TEMPORAL DYNAMICS AND MECHANISMS OF OSCILLATORY PATTERN REINSTATEMENT IN HUMAN EPISODIC MEMORY

by SEBASTIAN MICHELMANN

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
March of 2018

Abstract

A fundamental question in the investigation of episodic memory is how the human brain represents information from the past.

This thesis introduces a new method that tracks content specific representations in rhythmic fluctuations of brain activity (i.e. brain oscillations). It is demonstrated that a frequency band centred at 8 Hz carries information about remembered stimulus content. This is shown in human electrophysiological recordings during episodic memory formation and retrieval.

Strong and sustained power decreases consistently mark this 8 Hz frequency band; successful memory encoding and retrieval are associated with power decreases in low frequencies (<30 Hz) throughout this thesis and in numerous former studies. The presented results link power decreases to the reinstatement of oscillatory patterns in sensory specific areas for the first time and therefore implicate them in the representation of information.

Finally, the temporal dynamics of recollection are investigated by tracking information from distinct sub-events in continuous episodic memories. In behavioural and neural data, memory replay is faster than perception and takes place in a forward direction. Herein, fragments of fine-grained temporal patterns are reinstated; yet, subjects can skip flexibly between sub-events. Leveraging oscillatory mechanisms to track information can therefore identify episodic memory replay as a dynamic process.

Acknowledgements

There are two people that I want to thank above all for their support and for their influence on my thinking. Simon and Howard, my supervisors, have provided me with more than I could ask for in encouragement and guidance over the last four years. I enjoyed challenging and rigorous discussions with them, just as much as the personal conversations and occasional celebrations. Simon's optimism and venturousness have inspired me to think independently and believe in my ideas. The facility and fun he has in quickly apprehending complex topics make science look easy. He has helped me to demystify a field that is packed with complex methods and analysis approaches. Howard's mathematical rigor and his ability to identify the essence of methodological challenges have made an impression on me. He has encouraged me to pursue an understanding of the math behind the methods and taught me that zooming in can be equally important as zooming out.

I am grateful that I could be in this department. I had interesting discussions and got important advice from people in many different labs. I want to thank our CoOL cognitions and oscillations lab, the episodic memory lab, the adaptive memory lab and the members of the computational cognitive neuroimaging group. I specifically want to thank Bernhard Staresina for his support and guidance and Maria Wimber and Ian Charest for helpful suggestions and interesting discussions. Most of all, however, I thank Mathias Treder. Apart from my deep appreciation for him on a personal level, I cannot stress enough how much I benefited from his presence in the department. His arrival had a strong impact on my learning curve.

All my friends, old and new, deserve my gratitude for their help and support and importantly, simply for the good times. I feel privileged that I could meet so many exceptional, smart, interesting, caring, funny and genuine people here in Birmingham. Every place that I have left

so far has left me knowing that there are great people out there that I am proud to call my friends.

I want to thank my family for everything. They have challenged me to become the person that I am, while loving the person that I already was.

Finally, I want to acknowledge all the volunteers that participated in the following studies. Their patience, motivation and thoroughness were a cornerstone of this research.

Table of contents

List of illustrations

List of abbreviations

Analysis Of Variance	ANOVA
Associative Recognition	AR
Birmingham University Imaging Centre	BUIC
Circular Variance	CV
Cued Recall	CR
Cumulated Similarity	CS
Electroencephalogram	EEG
Event Related Field	ERF
Eye Movement Desensitization and Reprocessing	EMDR
functional Magnetic Resonance Imaging	fMRI
Fusiform Face Area	FFA
General Linear Model	GLM
Hertz	Hz
Independent Component Analysis	ICA
Inter Trial Phase Coherence	ITPC
intracranial Electroencephalogram	iEEG
Iterative Closest Point	ICP
Left Preauricular Point	LPA
linearly constrained minimum variance	lcmv
Magnet Resonance Imaging	MRI
Magnetencephalogram	MEG
Montreal Neurological Institute	MNI

Nasion	NAS
Pairwise Phase Consistency	PPC
Parahippocampal Place Area	PPA
Post-Traumatic Stress Disorder	PTSD
Reaction Time	RT
Region Of Interest	ROI
Representational Similarity Analysis	RSA
Right Preauricular Point	RPA
Saccade Evoked Potential	SEP
Single trial Phase Locking Value	S-PLV
Sir Peter Mansfield Imaging Centre	SPMIC
Situationally Accessible Memory system	SAM
Standard Deviation	S.D. (also s.d.)

