance of phase-coding for separating and organising memories (O'Keefe & Recce, 1993; Skaggs, McNaughton, Wilson, & Barnes, 1996). In this study, the same method that we previously developed was applied (Kerren, Linde-Domingo, Hanslmayr, & Wimber, 2018) for analysing data from a competitive associative learning paradigm (i.e., a proactive interference task), to directly test whether target and competing memories can be decoded at different phases along the hippocampal theta oscillation.

2. Method details

2.1. Participants

26 right-handed participants (18 female, 8 male) participated for financial or course credit compensation (mean = 24.1 years old, SD = 5.73) (youngest: 18, oldest: 33). They all had normal or corrected-to-normal vision and reported no history of neurological disorders. All experimental procedures in the present study were approved by and conducted in accordance with the University of Birmingham Research Ethics Committee (STEM). Written informed consent was obtained from participants before they took part in the experiment.

2.2. Material and setup

The material consisted of 72 images depicting animate and inanimate objects (equal number of mammals, birds, insects, and marine animals, electronic devices, clothes, fruits, and vegetables), and 72 images depicting indoor and outdoor scenes, taken from the BOSS database (Brodeur et al., 2010) and online royalty-free databases. Stimulus selection was motivated by previous success at distinguishing these categories using multi-variate pattern analysis (Carlson et al., 2013). In addition to the material used for the experiment, 3 images were used for demonstrative purposes. Images from both object and scene classes were pseudo-randomly split for each participant into 6 sets, so that each set consisted of 20 objects and 20 scenes, 10 animate and 10 inanimate, 10 indoor and 10 outdoor scenes. Each set constituted one learning block. Also, a list

of 144 action verbs, similar to those used in Linde-Domingo et al. (2019), was generated for the experiment, serving as cue words in the cued recall task. These words were randomly assigned to images and conditions for each participant.

2.2.1. Paradigm

Participants received instructions about the task and first performed one short practice block. All participants (see exclusion above) then performed 6 experimental blocks (40 encoding trials and 72 (24x3) retrieval trials per block), each consisting of an associative learning phase, a distractor task, and a retrieval test with 3 repetitions per target item (Figure 1). At encoding, participants were asked on each trial to encode a word together with an image associate. In the competitive condition (CC), a word was encoded together with two associates, separated by at least three intervening trials. The instruction was to always memorize the most recent associate that was presented together with a given word for the following memory test. Therefore, the second associate in the CC always served as the target, with the previously learned first associate (i.e. competitor) assumed to elicit proactive interference. In the non-competitive single exposure condition (NC-1), a word was encoded together with only one associate, and the same word was never repeated during encoding. This condition served as the behavioural baseline for measuring the effect of proactive interference on memory performance (i.e., having previously encoded a competing associate compared with only one associate). In the non-competitive double exposure condition (NC-2), participants encoded a word together with the same associate twice. This condition served as the neurophysiological baseline and was specifically designed to

control for neural effects induced by the repetition of the word cue (including but not limited to repetition suppression) (Epstein, Parker, & Feiler, 2008; Kristjansson & Campana, 2010). The conditions were shown in pseudo-randomised order within each block, where 1/3 of the associations were in NC-1, 1/3 in NC-2, and 1/3 in CC [NC-1 = 8 trials, NC-2 = 8 trials, CC = 8 trials]. The order of the trials belonging to those 3 conditions was pseudo-randomized such that the average serial position of each condition within a block was equal. Images were randomly assigned to these conditions for each participant, with the only constraint that in the CC, the associates needed to be from different image categories (one object and one scene, split such that on half of the CC trials, the target was an object and the competitor a scene image, and vice versa for the remaining half).

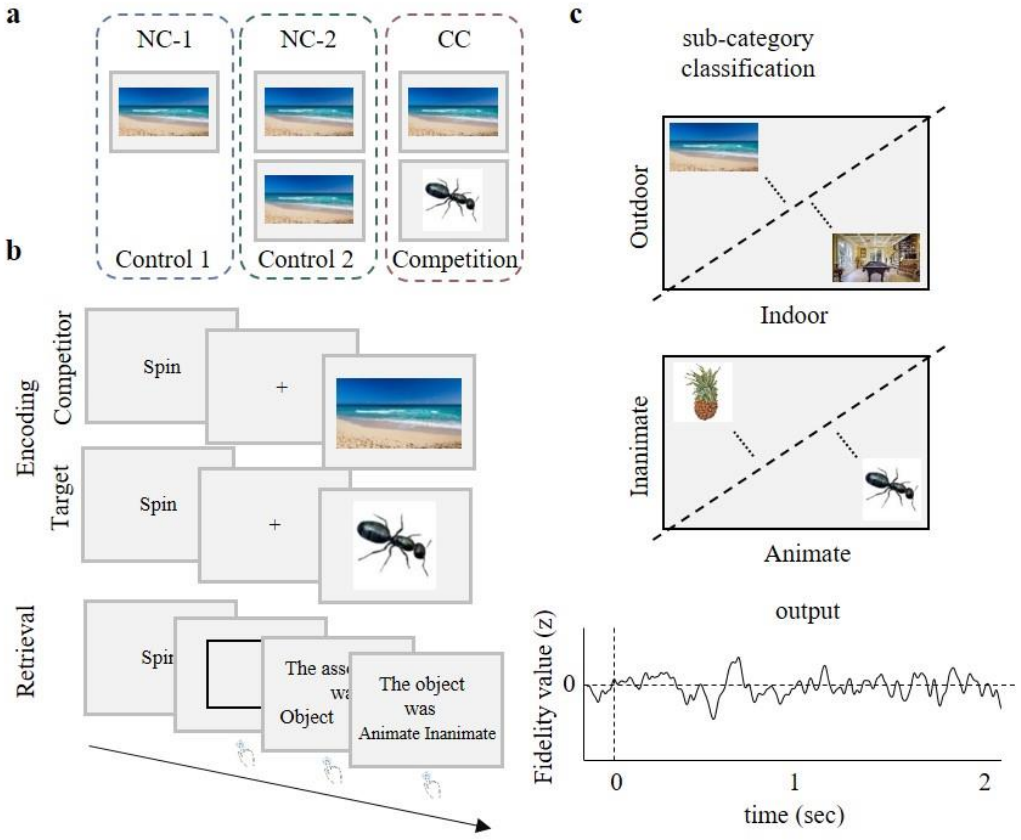

Figure 1. **Paradigm and rationale for decoding analyses. A)** The experiment consisted of 3 different conditions. In Control 1 (NC-1), subjects encoded the word together with one association. In Control 2 (NC-2), subjects encoded the word with the same association twice, and in the competitive condition (CC), subjects encoded the word together with two different associations. **B)** At encoding, subjects were instructed to create a vivid image representing the word-image association, and to constantly update their memory with the most recent associate to each word. At retrieval, the word prompted subjects to remember the most recently encoded associate to each word. **C)** Classification on both supra and sub category level was conducted (sub-category level shown here) to investigate reactivation of target and competitor memories at retrieval. The output of interest was the fidelity values at each trial.

A learning trial consisted of a jittered fixation cross (between 500 and 1500ms), a unique action verb (1500ms), a fixation cross (between 1000 and 1500ms), followed by a picture of an object or scene that was presented in the center of the screen for 4 seconds. The task was to come up with a vivid mental image that linked the image and the word presented in the current trial. As soon as they had a clear association in mind, participants pressed the right-thumb key on the button box. Participants were aware of the later memory test.

A distractor task followed each learning phase. Here participants had to respond if a given random number (between 1 and 99) presented on the screen was odd or even. They were instructed to accomplish as many trials as they could in 45 seconds and received feedback about their accuracy at the end of each distractor block.

After the distractor task, participants' memory for the 40 verb-object associations learned in the immediately preceding learning phase was tested in pseudo-random order, with three repetitions per association. The order was pseudo-randomized such that all associations were first tested once in random order, before testing all associations again in a new random order, and again a third time. No significant difference in serial position resulted when comparing CC and NC-2 items [t(23) = -

1.69, p = .1], CC and NC-1 items [t(23) = -1.18, p = .25], or NC-1 and NC-2 items [t(23) = -.43, p = .66]). One repetition consisted of 24 trials (8 of each condition), with the constraint that a trial could only occur once per repetition. Each trial consisted of a jittered fixation cross (500-1500ms), followed by one of the words as a reminder cue for the association. Participants were asked to bring back to mind the most recent associate of this word as vividly as possible. The cue was presented on the screen for 500ms and thereafter a blank screen with a black empty frame was presented. To capture the particular moment when participants consciously recalled a specific association, they were asked to press the right-thumb key as soon as they had a complete image of the associated memory in mind. If they did, the frame flashed once. After 4 s of a blank screen, a question appeared on the screen asking whether the retrieved item was an object, a scene, or they were unable to remember. The object and scene options randomly shifted between the left and right sides of the screen. If the participant did not remember the association, they were told to press the left-thumb button. If participants selected "object" or "scene", a follow-up question appeared (dependent on the response to the first question), asking whether the retrieved associate was an inanimate or animate object, or whether it was an indoor or outdoor scene.

In NC-1 and NC-2, there were an equal number of objects and scenes (4 objects and 4 scenes, but for NC-2 shown twice, therefore being 8 objects and 8 scenes), whereas in CC there were 8 objects and 8 scenes. In total this summed up to 40 trials for one block of learning (NC-1 = 8 trials, NC-2 = 16 trials, CC = 16 trials). At retrieval, only 24 trials were left (8 for NC-1, 8 for NC-2 and 8 for CC).

The experiment was set up via custom-written MATLAB 2016a (©The Mathworks, Munich, Germany) code using functions from the Psychophysics Toolbox Version 3 (Brainard, 1997). The presentation was projected onto a screen located 1.5 meters away from the participant, using Windows 64 bit.

3. Analysis

The general hypotheses and analysis steps were pre-registered and can be found on OSF.

After excluding participants based on the criteria stated in the pre-registered material, 24 participants (16 female, 8 male) were left, with an average age of 24.5 (SD = 5.73). For the magnetoencephalography (MEG) analysis, the same 24 participants as for the behavioural analysis were analysed. All participants performed all six blocks except for two, for whom time limit and button box errors occurred, resulting in only five blocks for these participants for MEG analysis.

3.1. Statistical analysis

In view of the pre-registration on OSF, all statistics were one-sided and conducted on a group level. If needed, non-parametric cluster-based permutation tests were conducted. For all analyses where a specific frequency band had to be preselected, the

analyses were limited to 4 to 8 Hz, as this was the frequency range of interest in the pre-registration.

4. MEG data analysis

The MEG was recorded at the Centre for Human Brain Health (CHBH), Birmingham, UK, using an Elekta Neuromag TRIUX system, with 306 channels (204 planar gradiometers and 102 magnetometers) with a 64-channel electrode cap, sampled at 1000Hz (Elekta, Stockholm, Sweden). As a sanity check, to evaluate the similarity between MEG and EEG topographies regarding the typical conflict theta over frontal MEG sensors and frontal EEG electrodes (Cavanagh & Frank, 2014). EEG was recorded in the initial 10 participants, but this data is not reported in any of the analyses in this chapter. The experiment was shown on a projector screen using a PROPixx projector (VPixx Technologies, Saint-Bruno, Canada) with a 1440Hz refresh rate, and the response from participants was collected using button response boxes (fMRI Button Pad (2-Hand) System, NAtA Technologies, Coquitlam, Canada). Only gradiometer data were included in the analyses of this chapter.

Individual anatomical data was acquired via magnetic resonance imaging (MRI) at the CHBH (3T Achieva scanner; Philips, Eindhoven, the Netherlands) with an MPRAGE sequence covering the whole head at $1mm^3$ resolution).

4.1. Preprocessing

Preprocessing was done using the FieldTrip toolbox (Oostenveld et al., 2011) and custom-written MATLAB code. A Butterworth high-pass filter of 1Hz, a low-pass filter of 200Hz, and a band-stop filter (50Hz; 100Hz, and 150Hz), were applied to the data. Data were then divided into trials from 1000ms pre-image to 2000ms post-image onset for encoding, and 1000ms pre-cue to 2000ms post-cue for retrieval. An automatized trial and component rejection was applied to the data. In a first step, data were high-pass filtered to 100Hz, and trials that were further away than 4 times the median absolute deviation from the amplitude distribution were automatically removed. In a second step, ICA was used to detect artifacts to be removed in the data. To this end, MEG data were downsampled to 250Hz, and only the first 1.5 seconds segments were used for ICA. Additionally, only one third of the trials was used to find components to exclude. The unmixing matrix was then applied to the non-downsampled data. To remove blink and cardiac artifacts, a template was created by running an independent component analysis on the six first participants (Gross et al., 2013). The components for the two artifacts were then averaged independently and used as a template for all participants. For the blink template, the algorithm correlated the topography of every component with the blink template and then found the best match, and subsequently removed it. For the cardiac component, the algorithm filtered the component time-series to between 10 and 20 Hz and correlated the peak amplitude of each trial with the cardiac time-series template (topography varies wildly so can't be used in the same way as for the blink). Any component which correlated to a greater extent than 4 times the median absolute deviation was classed as a cardiac component. Visual inspection of the removed components was then made. After these components were removed, an additional automatized artifact rejection was conducted, similar to

the initial step. The same procedure as in the first rejection round was followed but was done on channels instead of trials. Again, data were high-pass filtered to (i) 100Hz and (ii) 1Hz, and the channels that were further away than 3 times the median absolute deviation from the distribution were rejected. We chose three times the median absolute deviation for the last two rounds of rejection, to be somewhat more conservative. After components and trials were removed, all trials were again visually inspected, and trials still containing artifacts were manually removed. On average 206 out of 240 trials were kept (min = 170, max = 226, SD = 13) for encoding. On average 371 out of 432 trials were kept (min = 309, max = 400, SD = 22) for retrieval. Bad channels were interpolated using the triangulation method.

4.2 Time-frequency decomposition

The spectral difference between the CC and baseline (NC-2) during retrieval was calculated by first down-sampling the data to 200Hz and then convolving the raw data (-500ms pre-cue to 2000ms post-cue) with a complex Morlet wavelet of 6 cycles in steps of 50ms from 1 to 20 Hz for each condition. The combined power of each planar channel was then computed by summing the vertical and horizontal gradients. To test for significant differences between conditions, a non-parametric cluster-based permutation test was conducted. First, paired-samples t-tests were computed between the two conditions to investigate the difference in spectral power. To account for the multiple comparisons problem, the t-statistics for each time point (0ms to 1000ms post-cue), frequency band (4 to 8 Hz), and electrode were subjected to nonparametric cluster-based permutation testing, as implemented in the FieldTrip software. The

threshold for the statistical testing was set to an alpha level of 0.05. The minimum number of neighbouring channels that were considered a cluster was set to two. T-values above the threshold of 0.1 were then summed up, and compared against a distribution where condition labels were randomly assigned 3000 times with the Monte-Carlo method, following the standard method implemented in FieldTrip.

4.3. Multivariate pattern analysis

In order to attenuate unwanted noise, a Gaussian window with a full-width at half maximum (FWHM) in the time-domain of 40ms was applied to the signal. Each trial was then baseline corrected by subtracting the mean (-400 to -100 pre-image at encoding and pre-cue at retrieval) from all trials per sensor. Two separate Linear Discriminant Analysis (LDA) based classifiers were then trained on discriminating picture class, i.e. animate vs inanimate for objects, and indoor vs outdoor for scenes (Figure 1C). This important feature of the design allows us to have classifiers that provide independent measures of target and competitor reactivation. Training for both classifiers was done on the average encoding data from 100 to 650 ms post-image onset (time window of best decoding at encoding), and tested on cue onset at retrieval (200ms pre-cue up to 2000ms post-cue). As classifier features, the MEG gradiometer sensor patterns were used (pre-processed signal amplitude on each of the 204 channels), independently per participant and at each time point. To obtain a stable model for training, four random trials within the same class of the training set were repeatedly (20 times) averaged (Grootswagers, Wardle, & Carlson, 2017). To test for reactivation of target and competing memories, trials in the competitive condition were

split into those where the target was an object and the competitor was a scene, and vice versa and the corresponding object and scene classifiers were then used to separately indicate evidence for target and competitor reactivation on every single trial. As the training and testing data came from encoding and retrieval respectively, cross-validation was not necessary. The LDA, as used here, reduces the data from 204 channels into a single decoding time course per trial, and we used these single-trial, time-resolved outputs of the classifier as an index of memory reinstatement. More specifically, during training, the classifier found the decision boundary that could best separate the patterns of activity from the two classes (animate vs inanimate, or indoor vs outdoor) in a high-dimensional space. The classifier was then asked to estimate whether the unlabelled pattern of brain activity in any given retrieval trial and at each time point was more similar to one or the other class. This training-test procedure was repeated until every single retrieval trial had been classified. To avoid overfitting, the covariance matrix was regularized using shrinkage regularization, with the lambda set to 1 (Blankertz et al., 2011). On a single-trial level, the output of interest was the fidelity value, which indicates at any given time point, the distance to the decision boundary in a high-dimensional space, and presumably how confidently the classifier predicts that the pattern of brain activity belongs to one or the other of the two classes (Carlson et al., 2014).

4.4. Determine peak frequency of fidelity values using MODAL

To determine the dominant frequency of our single-trial classifier fidelity values, a plateau based band-pass filter based on an a priori defined frequency range was used,

and the change in phase per unit time was investigated, using the algorithm "MODAL" (Watrous et al., 2017). To replicate previous findings (Chapter 2), this analysis was run for the NC-1 condition as this most closely resembled the condition in (Kerren et al., 2018) of retrieving a single target. For each participant, the fidelity values from 200ms to 1200ms post-cue were averaged across trials, and before conducting a Fast Fourier Transformation (FFT) tapered with a Hann window. In line with the hypothesis and preregistration, data were then band-pass filtered in the theta band and transformed into an analytic signal using the Hilbert transform, and the phase angle time series was extracted. To obtain the frequency and phase at each sample, frequency sliding was applied to the data as follows: (sampling frequency*diff(unwrap(signal))/(2*pi)). "Phase slips" were attenuated by applying median filters with different length in the time domain (50ms to 350ms) to the data, wherefrom the median was taken (Cohen, 2014). Frequencies that did not exceed the 1/f-fitted line (within the already defined frequency range) were then excluded, which resulted in a vector for each participant containing the frequencies and time points exceeding the background spectrum where an oscillation was present. The average probability across time (200 to 1200ms post-cue, as in all other analyses using classifier output) for observing an oscillation in a given frequency between 4 and 8 Hz was then calculated (Figure 3B).

To test the significance of the frequency transformed data, one random label classifier per participant were randomly picked 500 times. For each random label classifier, the same procedure as for the real data to obtain an estimated chance distribution was followed. The significance was tested for each frequency in two ways. First, between 4 and 8 Hz, the estimated chance distribution was subtracted from the real value and

subsequently divided by the standard deviation of the estimated chance distribution. This gave a z-value, which was compared to the threshold of 2.33 at p = .01, correcting for 5 multiple comparisons (4-8). Secondly, the 95th percentile of the estimated chance distribution was calculated and compared that to the real data. Again, this procedure replicated the one used in Kerren et al. (2018).

4.5. Source Analysis

Individual MRI scans (anatomical data) were available for 18 participants, and for the remaining 6 participants the standard head model as implemented in FieldTrip was used. Anatomical data were aligned based on the sensor position obtained from a Polhemus system (Colchester, Vermont, USA). This was done in three steps: the first step was done manually by adjusting the alignment between the anatomical data and the head position; the second was done by the Iterative Closest Point algorithm implemented in the Fieldtrip toolbox (Oostenveld et al., 2011); the third step was again done manually to check that the alignment worked, and adjust if it did not. The realigned model was used to reslice and segment the brain to make the axes of the voxels consistent with the head position and subsequently to extract the brain compartments. The segmented brain was then used to create a forward model (head model). In our case, we used a semi-realistic forward model (Nolte, 2003). The forward model was used to create the source model (lead field), where for each grid point the source model matrix was calculated with a 1 cm resolution, and the virtual sensors were placed 10mm below the cortical surface, and subsequently warped into each brain. In total each participant had 3294 virtual sensors. A linear constrained

minimum variance (lcmv) beamforming approach (Gross et al., 2001) was then used to reconstruct the activity of virtual hippocampal channels in source space. Five percent regularisation was used. The filters were computed on 5Hz band-pass filtered data, from 0ms to 2000ms after cue onset (baseline-corrected [-400 to -100]). The filters were obtained from all data for all conditions, ensuring that the same filters were used for all our contrasts in source space. The left and right hippocampus were chosen as a Region of Interest (ROI), derived from the AAL atlas (Oostenveld et al., 2011). The individual MRI and the atlas were interpolated using the interpolation method *nearest*, finding the nearest region in the AAL atlas based on the Euclidean distance. Having created such masks for the left and right hippocampus, single-trial time courses were extracted from the corresponding virtual sensors, and activity averaged across those sensors in each ROI, resulting in one activity time course per trial for left and right hippocampus, respectively.

4.6. Phase-amplitude coupling between MEG data and fidelity values

The modulation index (MI) was calculated to test for a relationship between the fidelity values and a hippocampal oscillation. The signal was transformed to an analytic signal using the Hilbert transform and the phase angle time series was extracted. Each complex value data point was then point-wise divided by its magnitude, resulting in a 4D-matrix of phase values, containing trials*channels*frequencies*time. Phase values at 5Hz were then divided into 20 adjacent bins, ranging from – pi to pi. To link classifier-based memory reinstatement indices to the hippocampal phase, the z-scored amplitudes (fidelity values) of the classifier output from corresponding time points

were sorted into their respective phase bins, and the average amplitude of each phase bin was then calculated. Following this binning procedure, the modulation index was calculated by comparing the resulting phase distribution of classifier fidelity values to a uniform distribution (using the mean fidelity across bins). The Kullback-Leibler (KL) distance was then calculated using the equation in (Tort et al., 2010); point-wise multiplying the log-transformed quotient of the real data and the uniform distribution with the real data.

A baseline was computed by running the same analysis as described above, with the only difference being that the fidelity value trials were randomly assigned to the phase time series within participants. The phase time series of a given trial was paired with the classifier output from a random trial and subsequently binned in the same way as the correctly paired data. The trial reassignment was repeated 250 times, during which the MI was calculated for each iteration. After this, the threshold was created based on the 95th percentile across iterations. To test for significance, the real modulation index for each participant was compared against this participant's 95th percentile using a paired-samples t-test.

4.7. Extraction of fidelity values

The second measure of phase locking of target and competing memories was obtained by following the same procedure as in Kerren et al. (2018). For each trial, the highest fidelity value between 200 and 2000ms was extracted. The same rationale as in

Chapter 2 was used to exclude the first 200 ms after cue onset, namely to limit the influence of an early ERP peak in the results, while at the same time excluding time points where no memory reinstatement would reasonably be expected. Only peaks that exceeded the 95th percentile of a baseline were extracted. The baseline was created by randomly shuffling the labels of the classifier and running it 25 times. For each randomisation and each trial, the fidelity values were z-scored, resulting in an m-by-n matrix with each random label classifier and all times points. The 95th percentile of these values across trials functioned as a baseline from which the fidelity value needed to exceed to be included for further analysis. Each fidelity value also needed to stay above this value for at least 50ms (see Figure 4A for distribution of extracted fidelity values). These steps were taken to ensure that the extracted signals of interest were not only noise (see: Linde-Domingo et al. (2019), for a parallel procedure). Following this extraction procedure, the raw data were projected to source space following the same procedure as above. For each trial, the trial time of the maximum fidelity value was found for target and competitors separately and the corresponding phase value from the right and left hippocampal 5Hz oscillation was extracted. Importantly, the values were not averaged, which resulted in two complex numbers representing the phase of the two time points for each trial. To test for a significant phase difference between target and competitor memories, the pairwise phase difference using the equation above was calculated. This was done both on average across the three retrieval repetitions, but also for each repetition at retrieval. The difference was calculated separately for right and left hippocampus, and a p-value of .025 (.05/2) was therefore considered significant, controlling for multiple comparisons

4.8. Absolute phase difference between fidelity values for targets and competitors

To calculate the pairwise phase difference between the two conditions, the data were transformed back to their complex numbers. Once in the complex numbers, the absolute value or complex modulus of the complex number was obtained by using the Pythagorean Theorem and taking the square root of the sum of the squares of the parts. The amplitude was arbitrarily set to 1. The vectors were then point-wise divided with the complex modulus. And finally, to subtract the angles from each other, one vector was rotated by multiplying it with the other's complex conjugate. To obtain the angle, the inverse tangent of the ratio was taken of the product between the vectors. In MATLAB the following code was used:

$$\theta = angle(\exp(1i \times \boldsymbol{X}) .\div \exp(1i \times \boldsymbol{Y}))$$

where θ is the vector of angle difference.

5. Results

5.1. Memory performance

5.1.1. Lowest memory performance was obtained in the competitive condition

At retrieval, participants could maximally respond three times per trial. First, they indicated with a single button press the time point when they retrieved the association.

After that, the first question asked whether the associated image was an object or a scene, and finally, a third question asking whether it was an indoor or outdoor scene or animate or inanimate object, depending on the previous response. To count as a hit, participants needed to correctly respond to both follow-up questions. Thus, a hit was equalled to the strongest form of retrieval, similar to source recollection. To test for statistical differences between the three main conditions, a one-factorial repeated-measures ANOVA was conducted with memory accuracy as the dependent variable. The Greenhouse-Geisser corrected analysis showed a significant main effect of condition, $F(1.453, 46) = 57.99$, $p < .05$. Two planned comparisons followed up this main effect. Most importantly, there was a significant difference between CC (M=53.18%, SD = 19.20%) and NC-1 (M = 57.71%, SD = 13.98%), $t(23) = 1.80$, $p < .05$ (Figure 2A). This confirms the first behavioural hypothesis that retrieving a memory is more difficult when having encoded two different associates compared to when only having encoded one. For later brain-behaviour correlations, an interference score for each participant was calculated by subtracting memory performance in the CC from the NC-1 condition. Secondly, a significant difference was also obtained between NC-1 and NC-2, $t(23) = -11.91$, $p < .05$, demonstrating that not surprisingly, a second encoding exposure improved performance compared to a single exposure.

5.1.2. Participants picked the competing association more than could be explained by chance

Another behavioural index of competition is a participant's bias to select the specific sub-category (i.e., animate/inanimate for objects, and indoor/outdoor for scenes) of the

competitor during the third question of the staged retrieval. This index is called the intrusion score. First, this bias was averaged across all three retrieval repetitions. On trials where participants made an error (i.e., selected "object" though the target was a scene and vice versa), they showed a significantly above-chance (i.e., above 50%) probability of selecting the sub-category of the competitor, $t(23) = 10.48$, $p < .05$, suggesting that their errors were not random but biased towards the category of the second associate to the cue word (Figure 2B, left). The number of intrusions did not differ between first and third repetition, $t(23) = -.19$, $p = .85$, which is to be expected as no feedback was given (Figure 2B, right).

5.1.3. Memory performance correlates with number of intrusions and the amount of interference

Memory performance, calculated as percentage hits across all repetitions in the CC, significantly correlated with the number of intrusions participants experienced at retrieval (Figure 2C). The average number of intrusions across retrieval attempts correlated negatively with memory performance, $r = -.715$, $p < .05$. Similarly, the interference index correlated negatively with memory performance, $r = -.69$, $p < .05$.

For forthcoming analyses, both on behavioural and neurophysiological level, a median split across participants on interference was also conducted to compare high and low interference participants.

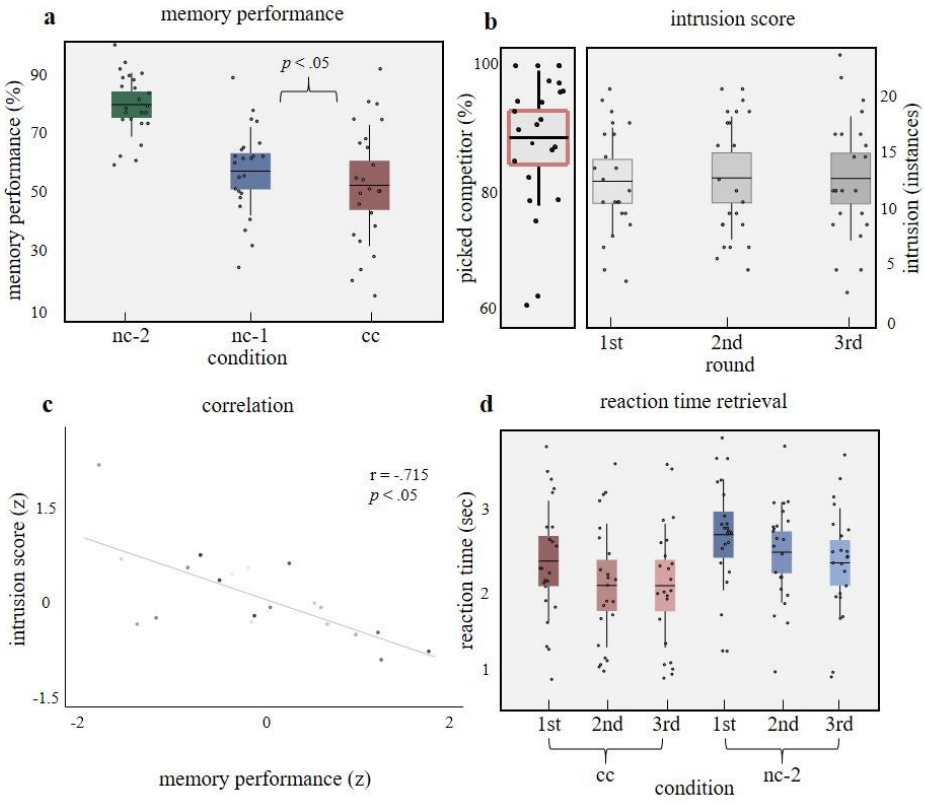

Figure 2. **Behavioural results. A)** Subjects were on average very good at remembering in the nc-2 condition, in which they associated a word together with the same image twice. The competitive condition was most difficult and subjects were significantly less successful in cc compared to nc-1, indicating proactive interference. **B)** To the left: Across retrieval attempts, participants picked the competitor memory significantly more than would be expected by chance. To the right: The number of intrusions did not decrease across rounds, but were significantly negatively correlated with memory performance, as shown in **C**. **D)** Reaction time at retrieval showed that across rounds subjects responded faster in both conditions.

5.2. Reaction time (RT)

5.2.1. At retrieval, NC-1 had slowest reaction times.

At retrieval, participants were asked to indicate with a button press when in time the association came back to mind. Three participants were excluded (2 for low memory performance, following the exclusion criteria in pre-registration, and 1 for not

indicating when they had retrieved the association). Only trials where participants answered correctly on both of the follow-up questions were included. To test for differences in RT across conditions, a repeated-measures ANOVA was conducted with the two conditions (CC and NC-1) and all three repetitions. Significant main effects for both condition and repetition were obtained, (condition, $F(1, 22) = 17.549$, $p < .05$; repetition, $F(2,21) = 9.952$, $p < .05$) (Figure 2D), indicating that participants were slower at retrieving the correct target memory in the competitive condition.

5.3. MEG results

5.3.1. Theta power increase for CC compared to NC-2.

In line with the conflict processing literature, increased theta power (4-8Hz) over mid-frontal sensors in the CC compared to the NC-2 conditions was expected (Cavanagh & Frank, 2014; Ferreira, Marful, Staudigl, Bajo, & Hanslmayr, 2014; Hanslmayr, Staudigl, Aslan, & Bauml, 2010). A non-parametric cluster-based permutation test revealed a significant difference between retrieving associations in the CC compared to NC-2 (clusterstat = 168.96, $p < .05$), ranging from cue onset to 300 ms post-cue (Figure 3A). Theta power has been shown to be strongest in the first round of retrieval-practice (Hanslmayr et al., 2010). Therefore, a possible reduction in theta power from first and last repetition was tested for, but no significant cluster surviving cluster-based correction was found.

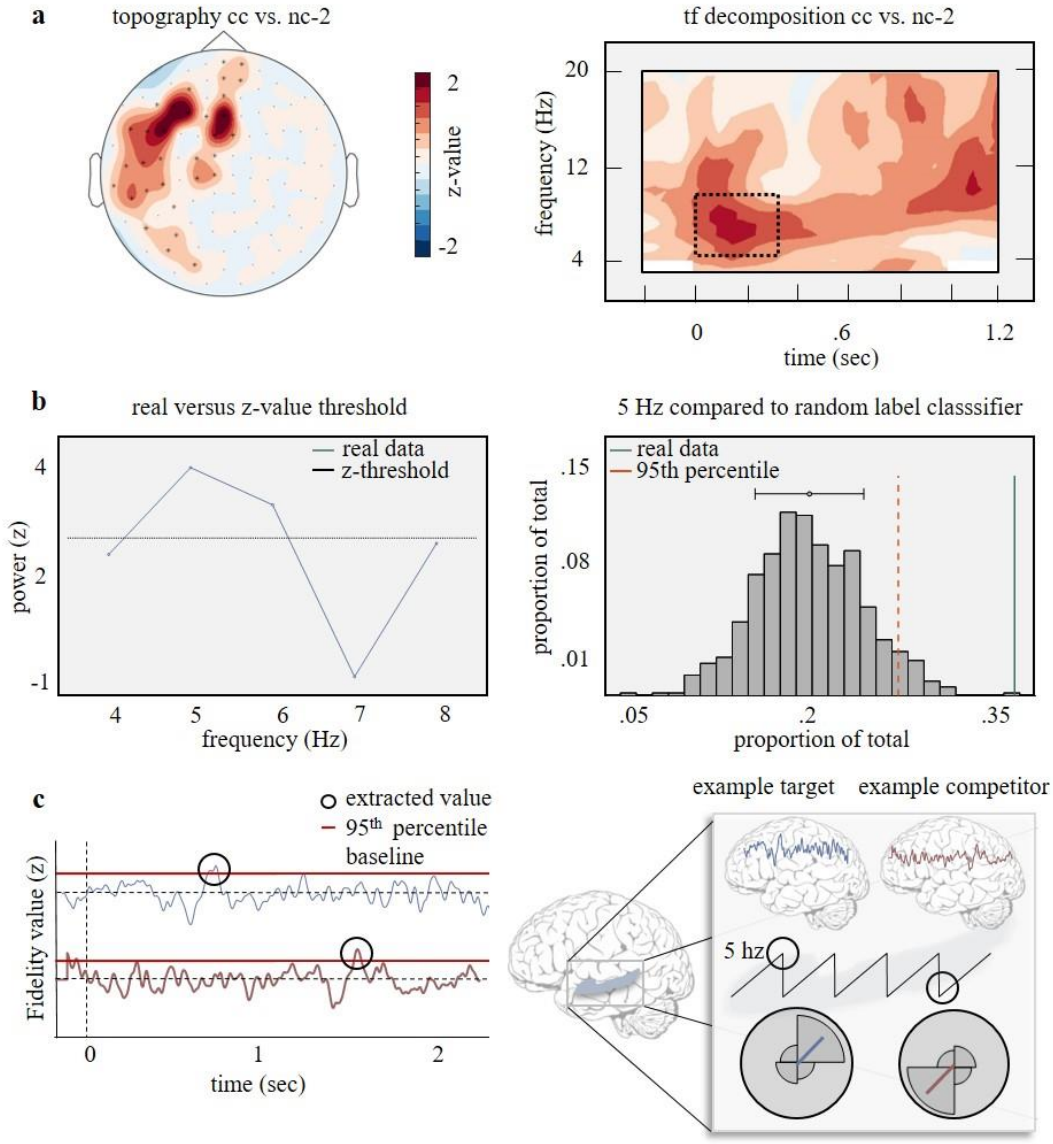

Figure 3. **Time-frequency results, peak frequency and rationale for classifier-to-theta coupling. A)** Topoplot showing average across significant time points (4-8Hz) from a non-parametric cluster-based permutation test between CC and NC-2 conditions. Strongest effect in frontocentral channels. Time-frequency plot averaged across significant channels from the permutation test showing the strongest frequency at 5-6Hz and approximately 200ms after cue onset. **B)** 5Hz was the strongest frequency across subjects, here compared to a baseline based on randomly shuffling the labels before classifying the same data. Blue line is the real data, black dotted line is the z-value threshold at p = .01 (0.05/5). In second panel of **B** the histogram is the estimated chance distribution, the red dotted line the 95th percentile of the randomly shuffled classifier and the green line the real data. **C)** Rationale behind the extraction of fidelity value peaks. At each trial peaks the strongest peak exceeding a constructed baseline was extracted for target and competitors separately. The time points of these peaks were used to extract the phase of the hippocampal 5Hz oscillation with the hypothesis being that target memories are reactivated at a different phase compared to competitor memories.

5.3.2. Fidelity values fluctuate at theta frequency

To find out whether target memories fluctuate at a slow oscillation, the same procedure as in (Kerren et al., 2018) was followed and the change in phase per unit time was investigated (Cohen, 2014). This analysis clears the data from frequencies not surviving the linear fit in log-log space, meaning that maximum frequency is only calculated on time points where an oscillation is present in the data. This analysis revealed that 5 and 6 Hz were significant, with a peak at 5Hz (Figure 3B). This analysis replicates the previous results showing that classifier-based indices of memory reactivation rhythmically fluctuate (Kerren et al, 2018). Even though the peak frequency differed (8Hz in the previous study), this confirms that memory retrieval fluctuates in the theta frequency range, and 5Hz was therefore used as the main frequency of interest for further source analyses.

5.3.3. Fidelity value extraction

The next main set of analyses was set out to test for a phase difference in the coupling of target and competitor reactivations to the hippocampal phase. To test for such coupling, the same procedure as in Kerren et al (2018), and previously registered in the pre-registration, was followed. The first measure of interest was the modulation index (MI) (Tort et al., 2010). This index shows whether there is a significant phase-modulation after having calculated the phase-amplitude coupling between the phase of the hippocampal 5Hz oscillation and the amplitude of the fidelity values. Phase

difference was measured between 200 and 1200ms post-cue at retrieval, but no evidence that target or competitor memories were modulated by the hippocampal phase was found (MI real data hippocampus right and left collapsed: mean = .0146, SD = .0025; estimated chance distribution: mean = .0156, SD = .0018, $p > .05$).

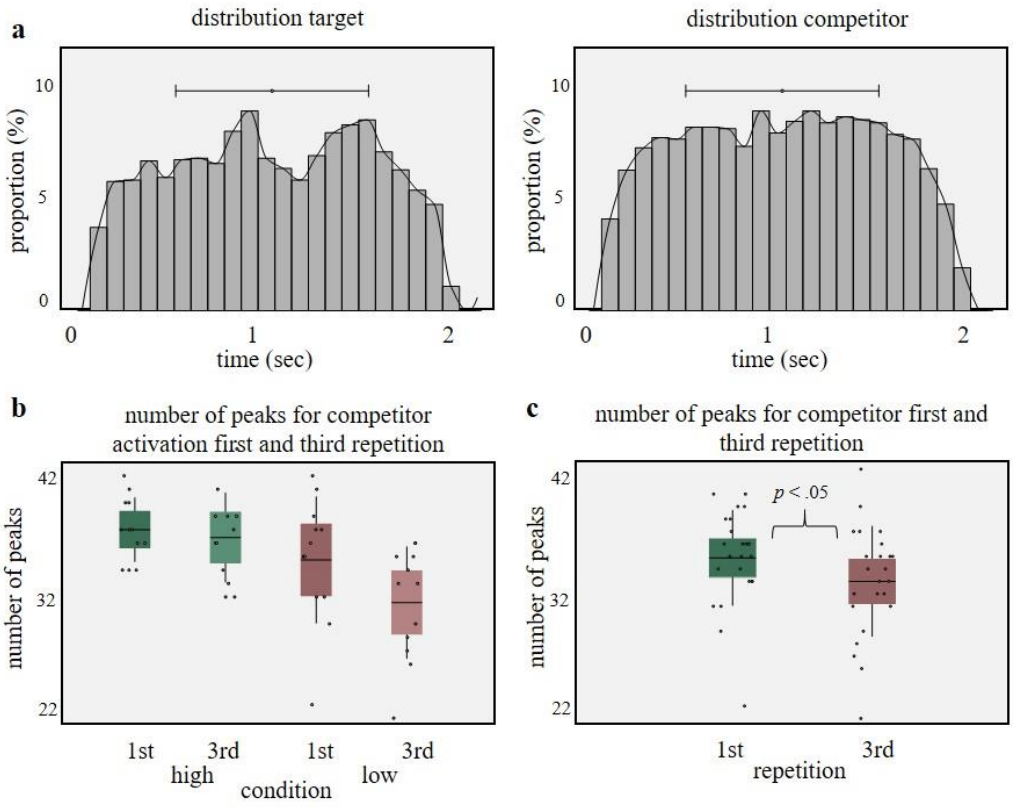

Figure 4. **Distribution and density of fidelity values. A)** The distribution of the extracted peaks showed an expected result for target memories with most peaks extracted in the time window of recollection (500-1500ms). For competitors it was more widespread. **B)** Significant main effects of condition and rounds were revealed. **C)** When collapsing over interference, a significant decrease in number of fidelity peaks for competitor from first to third repetition was shown ($p = .02$)

Analysis of RTs suggests that participants retrieved the target considerably later than 1200ms, with substantial variation across trials. Thus, it was decided (diverging from Kerren et al, 2018) to investigate the phase distribution of fidelity peaks instead of using all fidelity values (Figure 3C).

Based on these considerations, a fidelity-to-phase coupling analysis taking into account only the peak fidelity value(s) from every single trial that surpasses a given noise threshold was therefore conducted (see Kerren et al., 2018; Linde-Domingo et al., 2019), which was assumed to reflect the time point of maximal neural memory reactivation. The hypothesis remained unchanged, that is, it was still expected that target and competitor memories are reactivated at a different phase of the hippocampal 5Hz oscillation. The peak detection algorithm used here identifies fidelity peaks that exceed the noise baseline for at least 50ms, and identified peaks in on average 97% of the CC trials (max = 100%, min = 59%) for the target classifier (mean number of peak = 34.06, SD = 3.58), and 96% of the trials (max = 100%, min 65%) for the competitor classifier (mean number of peak = 33.97, SD = 3.44). In Figure 4A, all extracted fidelity peaks for target and competitors are plotted, showing that peaks were quite widely distributed for both targets (mean = 1092ms, SD = 459ms) and competitors (mean = 1085ms, SD = 460ms), both within the time window when recollection is thought to occur (Wimber, Staresina 2019). Before investigating their coupling to the theta rhythm, the behaviour of the fidelity peaks across conditions and repetitions was of interest. If competitors are weakened, across repetitions, it is expected that a significant down-regulation of competitor fidelity values is present between the first and third round (Rafidi, Hulbert, Brooks, & Norman, 2018). To test for such a decrease, a repeated-measures ANOVA with the factors interference (high and low) and rounds (first and third), revealed main effects for interference ($F(11) = 8.50$, $p = .01$) and rounds ($F(11) = 9.67$, $p = .01$), but no interaction ($F(11) = 2.48$, $p = .14$). A planned comparison followed up the main effect of rounds, collapsed over interference, and showed that the number of peaks in repetition one (mean = 34.9, SD

= 3.81) significantly differed from the number of peaks in the third repetition (mean = 33, SD = 4.42), t(23) = 2.44, p = .02 (Figure 4C). These results further corroborate previous studies showing a down-regulation of competition across retrieval attempts (Rafidi et al., 2018).