Chapter 1 – General Introduction

Theoretical background

Episodic Memory

Episodic memory is the memory for our personal experiences and their temporal relationship; it can loosely be described as mental time travel (Tulving, 1972, 1993). Practically, our episodic memory allows us to access information that is no longer present in the world around us. The related term memory reinstatement describes the evoking of information from the past and is also referred to as recollection (Yonelinas, 2002). This recollection of information can be investigated with a cued-recall paradigm (e.g. Fisher & Craik, 1977). In this, a cue is associated with a target, later only the cue is presented and the target must be remembered.

Importantly, episodic memory needs to be distinguished from working memory, which is the ability to temporarily maintain and manipulate information (Baddeley, 2003). In experiments that investigate episodic memory, a short distractor task is therefore included between the learning and recollection of information. It engages working memory in a task irrelevant way and prevents confounding working memory processes from influencing the episodic retrieval.

Brain Oscillations

The notion of brain oscillations refers to rhythmical fluctuations in the ongoing local field potential which can be measured with electroencephalography (EEG) and magnetencephalography (MEG) (e.g. Long, Burke, & Kahana, 2014; Staudigl & Hanslmayr, 2013; Wimber, Maaß, Staudigl, Richardson-Klavehn, & Hanslmayr, 2012; Zhang et al., 2015). Brain oscillations have been linked to cognitive functions in numerous ways (Buzsáki, 2006); importantly decreases in the amplitude of oscillations in the alpha and beta (8-30 Hz)

frequency band, play an important role in episodic memory formation and retrieval (Hanslmayr & Staudigl, 2014; Long et al., 2014; Zion-Golumbic, Kutas, & Bentin, 2010).

Information via Desynchronization

The information via desynchronization hypothesis (Hanslmayr, Staudigl, & Fellner, 2012) makes a clear prediction about the function of power decreases in brain oscillations and provides an explanation of their role for episodic memory. According to this framework, power decreases are crucial for the reinstatement of information-rich content in the neocortex. Specifically, the theory states that observed power decreases reflect a desynchronisation of activity in neural populations. In line with information theory (Shannon & Weaver, 1949), this desynchronization is crucial to maximize the information that a system can represent. Synchrony between neural assemblies, on the other hand signifies redundancy because several neural units fire in the same way.

Desynchronization, which is marked by power decreases, is also relevant for perception (Harris & Thiele, 2011; Jensen & Mazaheri, 2010; Marshall, Bergmann, & Jensen, 2015). Importantly, several studies implicate the phase of the alpha frequency band (8-12 Hz) (Klimesch et al., 1996) in the organization of incoming information, when oscillatory power decreases. Specifically, there is evidence that fluctuations at 7/8 Hz, rhythmically sample a continuous input stream during perception (Hanslmayr, Volberg, Wimber, Dalal, & Greenlee, 2013; Jensen, Bonnefond, & VanRullen, 2012; Landau & Fries, 2012; VanRullen, Busch, Drewes, & Dubois, 2011; VanRullen, Carlson, & Cavanagh, 2007).

This link between decreases in oscillatory power and the organization of information via the phase of the alpha frequency band could therefore generalize to episodic memory: Power decreases could be a general mechanism that allows for the representation of rich

information that is organized by oscillatory phase, during perception and in memory. Patterns of activation that are measured in oscillatory phase during encoding could therefore reappear during episodic memory reinstatement. This should happen in those frequency bands that are marked by power decreases.

Representational Similarity Analysis

Representations of content can be investigated with representational similarity analysis (RSA) (Kriegeskorte, 2008). RSA uses similarity in patterns of neural activity and assesses its structure. If neural data codes for representation-specific information, then measurements of activity that correspond to the same representation will be more similar to each other than patterns that correspond to different representations. This code can take multiple forms, e.g. spatial activation patterns, temporal activation patterns or patterns in spectral power. In functional magnetic resonance imaging (fMRI), the representational structure of spatial patterns has been used, for instance, to reveal which regions code for detailed events and which regions code for multi-event narratives (Collin, Milivojevic, & Doeller, 2015). In electrophysiology, this method can now answer questions like *"which frequency band codes temporal information and when?"* In summary, the key ingredients for RSA are a measure of similarity and repeated measurements of neural activity that belong to the same representation.

Coding of Representations in Oscillations

Representational similarity analysis can therefore be combined with measures of similarity in electrophysiology. Specifically, repeated measurements of electrophysiological activity can be compared in their similarity with the use of adequate similarity metrics. The structure of similarity between measurements can then be assessed to reveal, whether neural activity

codes for representations. These similarity metrics can be tailored to compare oscillatory activity that is confined to a certain frequency band. This will make it possible to investigate individual representations and their temporal structure, coded in power decreases of neural oscillations. An established set of measures that assess similarity in neural oscillations are measures of phase coherence (Lachaux et al., 2000; Mormann, Lehnertz, David, & E. Elger, 2000; Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1996; Vinck, van Wingerden, Womelsdorf, Fries, & Pennartz, 2010). These measures should lend themselves to the investigation of content specific representations that are coded in oscillatory patterns via representational similarity analysis.

Aims of this research

This research aims to observe memory reinstatement in neural patterns that are measured with electrophysiology. The investigation of memory representations with these temporally resolved recordings will make it possible to draw conclusions about temporal dynamics and mechanisms of episodic memory replay. Importantly the main prediction of the information via desynchronization hypothesis will be tested: Episodic memory reinstatement of information-rich content will result in strong power decreases; information about the stimulus content should be present in frequency bands that display these power decreases. A memory-paradigm to investigate these predictions will elicit reinstatement of information-rich stimuli. A cued-recall paradigm will separate the perception of content from its reinstatement in memory. Herein, target stimuli will be as naturalistic as possible and have an inherent temporal dynamic. Cues, on the other hand, will be simple words.

In chapter 2, an EEG study will investigate content specific temporal patterns of activity as they are reinstated in a purely memory driven way. To this end participants will associate dynamic naturalistic stimuli with simple word-cues. This will be realised in two modalities using short video-clips and sound-clips. Upon presentation of the word-cues, participants will reinstate vivid representations of the dynamic stimuli from memory. A direct prediction from the information via desynchronization hypothesis is that the reinstatement of these information-rich stimuli will elicit strong power decreases in the EEG recordings. Importantly oscillations that are marked by these power decreases should contain patterns that are specific to the content held in memory. This prediction will be tested with a new method that combines oscillatory phase coherence with representational similarity analysis.

Chapter 3 will reanalyse the dataset from chapter 2 to test, whether content specific patterns of naturalistic stimuli are maintained during episodic memory formation. In the experiment, the association with a cue takes place after the presentation of the naturalistic stimulus. This creates a time interval of association, in which the naturalistic stimulus is absent but relevant. The information via desynchronization hypothesis predicts again that strong power decreases will be associated with successful memory formation; those are indicative of information about the stimulus content. Representational similarity of oscillatory patterns will be assessed in order to test, whether the association with a dynamic stimulus, yields a detectable representation of that absent stimulus.

Chapter 4 will finally leverage the results from chapter 2 and 3 and track memory replay throughout continuous episodes that consist of distinct sub-events. This will clarify the temporal dynamics of memory replay, such as its direction and speed. In a behavioural experiment and an MEG study subjects will associate word-cues with distinct scenes that

form continuous videos (named video-episodes). Later participants will be asked upon the presentation of the word-cue, in which exact scene they have learned an association. Behaviourally the reaction time to this scene-response will determine the direction and speed of memory replay: It should take longer to recall associations that were formed later in the video-episode, if replay is forward. The distance between reaction times for early and late associations will indicate the speed of replay. In the MEG study, response intervals will be fixed; in the response interval, content specific patterns from encoding will be tracked. Representational similarity analysis of oscillatory patterns should replicate findings of content specific reinstatement in a frequency band that is marked by strong power decreases. Tracking those patterns will make it possible to statistically test the direction of replay, which could either be forward or backward. Finally the distance between different patterns will be assessed. It is hypothesised that memory replay is a flexible mechanism that allows participants to jump between salient boundaries in the continuous video-episodes. Therefore the skipping between scenes should take place on a faster time-scale than the replay of individual scenes. This will be tested by statistically comparing the time of reinstatement between distinct patterns from encoding.

Chapter 5 will summarise and discuss these findings. Finally a first outlook is given, on new questions that arise from this research.

Chapter 2 – The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans

Preface

The following study was designed to test the prediction that decreases in oscillatory power contain information about memory content. Signatures of memory content should be detectable in desynchronizing frequency bands when human volunteers recall information-rich stimuli from memory. This research was published in near identical form in PLOS Biology under the title: *The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans* (Michelmann, Bowman, & Hanslmayr, 2016).

Contributions

The experiments were conceived and designed by SM and SH. SM performed the experiments.

All data analysis was performed by SM under supervision of SH, the manuscript was written by SM under supervision of SH and HB.