5.3.4. Hippocampal theta oscillation clocks the reactivation of target and competitor memories

Next, the difference in theta phase between reactivation peaks identified for target and competitor memories was directly investigated (Figure 5). The time points of fidelity values exceeding a 95th percentile baseline for each trial of target and competitor classifiers were extracted. The phase angle of the hippocampal 5Hz oscillation was then found for each fidelity value and the difference between target and competitor phase was taken. As mentioned in the result section, the difference was calculated separately for the right and left hippocampus, and a p-value of .025 (.05/2) was therefore considered significant, controlling for multiple comparisons. Across participants, phase-clustering for target and competitors separately was not expected to be found. However, a within-participant contrast between the mean target phase and the mean competitor phase was hypothesised to significantly differ from each other (Norman et al., 2007). Conducting a pair-wise difference, in the first repetition, between each participant's mean target and mean competitor reactivation phase revealed a borderline difference from a uniform distribution (Rayleigh statistics, z(23) = 3.47, p = .03), with a mean angle of -137 degrees (Figure 5A).

In the second repetition, contrasting to a uniform distribution, a pair-wise difference between each participant's mean target and mean competitor reactivation phase revealed a significant phase difference (Rayleigh statistics , $z(23) = 4.04$, $p = .02$; mean angle = -99 degrees). These results show that when subtracting the phase of the competitor from the phase of the target within participants, a significant phase difference is observed, indicating that target and competitor memories are reactivated at approximately 99 degrees different from each other along the hippocampal 5Hz oscillation.

For left hippocampus in the third repetition, no significant phase difference between target and competitor memories was found (Rayleigh statistics, $z(23) = .99$, $p = .37$; mean angle = 121 degrees). No effect of phase difference was found in the right hippocampus.

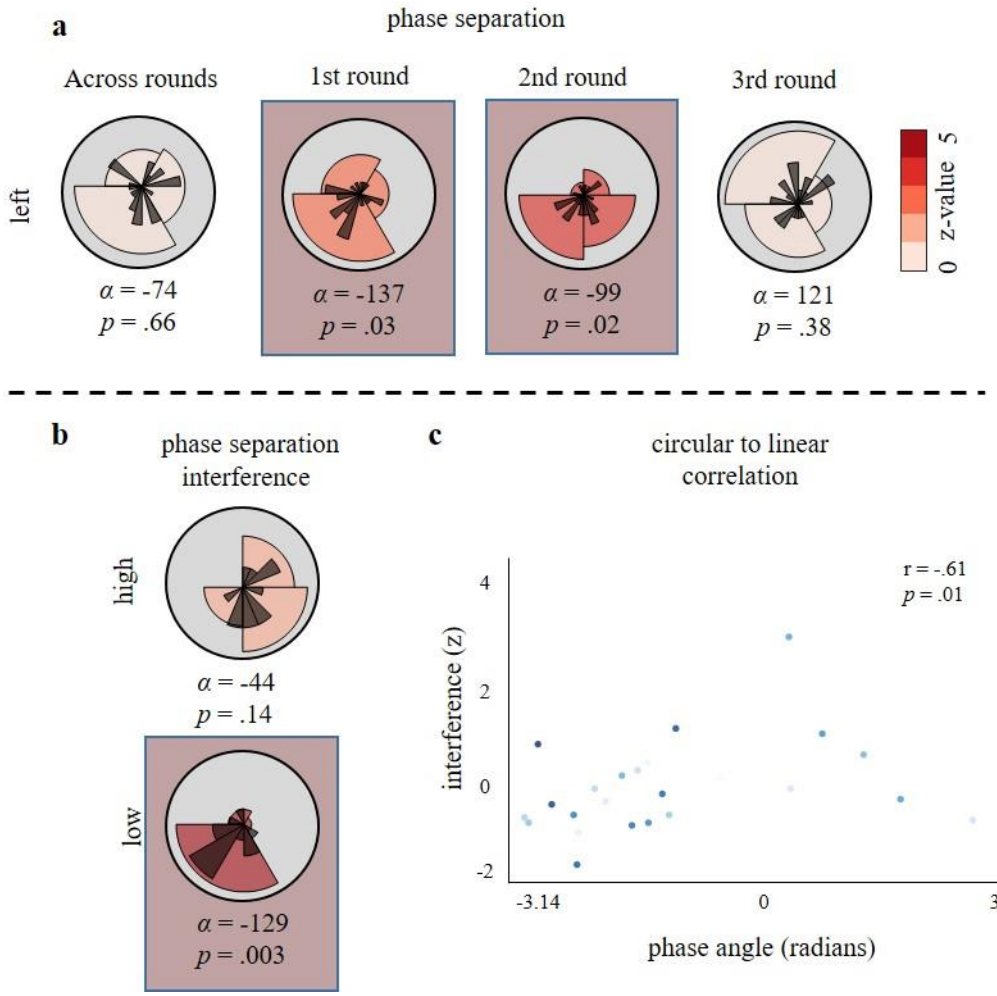

Figure 5. **Phase difference results on average and for each round. A)** When subtracting target phases from competitor phases within each subject, a significant phase difference was found for left hemisphere for first and second but not third repetition, with a mean angle of -137 degrees difference for first repetition and -99 degrees for second repetition. Averaged across all three rounds, no significant difference was found. **B)** When splitting the difference on interference in the second round, a significant clustering for low levels of interference with larger phase difference than for high interference was found. **C)** Interference also significantly correlated with the phase difference between target and competitor memories, such that the further away the two memories were from each other, the lower the interference. Z-value indicate strength of Rayleigh statistics test.

5.3.5. Phase difference shows a functional relationship to behaviour

Next, a functional relationship between the phase difference in the left hippocampus and behavioural measures was investigated, with the assumption being that the greater the difference was between target and competitor memories, the less amount of

interference participants experienced. To test for this, the sample was median split based on participants who had high and low interference, to then investigate how the absolute phase difference between target and competitor reactivation behaved in these two sub-groups. This was done only for the second repetition (Figure 5B), in which a significant phase difference across the entire participant sample was found, controlling for multiple comparisons (see the previous section).

Splitting the phase difference in the second repetition based on the interference score revealed a strong phase difference for low interference, $z(11) = 5.44$, $p = .004$ (mean angle -129 degrees), but not for high interference, $z(11) = 2.03$, $p = .13$ (mean angle -44 degrees) (Figure 5B). Additionally, phase difference was correlated with interference, ($r = -.61$, $p = .01$), such that participants with relatively lower levels of behavioural interference show larger phase difference between the two overlapping memories (Figure 5C).

These results are in line with the hypothesis and show that the phase difference between target and competitor reactivations, as identified in the entire participant sample is more pronounced in participants with low levels of interference. Although a phase difference on a group level indicates that participants shared the same difference between target and competitor memories, the most important finding is that participants with low levels of interference also on average had a greater phase difference between target and competitor memories.

5.3.6. Participant with large decline in competitor reactivations show significant phase difference

The final analysis related neural indices of competitor down-regulation to measures of phase separation. To explicitly test whether participants who were able to down-regulate competitors from first to the third repetition had a relationship to the amount of phase difference between competing memories, phase difference was correlated with the decline in competitor fidelity peaks from first to the third round. A correlation was found ($r = .52$, $p = .04$), meaning that the further away target and competitors are from each other in theta phase, the greater was the decline and thus the less interference a participant experienced.

6. Discussion

In attempting to remember a specific experience, one critical task for the brain is to shield target retrieval against competing memories. Based on neural network simulations (Norman, Newman, et al., 2006) showing how memory competition can be resolved by the relative up- and down-regulation of relevant and irrelevant memories by slow oscillations during retrieval, this study sought to test the hypothesis that phase-coding along the hippocampal theta rhythm acts as a separating mechanism for overlapping competing memories.

First, previous findings (Chapter 2 and 3) showing that neural indices of target memory reactivation fluctuate following a slow oscillation were replicated (Figure 3B). More importantly, the preferred phase along the hippocampal theta oscillation differs depending on memory relevance, with target and competitor memories being separated by on average 100 degrees, at least in the first two rounds of retrieval (Figure 5A). These results provide new insights into the powerful computations along the hippocampal theta oscillation for relevance determination aiding successful memory retrieval.

Participants encoded words together with one or two associates, either depicting an object or a scene (Figure 1). In the critical competitive condition (CC), one word was paired with two different associates during encoding, and participants were instructed to always update their memory to recall the most recent associate to a given word cue in the later memory test. In this way, the amount of proactive interference (PI) the first associate elicited on the second associate could be measured. Behavioural performance (Figure 2) showed that participants had more difficulty retrieving the target memory in the competitive condition, compared to having only encoded the word together with one associate. This observation suggests that the firstl encoded association interfered at the time of retrieval of the target memory.

During retrieval it is assumed that the brain "punishes" competing memories by down-regulating their activation state, making them less interfering on future retrieval attempts (Anderson, 2003; Anderson et al., 1994; Kuhl, Rissman, Chun, & Wagner, 2011; Waldhauser, Johansson, & Hanslmayr, 2012; Wimber, Alink, Charest,

Kriegeskorte, & Anderson, 2015). In terms of electrophysiology, the most prominent finding linked to mnemonic interference is an increased frontal midline theta power effect for conditions with high competition versus low competition (Ferreira et al., 2014; Hanslmayr et al., 2010). In the present study, this theta increase when contrasting our high-competition CC with the low competition NC-2 condition was replicated (Figure 3A). This increased frontal midline effect in the interference condition is thus consistent with previous studies on mnemonic competition (Ferreira et al., 2014; Hanslmayr et al., 2010), as well as studies linking mid-frontal theta to a general error monitoring and conflict resolution process (Cavanagh & Frank, 2014).

Having established that the paradigm induced interference on a behavioural level, and a corresponding increase in theta power on a neurophysiological level, next up was the main hypothesis that successful memory retrieval under the influence of proactive interference is supported by strong phase separation of target and competing memories along the hippocampal theta rhythm. Although it is assumed that hippocampus automatically separates neural representations at encoding (Marr, 1971), some overlap will remain that consequently can cause interference, as shown in numerous behavioural studies (e.g., Anderson, 2003). To successfully retrieve the relevant memory, the brain needs to find ways of separating the co-activated competing memories (Norman & O'Reilly, 2003). One proposal of how this is achieved is by neural differentiation. In contrast to the automatic pattern separation process (O'Reilly & Rudy, 2001), differentiation is less automatic, and instead driven by the competition between neural representations (McClelland et al., 1995). As soon as the brain detects competition between the co-active representations, learning will take place and act on

the representations to pull them apart for a successful response. This adaptive process has previously been assigned to the neocortex (O'Reilly & Rudy, 2001), but more recently, evidence has shown that hippocampus is the main candidate for achieving this differentiation process, too (Favila et al., 2016; Ketz et al., 2013; Schlichting et al., 2015), due to the sparse activity via relative high inhibition in this region (Barnes, McNaughton, Mizumori, Leonard, & Lin, 1990; Norman & O'Reilly, 2003). Neural differentiation is assumed to be further modulated during retrieval by the strength of the competing memories (Ritvo et al., 2019). Only when memories are moderately active compared to the target memory, will they be down-regulated and differentiated from the target memory (Newman & Norman, 2010). Importantly, these theories have been simulated in the neural network model tested in the present study (Norman et al., 2007). In this model, it is assumed that the time-varying level of inhibition along the phases of a slow oscillation will have a differential effect on target and competitor memories, to ultimately solve memory interference. One of the main predictions of this model is that (the moderately active) competing memories will be active at a specific, low-inhibition phase of the slow oscillation, phase-shifted from the phase at which the strongest target activation should be observed. Moreover, stronger phase separation can be expected across repeated attempts to recall the target memory if differentiation is successful.

To test these basic assumptions, the same decoding approach as in our previous paper was used (Kerren et al., 2018), in which a neural index of memory reinstatement was obtained separately for target and competitors using a Linear Discriminant Analysis (LDA) (Figure 1C). The classifiers were trained on maximal accuracy at encoding and

tested at each sample point (1 ms resolution) throughout the retrieval period, starting at cue onset. On each trial, the time points of maximal target and competitor memory reinstatement were determined, and on each single recall trial, the phase difference at which these reactivations occurred, measured in phase units of the hippocampal theta oscillation was calculated. This analysis revealed a phase-shift of approximately 100 degrees (Figure 5A) between target memories and competitors in the first two retrieval repetitions (only seconds repetition surviving a control for multiple comparisons), suggesting that competing memories can be separated within a theta cycle, much in line with the mechanistic "phase-coding" model explaining how different representations are allocated to different theta phases depending on their relevance for the current task (Lisman & Idiart, 1995; Lisman & Jensen, 2013).

In a recent study (Kunz et al., 2019), iEEG was analysed whilst participants were conducting a goal-directed spatial memory task. Results showed that multiple goal locations were organised across multiple phases along the hippocampal theta oscillation. Other recent work has suggested a role for the hippocampal theta oscillation in organising the contents of working memory when multiple representations have to be maintained online (Bahramisharif et al., 2018; Fuentemilla et al., 2010; Lisman & Idiart, 1995). Evidence is thus accumulating that the hippocampus acts as an adaptive modulator for representations that need to be separated for maintenance or retrieval. Consequently, it was hypothesised in the present study that the phase difference between target and competitors we observed would be further modulated by behaviour, such that participants with less interference would demonstrate stronger phase separation. By splitting the phase difference

between target and competitor memories based on the amount of interference a participant experienced, it was shown that participants who experienced low levels of interference, also showed a highly consistent phase separation angle (Figure 5B). Less consistent phase separation was found for participants who experienced high levels of interference. It is important to emphasise that also a significant phase separation on a group level was found, but the median split on behaviour revealed that clustering was more pronounced in the low interference group, while the high interference group showed less consistent clustering. It is also interesting to note that this more variable phase separation in the high interference group was accompanied by a numerically smaller phase distance (i.e., average separation angle) between target and competitor reactivations. It thus appears that in participants experiencing high levels of interference, targets and competitors are reactivated both closer together in phase space, and at more variable time points along the hippocampal theta rhythm.

This finding is corroborated by another analysis correlating competitor attenuation and phase difference. The decline in competitor reactivations significantly correlated with the size of the phase difference, such that participants with a separation angle closer to 0 showed less of a decline. Low susceptibility to interference in our study is thus associated with a larger and more consistent phase distance of the target and competing memories.

Theta-phase models (Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007) of the hippocampus suggest that encoding and retrieval are separated by approximately 180 degrees, with one half of the cycle being devoted to encoding and the other half

to retrieval. Such phase-specific circuit operations have been demonstrated in rodents numerous times (e.g., Douchamps et al., 2013), and recently evidence was shown for a similar process separation in human participants using scalp EEG (Kerren et al, 2018). Interestingly, such strong phase separation along the oscillation is consistent with the model (Norman, Newman, et al., 2006; Norman et al., 2007) where targets should be maximally decodable during the opposite half of the cycle (i.e., the high inhibition half) compared to competitors (low inhibition half). However, it is unclear how this finding can be reconciled with the fact that both target and competitor reactivations are mnemonic/retrieval related activations, and should thus both fall onto the retrieval half of the theta cycle. Future modelling work is needed to reconcile this apparent theoretical conflict between models.

Another theoretical possibility to overcome interference that has been proposed in the literature is through integration, leading to the two related experiences being stored as one unified representation in the brain (DeVito, Lykken, Kanter, & Eichenbaum, 2010; Grande et al., 2019; Horner, Bisby, Bush, Lin, & Burgess, 2015). In computational frameworks, integration is assumed to either happen when the initial memory is strongly encoded (O'Reilly & Rudy, 2001), or when the overlapping memory (competitor) is very strongly activated together with the target memory (Ritvo et al., 2019). Both differentiation and integration can reduce competition, but through very different underlying mechanisms. Instructing participants to form associations between memories that share the same cue allows new experiences to be integrated with the old memories, and such integration has been shown to behaviourally reduce interference (Anderson & McCulloch, 1999). Although not explicitly instructed in the

present study, it is possible that those participants who experienced less interference behaviourally used a strategy of integrating the target memory into the previously created memory trace of the competitor. More specifically, when the word cue is shown the second time at encoding, the hippocampus might prompt for retrieval of the already existing memory trace, and when a new association is shown, instead of separating this association from the previously laid-down trace, hippocampus integrates them (Guzowski, Knierim, & Moser, 2004; O'Reilly & Rudy, 2001). In light of the present results, those participants who formed an integrated trace at encoding would not experience interference at retrieval, and thus should show no strong phase separation of targets and competitors. This is, however, the opposite of what was observed in the present study, in which the strongest phase difference was found in the low interference group. Thus, the results are unlikely to be explained by participants overcoming interference by integrating the competing memories at encoding, and rather support an active down-regulation view of retrieval (Anderson, 2003; Anderson et al., 1994; Hulbert & Norman, 2015; Ritvo et al., 2019; Wimber et al., 2015).

In the pre-registration, the hippocampal phase as a modulator for memory reinstatement and ultimately tracking the relevance of memories by calculating the MI between 200 and 1200 ms was described. This would have been an exact replication of the approach used in the previous study (Kerren et al., 2018). However, no significant results were obtained using this method. It is argued that the peak detection method, as also used in Kerren et al. (2018) and Linde-Domingo et al. (2019) might be more sensitive in that it only includes the time points with maximum fidelity,

presumably less affected by general noise in the classification than taking the entire fidelity time course into account.

The fidelity peak detection approach was also used successfully in the previous paper to investigate the theta phase around classifier peaks (Kerren et al., 2018). Specifically, in that study, a method was developed in which epochs aligned to time points when the hyperplane best separates the two classes of neural patterns were created (i.e., retrieval of animate vs inanimate objects). In this way, it could be shown that the main difference between animate and inanimate object recall was derived from sources in the anterior temporal lobe, an area known to be engaged in semantic processing (Patterson et al., 2007). Using a similar approach, Dijkstra, Ambrogioni, and van Gerven (2019) realigned their data based on the time points when the classification of imagined mental images peaked, and could in that way enhance the effect of imagery activation on a source level, demonstrating the validity and power of single-trial level fidelity values. In the previous study, realigning the raw data based on the fidelity peaks also revealed a significant phase clustering approximately 200-250ms before the classifier detected maximal memory reinstatement. It was argued in that study that these results show how the hippocampus is reactivating the relevant memory index at an "optimal" phase, to then send signals to neocortical areas for reinstatement with an approximate temporal delay of 250ms (Marr, 1971; McClelland et al., 1995; Norman & O'Reilly, 2003; O'Reilly et al., 2014; Teyler & DiScenna, 1986; Teyler & Rudy, 2007). In the present study, these findings are expanded to show that the hippocampal theta rhythm not only clocks the reinstatement of single memory associations but also separates the reinstatement of multiple competing memories in time.

In sum, the present study shows evidence for how the hippocampal theta rhythm modulates and tracks the reinstatement of neural patterns representing target and competitor memories. Successful retrieval of target memories was related to a phase separation along the hippocampal theta oscillation during early recall attempts. These results are in line with a neural network model suggesting a central role for slow oscillations in resolving mnemonic competition and expand previous evidence for a role of the hippocampal theta oscillation in separating overlapping memories to the domain of long-term memory.

7. Acknowledgements

This work was supported by a fellowship from Stiftelsen Olle Engkvist Byggmästare awarded to M.W. and C.K., and a Starting Grant from the European Research Council awarded to M.W. (ERC-2016-STG-715714).

8. Authors contributions

Conceptualization, C.K. and M.W.; Methodology, C.K., B.G. and M.W.; Investigation C.K..; Formal Analysis, C.K. .; Writing – Original Draft, C.K. and M.W.; Writing – Review & Editing, C.K., B.G. and M.W.; Visualization, C.K.; Funding acquisition, C.K. and M.W.

Chapter 5 - Episodic Memory Reinstatement in Neocortex at the Time of Hippocampal Sharp-wave Ripples.

*At the time of thesis submission, this chapter represents preliminary analyses (*Casper Kerrén, Jürgen Fell, Maria Wimber and Bernhard Staresina*).*

Abstract

Hippocampal sharp-wave ripples (SPW-Rs) are a fundamental neural signature of memory consolidation and replay. Recent evidence also suggests an important role in the active, awake retrieval of episodic memories, with the density of SPW-Rs being higher, and the coupling between cortex and MTL being stronger when participants are successful at retrieving previously encoded content. The present study set out to replicate these findings by analysing data directly recorded from the hippocampus and adjacent regions together with neocortical areas whilst patients were participating in a source memory recognition task. The results provide support for recent discoveries, further showing that the density of SPW-Rs is higher and neocortical reinstatement is significantly stronger when correct recognition judgements were accompanied by source recollection, compared unsuccessful attempts. These results identify the hippocampal SPW-R as a central neurophysiological component signifying early stages of the retrieval process and mediating the interplay between the hippocampus and neocortical regions that may lead to retrieval-mediated memory stabilisation.

1. Introduction

To recover previously encoded information, hippocampus and cortex need to be coordinated, and in rodents, this communication has been shown to depend on the timing of hippocampal sharp-wave ripples (SPW-Rs). Specifically, the reactivation of cell assemblies that were originally active during encoding is accompanied by SPW-Rs (Diba & Buzsaki, 2007; Lee & Wilson, 2002). SPW-Rs are composed of sharp-waves (slow, large-amplitude, excitatory oscillations at around ~3Hz [SPW]) transmitted from CA3 to CA1 and ripples (fast-oscillations at around 80-120Hz in human subjects and 110-250Hz in rodents originating in CA1) (Buzsaki, Leung, & Vanderwolf, 1983; Csicsvari, Hirase, Mamiya, & Buzsaki, 2000). The quantity of SPW-Rs is largest during non-rapid eye-movement sleep, and when rodents eat and drink (Buzsaki et al., 1983), and is suggested to support the replay of events for memory consolidation and to plan future behaviour and decisions (Buzsaki, 2015). SPW-Rs are important for spatial navigation, where especially long-duration SPW-Rs increase in situations demanding spatial memory, which some authors argue could potentially be due to the recruitment of more neurons representing relevant spatial information (Fernandez-Ruiz et al., 2019). Thus, SPW-Rs are an important hippocampal feature necessary for successful episodic memory consolidation and retrieval in rodents. In human subjects, SPW-Rs have been implicated in memory processes too, mainly during sleep (Axmacher, Elger, & Fell, 2008; Bragin, Engel Jr, Wilson, Fried, & Buzsáki, 1999; Buzsaki, 1996). These results have gained further support by showing that selectively eliminating SPW-Rs during the sleep following a learning period impairs memory (Girardeau, Benchenane, Wiener, Buzsaki, & Zugaro,

2009). In human subjects, only three studies have investigated SPW-Rs in episodic memory retrieval. In these studies it was found that the density of SPW-Rs is higher for successful compared to unsuccessful memory retrieval (Norman et al., 2019); that ripples are accompanied by bursts of single-unit activity (Vaz, Wittig, Inati, & Zaghloul, 2020); and that coupled SPW-Rs between medial temporal lobe (MTL) and temporal association cortex precede cortical reinstatement of verbal episodic memory (Vaz et al., 2019). These results indicate that the role of SPW-Rs in episodic memory is not limited to offline periods and sleep. In this chapter, preliminary results are provided testing the assumption that the number of SPW-Rs would be higher, and cortical memory reinstatement around the time of hippocampal SPW-R events would be stronger, for trials on which participants correctly recollect visual information associated with a word cue during encoding (source correct, SC), compared to no such source recollection (source incorrect, SI).