Abstract

Reinstatement of dynamic memories requires the replay of neural patterns that unfold over time in a similar manner as during perception. However, little is known about the mechanisms that guide such a temporally structured replay in humans, because previous studies either used unsuitable methods or paradigms to address this question. We here overcome these limitations by developing a new analysis method to detect the replay of temporal patterns in a paradigm that requires participants to mentally replay short sound or video clips. We show that memory reinstatement is accompanied by a decrease of low frequency (8 Hz) power, which carries a temporal phase signature of the replayed stimulus. These replay effects were evident in the visual as well as in the auditory domain and were localized to sensory specific regions. These results suggest low frequency phase to be a domain general mechanism that orchestrates dynamic memory replay in humans.

Introduction

Episodic memories are dynamic, multisensory events that are coded in our memory system. If you remember the last time you had dinner at your favourite restaurant you will probably recall the person you were with, the music playing in the background and the smell and taste of that delicious food. Whenever we re-experience episodic memories this way, the events unravel in front of our mind in a temporal order. Even subparts of these episodes, such as the movement of lips in a conversation or parts of the background melody, have an inherent temporal dynamic to them. Given this abundance of temporal structure in our memories, it is rather surprising how limited our understanding is as to how human brains orchestrate such dynamic memory replay. Here we address this question for the first time to our knowledge, and identify a neural mechanism that carries the temporal signature of individual dynamic episodic memories. By cuing dynamic memories of auditory and visual content, we were able to detect the presence of phase patterns in the electroencephalographic (EEG) signal which indicate the replay of individual auditory or visual stimuli in memory. Temporal signatures were carried by a frequency that was markedly similar in two sensory domains (~8 Hz), they appeared in sensory-specific regions, and were related to decreases in power in the same frequency.

Previous findings suggest that perception is not continuous but, instead, is rhythmically sampled in discrete snapshots guided by the phase of low alpha (~7-8 Hz) (Hanslmayr et al., 2013; Landau & Fries, 2012; VanRullen et al., 2007), which suggests a pivotal role of low alpha phase for providing a temporal structure during perception (Canavier, 2015; Hanslmayr, Staresina, & Bowman, 2016; Jensen et al., 2012; VanRullen et al., 2011; Watrous, Deuker, Fell,

& Axmacher, 2015). Accordingly, recent studies showed that low frequency phase carries reliable information about stimulus content (Ng, Logothetis, & Kayser, 2013; Schyns, Thut, & Gross, 2011). This key role of oscillatory phase during perception makes it a prime candidate to also organize the replay of neural representations in episodic memory, which is an untested prediction to date.

A ubiquitous electrophysiological signature of successful memory processing is a pronounced power decrease in low frequencies, especially in alpha (Burgess & Gruzelier, 2000; Klimesch, 1997, 1999; Long et al., 2014; Zion-Golumbic et al., 2010). On a theoretical level alpha power decreases affect neural processing in two ways. Firstly, they promote increased neural activity as reflected by increased neural firing rates and increased BOLD signal (Haegens, Nacher, Luna, Romo, & Jensen, 2011; Klimesch, 2012; Zumer, Scheeringa, Schoffelen, Norris, & Jensen, 2014). Importantly, even when alpha power is decreased its phase still rhythmically modulates firing rates (Jensen et al., 2012). Secondly, alpha power decreases reflect a relative de-correlation of neural activity, and thereby index an increase in information coding capacity (Hanslmayr et al., 2016). Accordingly, a mechanism by which alpha power decreases allow for the temporal organization of information via phase has been proposed in perception (Jensen et al., 2012), however, whether memory replay is guided by a similar mechanism is an open question (Hanslmayr et al., 2016).

The reinstatement of neural patterns in memory can be detected with multivariate analysis methods like representational similarity analysis (Kriegeskorte, 2008) (RSA). This approach has been successfully applied in functional magnetic resonance imaging (fMRI) (Staresina, Henson, Kriegeskorte, & Alink, 2012; Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015), EEG/MEG (Jafarpour, Fuentemilla, Horner, Penny, & Duzel, 2014; J. D. Johnson, Price,

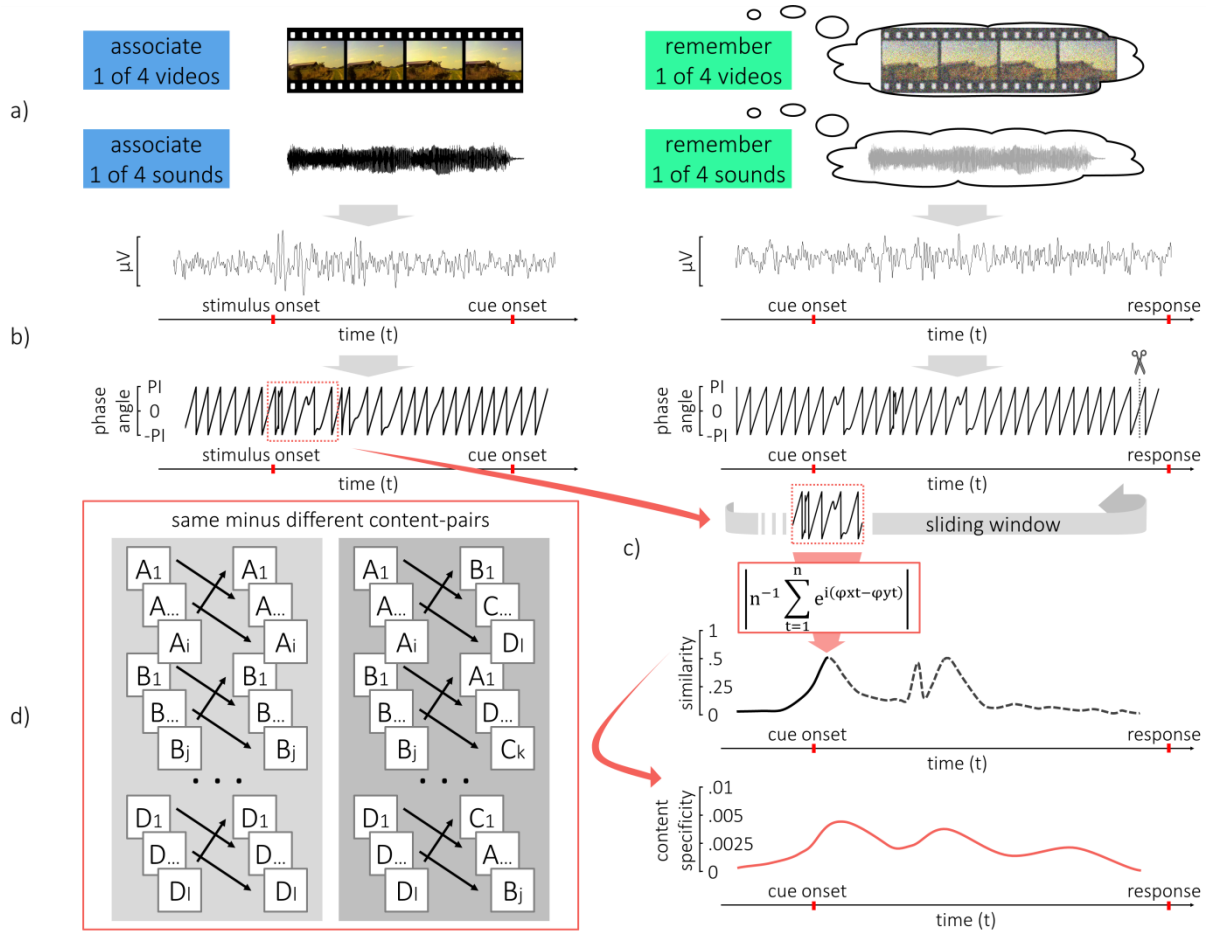
& Leiker, 2015; Kurth-Nelson, Barnes, Sejdinovic, Dolan, & Dayan, 2015; Ng et al., 2013; Staudigl, Vollmar, Noachtar, & Hanslmayr, 2015; Wimber et al., 2012) and intracranial EEG (iEEG) (Yaffe et al., 2014; Zhang et al., 2015). However, even though some previous studies were able to decode information from oscillatory patterns, the mechanism by which oscillations carry mnemonic information remains completely unclear. This is because most prior studies either settle for classification of reactivated memories and thus do not aim for mechanistic explanations of memory replay, or because they use static stimuli and analysis procedures.

We overcome these central limitations by testing whether a temporal signature that is present, while a video or a sound clip is perceived is actively reproduced by the brain during retrieval. By temporal signature we mean a sequence of electrophysiological activity that is specific to an individual stimulus. To this end, we test the mechanistic hypothesis that low frequency power decreases are linked with the reinstatement of such stimulus-specific phase patterns. A paradigm was used where memories of dynamic content are cued by a static word (see figure 102 a-d). In a visual and in an auditory condition, we asked subjects to watch (or listen to) 3 second long video or sound clips, and then to associate the respective stimulus with a word. Importantly only four videos/sounds were repeatedly associated with different words. In the retrieval block we then only presented the word cue (or a distractor word) under the instruction to vividly replay the associated video or sound. Note that there was no overlap in sensory input between the video/sound and the word, enabling us to investigate purely memory driven reinstatement of temporal signatures.