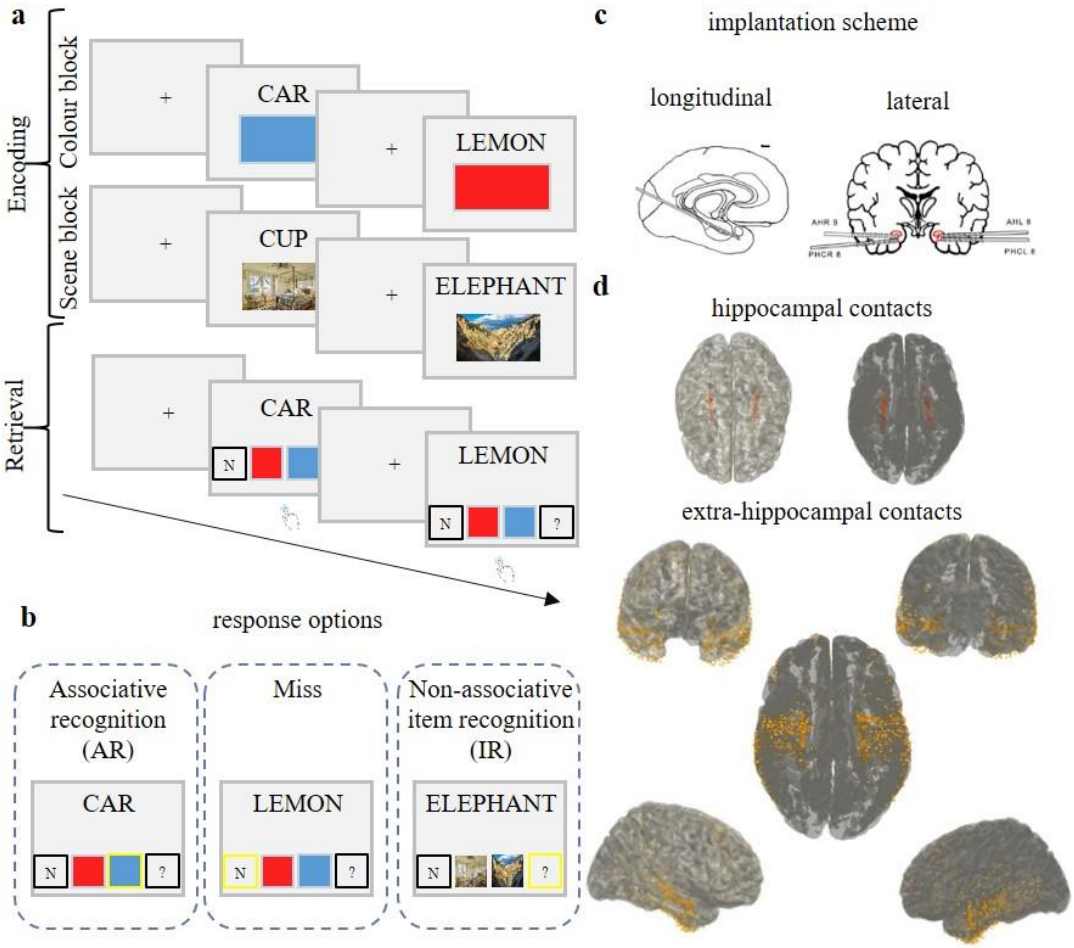

***Figure 1*. Paradigm and electrode coverage. A)** At encoding, subjects learned words and images. The images were of colours (red and blue) or scenes (indoor and outdoor). **B)** At retrieval, all 50 associations from encoding were shown, together with 25 new associations. The task for the subject was to judge whether the association had been seen before or not. **C)** Implementation scheme, showing the longitudinal implementation scheme to the left and lateral to the right. **D)** Coverage of all 12 patients, on top showing the hippocampal contacts used and extra-hippocampal contacts used at bottom.

2. Method details

2.1. Participants

A total of 12 patients (6 females) with a mean age of 33 years (SD = 9.3 years) suffering from pharmaco-resistant epilepsy participated in the experiment. Informed consent was obtained from all participants, and the study was approved by the ethics committee of the Medical Faculty of the University of Bonn.

2.2. Experimental procedures

The same experimental procedure was used in the present study as in Chapter 3, but for completeness, it is included here, too.

Participants were sitting upright in a sound-attenuated room in front of a laptop computer at an approximately 50-cm distance and partook in an associative learning paradigm (Figure 1), where each block consisted of an encoding phase, a distracter phase, and a retrieval phase. During encoding, a German noun was paired with either a colour or a scene, depending on which run they were in. Colour and scene runs alternated, wherein colour runs, the noun was either paired with a red or blue square, and in scene runs with either an indoor or an outdoor image. The encoding task was to associate the word and the picture stimulus (colour or scene) by imagining, as vividly as possible, the entity described by the noun together with the colour (e.g., a red lemon) or scene (e.g. an elephant in the mountains), and to rate the plausibility of that self-created image. Participants were given a maximum of 3 seconds to make their plausibility judgment. Each trial was preceded by a jittered intertrial interval (700–1300 ms, mean = 1000 ms) during which a fixation cross was shown in the centre of the screen. Participants' button press terminated the trial. During retrieval, participants

were presented with 50 previously seen words intermixed with 25 novel words, together with four choice options. The task was to indicate, with a single button press, whether the word was new ('N' response), whether it was old but the target association could not be retrieved ('?' response), or whether the word was old and the target association was also remembered (in which case the correct colour or scene response was to be selected). Responses were given in a self-paced manner but had an upper time limit of 5 seconds. Again, each trial was terminated with the button press and was preceded/followed by a jittered intertrial interval (700– 1300 ms, mean = 1000 ms) showing a fixation cross. Each run lasted approximately 9 minutes. Eight participants completed all six runs and three participants completed five runs. Two participants completed the six runs in two different sessions.

2.3. Implementation of depth electrodes

Intracranial electroencephalography (iEEG) data were referenced to linked mastoids and recorded from medial temporal lobe regions, hippocampus and, additional cortical regions with a sampling rate of 1 kHz (bandpass filter: 0.01Hz to 300Hz) (see Figure 1C and D). The depth electrodes in the hippocampus were implanted stereotactically during presurgical evaluation following two different schemes. For 8 of the patients, depth electrodes were implanted along the longitudinal axis of the hippocampus, whereas 4 patients had depth electrodes implanted laterally via the temporal lobe (see Figure 1C). Patients received anticonvulsive medication (plasma levels within the therapeutic range). The drug regimen at the time of recording for each participant is listed in Supplementary Table 1. To determine localisation of electrode contacts,

anatomy and pair-wise coherence were used, and for analysis purposes, only contacts in the non-pathological hemisphere were included. Participants also had additional electrodes placed on the scalp at position Cz, C3, C4 and, Oz, according to the 10-20 system. These electrodes were later removed from further analyses.

2.4. Electrode selection

The localisation of electrode contacts was decided based on several different criteria. First, two researchers independently evaluated the post-implantation MRI scans and made notes of the electrode contacts implanted in the hippocampus, with only contacts agreed on between the two researchers kept for later analysis. Secondly, the pairwise channel coherence (4–8Hz) at retrieval was calculated, with the assumption being that electrode contacts within the same region would show strong coherence (Mormann et al., 2008; Staresina et al., 2012). Thirdly, the event-related potential (ERP) of each electrode contact was calculated, again with the assumption that contacts in the same area would show similar ERPs. These localisation criteria resulted in a total of 75 contacts across 12 patients (mean = 6.25, SD = 2.2 electrodes) judged as placed in the hippocampus. For extra-hippocampal regions, 665 electrode contacts (mean = 55.4, SD = 21.1 contacts) across participants were included in subsequent analyses (Figure 1D).

2.5. Preprocessing

Data processing was performed with FieldTrip (Oostenveld et al., 2011) and standard MATLAB functions as well as custom-written MATLAB code. Line-noise was removed using bandstop filters (2Hz width) centred at 50, 100, 150, and 200Hz. Data were after this re-referenced based on the common trimmed average (using the MATLAB function *trimmean*), trimming away 10% of the highest and lowest outliers before subtracting. This was followed by an automatic artifact rejection. For this step, the data were z-transformed based on three different metrics: absolute amplitude, gradient (the amplitude difference between two adjacent time points), and the amplitude of the same data 250Hz high-pass filtered. The latter measure was included to identify epileptogenic spikes. If a time-point exceeded a z-score of 6 in any of the measures or a z-score of 4 for a conjunction of gradient and high frequency or amplitude and amplitude for high-pass filtered data, it was marked as an artifact. An additional 50ms on each side was marked as artifact and thus not included in any further analysis steps.

2.6. Ripple detection

For each hippocampal channel, ripple events were detected using established detection algorithms (Molle, Bergmann, Marshall, & Born, 2011; Molle, Marshall, Gais, & Born, 2002; Staresina et al., 2015). In brief, data were filtered between 80 and 120Hz, and only events between 38ms to 500ms in duration were considered for extraction. Additionally, to be considered for extraction, the event had to continue for a minimum of three cycles based on the raw signal. The density of extracted ripples was calculated as follows. Each event was divided by the length of the trial that it belonged to, up

until the reaction time. As no ripple events could be detected when an artifact was present, the length of any artifact was subtracted from the trial length before dividing. The average number of ripples could be considered the frequency at which the ripple occurred.

2.7. Isolation index

Previous studies have shown that ripple events occur locally (Khodagholy, Gelinas, & Buzsáki, 2017; Vaz et al., 2019). In the present study, it was therefore of interest to quantify the number of ripples that occurred in isolation in hippocampal contacts. For each ripple, the conditional probability that a ripple was detected in more than one hippocampal electrode contact was investigated. Thus, given a ripple in one contact, how likely is a ripple identified in other electrode contacts within a narrow time window? This isolation index was calculated for three different time periods (1 ms, 10 ms, and 50 ms ± the time of the event).

2.8. Multivariate pattern analysis

The raw iEEG time series were realigned (i.e., epoched) based on both the trial onset at encoding, and based on ripple events at retrieval. This resulted in two datasets, where the encoding dataset functioned as our training set (-500ms pre-cue to 2000ms post-cue), and the ripple-aligned data (-1000ms pre-ripple to 1000ms post-ripple) as our testing set. Hippocampal contacts were excluded from the neocortical classification analysis. Both the encoding and retrieval data were down-sampled to

200Hz, smoothed with a 100ms moving average (using the MATLAB function *smoothdata*) and baseline-corrected (-200 to -50ms before cue onset). For the ripple-aligned data, we found the pre-cue baseline for the trial corresponding to the extracted ripple. For encoding, all trials were used, splitting the data into two classes, colour and scene, that the classifier was trained on. For retrieval, the classification performance of colour vs scene was tested separately for source correct and source incorrect trials. Before running the classification, both the training and test datasets were z-scored separately. An LDA was used, training and testing on each time point with the shrinkage parameter set to 1, for optimal regularisation. Since training and testing was performed on independent datasets, no cross-validation was performed.

3. Results

3.1. Ripple density

Extracted ripples showed the typical characteristics described in the literature (Figure B and C). The density of ripples was higher when patients correctly retrieved the source (SC; mean = .26Hz, SD = .11Hz) at retrieval as compared to incorrect retrieval (SI; mean = .20Hz, SD = .07Hz), $t(11) = 3.67$, $p = .0037$; see Figure 2D). This pattern was consistent across all 12 patients, showing evidence for a previously reported result in which patients were performing a free-recall task (Vaz et al., 2019).

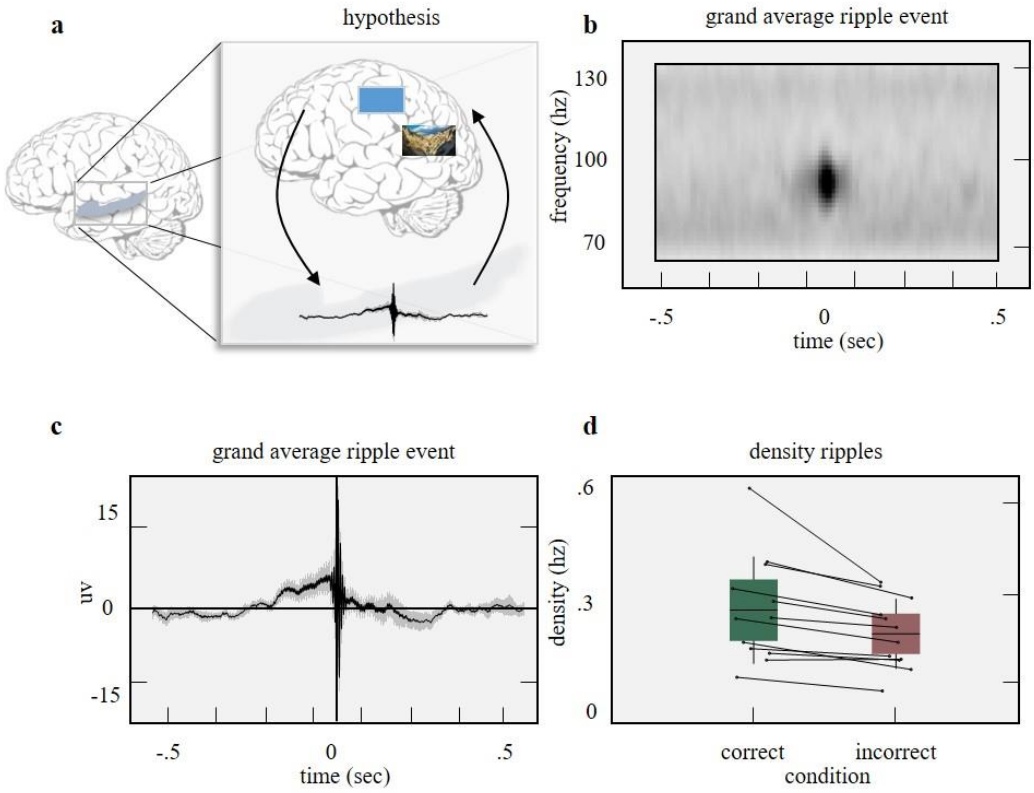

***Figure 2*. Rationale and results for extracted ripples. A)** Around the time of a ripple event, an interplay between neocortical areas and hippocampus is assumed to take place, with neocortical regions biasing processing of information in hippocampus. Hippocampal SPW-Rs are then assumed to trigger the reinstatement of information in neocortical areas. **B)** Showing a time frequency plot with the grand average frequency signature of all detected ripples across all subjects, with a mean of approximately 90Hz **C)** The grand average ripple time-locked to maximum amplitude at time zero shown. **D)** The density of ripples was significantly higher for source correct versus source incorrect retrieval trials. This pattern was seen for each single subject.

3.2. Majority of ripples occurring in isolation

An isolation index was calculated for each channel, to investigate whether ripple events occurred in isolation, or adjacent electrode contacts all showed evidence for an event simultaneously. The isolation index was calculated for each patient using three different time windows (±1ms, ±10ms, and ±50ms) in which a ripple could occur in

any other electrode contact that was included for ripple extraction. This was done to understand whether adjacent contacts picked up the same signal, or if the ripple events only occurred locally. When only allowing for 1ms to pass between a ripple in contact one and the rest of the used contacts, 69.3% (SD = 15.8%) of the ripples occurred in isolation. Instead allowing for 10ms between events decreased the number of ripples occurring in isolation to 65.9% (SD = 16.8%). Finally, when allowing for 50ms between events, the percentage of ripples occurring in isolation was 51.7% (SD = 18.5%). Thus, as expected, the isolation index decreases when increasing the co-incidence time window, but 50% of ripples occurred in isolation, suggesting a rather local phenomenon.

3.3. Neocortical reinstatement surrounding ripple events

The three subsequent analyses were aimed at directly investigating neural reinstatement (Figure 2A). To this end, data were realigned based on hippocampal ripple events, with the onset of a ripple at time zero. Importantly, data were aligned based on hippocampal ripple events, but all hippocampal contacts were excluded from the decoding analyses investigating neocortical reinstatement. An LDA classifier was trained at each time-point at encoding (-500 to 2000ms), time-locked to cue onset, and tested at each time point surrounding the onset time of ripples at retrieval (-1000 to 1000ms). The main analysis of interest tested for significantly more ripple-bound memory reinstatement for source correct than source incorrect trials. Figure 3A shows that indeed a significantly higher decoding accuracy for source correct than source

incorrect trials was found (significance tested using a non-parametric cluster-based permutation procedure, $p = .02$), providing the first evidence that ripples are related to neocortical reinstatement. Next, based on these findings, it was of interest to understand whether the reinstatement was exclusive to neocortical regions, or if hippocampal contacts showed a similar pattern. Therefore, all neocortical contacts were excluded, and only the same contacts as were used for extracting the ripple events were included. Using a non-parametric cluster-based permutation procedure, no significant difference between source correct and incorrect when only including hippocampal electrodes was found, suggesting that ripples are a hippocampal signature for reinstatement in neocortical areas rather than in hippocampus (Figure 3B). This was also true when contrasting source correct to chance level (not shown in Figure 3). Lastly, based on recent findings showing that ripple events recruited more neurons in a spatial navigation task (Fernandez-Ruiz et al., 2019), and that ripple events are influenced by cortical input to the hippocampus, it was hypothesised that a cortical-hippocampal-cortical information loop could take place, with the ripple event at time zero, changing the dynamics of the retrieved material from coarse-grained (supra-category decoding) to a more fine-grained reinstatement (sub-category decoding). To this end, the classifier was trained and tested on red versus blue and indoor versus outdoor separately. The results of the two classification analyses were then collapsed before running a non-parametric cluster-based permutation between SC and SI. This test did not reveal a significant result (Figure 3C). Similarly, contrasting SC to chance level did not reveal a significant reinstatement (not shown in Figure 3).

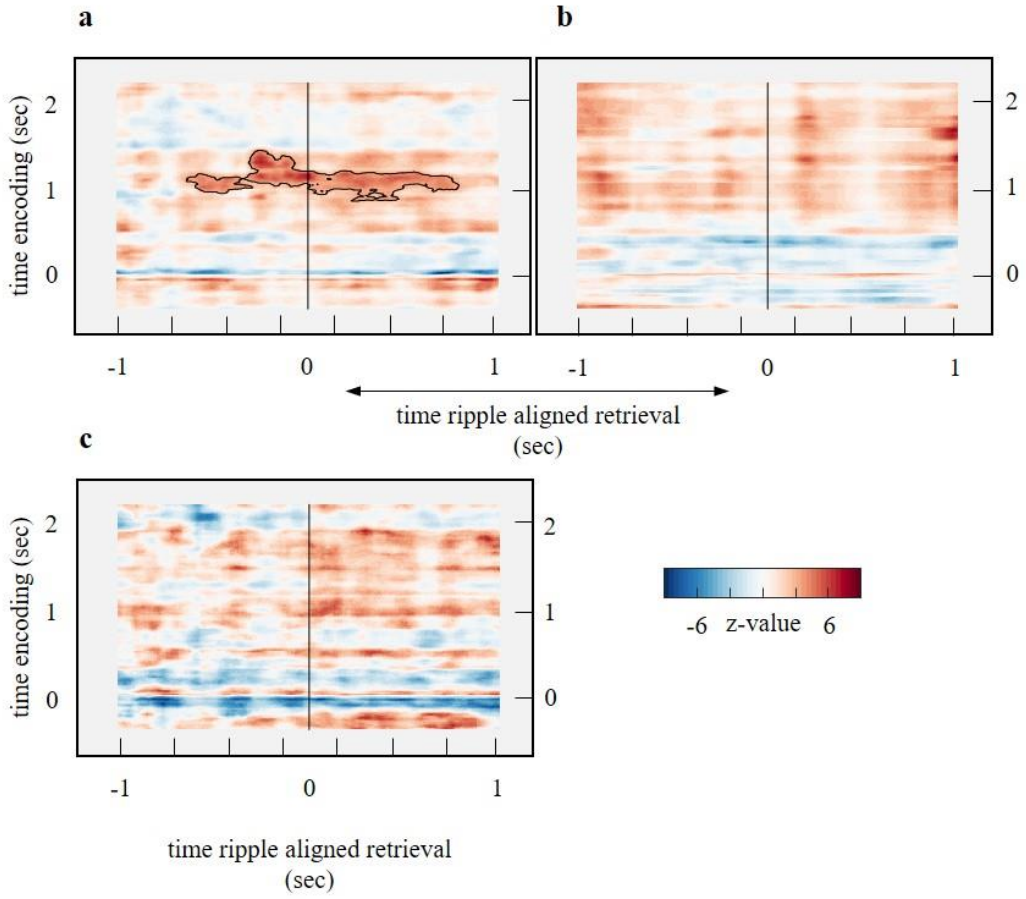

***Figure 3*. Decoding analyses. A)** Decoding colours versus scenes, training at encoding from cue onset, and testing at ripple-aligned data. Results showing significant reinstatement for source correct versus source incorrect both before and after hippocampal ripple events. **B)** No significant reinstatement around ripple events when only including same hippocampal channels as the events were extracted from. **C)** Decoding sub-category classifiers (red versus blue and indoor versus outdoor) SC versus SI showed no significant performance around ripple events.

4. Discussion

Until recently, the involvement of ripples in the active retrieval of episodic memories has not been a major topic of interest. In the rodent literature, ripples are a key mechanism in consolidation and replay, during offline states after learning, when the animal is resting or sleeping (Buzsaki, 2015). Similarly, ripples in human subjects have

foremost been studied when patients are asleep or in offline states (Diekelmann & Born, 2010). From these studies, a possible relationship between neocortical activity and ripples has been sketched. Spontaneous neocortical activity is hypothesized to signal hippocampus, which in turn, if the spontaneously elicited patterns are similar enough to reawaken the stored index in the hippocampus, compressed information about the memory will be sent back to widespread regions of the cortex, where the memory is being reinstated (Ji & Wilson, 2007; Rothschild, Eban, & Frank, 2017). In rodents, ripples have been implicated in spatial working-memory tasks, where the duration of ripples is important for correctly retrieving the spatial information (Fernandez-Ruiz et al., 2019). Only very recently, studies in human subjects have demonstrated the importance of ripples in episodic memory retrieval (Y. Norman et al., 2019; Vaz et al., 2019; Vaz et al., 2020). Furthermore, in rodents, ripples are preceded by neocortical activity, which in turn influences hippocampal activity around the time of the ripples event (Rothschild et al., 2017). In the work presented in the present chapter, the assumption of ripples being an important building block in episodic memory retrieval was tested, with the hypothesis that correct trials were related to a higher density of ripples and strong memory reinstatement. Evidence for both of the hypotheses was found, in that successful episodic memory retrieval was accompanied by more ripples (Figure 2B), and significantly more neocortical reinstatement (Figure 3A), than unsuccessful source retrieval. The result thus provides further evidence for hippocampal ripples being important for memory retrieval.

To extract ripples from the data, established ripple detection algorithms were used (Molle et al., 2011; Molle et al., 2002; Staresina et al., 2015). Ripple events between

38 ms and 500 ms were detected. The higher cut-off was selected to account for a recent finding suggesting that long-duration ripples are linked to episodic memory retrieval (Fernandez-Ruiz et al., 2019). The density of ripples detected was calculated for trials in which participants correctly retrieved the association and the source, and for trials in which they did not. A significant difference between trials on a group level was found, with each participant showing more ripple events in correct versus incorrect trials (Figure 2B). Next, the extent to which ripple events occurred in isolation was quantified. The results suggested that a majority of the events were isolated to one electrode contact when allowing for 10ms before and after that event to detect ripple events in other contacts. Even when increasing this time window to 100ms around the event, only approximately 50 percent of the ripples occurred in isolation. In a previous study, it was shown that only about 10-40% of ripple events meet a statistical threshold for being considered important for the replay of previously encoded events (Joo & Frank, 2018; Tingley & Peyrache, 2020). Thus, potentially, the detection algorithm finds ripples with different relevance for replay, and a more sensitive extraction, taking into consideration the parameters discussed in Tingley and Peyrache (2020) might have shown a different result.

The next step was to realign the data with the ripple maximum at time zero. The classifiers were trained at each single time point following cue onset at encoding and tested it at each time point at retrieval around the time of ripple onset. Based on the recent findings showing that hippocampal ripples are important for episodic memory retrieval (Norman et al., 2019; Vaz et al., 2019; Vaz et al., 2020), the assumption was that increased neocortical reinstatement would be found in the surround of the ripple

events, and that reinstatement would be stronger for successful than unsuccessful source memory retrieval. One possibility was that the ripples initiated the transfer of information between the hippocampus and neocortex, which would result in neocortical reinstatement following the ripples events (Ji & Wilson, 2007; Rothschild et al., 2017; Vaz et al., 2019). A second possibility was that neocortical activity precedes ripple events, which in turn precede the second wave of neocortical activity (Rothschild et al., 2017; Swanson et al., 2020). The results seem to support the latter option; significant reinstatement was observed both before and after ripple events (Figure 3A). An intriguing explanation of this time course is that neocortical regions are biasing hippocampal ripples, which in turn send information to other areas of neocortex for full reinstatement (Swanson et al., 2020). Potentially, the reinstatement before the ripple events could reflect the processing of the cue, and the reinstatement after the ripple events the processing of memory (Staresina & Wimber, 2019). Further analyses are planned on the current dataset to fully investigate this theorized cortical-hippocampal-cortical loop.

Through sparse coding and a high learning rate, the hippocampus is foremost known to encode individual events (McClelland et al., 1995). In fact, according to the hippocampal indexing theory information in the hippocampus is only stored as an index to the episode unique patterns in the neocortex (Teyler & DiScenna, 1986; Teyler & Rudy, 2007). Thus, the hippocampus in itself does not contain the content of the encoded episode, instead, an index is stored that points to the relevant regions for the content to be accessed. However, accumulating evidence suggests that area CA1 might be able to also extract regularities from the environment, in a similar manner to

what has been assumed for neocortex (Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014; Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Schapiro, Turk-Browne, Norman, & Botvinick, 2016). More specifically, Schapiro et al. (2017) showed that through overlapping representations in CA1, the monosynaptic pathway from EC to CA1 could represent statistical regularities across episodes. Given these results, it was expected that there was at least a possibility for categorical reinstatement to be evident also in the hippocampus, and foremost in region CA1. Therefore, the same decoding analysis as above was conducted, classifying objects versus scenes, but this time only including the same hippocampal channels as for extracting the ripples (Figure 3B). No significant difference between SC and SI was found, neither a difference between SC and chance level. This null finding is instead in support of the hippocampal indexing theory, which, if correct would not assume reinstatement of category-based patterns in hippocampus. However, several other possibilities could also explain the result. First, ripples might be promoting the transfer of information to neocortical regions, and therefore only reinstatement will be found in neocortical regions. Second, the number of electrode contacts implanted in hippocampus were too few to obtain enough statistical power, or third that the features for classification were not variable enough. Another potentially more interesting explanation is that the electrode positions were too variable for this categorical classification. In a study similar to the present (Staresina et al., 2012), no evidence for hippocampal reinstatement was found. In this study, the authors argued that due to pattern completion and pattern separation taking place in different subfields (Marr, 1971; McClelland et al., 1995; Yassa & Stark, 2011) it might be necessary to subdivide hippocampus to avoid picking up these signals simultaneously, which will prevent

detection of reinstatement (Staresina et al., 2012). In the present study, contacts along the whole longitudinal axis of the hippocampus were used, likely covering several subfields. One way of potentially circumventing this problem would be to use representational similarity analysis (RSA), with the assumption that some electrodes (close to DG) should pick up episode-unique reinstatement (e.g., diagonal in an encoding-retrieval similarity, RSA-type analysis), due to each episode being coded as distinct. On the contrary, in CA1, instead, category-unique reinstatement would be evident. Future analyses are planned to evaluate a similarity-based approach.

The role of ripples also in reinstatement on a sub-category level was also investigated. One specific hypothesis tested here was that sub-category classification would reveal a later onset of neocortical reinstatement, because it requires access to more fine-grained and detailed information, in line with previous literature (Linde-Domingo et al., 2019). This analysis did not reveal a significant difference between correct trials and incorrect trials for the sub-category classifiers collapsed (Figure 3C). It is difficult to pinpoint the exact reason to why this analysis did not show a significant effect. One explanation could be the that the coverage of neocortical areas for some patients included areas more responsive to scene processing than colour processing, such as the parahippocampal gyrus (Aminoff, Kveraga, & Bar, 2013; Staresina et al., 2012) and retrosplenial cortex (RSC; (Epstein, 2008)), as compared to more occipital areas (Bartels & Zeki, 2000), and vice versa. A second explanation could be that the role of the ripples is to initiate the memory reinstatement in neocortical areas on a more coarse category-based level. If this is the case, the ripple-aligned data would not contain the more fine-grained information of the sub-categories. However, this explanation would

go against theories which assign the role of hippocampus particularly to scene processing (Dalton, Zeidman, McCormick, & Maguire, 2018; Zeidman, Mullally, & Maguire, 2015), meaning that it would be possible to obtain significant reinstatement using the sub-category classification approach. Future analyses are planned to evaluate the specific localisation of neocortical contacts and their relationship to classifier performance on a sub-category level.

In the present study evidence for the involvement of hippocampal SPW-Rs in episodic memory retrieval in human subjects is shown. Correctly remembering an episode is accompanied by a higher density of SPW-Rs. Furthermore, when realigning the data based on hippocampal SPW-R events, the study first replicates previous findings, but also moves beyond memory reinstatement in free recall and verbal materials, and sheds light on the time course of reinstatement around ripples, and the functional role for memory.

5. Acknowledgements

This work was supported by a fellowship from Stiftelsen Olle Engkvist Byggmästare awarded to M.W. and C.K., and a Starting Grant from the European Research Council awarded to M.W. (ERC-2016-STG-715714).

6. Authors contributions

Conceptualization, C.K., B.S. and M.W.; Methodology, C.K., B.S. and M.W.; Investigation C.K and B.S.; Formal Analysis, C.K. and B.S. .; Writing – Original Draft, C.K. and M.W.; Writing – Review & Editing, C.K., B.S. and M.W.; Visualization, C.K.; Funding acquisition, C.K. and M.W.

7. Supplementary material

Table 1: Participant's drug regimen at the time of recordings

Participant	Anticonvulsant	Antidepressant
p01	Clobazam, Valproat	-
p02	Clobazam, Valproat	-
p03	Lamotrigin, Levetiracetam	-
p04	Lacosamide, Levetiracetam	-
p05	Lamotrigin, Levetiracetam	-
p06	Lamotrigin, Oxcarbazepin	-
p07	Lamotrigin, Oxcarbazepin	Citalopram
p08	Lamotrigin	-
p09	Levetiracetam	-
p10	Lamotrigin	Sertralin
p11	Levetiracetam, Oxcarbazepin	-
p12	Oxcarbazepin	-

Chapter 6 - General discussion

This doctoral thesis is focused on the temporal dynamics of episodic memory processes, with the main aim of linking the timing of neural reinstatement during memory recall to neural rhythms. On the one hand, reinstatement was linked to the hippocampal theta rhythm, and on the other to hippocampal SPW-Rs. First, the aim was to link memory reinstatement to the phase of the hippocampal theta rhythm. To have optimal coverage for finding brain-wide mnemonic patterns, scalp EEG was used in the first study. The second study complemented this first study, by recording intracranial EEG directly from the hippocampus and surrounding neocortex. The third study then directly built on the findings from the first one, to test a model suggesting that the brain assigns overlapping memories to different phases along the hippocampal theta rhythm. Lastly, in the final study, it was tested whether neocortical reinstatement is also linked to faster rhythms, and specifically whether it occurs around the time of hippocampal SPW-Rs, which have previously been shown to be involved in offline replay and consolidation.

In this Chapter, the principal findings for each of the Chapters will briefly be summarised. Thereafter the results will be anchored in the existing body of literature, discussing the major consistencies and discrepancies both concerning other work, but also within this thesis. The last section of this General discussion will then, based on the empirical results in this thesis, propose a few ideas that can be tested in future projects.

1. Principal findings

1.1. The rhythmic modulation of encoding and retrieval

Brain imaging methods allowing for analysing human brain data with a high temporal resolution, such as EEG and MEG, have been used in the field for a long time. However, little is still known about the precise temporal dynamics of episodic memory retrieval. A biophysical theta-based model was developed to explain how the hippocampus can quasi-simultaneously encode new and retrieve old information, processes that arguably rely on incompatible computations (Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007). The model (Figure 1, General Introduction) builds on findings from studies in rodents showing that opposing phases of the rat's hippocampal theta rhythm are related to LTP and LTD, respectively (Huerta & Lisman, 1993; Pavlides et al., 1988). Encoding an event is assumed to be related to strong input from EC to the hippocampus. This strengthening of input pathways, together with the strong input of orthogonalised patterns from DG to CA3, lower activity in CA3, between CA3 and CA1, and weak output from hippocampus back to neocortical areas, sets the system in a state that is optimal for creating new patterns of the incoming information (i.e., pattern separation). During retrieval, on the other hand, the reverse pattern is true. Input from EC to the hippocampus is weaker, and activity in connections within CA3, and between CA3 and CA1, is strong. This together with a strong output from the hippocampus is the optimal configuration for allowing formed (at encoding) cell assemblies to fire together and thus generating the output necessary for retrieving an event (i.e., pattern completion). Importantly, the hippocampal system

is assumed to constantly fluctuate between these two states on a fast time scale, and this flip is assumed to be orchestrated by the theta rhythm (Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007).

The studies in Chapter 2 and 3 tested this basic assumption of a mnemonic flip to provide direct evidence for the model in humans. There were a several substantial differences between the two studies, most notably, of course, the two different recording methods that were used, capturing neural signals on two distinct anatomical scales (scalp EEG and iEEG), and in different populations (healthy young population and epileptic patients, respectively). Other significant differences include the use of a cued recall task (Chapter 2) versus a source memory recognition paradigm (Chapter 3), and the use of animate vs inanimate material for classification in Chapter 2, whereas Chapter 3 used objects and scenes. It is important to emphasize, however, that to replicate the findings in Chapter 2, identical settings regarding the analysis steps were used in Chapter 3. Data were analysed by the means of a modulation index, in which the coupling between the phase of the hippocampal rhythm and the amplitude of the classifier-based neural index of memory retrieval was calculated. In both studies, we found a significant modulation of the representations active during encoding and retrieval. A contrast between the two task phases could then be computed by extracting and comparing the phase at which we found the maximum fidelity for encoding and retrieval, respectively. This indicated a phase difference of approximately 180 degrees for Chapters 2 and 3. The results thus show good agreement despite the above-mentioned differences and suggest an important role of the hippocampal theta oscillation in clocking neocortical activity that is relevant for encoding or retrieving

episodic memories. Overall, the results are consistent with computational models suggesting that the hippocampal circuits are optimally geared towards encoding novel incoming information, versus retrieving previously-stored internal information, at distinct phases of the theta rhythm.

1.2. Interference along the hippocampal theta rhythm

Building on the finding that target memories can be decoded from brain activity at particular phases in the theta rhythm, Chapter 4 then tested the assumptions of another computational model related to the theta rhythm (Norman et al., 2007), this one using transitions in theta phase to resolve interference between overlapping memories. In Chapter 4 it is believed that the computations at different phases of the rhythm will support successful memory retrieval for different reasons. The model capitalizes on the known properties of slow oscillations to raise or lower the threshold for local neurons to fire and that LTP and LTD are induced at different phases of the oscillation. At the high inhibition phase of a slow oscillation, only the most relevant neurons (i.e., those that receive the strongest presynaptic input) will fire, while at the opposite low inhibition phase, additional units will surpass the firing threshold, producing more "noise" in the firing patterns (Lisman & Idiart, 1995; Lisman & Jensen, 2013). In the model tested here, the transition from high to lower levels of inhibition is used to identify and strengthen overly weak nodes of the target memories (via LTP), whereas the opposite phase (i.e., the transition from low to higher levels of inhibition) acts to identify and weaken overly strong nodes of competing memories (via LTD). The

present study used a variant of a classic AB-AC proactive interference paradigm, where subjects were to encode a word together with either one or two images.

Following the same rationale as in Chapters 2 and 3, a neural index of memory reactivation was obtained from multivariate pattern classification of image category during retrieval, independently for target and competitors. The most important metric in Chapter 4 was again the degree to which the hippocampal theta oscillation modulated the classifier-based neocortical patterns during recall. If the model is correct, target and competitors should be preferably active at different phases of the rhythm.

Indeed, when subtracting the maximum phase of competitor reinstatement from that of target memories, a significant difference of 100 degrees was obtained, for the second round of retrieval only. Importantly, the phase difference between memories was correlated with our behavioural index of interference, such that a stronger phase difference between target and competitor memories was found in subjects who showed low interference behaviourally. The phase difference also correlated with the amount of interference, with low interference being related to strong phase separation. These results demonstrate that the assignment of overlapping memories to distinct phases of the theta rhythm is functionally relevant to memory performance, in line with a role of theta phase in reducing mnemonic competition (Norman et al., 2007).

1.3. Neocortical reinstatement surrounding hippocampal SPW-Rs

Across three studies, evidence for the intricate link between memory replay and slow hippocampal rhythms has been shown. In these studies, neural indices of memory retrieval were obtained from multivariate pattern classifiers. Such classifiers use the patterns of brain activity that best differentiate between the relevant classes they are asked to distinguish, and the diagnostic information supporting the classification is likely high-dimensional and can span several brain regions and frequency bands. Previous studies have suggested that high-frequency rhythms in the gamma band represent the firing of cell assemblies that code for the content of mental representations (Nyhus & Curran, 2010; Tallon-Baudry & Bertrand, 1999), making gamma a prime candidate for providing diagnostic information for pattern classifiers. Moreover, high-frequency ripples have been identified as a mediator of information during replay (Buzsaki, 2015), again suggesting a central role of high-frequency activity for retrieval. Motivated by these findings, it was the goal of the last study to identify the relationship between memory reinstatement and fast neural rhythms and particularly sharp-wave ripples.

Previous studies on ripples have shown a cortical-hippocampal-cortical relationship. Activity in neocortical areas influences the processing of memory reactivation during ripples events in the hippocampus, which in turn, affect the reinstatement of the memory in the neocortex (Ji & Wilson, 2007; Rothschild et al., 2017). While early studies in human subjects focused on ripples during sleep and offline periods (Axmacher et al., 2008), recent studies have shown evidence for the importance of SPW-Rs during active memory retrieval. In one study, a higher density of ripples was found when a subject correctly remembered as compared to incorrectly remembered

an item in a free-recall task (Norman et al., 2019). In another study, the coupling between MTL and cortical ripples was stronger when subjects correctly remembered (Vaz et al., 2019). In Chapter 5, the assumption of ripples being an important building block in episodic memory retrieval was tested, with the specific hypothesis that correct source memory is related to a higher density of ripples and strong memory reinstatement.

First, a significant difference between the number of ripples for correct and incorrect source retrieval was found. Second, results from a ripple-locked pattern classification analysis showed strong reinstatement in neocortical areas for correct trials surrounding the ripple onset. Although preliminary at the time of submission, these results provide supporting evidence for the relationship between hippocampal ripples and neocortical reinstatement, paving the way to a new and exciting area of research regarding the functional role of ripples in online and offline memory stabilization.

2. Converging and non-converging findings across studies

This section will discuss how the results across the four studies provide converging evidence for the computations along the hippocampal theta rhythm, and whether these computations are exclusive to the hippocampal formation. Second, the peak frequency of theta is discussed, as a major area of non-convergence across chapters. This will be followed by a discussion of how theta-locked classification can be compared and related to other commonly used approaches such as theta-locked gamma and alpha/beta desynchronisation.

2.1. Phase-coding mechanisms and theta oscillations

The rationale for selecting the hippocampal theta oscillation as a modulator of neocortical activity is based on a large base of research on the specific characteristics of this neural oscillation and its involvement in episodic memory, as has been described in the general introduction of this doctoral thesis.

One intriguing question regarding phase-coding along the theta oscillation in general, and our results specifically, is whether the computations investigated in these studies are exclusive to the hippocampal theta oscillation, or whether it would be possible that the neocortical theta could show similar results? As previously mentioned, extensive work on the hippocampal theta oscillation and the ability of the hippocampus to learn associations in a one-shot manner (McClelland et al., 1995), inspired us to focus on

this region. In rodents, theta phase coding has foremost been investigated in hippocampal and adjacent regions. This theta has been shown to both phase-lock cells in and around the hippocampus, but also in the medial prefrontal cortex (Hyman, Hasselmo, & Seamans, 2011) and amygdala (Seidenbecher, Laxmi, Stork, & Pape, 2003). In human subjects, relatively few studies have shown evidence of how information (assumed to be coded in high frequencies) is coded along the phase of the theta oscillation (cross-frequency coupling [CFC]) (Bahramisharif et al., 2018; Canolty et al., 2006; Friese et al., 2013; Kaplan et al., 2014; Köster, Friese, Schöne, Trujillo-Barreto, & Gruber, 2014). However, these studies suggest that the neocortical theta oscillation might show a similar kind of characteristic as the hippocampal theta oscillation. For example, evidence was recently shown for the mechanistic Lisman-Idiart-Jensen (LIJ) model first proposed in the mid-90' by Lisman and Idiart (1995). In this study (Bahramisharif et al., 2018), three-letter sequences could be decoded from gamma bursts (mainly in the occipital and temporal lobe), which again were sequentially coupled to the neocortical theta oscillation in the same regions. These results indicate a role of theta in separating and organising information, in the same fashion as the hippocampal theta. In another study, successful episodic encoding was related to posterior gamma oscillations that were phase-locked to a frontal theta oscillation (Friese et al., 2013). These results are similar to that of Heusser et al. (2016), but the theta oscillation picked up in Bahramisharif et al. (2018), Friese et al. (2013) and Canolty et al. (2006) was neocortical, whereas, in Heusser et al. (2016), the only region that stood out when source-localising the effect was hippocampus. Moreover, one study investigated working memory (Bahramisharif et al., 2018), and Heusser et al. (2016) and Friese et al. (2013) investigated episodic memory encoding. Despite

these differences, all studies nevertheless found a similar pattern of a fast gamma rhythm locking to, and being sequentially organised by the phase of, a slower theta oscillation. This set of empirical evidence is thus consistent with a more general role of slow oscillations in the theta range in organizing the representations of mental contents in time, beyond the hippocampus.