We hypothesize that we will find content-specific temporal signatures in those frequency bands that show pronounced power decreases during episodic memory retrieval (Hanslmayr

et al., 2016). Applying the logic of RSA to measures of phase-based similarity, we designed a new method that can detect content-specific signatures in neural time series, where the exact onset of replay is not known (see figure 101), which is the case in our retrieval phase. To our knowledge this is the first time that a method can use oscillatory phase patterns to decode content from activity that is not time-locked. We assess reinstatement in the auditory and in the visual modality in order to validate our novel, dynamic RSA method and to test for a domain general memory replay mechanism. Whole brain activity was measured via high density EEG and individual MRIs were collected to increase the fidelity of source localization.

Figure 101: Detection of content specific reinstatement of temporal patterns



Legend to figure 101: Detection of content specific reinstatement of temporal patterns

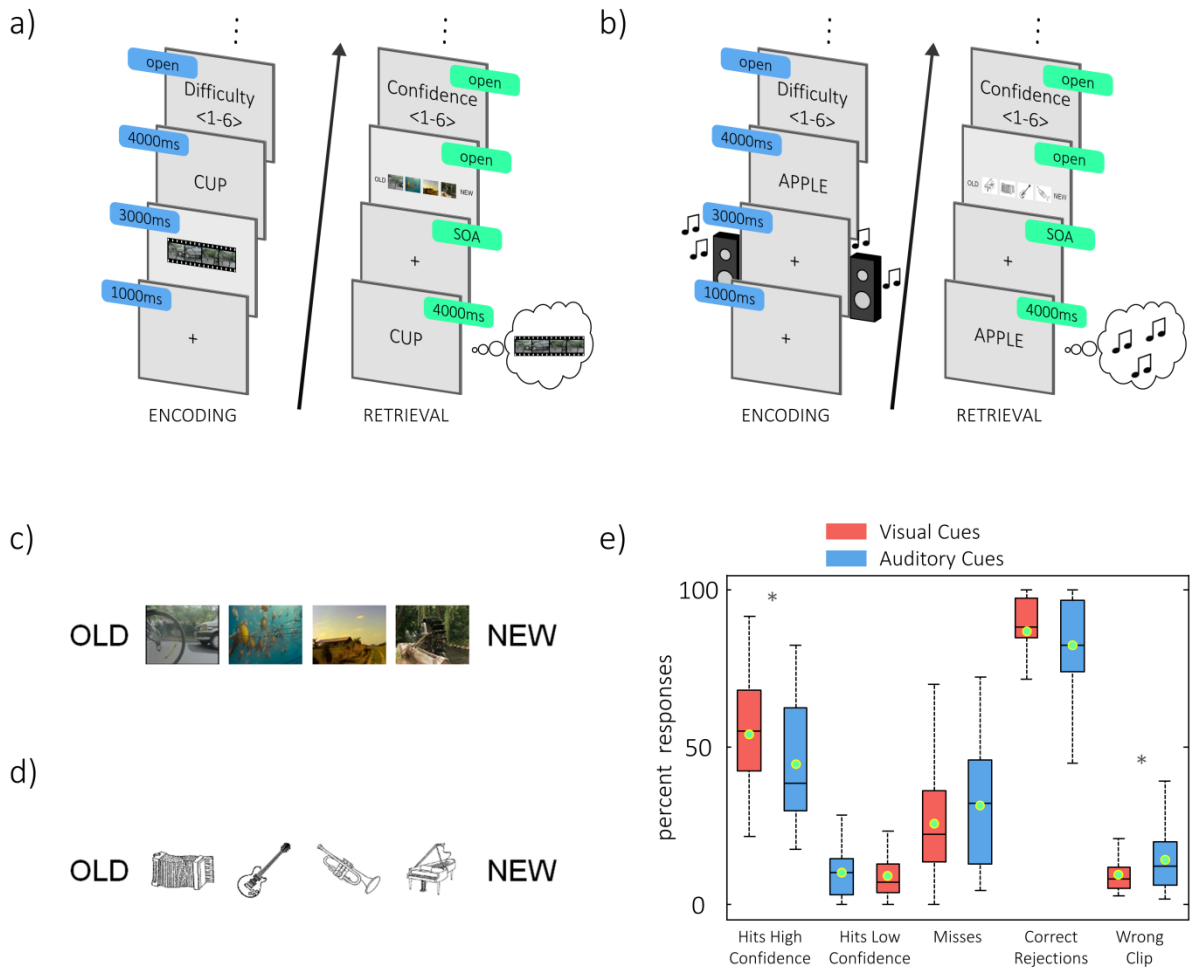
(a) During encoding subjects associated 1 of 4 videos with a different cue word in every trial in the visual condition, or they associated 1 of 4 sounds with different cue words in the auditory condition (a, left). During retrieval subjects only saw the static word-cue on the screen and were asked to recall the corresponding dynamic stimulus (a, right). (b) At every electrode the oscillatory phase for a frequency of interest was extracted from the EEG activity. A time window from encoding was then selected and the time course of phase in this window was compared to retrieval. (c) A sliding window was used to assess the similarity, based on the constancy of phase angle differences over time (Single Trial Phase Locking Value (Lachaux et al., 2000; Mormann et al., 2000)*). This measure made it possible to assess similarity between single trials where the strength of similarity ranged from 0 to 1, even though an oscillatory pattern was compared. Therefore similarity could be averaged across trials, time and participants. To avoid confounds from the response and the response scale, the data was cut at the end of the retrieval trial and the window was slid out, back into the prestimulus interval. This was done for trial-pairs of same content (e.g. learning A, remembering A) and for trial-pairs of different content (e.g. learning A, remembering D). (c-d) The difference in similarity between pairs of same and pairs of different content was interpreted as evidence for content specific reinstatement of temporal patterns.*