The theta-based model that was tested in Chapters 2 and 3 is explicitly linked to the theta oscillation in the hippocampus and its role in synchronizing sub-sets of the hippocampal circuits (Hasselmo et al., 2002). This rationale makes sense given the very unique characteristics of the subregions of the hippocampus, with pattern separation being strongly associated with Dentate Gyrus and pattern completion with CA3 (Yassa & Stark, 2011). The model tested in Chapter 4, on the other hand, is agnostic as to whether the slow oscillation regulating the level of inhibition is neocortical, hippocampal, or both (Norman et al., 2007). In fact, in their modelling work, the authors explicitly make a comparison to the dynamics of the hippocampal theta oscillation described in Chapter 2 and 3 but also argue that the pruning of memories at retrieval is more likely to take place in neocortical areas (Norman, Newman, et al. (2006), as also theorised elsewhere (Antony et al., 2017). Evidence from fMRI studies has provided mixed results regarding the locus of the representational changes and neural differentiation that result from repeated learning or repeated retrieval (Favila et al., 2016; Schlichting et al., 2015). Of special interest for the current question, Schlichting et al. (2015) report evidence suggesting that the posterior hippocampus together with anterior mPFC is more involved in separating memories, whereas anterior hippocampus together with posterior mPFC is more

involved in integrating memories. Although fMRI studies are not directly analysing theta oscillations, they can indicate the location where representational changes take place. Thus, results from fMRI studies seem to suggest that a differentiation process can take place both in the hippocampus and in the neocortex.

There is relatively scarce evidence for theta phase coding in cortical areas, and it is, therefore, difficult to conclude as to whether there is a fundamental difference between theta phase coding in neocortical areas and the hippocampus. Together with the results from Chapters 2, 3, and 4, the existing empirical evidence from both electrophysiological data (e.g., Heusser et al., 2016, Bahramisharif et al., 2018) and fMRI data rather point to a more general mechanism of slow oscillations in organising memories. An interesting avenue for future work is to investigate whether the temporal organization of mnemonic information remains stable or changes when comparing the organization relative to the hippocampal theta rhythm with that along other, neocortical slow rhythms.

2.2. The inconsistency of peak frequency

The peak frequency at which the hippocampal phase-modulated memory-related patterns, though always in the theta range, strongly differed between Chapter 2 (8Hz), Chapter 3 (2Hz), and Chapter 4 (5Hz). How can this difference be explained?

One potential explanation is that the different tasks in Chapters 2, 3, and 4 calls for a different pace of the slow oscillation. In Chapter 2, neural patterns belonging to

animate and inanimate objects were the input classes to the decoding analysis. In Chapter 3, objects were contrasted with scenes, and in Chapter 4, two within category classifiers were used to find differences between animate and inanimate objects and indoor and outdoor scenes. Additionally, in Chapter 4, a competitive memory association task was used. Although the condition in which subjects only associated the word with one image (NC-1) was analysed resulting in 5Hz being the peak frequency, the overall task demands might have slowed the frequency down to adjust to the higher selection demands and keeping the separation of competing material along the theta oscillation intact. Such slowing has been suggested in both the working-memory (Bahramisharif et al., 2018; Lisman & Idiart, 1995; Lisman & Jensen, 2013) and the spatial navigation literature (Kunz et al., 2019). In working-memory models, the frequency of the slow oscillation theoretically limits the number of high-frequency bursts that can fit in a single (theta) cycle (Lisman & Idiart, 1995; Lisman & Jensen, 2013). That means that if the theta oscillation slows down, more information can be slotted in. The difference in the peak frequencies across studies, which identified at 8Hz in Chapter 2 (single association cued recall) and Chapter 4 (selective recall amongst two competing associates) could thus be explained by a functional theta slowing, with the slower oscillation in Chapter 4 allowing for two stimuli to be allocated distinct slots along the cycle of the hippocampal oscillation. This conclusion is speculative at this point, and empirical work would need to test how the dominant theta frequency changes systematically with an increasing number of to-be-encoded or to-be-recalled memories.

Another potential explanation is the differences in methodology and analysis steps compared to other studies using gamma bursts as a proxy for memory reinstatement (Griffiths, Martín-Buro, Staresina, & Hanslmayr, 2020; Heusser et al., 2016; Köster et al., 2014; Nyhus & Curran, 2010). In many of these studies, the hippocampal raw data is analysed to find the peak frequency with which gamma bursts are coupled to (Aru et al., 2015). In Chapter 3, this more common procedure was used, in which the raw data was analysed instead of the fidelity values. This resulted in 2Hz as being the peak frequency, which is a frequency that has been theorised to be the equivalent of the rodent 4-10Hz theta, important for spatial navigation and memory (Jacobs, 2014). In Chapters 2 and 4, peak frequency was determined based on the average fidelity values, not on the actual raw data. The logic behind this is that the interest lied in the modulation of the memory index, where the assumption was that the strongest frequency should show the underlying oscillation responsible for this modulation. Although source analysis did suggest sources in the MTL and interlinked posterior-medial recollection network (Ritchey & Cooper, 2020), Chapter 3 then sought to more directly establish the link to the hippocampus by applying the same analyses to intracranial recordings from hippocampal and neocortical electrodes. To establish such a link, we decided that it was more appropriate to first determine the most dominant oscillation in the hippocampal recordings (using IRASA [see: Wen and Liu, 2016]), and to then analyse its ability to modulate neocortical, memory-related patterns. The dominant frequency identified this way in Chapter 3 was indeed slower than the theta rhythm seen over the scalp recordings in Chapters 2 and 4. It is noteworthy that a secondary peak was found around 8Hz (see Figure 2E, Chapter 3), although not showing a significant MI index (see Figure 3A and B, Chapter 3), this still indicates

that a faster oscillation is still present in the recordings from hippocampal electrodes. It is thus possible that the hippocampus proper oscillates at a slower rhythm in humans than in rodents (Jacobs, 2014), but that a faster oscillation is also present, and that the latter possibly synchronizes more strongly with neocortical networks (Hebscher, Meltzer, & Gilboa, 2019). This would be in agreement with a theory arguing for one fast and one slow theta oscillation dominating in the hippocampus during spatial navigation and episodic memory retrieval (Jacobs, 2014; Lega et al., 2012; Pastotter & Bauml, 2014). In fact, in a recently published article both a slow (~3Hz) and a fast (~8Hz) oscillation were found when analysing hippocampal activity whilst subjects performed a virtual spatial navigation task (Goyal et al., 2020). If it is true that a faster oscillation synchronizes more strongly with neocortical networks, then EEG and MEG might be more likely to pick up the modulation of reinstated memory content in areas that are very remote from the hippocampus, areas that might naturally oscillate at a faster frequency. Again, more empirical work is needed to directly test these ideas.

Across the three experiments, differences in the anatomical scale in which data are recorded at and the specific analytic procedure might have determined the peak frequency. However, differences might also be further explained by the presence of two different hippocampal thetas and lastly by varying degrees of selection demands, which would result in a theorised slowing of the theta.

2.3. The relationship between theta-locked classification and other commonly used approaches

A new method was used in Chapters 2-4, developed in Kerren et al. (2018), to link memory reinstatement to the phase of a slow oscillation. One of the main advantages of this method is the ability to track reactivation on a millisecond time scale when the reinstatement time course is unknown. Importantly, two different principle approaches were used across the chapters. In Chapters 2 and 3 a modulation index was calculated, which bins every single fidelity value according to its respective theta phase bin. In Chapter 4, a complementary method was used that only considered classifier peaks that exceeded a given noise threshold, and then investigated whether these peaks systematically occurred at a given theta phase. One question regarding this new method concerns its relationship to other, frequently used approaches. As has been described before, previous studies have on the one hand used gamma power to theta phase coupling to investigate both working memory and episodic memory (Axmacher et al., 2010; Fries, 2005; Lega, Burke, Jacobs, & Kahana, 2016), and alpha/beta desynchronization on the other (Griffiths et al., 2020; Griffiths et al., 2019; Parish et al., 2017). Do these methods provide similar information, such that theta-linked memory reinstatement can be seen as functionally similar to theta-linked alpha/beta or gamma power?

Evidence indicates that synchronised gamma in itself binds event-related information representing mental content and that gamma power coupled to the phase of theta is crucial for successful working memory, spatial memory and episodic memory encoding and retrieval (Axmacher et al., 2010; Griffiths et al., 2019; Lega et al., 2016). Work in rodents suggests that gamma code can functionally be further split into two distinct gamma frequencies, with a faster gamma rhythm more dominant during

encoding, and a slower gamma rhythm more dominant during retrieval (Zheng, Bieri, Hsiao, & Colgin, 2016). These two types of gamma have been further shown to have different sources, with the slow gamma driven by input from CA3 and fast gamma driven by input from EC (Colgin et al., 2009). A recent study showed the first evidence of this distinction in human subjects (Griffiths et al., 2019). In this study, intracranial EEG was recorded whilst subjects conducted an associative memory task. Episodic memory formation was tightly connected to power increase in fast gamma, whereas episodic memory retrieval was related to slow gamma power increase (Griffiths et al., 2019). This provides evidence for the idea of fast gamma allowing for spike time-dependent plasticity at encoding (Bi & Poo, 1998; Jutras, Fries, & Buffalo, 2009), whereas slow gamma allows for pattern completion at retrieval (Colgin et al., 2009). Interestingly for Chapter 2 and 3 of the present thesis, slow and fast gamma might provide of optimal states that put hippocampus in an encoding or retrieval mode (Colgin, 2016; Colgin et al., 2009; Zheng et al., 2016). At encoding, information needs to flow into the hippocampus, and thus a strong connection between EC and hippocampus is crucial (Colgin et al., 2009). This is suggested to be facilitated by fast gamma oscillations (Axmacher, Mormann, Fernandez, Elger, & Fell, 2006). On the contrary, at retrieval a strong connection needs to be built between CA3 and CA1 for the hippocampus to output pattern completed activity to other regions, something proposed to be achieved by slow gamma (Colgin et al., 2009). Chapters 2 and 3 leveraged the idea of the different phases of the theta oscillation determining the state of the hippocampal system, more so than the idea of slow and fast gamma. However, these are by no means mutually exclusive. Gamma oscillations occur simultaneously with theta oscillations (Buzsaki et al., 1983), and have been proposed to play the main

role in theta phase precession, described before (Jensen & Lisman, 1996). As described before, the state that the hippocampus is in might be dependent on the gamma oscillation that is nested in the theta phase (Colgin et al., 2009). Thus, encoding and retrieval state might be determined by the phase of theta, but also slow and fast gamma oscillations potentially nested in the different phases of theta. An outstanding question that follows is whether the method used in Chapters 2-4 is picking up information in gamma frequency (from hippocampus or neocortical areas), or whether another frequency in neocortical regions might be carrying the important information which is coupled to the hippocampal theta phase.

The involvement of neocortical gamma oscillations and the relationship to theta phase in memory has been shown in a few studies investigating episodic memory. In one study, using scalp EEG it could be shown that the processing of subsequently remembered items was related to stronger posterior gamma power to frontal cortex theta phase than subsequently forgotten items (Friese et al., 2013). In a direct follow-up study, a similar theta/gamma phase-amplitude coupling was found also for episodic memory retrieval (Köster et al., 2014). In another study, ECoG data was recorded and showed strong gamma-power-to-theta-phase coupling in cortical areas during different behavioural tasks. Similar posterior fast oscillation to the frontal theta phase has been shown in the working memory literature (e.g., Berger, et al., 2019). Although three studies provide evidence for a role of theta/gamma phase-amplitude coupling, the relative scarcity of studies in neocortical areas still makes it difficult to conclude as to what exact role this phenomenon has in human episodic memory.

Other frequencies that have shown important roles in episodic memory is alpha (8-12Hz) and beta (12-30Hz). The Synch/deSynch neural network model proposes that cells in neocortex need to break out of the alpha/beta oscillation to represent a stimulus, resulting in desynchronization of alpha/beta, at the same time as cells in hippocampus need to synchronise in theta frequency (Parish et al., 2017). Evidence for this model was recently shown in an iEEG study, in which participants completed an associative memory task (Griffiths et al., 2019). Results showed that during encoding, alpha/beta desynchronization in neocortex preceded theta/gamma synchronisation in the hippocampus. Intriguingly, at retrieval, the reverse pattern was true, where theta/gamma synchronisation preceded alpha/beta desynchronization. In a continuation of this study, MEG was recorded whilst participants took part in a task in which they needed to bind temporally separated information together (Griffiths et al., 2020). The results in this study showed that the representation of individual items was related to an alpha/beta desynchronization at encoding and retrieval, and the binding between the items increased theta/gamma phase-amplitude coupling in the hippocampus. The authors concluded from these findings that alpha/beta desynchronization in the neocortex is representing stimuli, whereas the hippocampal theta/gamma phase-amplitude coupling is crucial for the binding between events.

In light of the above-mentioned studies, both alpha/beta as well as gamma power show properties of representing stimuli. This creates a complex picture as to which frequency a researcher should look for representations of memory, and which frequency that carries most information about that specific memory. The algorithm used in Chapter 2-4 is agnostic to the frequency at which the dissimilarity between

mental contents was found. This is also one of the main reasons as to why the method is powerful; it can reveal more details of mental content as classifiers trained on a wide frequency range are agnostic about where the diagnostic information is coming from. This is important given the above-mentioned conundrum. The classifier-based method is not isolated to any frequency range, and gamma as well as alpha/beta activity should be possible to pick up if the diagnostic information lies in those frequencies. One obvious drawback of the classifier-based procedure is that researchers are completely blind as to where the diagnostic information is coming from. In future studies, it would be interesting to look more closely into the agreement and discrepancies using the different methods.

In sum, the results in Chapters 2, 3, and 4 seem to suggest that the classifier-to-theta coupling method used throughout this thesis is conceptually similar to gamma-theta methods used in other studies. As no studies today have investigated the role of theta/alpha/beta phase-amplitude coupling in episodic memory, it is difficult to say whether the method is also equivalent to this. Notwithstanding, the method used in Chapters 2 to 4 has the advantage that it is not limited to information that can be decoded from one or the other frequency band, which could potentially mean that the information about mental content is coded in alpha, beta or gamma power. It would be helpful if future studies could explicitly compare these different methods within one study, to determine whether they arrive at the same conclusions.

3. A proposed time course of episodic memory retrieval

Based on the results in the present doctoral thesis, an integrative proposal of a possible retrieval time-course is here put forward.

In the context of this thesis, a method was developed that allows us to detect rhythmicity in the recall signatures that arguably represent reactivated content. Across chapters, this type of analysis consistently demonstrated that memory retrieval is indeed is a fluctuating process. The theta rhythmicity of the memory-related neural patterns in itself is a novel, and one of the most central insights, that can be derived from the work in this thesis regarding the time course of memory recall. How is this fluctuation, presumably in neocortical patterns, then linked to the hippocampal theta rhythm? The thesis also offers some clues as to this question. In Chapter 2, we were interested whether we could observe a signature of the process that sets of the neocortical reinstatement. We ,therefore, realigned the raw data based on the time of the fidelity value peaks, and could thereby show a strong phase-locking effect taking place approximately 200-250ms before the maximal evidence for memory reinstatement. This theta phase-locked signal was further shown to be related to the hippocampus using source analysis, in line with the hippocampus as a region initiating the retrieval process, before signals are sent to other regions, in which the content is activated 200-250ms later. In line with these results, several recent studies have found a similar temporal delay between the hippocampus and neocortical regions (Griffiths et al., 2019; Rutishauser et al., 2015; Staresina & Wimber, 2019). In one study (Griffiths et al., 2019), neocortical activity preceded the hippocampal activity at encoding, but at test, this direction flipped, with a delay between the regions being approximately 250ms. Thus, accumulating evidence is supporting the long-standing

idea of an interaction between the hippocampus and neocortex to reinstate patterns belonging to previously encoded memories (Marr, 1971; McClelland et al., 1995; Norman & O'Reilly, 2003). It is interesting to note that the time delay between the phase-locking effect in hippocampus and the reinstatement in neocortical areas corresponds roughly to one theta cycle at 4 Hz, or two cycles at 8 Hz, possibly suggesting that a cue-related phase reset in the hippocampus can be followed by detectable read-out from the hippocampus in neocortex within 1-2 theta cycles. The exact frequency has been discussed before, and it suffices to here settle with the range being within what is normally considered theta. Together with the fluctuating reinstatement described in Chapters 2, 3, and 4, one can then imagine a gradual build-up of the memory trace across theta cycles, increasing at each time point that is considered optimal for transferring information from hippocampus to neocortex (Raghavachari et al., 2001). This means that each time the theta cycle reaches that optimal phase, the hippocampus can send information relevant for the reinstatement of that particular memory. Such a mechanism allows for an organised output from the hippocampus such that the retrieval process does not interfere with sensory processing, which is thought to be optimal at the opposite phase (Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007). This transfer of information needs a precise and coherent temporal order, something that can be achieved by the means of SPW-Rs, which have been implicated in triggering the transfer of material between hippocampus and neocortex, especially during consolidation and sleep (Diekelmann & Born, 2010). More specifically, it has until recently been unknown whether SPW-Rs only reflect a subconscious mechanism, or whether they also can trigger conscious recollection. In Chapter 5, it is shown that hippocampal SPW-Rs events surround neocortical memory

reinstatement, consequently putting ripples at the centre of importance in transferring compressed information to neocortex also during awake memory retrieval. What is currently unknown is whether SPW-Rs are phase-locked to the same theta phase as we have shown evidence for in Chapter 2 to 4, and will be further discussed under the *Future Studies* section.

The 200-250ms delay between the hippocampus and neocortical areas should not be confused with the actual time after cue onset when the memory is being reinstated in neocortical areas. In the neocortex, evidence for the reactivation of a particular memory trace is being added together, and once a threshold is passed, a conscious reinstatement takes place in which the subject can retrieve the encoded memory together with important contextual details (Yonelinas, 2002). This neocortical recollection process is assumed to occur around 500ms after the cue has been presented (Staresina & Wimber, 2019), and can, depending on the task and experimental settings, continue until the time point at which the subject makes a decision.

With the results from the present thesis, a timeline of memory retrieval can be proposed. Approximately 200ms after a visual cue has been presented, it reaches the hippocampus (Staresina & Wimber, 2019). At around 250ms after cue onset, an optimal phase for retrieval is being set, in which phase the index of the target memory is reactivated. The reactivation of the target memory index continues until enough "parcels" of information have been sent from the hippocampus to the neocortex. The data in the present thesis suggest that each parcel that can be detected in the neocortex is preceded by a consistent phase in the hippocampus, suggesting that the hippocampus

keeps clocking the retrieval process beyond the initial phase reset. In the neocortex, evidence is being accumulated until the subject experiences a conscious recollection at around 1000-1500ms and a mnemonic decision can be reached.

4. Future research

The three first projects presented in this thesis have furthered our understanding of the different functional roles of the theta rhythm in memory retrieval, with a specific focus on the role of phase coding as a mechanism of separating temporally overlapping cognitive operations (encoding/retrieval, Chapters 2 and 3) and overlapping mnemonic representations (Chapter 4). In addition to this, the fourth project (Chapter 5) suggests a new intriguing role of hippocampal ripples in coordinating memory reinstatement during active wake retrieval.

In the same way that some questions have been answered and light has been shone on interesting cognitive processes in memory retrieval, it has equally stimulated new questions.

4.1. How many competing items can be assigned a slot along the theta oscillation?

Empirical evidence in Chapter 4 suggests that the hippocampal theta rhythm can temporally separate reinstatement of two items along the oscillation, items that are linked by a common reminder. In each cycle, patterns for two memories are being reinstated, but at different phases. This mechanism is supposed to resolve competition and has been described in neural network models (Norman et al., 2007). One interesting question that follows from the findings in the working-memory (Bahramisharif et al., 2018) field and the studies in the present thesis is whether it would be possible for the hippocampus to separate more than two memories along the oscillation. For example, if subjects were simply told to encode four subsequent events, the assumption would be, based on the phase-coding model (Jensen & Lisman, 2005), that the relative order would be kept for upcoming retrieval attempts (Heusser et al., 2016). However, if you instead let subjects encode four events, but later at retrieval instructed them to only retrieve one of them, the hypothesis would be, based on the neural network model by Norman et al. (2007), that the memory that they are instructed to retrieve would, across retrieval attempts, be reinstated at the optimal phase of retrieval. The three competing memories would either be chunked and reactivated together at a competitor phase or kept at a similar distance as during encoding. The results of such a potential study could provide new knowledge about on the one hand whether all irrelevant, competing memories would be chunked together, or remain phase-separated, and on the other, whether the reactivation order during

retrieval would rather reflect sequences from learning or relevance for the retrieval goals.

4.2. Can the post-encoding period act as mini-consolidation?

In Chapter 5, the density of hippocampal ripples was greater for source correct compared to source incorrect trials in an active retrieval task. Given that this is one of the first studies investigating hippocampal ripples in human episodic memory, several questions remain unanswered. First, is it possible that the "offline" period after an encoding block is similarly involved in the rapid transferring of information by the means of SPW-Rs (Buzsaki, 2015). Secondly, if this is the case, is the transferring related to a phase of a slow hippocampal oscillation? That is, does the post-encoding phase act as a "mini-consolidation" period, with the replay taking place at specific time points? Partial evidence for this view can be found in the literature. In studies investigating consolidation, a nap after a learning session improves memory performance (Lahl, Wispel, Willigens, & Pietrowsky, 2008; Rasch, Buchel, Gais, & Born, 2007). Furthermore, stored representations are being reactivated during awake replay in rats (Karlsson & Frank, 2009), and evidence for this replay has been found in human subjects, too (Tambini, Ketz, & Davachi, 2010; Zhang, Fell, & Axmacher, 2018). In the latter study, patients had an afternoon nap post-encoding, which contained both sleep and awake periods. The authors extracted stimulus-specific gamma activity (30-90Hz) and used representational-similarity analysis (RSA) to investigate whether learning-related activity patterns re-emerged during post-encoding awake and nap periods. They found that both early and late encoding representations

were reinstated during these post-encoding stages. However, spontaneous replays during awake post-encoding did not predict later memory performance. In another study, it was investigated how prior knowledge can affect post-encoding activity (Liu, Grady, & Moscovitch, 2018). The authors found evidence for a continuous interplay between the hippocampus, vmPFC, and anterior temporal pole during post-encoding, at the same time as posterior perceptual regions rested. They interpreted this as internal, off-line processing of recently encoded material. In Chapter 5, the period between encoding and retrieval was a minimum of one minute long, in which patients could interact with the experiment leader. In this time window, the hippocampus might take the opportunity to consolidate memories, as the brain is not occupied by memory encoding. This means that whenever there is a window in which the subject can focus attention internally, the hippocampus might this time to replay memories and thereby facilitate their long-term storage (Mednick, Cai, Shuman, Anagnostaras, & Wixted, 2011). To investigate whether patterns from encoding re-emerge in this short time window post-encoding, the same method as in Study 4 can be used, in which realignment of the data is based on ripple events post-encoding. A contrast between encoding blocks of high subsequent memory effects can then be contrasted with encoding blocks of lower subsequent memory effects with the assumption being that it is easier to decode brain patterns for one or the other class in the high memory blocks. If stronger evidence for previously encoded events is found in the high memory group, this would potentially mean that during this short time windows directly following encoding, the brain goes off-line and consolidates what has recently been encoded. Secondly, to investigate whether ripple events are related to a specific phase of a slow oscillation post-encoding, a similar method as in this thesis can again be used. The

maximum amplitude of the ripple can be coupled to a slow oscillation. If the ripples are locked to a certain phase of the oscillation, there is an intriguing possibility that partly of what is observed in Chapter 2-4 reflects a theta to ripple phase-amplitude coupling. In the next paragraph, this relationship is further investigated.

4.3. Are SPW-R locked to the phase of the theta oscillation?

Throughout this thesis, evidence has been provided suggesting that the hippocampal theta oscillation is organising memory retrieval. Chapters 2, 3 and 4 consistently demonstrate that memory reinstatement is clocked by a theta oscillation likely originating in the hippocampus, while Chapter 5 shows that reinstatement in neocortical areas temporally surrounds SPW-Rs events originating from the hippocampus. An important missing link is the timing of ripple events relative to a slow oscillation. More specifically, it is currently unknown whether hippocampal SPW-R are locked to the hippocampal theta oscillation similarly as fidelity values and gamma power. As described before, SPW-Rs are thought to promote the transfer of information from the hippocampus to neocortex during consolidation and active recall. Similarly, theta oscillations have been shown to facilitate the transfer of information between brain regions that are synchronized in this rhythm, with gated theta opening up for time points optimal for information transfer (Raghavachari et al., 2001). In Chapter 2 to 4, evidence for gated theta is shown in that there is an optimal time point for retrieval of target memories when the dynamics in hippocampus promote LTD rather than LTP (Hasselmo et al., 2002; Kunec et al., 2005; Manns et al., 2007). Based on this evidence, one can hypothesise that theta is providing the distinct time windows

at which optimal transfer of information by the means of SPW-Rs to other regions can be achieved. Furthermore, encoding an event should preferably lock ripples to one phase of theta, whereas retrieval of an already encoded event should lock ripples to another phase.

Evidence that hippocampal spiking is locked to theta has been shown both in models and in empirical studies investigating replay, with the idea being that the replay of spike sequences from encoding supports consolidation (Carr, Jadhav, & Frank, 2011; Drieu, Todorova, & Zugaro, 2018; Nicola & Clopath, 2018). Recently, a study showed that ripples underlie bursts of spiking-activity in human episodic memory, with each ripple being accompanied by an increase of single-unit bursting activity (Vaz et al., 2020), pointing to a potential similar mechanism also in episodic memory retrieval. In three previous studies (rodents: Logothetis et al. (2012); human subjects: Axmacher et al. (2008); Staresina et al. (2015)), ripples are phase-locked to slow oscillations, but not theta. In one, ripples were phase-locked to hippocampal delta activity. Axmacher et al. (2008) interpreted this locking as a possible mechanism for how the hippocampus and neocortex are communicating for transferring of information. However, the phase along the delta oscillation with which the ripples were coupled did not predict the transmission of ripples between the hippocampus and cortex. In the second study, ripples were found to be clustered in spindles (12-16Hz), which in turn were clustered in the hippocampal slow-oscillations (~.75Hz) (Staresina et al., 2015). This hierarchical clustering of three different oscillations was found to promote information transfer between hippocampus and neocortical areas. Furthermore, closed-loop experiments have shown that such coupling between SPW-R and slow waves

influences consolidation (Maingret, Girardeau, Todorova, Goutierre, & Zugaro, 2016), which consequently could result in neocortical oscillations going into an optimal state in which neocortical areas can listen to hippocampus and promote the communication between neocortex and hippocampus via SPW-Rs. Taken together, previous studies investigating ripples' relationship to other oscillations have found evidence for ripples being nested in spindles and delta activity. However, no previous study has shown a relationship between ripple events and the hippocampal theta.

One potential caveat in this research question is that SPW-Rs are suppressed during extrinsic behaviour (Vandecasteele et al., 2014), at the same time as theta is being upregulated. This means that during time points when subjects engage in encoding activity, SPW-Rs are not occurring in the same way as during intrinsic behaviour such as retrieval activity, and it could, therefore, be difficult to assess the question of whether SPW-Rs are differently phase-modulated by the hippocampal oscillation at encoding and retrieval. Notwithstanding, it would still be interesting to investigate the relationship between these two neural substrates during retrieval, to better understand how and when information is being transferred from the hippocampus to neocortical regions.

In sum, ripples are coupled to slow oscillations, but not theta, during replay and consolidation, but none of the above-mentioned studies could show a direct link between transferring of information and the ripple-to-slow-oscillation coupling during episodic memory retrieval.

5. Concluding remarks

During the past decades, the understanding of the brain circuits involved in episodic memory has grown enormously. In the late 1950s, at the same time as Scoville and Miller highlighted the pivotal role of the hippocampus in episodic memory, Karl Lashley was searching for the engram. It is truly amazing how fast it has developed since. The most recent advances have been fuelled by the development of novel analysis methods, including multivariate pattern analysis, which has enabled researchers to isolate memory-related patterns of brain activity that used to be "hidden" in imaging and electrophysiological experiments in humans. In the present thesis, these methods were thoroughly used and taken advantage of. This led to new important findings regarding, on the one hand, the role of phase-coding along the hippocampal theta rhythm in episodic memory, and the relationship between neocortical memory reinstatement and hippocampal ripple events on the other. The results indicate that the neural signatures of encoding and retrieval are fluctuating processes that both follow a slow hippocampal oscillation. Furthermore, at the time of retrieval, the hippocampal theta oscillation adaptively separates relevant from irrelevant memories. Lastly, reinstatement in neocortical areas temporally surrounds hippocampal ripple events. Together, these findings bridge gaps between rodent literature, computational models, and studies in human subjects. However, they also motivate several new and exciting questions, and hopefully, the work in this doctoral thesis will stimulate further investigation of these two hippocampal signatures in human episodic memory.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A Learning Algorithm for Boltzmann Machines*. *Cogn Sci, 9*(1), 147-169. doi:10.1207/s15516709cog0901_7

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proc Natl Acad Sci U S A, 91*(15), 7041-7045.

Amaral, D. G., & Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience, 31*(3), 571-591. doi:10.1016/0306-4522(89)90424-7

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends Cogn Sci, 17*(8), 379-390. doi:10.1016/j.tics.2013.06.009

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*(4), 415-445. doi:https://doi.org/10.1016/j.jml.2003.08.006

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *J Exp Psychol Learn Mem Cogn, 20*(5), 1063-1087.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 608-629. doi:10.1037/0278-7393.25.3.608

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends Cogn Sci, 21*(8), 573-576. doi:10.1016/j.tics.2017.05.001

Aru, J., Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., . . . Vicente, R. (2015). Untangling cross-frequency coupling in neuroscience. *Curr Opin Neurobiol, 31*, 51-61. doi:10.1016/j.conb.2014.08.002

Axmacher, N., Elger, C. E., & Fell, J. (2008). Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain, 131*(Pt 7), 1806-1817. doi:10.1093/brain/awn103

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proc Natl Acad Sci U S A, 107*(7), 3228-3233. doi:10.1073/pnas.0911531107

Axmacher, N., Mormann, F., Fernandez, G., Elger, C. E., & Fell, J. (2006). Memory formation by neuronal synchronization. *Brain Res Rev, 52*(1), 170-182. doi:10.1016/j.brainresrev.2006.01.007

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In *Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 28-61). New York, NY, US: Cambridge University Press.

Bahramisharif, A., Jensen, O., Jacobs, J., & Lisman, J. (2018). Serial representation of items during working memory maintenance at letter-selective cortical sites. *PLoS Biol, 16*(8), e2003805. doi:10.1371/journal.pbio.2003805

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science, 319*(5870), 1640-1642. doi:10.1126/science.1152882

Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W., & Lin, L. H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Prog Brain Res, 83*, 287-300. doi:10.1016/s0079-6123(08)61257-1

Bartels, A., & Zeki, S. (2000). The neural basis of romantic love. *Neuroreport, 11*(17), 3829-3834. doi:10.1097/00001756-200011270-00046

Berg, R. W., & Kleinfeld, D. (2003). Rhythmic whisking by rat: retraction as well as protraction of the vibrissae is under active muscular control. *J Neurophysiol, 89*(1), 104-117. doi:10.1152/jn.00600.2002

Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience, 18*(24), 10464-10472.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Muller, K. R. (2011). Single-trial analysis and classification of ERP components--a tutorial. *Neuroimage, 56*(2), 814-825. doi:10.1016/j.neuroimage.2010.06.048

Bohbot, V. D., Copara, M. S., Gotman, J., & Ekstrom, A. D. (2017). Low-frequency theta oscillations in the human hippocampus during real-world and virtual navigation. *Nat Commun, 8*, 14415. doi:10.1038/ncomms14415

Bonnefond, M., & Jensen, O. (2012). Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Curr Biol, 22*(20), 1969-1974. doi:10.1016/j.cub.2012.08.029

Bragin, A., Engel Jr, J., Wilson, C. L., Fried, I., & Buzsáki, G. (1999). High-frequency oscillations in human brain. *Hippocampus, 9*(2), 137-142. doi:10.1002/(sici)1098-1063(1999)9:2<137::Aid-hipo5>3.0.Co;2-0

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One, 5*(5), e10773. doi:10.1371/journal.pone.0010773

Burgess, N., & O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus, 6*(6), 749-762. doi:10.1002/(sici)1098-1063(1996)6:6<749::Aid-hipo16>3.0.Co;2-0

Buzsaki, G. (1996). The hippocampo-neocortical dialogue. *Cereb Cortex, 6*(2), 81-92. doi:10.1093/cercor/6.2.81

Buzsaki, G. (2002). Theta oscillations in the hippocampus. *Neuron, 33*(3), 325-340.

Buzsaki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus, 25*(10), 1073-1188. doi:10.1002/hipo.22488

Buzsaki, G., Leung, L. W., & Vanderwolf, C. H. (1983). Cellular bases of hippocampal EEG in the behaving rat. *Brain Res, 287*(2), 139-171. doi:10.1016/0165-0173(83)90037-1

Buzsaki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat Neurosci, 16*(2), 130-138. doi:10.1038/nn.3304

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., . . . Knight, R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science, 313*(5793), 1626-1628. doi:10.1126/science.1128115

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. *J Vis, 13*(10). doi:10.1167/13.10.1

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. *J Cogn Neurosci, 26*(1), 132-142. doi:10.1162/jocn_a_00476

Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nat Neurosci, 14*(2), 147-153. doi:10.1038/nn.2732

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn Sci, 18*(8), 414-421. doi:10.1016/j.tics.2014.04.012

Chrobak, J. J., Lorincz, A., & Buzsaki, G. (2000). Physiological patterns in the hippocampo-entorhinal cortex system. *Hippocampus, 10*(4), 457-465. doi:10.1002/1098-1063(2000)10:4<457::AID-HIPO12>3.0.CO;2-Z

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

Cohen, M. X. (2014). *Analyzing Neural Times Series Data: Theory and Practice.*: MIT Press.

Cohen, M. X. (2014). Fluctuations in oscillation frequency control spike timing and coordinate neural networks. *J Neurosci, 34*(27), 8988-8998. doi:10.1523/JNEUROSCI.0261-14.2014

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. doi:10.1126/science.7414331

Colgin, L. L. (2016). Rhythms of the hippocampal network. *Nat Rev Neurosci, 17*(4), 239-249. doi:10.1038/nrn.2016.21

Colgin, L. L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., . . . Moser, E. I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature, 462*(7271), 353-357. doi:10.1038/nature08573

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 240-247. doi:https://doi.org/10.1016/S0022-5371(69)80069-1

Csicsvari, J., Hirase, H., Mamiya, A., & Buzsaki, G. (2000). Ensemble patterns of hippocampal CA3-CA1 neurons during sharp wave-associated population events. *Neuron, 28*(2), 585-594. doi:10.1016/s0896-6273(00)00135-5

Dalton, M. A., Zeidman, P., McCormick, C., & Maguire, E. A. (2018). Differentiable Processing of Objects, Associations, and Scenes within the Hippocampus. *J Neurosci, 38*(38), 8146-8159. doi:10.1523/JNEUROSCI.0263-18.2018

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychol Bull, 136*(1), 87-102. doi:10.1037/a0017937

DeVito, L. M., Lykken, C., Kanter, B. R., & Eichenbaum, H. (2010). Prefrontal cortex: role in acquisition of overlapping associations and transitive inference. *Learn Mem, 17*(3), 161-167. doi:10.1101/lm.1685710

Diba, K., & Buzsaki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nat Neurosci, 10*(10), 1241-1242. doi:10.1038/nn1961

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nat Rev Neurosci, 11*(2), 114-126. doi:10.1038/nrn2762

Dijkstra, N., Ambrogioni, L., & van Gerven, M. A. J. (2019). Neural dynamics of perceptual inference and its reversal during imagery. *bioRxiv*, 781294. doi:10.1101/781294

Douchamps, V., Jeewajee, A., Blundell, P., Burgess, N., & Lever, C. (2013). Evidence for encoding versus retrieval scheduling in the hippocampus by theta phase and acetylcholine. *J Neurosci, 33*(20), 8689-8704. doi:10.1523/JNEUROSCI.4483-12.2013

Drieu, C., Todorova, R., & Zugaro, M. (2018). Nested sequences of hippocampal assemblies during behavior support subsequent sleep replay. *Science, 362*(6415), 675-679. doi:10.1126/science.aat2952

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

Duvernoy, H., Cattin, F., & Risold, P. (2007). *The Human Hippocampus. Functional Anatomy, Vascularization and Serial Sections with MRI.* (B. Springer Ed. 4th ed.): Springer.

Eichenbaum, H. (2010). Memory systems. *WIREs Cognitive Science, 1*(4), 478-490. doi:10.1002/wcs.49

Eichenbaum, H., & Buckingham, J. (1990). Studies on hippocampal processing: Experiment, theory, and model. In *Learning and computational neuroscience: Foundations of adaptive networks.* (pp. 171-231). Cambridge, MA, US: The MIT Press.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci, 12*(10), 388-396. doi:10.1016/j.tics.2008.07.004

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two kinds of FMRI repetition suppression? Evidence for dissociable neural mechanisms. *J Neurophysiol, 99*(6), 2877-2886. doi:10.1152/jn.90376.2008

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

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nat Rev Neurosci, 12*(2), 105-118. doi:10.1038/nrn2979

Fernandez-Ruiz, A., Oliva, A., Fermino de Oliveira, E., Rocha-Almeida, F., Tingley, D., & Buzsaki, G. (2019). Long-duration hippocampal sharp wave ripples improve memory. *Science, 364*(6445), 1082-1086. doi:10.1126/science.aax0758

Ferreira, C. S., Marful, A., Staudigl, T., Bajo, T., & Hanslmayr, S. (2014). Medial prefrontal theta oscillations track the time course of interference during selective memory retrieval. *J Cogn Neurosci, 26*(4), 777-791. doi:10.1162/jocn_a_00523

Frankland, P. W., Josselyn, S. A., & Kohler, S. (2019). The neurobiological foundation of memory retrieval. *Nat Neurosci, 22*(10), 1576-1585. doi:10.1038/s41593-019-0493-1

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci, 9*(10), 474-480. doi:10.1016/j.tics.2005.08.011

Friese, U., Koster, M., Hassler, U., Martens, U., Trujillo-Barreto, N., & Gruber, T. (2013). Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. *Neuroimage, 66*, 642-647. doi:10.1016/j.neuroimage.2012.11.002

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Duzel, E. (2010). Theta-coupled periodic replay in working memory. *Curr Biol, 20*(7), 606-612. doi:10.1016/j.cub.2010.01.057

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsaki, G., & Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nat Neurosci, 12*(10), 1222-1223. doi:10.1038/nn.2384

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci, 15*(1), 20-25. doi:10.1016/0166-2236(92)90344-8

Goyal, A., Miller, J., Qasim, S. E., Watrous, A. J., Zhang, H., Stein, J. M., . . . Jacobs, J. (2020). Functionally distinct high and low theta oscillations in the human hippocampus. *Nature Communications, 11*(1), 2469. doi:10.1038/s41467-020-15670-6

Grande, X., Berron, D., Horner, A. J., Bisby, J. A., Duzel, E., & Burgess, N. (2019). Holistic Recollection via Pattern Completion Involves Hippocampal Subfield CA3. *J Neurosci, 39*(41), 8100-8111. doi:10.1523/JNEUROSCI.0722-19.2019

Griffin, A. L., Eichenbaum, H., & Hasselmo, M. E. (2007). Spatial representations of hippocampal CA1 neurons are modulated by behavioral context in a hippocampus-dependent memory task. *J Neurosci, 27*(9), 2416-2423. doi:10.1523/JNEUROSCI.4083-06.2007

Griffiths, B. J., Martín-Buro, M. C., Staresina, B. P., & Hanslmayr, S. (2020). Disentangling the roles of neocortical alpha/beta and hippocampal theta/gamma activity in human episodic memory. *bioRxiv*, 2020.2001.2022.915330. doi:10.1101/2020.01.22.915330

Griffiths, B. J., Parish, G., Roux, F., Michelmann, S., van der Plas, M., Kolibius, L. D., . . . Hanslmayr, S. (2019). Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. *bioRxiv*, 305698. doi:10.1101/305698

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J Cogn Neurosci, 29*(4), 677-697. doi:10.1162/jocn_a_01068

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., . . . Schoffelen, J. M. (2013). Good practice for conducting and reporting MEG research. *Neuroimage, 65*, 349-363. doi:10.1016/j.neuroimage.2012.10.001

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proc Natl Acad Sci U S A, 98*(2), 694-699. doi:10.1073/pnas.98.2.694

Gruber, W. R., Klimesch, W., Sauseng, P., & Doppelmayr, M. (2005). Alpha phase synchronization predicts P1 and N1 latency and amplitude size. *Cereb Cortex, 15*(4), 371-377. doi:10.1093/cercor/bhh139

Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron, 44*(4), 581-584. doi:10.1016/j.neuron.2004.11.003

Hanert, A., Weber, F. D., Pedersen, A., Born, J., & Bartsch, T. (2017). Sleep in Humans Stabilizes Pattern Separation Performance. *J Neurosci, 37*(50), 12238-12246. doi:10.1523/JNEUROSCI.1189-17.2017

Hangya, B., Borhegyi, Z., Szilagyi, N., Freund, T. F., & Varga, V. (2009). GABAergic neurons of the medial septum lead the hippocampal network during theta activity. *J Neurosci, 29*(25), 8094-8102. doi:10.1523/jneurosci.5665-08.2009

Hanslmayr, S., Staudigl, T., Aslan, A., & Bauml, K. H. (2010). Theta oscillations predict the detrimental effects of memory retrieval. *Cogn Affect Behav Neurosci, 10*(3), 329-338. doi:10.3758/CABN.10.3.329

Hargreaves, E. L., Rao, G., Lee, I., & Knierim, J. J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science, 308*(5729), 1792-1794. doi:10.1126/science.1110449

Hasselmo, M. E. (2005). What is the function of hippocampal theta rhythm?--Linking behavioral data to phasic properties of field potential and unit recording data. *Hippocampus, 15*(7), 936-949. doi:10.1002/hipo.20116

Hasselmo, M. E., Bodelon, C., & Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance

reversal of prior learning. *Neural Comput, 14*(4), 793-817. doi:10.1162/089976602317318965

Hasselmo, M. E., & Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Netw, 18*(9), 1172-1190. doi:10.1016/j.neunet.2005.08.007

Hasselmo, M. E., & Fehlau, B. P. (2001). Differences in time course of ACh and GABA modulation of excitatory synaptic potentials in slices of rat hippocampus. *J Neurophysiol, 86*(4), 1792-1802. doi:10.1152/jn.2001.86.4.1792

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. Oxford, England: Wiley.

Hebscher, M., Meltzer, J. A., & Gilboa, A. (2019). A causal role for the precuneus in network-wide theta and gamma oscillatory activity during complex memory retrieval. *Elife, 8*, e43114. doi:10.7554/eLife.43114

Heusser, A. C., Poeppel, D., Ezzyat, Y., & Davachi, L. (2016). Episodic sequence memory is supported by a theta-gamma phase code. *Nat Neurosci, 19*(10), 1374-1380. doi:10.1038/nn.4374

Horner, A. J., Bisby, J. A., Bush, D., Lin, W. J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat Commun, 6*, 7462. doi:10.1038/ncomms8462

Howard, M. W., & Eichenbaum, H. (2013). The hippocampus, time, and memory across scales. *J Exp Psychol Gen, 142*(4), 1211-1230. doi:10.1037/a0033621

Huerta, P. T., & Lisman, J. E. (1993). Heightened synaptic plasticity of hippocampal CA1 neurons during a cholinergically induced rhythmic state. *Nature, 364*(6439), 723-725. doi:10.1038/364723a0

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

Hyman, J. M., Hasselmo, M. E., & Seamans, J. K. (2011). What is the Functional Relevance of Prefrontal Cortex Entrainment to Hippocampal Theta Rhythms? *Front Neurosci, 5*, 24. doi:10.3389/fnins.2011.00024

Hyman, J. M., Wyble, B. P., Goyal, V., Rossi, C. A., & Hasselmo, M. E. (2003). Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to the peak of theta and long-term depression when delivered to the trough. *J Neurosci, 23*(37), 11725-11731.