Results

Behavioural performance

Behavioural results are shown in figure 102e. In the visual session, participants remembered on average 53.92% (s.d. = 17.56%) of the video-clips with high confidence (rating > 4), they further remembered 9.97% (s.d. = 7.62%) of the clips with low confidence, however in order to increase the signal to noise ratio, hits with a low confidence rating were not included in further analysis. In the auditory session, 44.44% (s.d. = 19.8%) of the audio-clips were remembered with high confidence, which was significantly less than in the visual condition (t_{23} = -2.81, P = 0.01). An additional 9.06% (s.d. = 6.9%) of the audio-clips were remembered with low confidence. In accordance, the number of misses showed a trend to be lower in the visual session (mean 25.66%, s.d. = 17.56%) than in the auditory session (31.46%, s.d. = 19.15%, t_{23} = -1.91, P = 0.07). Another trend was observed towards a better identification of distractor words in the visual session (t_{23} = 1.92, P = 0.07), where 86.88% (s.d. = 13.03%) of the distractors were correctly rejected, while subjects only identified 82.43% (s.d. = 15.86%) of the distractors correctly as new words in the auditory session. Keeping with the slightly better performance for visual compared to auditory memories, the wrong video clip was less frequently selected in the visual (9.4%, s.d = 6.33%) condition, compared to the wrong sound clip in the auditory session (14%, s.d. = 9.69%, t_{23} = -2.86, P < 0.01).

Figure 102: Experimental design and behavioural results



Legend to figure 102: Experimental design and behavioural results.

*(a-b) Trial sequences are shown for the visual (a) and the auditory session (b). In the encoding-block (a-b, left) participants were presented with a dynamic stimulus that played for 3 seconds and was immediately followed by a word-cue, which was presented for 4 seconds. During encoding subjects learned 120 associations between 4 repeatedly shown dynamic stimuli and 120 different words. At the end of every encoding trial, the perceived difficulty of the association was rated on a scale from 1 to 6. In the retrieval-block (a-b, right) subjects only saw the static word-cue and were asked to vividly recall the dynamic stimulus, which it was associated with. Cues from encoding were mixed with 60 new words that served as distractors. Note that during encoding, the word cue was shown after the dynamic stimulus, avoiding sensory overlap between encoding and retrieval. Subjects were then asked to indicate the stimulus they recalled. Response options (c-d) consisted of 4 small screenshots of the video-clips in the visual condition (c) and of 4 small instruments, representing the sounds, in the auditory condition (d). Additionally the response option "NEW" was presented to indicate that the word was not presented in the encoding block (distractor item), the response option "OLD" was available to indicate that subjects remembered learning the word, but could not recall the content it was associated with. After responding, subjects were asked to rate the confidence in their answer on a scale from 1 to 6 (a-b, right). e) Behavioural performance is plotted. Hits are trials in which the correct associate was remembered (i.e. video or sound). A rating of high confidence was considered a rating > 4. Misses were defined as those trials in which the associate was not remembered and as those trials, where a word-cue was wrongly named as a distractor. Boxes are 25^{th} and 75^{th} percentiles around the median, whiskers represent minimum and maximum, green points are arithmetic means. * $p<0.05$.*