Jacobs, J. (2014). Hippocampal theta oscillations are slower in humans than in rodents: implications for models of spatial navigation and memory. *Philos Trans R Soc Lond B Biol Sci, 369*(1635), 20130304. doi:10.1098/rstb.2013.0304

Jafarpour, A., Fuentemilla, L., Horner, A. J., Penny, W., & Duzel, E. (2014). Replay of Very Early Encoding Representations during Recollection. *The Journal of Neuroscience, 34*(1), 242-248. doi:10.1523/jneurosci.1865-13.2014

Jensen, O. (2006). Maintenance of multiple working memory items by temporal segmentation. *Neuroscience, 139*(1), 237-249. doi:10.1016/j.neuroscience.2005.06.004

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn Sci, 16*(4), 200-206. doi:10.1016/j.tics.2012.03.002

Jensen, O., & Lisman, J. E. (1996). Hippocampal CA3 region predicts memory sequences: Accounting for the phase precession of place cells. *Learning & Memory, 3*(2-3), 279-287. doi:10.1101/lm.3.2-3.279

Jensen, O., & Lisman, J. E. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends Neurosci, 28*(2), 67-72. doi:10.1016/j.tins.2004.12.001

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

Johnson, J. D., Price, M. H., & Leiker, E. K. (2015). Episodic retrieval involves early and sustained effects of reactivating information from encoding. *Neuroimage, 106*, 300-310. doi:10.1016/j.neuroimage.2014.11.013

Joo, H. R., & Frank, L. M. (2018). The hippocampal sharp wave–ripple in memory retrieval for immediate use and consolidation. *Nature Reviews Neuroscience, 19*(12), 744-757. doi:10.1038/s41583-018-0077-1

Jutras, M. J., Fries, P., & Buffalo, E. (2009). Gamma-band Synchronization in the Macaque Hippocampus and Memory Formation. *Journal of Neuroscience Nursing, 29*(40), 12521-12531. doi:doi:10.1523/JNEUROSCI.0640-09.2009

Kahana, M. J. (2012). *Foundations of human memory*. New York, NY, US: Oxford University Press.

Kaplan, R., Bush, D., Bonnefond, M., Bandettini, P. A., Barnes, G. R., Doeller, C. F., & Burgess, N. (2014). Medial prefrontal theta phase coupling during spatial memory retrieval. *Hippocampus, 24*(6), 656-665. doi:10.1002/hipo.22255

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

Keresztes, A., & Racsmany, M. (2013). Interference resolution in retrieval-induced forgetting: behavioral evidence for a nonmonotonic relationship between interference and forgetting. *Mem Cognit, 41*(4), 511-518. doi:10.3758/s13421-012-0276-3

Kerren, C., Linde-Domingo, J., Hanslmayr, S., & Wimber, M. (2018). An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval. *Curr Biol, 28*(21), 3383-3392 e3386. doi:10.1016/j.cub.2018.08.065

Ketz, N., Morkonda, S. G., & O'Reilly, R. C. (2013). Theta Coordinated Error-Driven Learning in the Hippocampus. *PLOS Computational Biology, 9*(6), e1003067. doi:10.1371/journal.pcbi.1003067

Khodagholy, D., Gelinas, J. N., & Buzsáki, G. (2017). Learning-enhanced coupling between ripple oscillations in association cortices and hippocampus. *Science, 358*(6361), 369-372. doi:10.1126/science.aan6203

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn Sci, 18*(4), 203-210. doi:10.1016/j.tics.2014.01.002

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res Rev, 53*(1), 63-88. doi:10.1016/j.brainresrev.2006.06.003

Köster, M., Friese, U., Schöne, B., Trujillo-Barreto, N., & Gruber, T. (2014). Theta–gamma coupling during episodic retrieval in the human EEG. *Brain Research, 1577*, 57-68. doi:https://doi.org/10.1016/j.brainres.2014.06.028

Kristjansson, A., & Campana, G. (2010). Where perception meets memory: a review of repetition priming in visual search tasks. *Atten Percept Psychophys, 72*(1), 5-18. doi:10.3758/APP.72.1.5

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proc Natl Acad Sci U S A, 108*(14), 5903-5908. doi:10.1073/pnas.1016939108

Kunec, S., Hasselmo, M. E., & Kopell, N. (2005). Encoding and retrieval in the CA3 region of the hippocampus: a model of theta-phase separation. *J Neurophysiol, 94*(1), 70-82. doi:10.1152/jn.00731.2004

Kunz, L., Wang, L., Lachner-Piza, D., Zhang, H., Brandt, A., Dumpelmann, M., . . . Axmacher, N. (2019). Hippocampal theta phases organize the reactivation of large-scale electrophysiological representations during goal-directed navigation. *Sci Adv, 5*(7), eaav8192. doi:10.1126/sciadv.aav8192

Kurth-Nelson, Z., Barnes, G., Sejdinovic, D., Dolan, R., & Dayan, P. (2015). Temporal structure in associative retrieval. *Elife, 4*. doi:10.7554/eLife.04919

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *J Sleep Res, 17*(1), 3-10. doi:10.1111/j.1365-2869.2008.00622.x

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., . . . Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp, 10*(3), 120-131.

Lashley, K. S. (1950). In search of the engram. In *Physiological mechanisms in animal behavior. (Society's Symposium IV.).* (pp. 454-482). Oxford, England: Academic Press.

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron, 36*(6), 1183-1194. doi:10.1016/s0896-6273(02)01096-6

Lega, B., Burke, J., Jacobs, J., & Kahana, M. J. (2016). Slow-Theta-to-Gamma Phase-Amplitude Coupling in Human Hippocampus Supports the Formation of New Episodic Memories. *Cereb Cortex, 26*(1), 268-278. doi:10.1093/cercor/bhu232

Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus, 22*(4), 748-761. doi:10.1002/hipo.20937

Levy, W. B. (1989). A Computational Approach to Hippocampal Function. In R. D. Hawkins & G. H. Bower (Eds.), *Psychology of Learning and Motivation* (Vol. 23, pp. 243-305): Academic Press.

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications, 10*(1), 179. doi:10.1038/s41467-018-08080-2

Lisman, J. E., & Idiart, M. A. (1995). Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science, 267*(5203), 1512-1515. doi:10.1126/science.7878473

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron, 77*(6), 1002-1016. doi:10.1016/j.neuron.2013.03.007

Liu, Z. X., Grady, C., & Moscovitch, M. (2018). The effect of prior knowledge on post-encoding brain connectivity and its relation to subsequent memory. *Neuroimage, 167*, 211-223. doi:10.1016/j.neuroimage.2017.11.032

Logothetis, N. K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H. C., . . . Oeltermann, A. (2012). Hippocampal-cortical interaction during periods of subcortical silence. *Nature, 491*(7425), 547-553. doi:10.1038/nature11618

Lubenov, E. V., & Siapas, A. G. (2009). Hippocampal theta oscillations are travelling waves. *Nature, 459*(7246), 534-539. doi:10.1038/nature08010

Macrides, F., Eichenbaum, H., & Forbes, W. (1982). Temporal relationship between sniffing and the limbic theta rhythm during odor discrimination reversal learning. *The Journal of Neuroscience, 2*(12), 1705-1717. doi:10.1523/jneurosci.02-12-01705.1982

Maingret, N., Girardeau, G., Todorova, R., Goutierre, M., & Zugaro, M. (2016). Hippocampo-cortical coupling mediates memory consolidation during sleep. *Nat Neurosci, 19*(7), 959-964. doi:10.1038/nn.4304

Manns, J. R., Zilli, E. A., Ong, K. C., Hasselmo, M. E., & Eichenbaum, H. (2007). Hippocampal CA1 spiking during encoding and retrieval: relation to theta phase. *Neurobiol Learn Mem, 87*(1), 9-20. doi:10.1016/j.nlm.2006.05.007

Marr, D. (1971). Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci, 262*(841), 23-81.

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. *Psychol Rev, 102*(3), 419-457.

McCloskey, M., & Cohen, N. J. (1989) Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In*: Vol. 24. Psychology of Learning and Motivation - Advances in Research and Theory* (pp. 109-165).

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences, 10*(10), 408-415. doi:https://doi.org/10.1016/0166-2236(87)90011-7

Mednick, S. C., Cai, D. J., Shuman, T., Anagnostaras, S., & Wixted, J. T. (2011). An opportunistic theory of cellular and systems consolidation. *Trends in Neurosciences, 34*(10), 504-514. doi:10.1016/j.tins.2011.06.003

Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. *PLoS Biol, 14*(8), e1002528. doi:10.1371/journal.pbio.1002528

Miller, K. J., Sorensen, L. B., Ojemann, J. G., & den Nijs, M. (2009). Power-law scaling in the brain surface electric potential. *PLoS Comput Biol, 5*(12), e1000609. doi:10.1371/journal.pcbi.1000609

Molle, M., Bergmann, T. O., Marshall, L., & Born, J. (2011). Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing. *Sleep, 34*(10), 1411-1421. doi:10.5665/sleep.1290

Molle, M., Marshall, L., Gais, S., & Born, J. (2002). Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *J Neurosci, 22*(24), 10941-10947.

Mormann, F., Osterhage, H., Andrzejak, R. G., Weber, B., Fernandez, G., Fell, J., . . . Lehnertz, K. (2008). Independent delta/theta rhythms in the human hippocampus and entorhinal cortex. *Front Hum Neurosci, 2*, 3. doi:10.3389/neuro.09.003.2008

Newman, E. L., & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual representations. *Cereb Cortex, 20*(11), 2760-2770. doi:10.1093/cercor/bhq021

Nicola, W., & Clopath, C. (2018). The Dance of the Interneurons: How Inhibition Facilitates Fast Compressible and Reversible Learning in Hippocampus. *bioRxiv*, 318303. doi:10.1101/318303

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys Med Biol, 48*(22), 3637-3652. doi:10.1088/0031-9155/48/22/002

Norman, K. A., Newman, E., Detre, G., & Polyn, S. (2006). How inhibitory oscillations can train neural networks and punish competitors. *Neural Comput, 18*(7), 1577-1610. doi:10.1162/neco.2006.18.7.1577

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychol Rev, 114*(4), 887-953. doi:10.1037/0033-295X.114.4.887

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol Rev, 110*(4), 611-646. doi:10.1037/0033-295X.110.4.611

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

Norman, Y., Yeagle, E. M., Khuvis, S., Harel, M., Mehta, A. D., & Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science, 365*(6454). doi:10.1126/science.aax1030

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neurosci Biobehav Rev, 34*(7), 1023-1035. doi:10.1016/j.neubiorev.2009.12.014

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus, 3*(3), 317-330. doi:10.1002/hipo.450030307

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus, 3*(3), 317-330. doi:10.1002/hipo.450030307

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

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends Cogn Sci, 6*(12), 505-510.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol Rev, 108*(2), 311-345. doi:10.1037/0033-295x.108.2.311

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci, 2011*, 156869. doi:10.1155/2011/156869

Parish, G., Hanslmayr, S., & Bowman, H. (2017). The Sync/deSync model: How a synchronized hippocampus and a de-synchronized neocortex code memories. *bioRxiv*.

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5*.

Pastotter, B., & Bauml, K. T. (2014). Distinct slow and fast cortical theta dynamics in episodic memory retrieval. *Neuroimage, 94*, 155-161. doi:10.1016/j.neuroimage.2014.03.002

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci, 8*(12), 976-987. doi:10.1038/nrn2277

Pavlides, C., Greenstein, Y. J., Grudman, M., & Winson, J. (1988). Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. *Brain Res, 439*(1-2), 383-387.

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. *Nat Neurosci, 12*(7), 919-926. doi:10.1038/nn.2337

Place, R., Farovik, A., Brockmann, M., & Eichenbaum, H. (2016). Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nat Neurosci, 19*(8), 992-994. doi:10.1038/nn.4327

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends Cogn Sci, 17*(5), 230-240. doi:10.1016/j.tics.2013.03.005

Poppenk, J., & Moscovitch, M. (2011). A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. *Neuron, 72*(6), 931-937. doi:10.1016/j.neuron.2011.10.014

Poppenk, J., & Norman, K. A. (2014). Briefly cuing memories leads to suppression of their neural representations. *J Neurosci, 34*(23), 8010-8020. doi:10.1523/JNEUROSCI.4584-13.2014

Rafidi, N. S., Hulbert, J. C., Brooks, P. P., & Norman, K. A. (2018). Reductions in Retrieval Competition Predict the Benefit of Repeated Testing. *Sci Rep, 8*(1), 11714. doi:10.1038/s41598-018-29686-y

Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., . . . Lisman, J. E. (2001). Gating of human theta oscillations by a working memory task. *J Neurosci, 21*(9), 3175-3183.

Rasch, B., Buchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science, 315*(5817), 1426-1429. doi:10.1126/science.1138581

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code* (M. Press Ed.): Cambridge, MA.

Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: insights from functional brain imaging. *Annu Rev Psychol, 63*, 101-128. doi:10.1146/annurev-psych-120710-100344

Ritchey, M., & Cooper, R. A. (2020). Deconstructing the Posterior Medial Episodic Network. *Trends in Cognitive Sciences, 24*(6), 451-465. doi:10.1016/j.tics.2020.03.006

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

Rizzuto, D. S., Madsen, J. R., Bromfield, E. B., Schulze-Bonhage, A., & Kahana, M. J. (2006). Human neocortical oscillations exhibit theta phase differences between encoding and retrieval. *Neuroimage, 31*(3), 1352-1358. doi:10.1016/j.neuroimage.2006.01.009

Rolls, E. T. (1996). A theory of hippocampal function in memory. *Hippocampus, 6*(6), 601-620. doi:10.1002/(SICI)1098-1063(1996)6:6<601::AID-HIPO5>3.0.CO;2-J

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

Roy, D. S., Park, Y.-G., Ogawa, S. K., Cho, J. H., Choi, H., Kamensky, L., . . . Tonegawa, S. (2019). Brain-wide mapping of contextual fear memory engram ensembles supports the dispersed engram complex hypothesis. *bioRxiv*, 668483. doi:10.1101/668483

Rugg, M. D., Johnson, J. D., Park, H., & Uncapher, M. R. (2008). Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspective. *Prog Brain Res, 169*, 339-352. doi:10.1016/s0079-6123(07)00021-0

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Curr Opin Neurobiol, 23*(2), 255-260. doi:10.1016/j.conb.2012.11.005

Rutishauser, U., Ross, I. B., Mamelak, A. N., & Schuman, E. M. (2010). Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature, 464*(7290), 903-907. doi:10.1038/nature08860

Rutishauser, U., Ye, S., Koroma, M., Tudusciuc, O., Ross, I. B., Chung, J. M., & Mamelak, A. N. (2015). Representation of retrieval confidence by single neurons in the human medial temporal lobe. *Nat Neurosci, 18*(7), 1041-1050. doi:10.1038/nn.4041

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *J Cogn Neurosci, 26*(8), 1736-1747. doi:10.1162/jocn_a_00578

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 372*(1711), 20160049. doi:10.1098/rstb.2016.0049

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus, 26*(1), 3-8. doi:10.1002/hipo.22523

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation

signatures in the hippocampus and prefrontal cortex. *Nat Commun, 6*, 8151. doi:10.1038/ncomms9151

Scoville WB, Milner B. 1957. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Psychol*.20:11–21

Seidenbecher, T., Laxmi, T. R., Stork, O., & Pape, H. C. (2003). Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. *Science, 301*(5634), 846-850. doi:10.1126/science.1085818

Semon, R. (1921). *The nmeme*. London: George Allen & Unwin.

Sirota, A., Csicsvari, J., Buhl, D., & Buzsaki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proc Natl Acad Sci U S A, 100*(4), 2065-2069. doi:10.1073/pnas.0437938100

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus, 6*(2), 149-172. doi:10.1002/(sici)1098-1063(1996)6:2<149::Aid-hipo6>3.0.Co;2-k

Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., . . . Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nat Neurosci, 18*(11), 1679-1686. doi:10.1038/nn.4119

Staresina, B. P., Fell, J., Do Lam, A. T., Axmacher, N., & Henson, R. N. (2012). Memory signals are temporally dissociated in and across human hippocampus and perirhinal cortex. *Nat Neurosci, 15*(8), 1167-1173. doi:10.1038/nn.3154

Staresina, B. P., & Wimber, M. (2019). A Neural Chronometry of Memory Recall. *Trends Cogn Sci*. doi:10.1016/j.tics.2019.09.011

Staudigl, T., Vollmar, C., Noachtar, S., & Hanslmayr, S. (2015). Temporal-pattern similarity analysis reveals the beneficial and detrimental effects of context reinstatement on human memory. *J Neurosci, 35*(13), 5373-5384. doi:10.1523/JNEUROSCI.4198-14.2015

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage, 65*, 69-82. doi:10.1016/j.neuroimage.2012.09.063

Swanson, R. A., Levenstein, D., McClain, K., Tingley, D., & Buzsáki, G. (2020). Variable specificity of memory trace reactivation during hippocampal sharp wave ripples. *Current Opinion in Behavioral Sciences, 32*, 126-135. doi:https://doi.org/10.1016/j.cobeha.2020.02.008

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci, 3*(4), 151-162.

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

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behav Neurosci, 100*(2), 147-154.

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

Tingley, D., & Peyrache, A. (2020). On the methods for reactivation and replay analysis. *Philos Trans R Soc Lond B Biol Sci, 375*(1799), 20190231. doi:10.1098/rstb.2019.0231

Tort, A. B., Komorowski, R., Eichenbaum, H., & Kopell, N. (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *J Neurophysiol, 104*(2), 1195-1210. doi:10.1152/jn.00106.2010

Tulving, E. (1972). Episodic and semantic memory. In *Organization of memory.* (pp. xiii, 423-xiii, 423). Oxford, England: Academic Press.

Tulving, E. (1999). On the uniqueness of episodic memory. In *Cognitive neuroscience of memory.* (pp. 11-42). Ashland, OH, US: Hogrefe & Huber Publishers.

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology, 53*(1), 1-25. doi:10.1146/annurev.psych.53.100901.135114

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus, 8*(3), 198-204. doi:10.1002/(SICI)1098-1063(1998)8:3<198::AID-HIPO2>3.0.CO;2-G

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373. doi:10.1037/h0020071

Tyler, L. K., Chiu, S., Zhuang, J., Randall, B., Devereux, B. J., Wright, P., . . . Taylor, K. I. (2013). Objects and categories: feature statistics and object processing in the ventral stream. *J Cogn Neurosci, 25*(10), 1723-1735. doi:10.1162/jocn_a_00419

Underwood, B. J., & Postman, L. (1960). Extraexperimental sources of interference in forgetting. *Psychological Review, 67*(2), 73-95. doi:10.1037/h0041865

Van Veen, B. D., van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng, 44*(9), 867-880. doi:10.1109/10.623056

Vandecasteele, M., Varga, V., Berenyi, A., Papp, E., Bartho, P., Venance, L., . . . Buzsaki, G. (2014). Optogenetic activation of septal cholinergic neurons suppresses sharp wave ripples and enhances theta oscillations in the hippocampus. *Proc Natl Acad Sci U S A, 111*(37), 13535-13540. doi:10.1073/pnas.1411233111

Vanderwolf, C. H. (1969). Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr Clin Neurophysiol, 26*(4), 407-418. doi:10.1016/0013-4694(69)90092-3

Vaz, A. P., Inati, S. K., Brunel, N., & Zaghloul, K. A. (2019). Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. *Science, 363*(6430), 975-978. doi:10.1126/science.aau8956

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

Vertes, R. P., & Kocsis, B. (1997). Brainstem-diencephalo-septohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience, 81*(4), 893-926. doi:10.1016/s0306-4522(97)00239-x

Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., & Gazzaley, A. (2015). Age-Related Changes in 1/f Neural Electrophysiological

Noise. *J Neurosci, 35*(38), 13257-13265. doi:10.1523/JNEUROSCI.2332-14.2015

Waldhauser, G. T., Braun, V., & Hanslmayr, S. (2016). Episodic Memory Retrieval Functionally Relies on Very Rapid Reactivation of Sensory Information. *J Neurosci, 36*(1), 251-260. doi:10.1523/jneurosci.2101-15.2016

Waldhauser, G. T., Johansson, M., & Hanslmayr, S. (2012). alpha/beta oscillations indicate inhibition of interfering visual memories. *J Neurosci, 32*(6), 1953-1961. doi:10.1523/JNEUROSCI.4201-11.2012

Wang, J. X., Rogers, L. M., Gross, E. Z., Ryals, A. J., Dokucu, M. E., Brandstatt, K. L., . . . Voss, J. L. (2014). Targeted enhancement of cortical-hippocampal brain networks and associative memory. *Science, 345*(6200), 1054-1057. doi:10.1126/science.1252900

Watrous, A. J., Miller, J., Qasim, S. E., Fried, I., & Jacobs, J. (2017). Phase-tuned neuronal firing encodes human contextual representations for navigational goals. *bioRxiv*.

Wen, H., & Liu, Z. (2016). Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. *Brain Topogr, 29*(1), 13-26. doi:10.1007/s10548-015-0448-0

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat Neurosci, 18*(4), 582-589. doi:10.1038/nn.3973

Wimber, M., Maass, A., Staudigl, T., Richardson-Klavehn, A., & Hanslmayr, S. (2012). Rapid memory reactivation revealed by oscillatory entrainment. *Curr Biol, 22*(16), 1482-1486. doi:10.1016/j.cub.2012.05.054

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

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language, 46*(3), 441-517. doi:https://doi.org/10.1006/jmla.2002.2864

Zeidman, P., Mullally, S. L., & Maguire, E. A. (2015). Constructing, Perceiving, and Maintaining Scenes: Hippocampal Activity and Connectivity. *Cerebral cortex (New York, N.Y. : 1991), 25*(10), 3836-3855. doi:10.1093/cercor/bhu266

Zhang, H., Fell, J., & Axmacher, N. (2018). Electrophysiological mechanisms of human memory consolidation. *Nat Commun, 9*(1), 4103. doi:10.1038/s41467-018-06553-y

Zheng, C., Bieri, K. W., Hsiao, Y. T., & Colgin, L. L. (2016). Spatial Sequence Coding Differs during Slow and Fast Gamma Rhythms in the Hippocampus. *Neuron, 89*(2), 398-408. doi:10.1016/j.neuron.2015.12.005