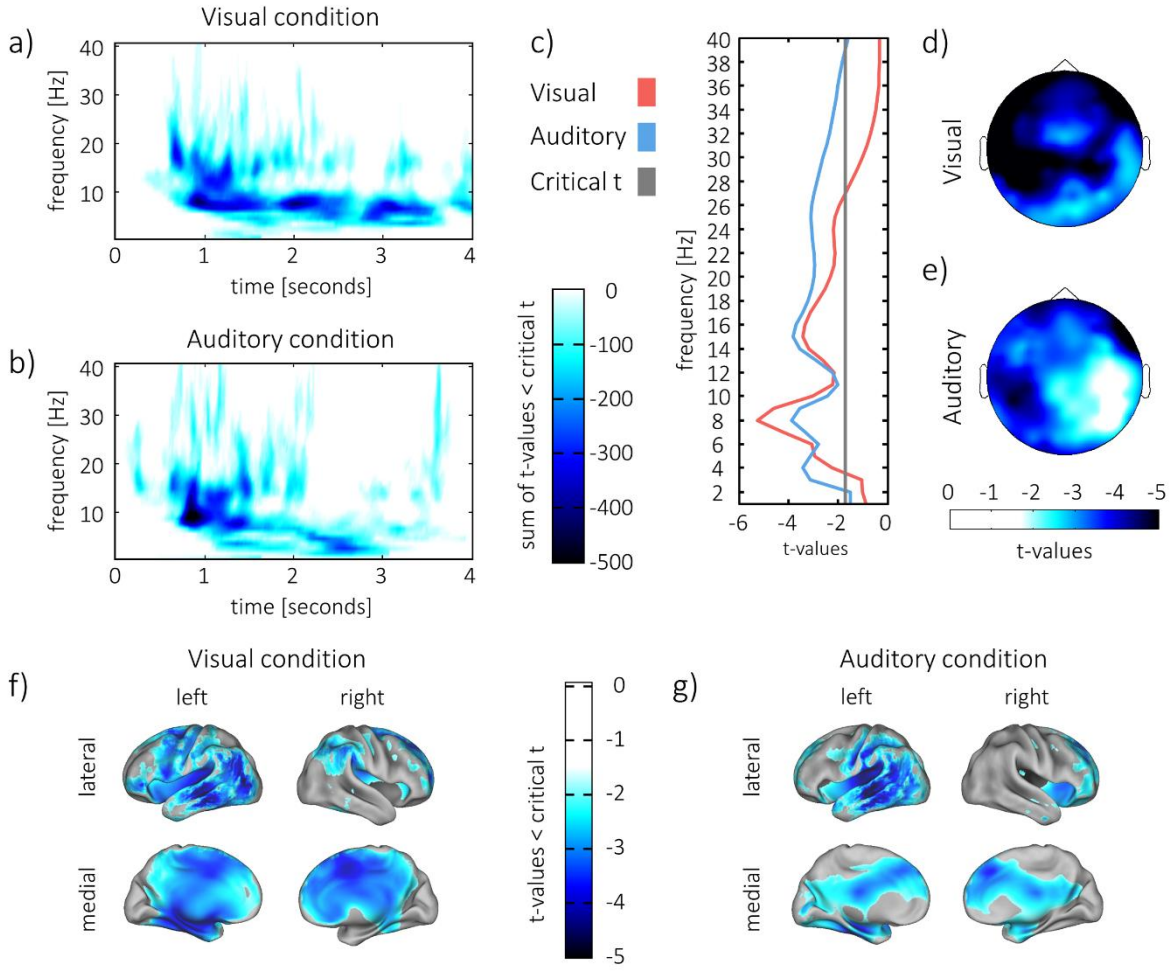
Successful memory is associated with low frequency power decreases

In order to identify oscillatory correlates of memory reinstatement, trials in which subjects were presented with a memory cue and strongly reinstated the content (i.e. High Confidence Hits) were contrasted with trials in which participants were presented with a distractor item and correctly indicated it as a new item (i.e. Correct Rejections). As expected, successful memory retrieval was associated with strong power decreases in the low frequencies (<30 Hz); power increases did not survive statistical testing, including the gamma frequency range (up to 140 Hz). The clusters that survived multiple comparisons correction (see methods) are shown in figure 103. Stronger power decreases for hits were obtained, when compared to correct rejections in the visual (figure 103a, $P < 0.001$) and in the auditory condition (figure 103b; $P < 0.001$). The same results emerged when a contrast was built between High Confidence Hits and trials in which subjects failed to remember the corresponding video- or sound-clip, that is, when they either failed to retrieve the correct associate, or judged an old item as new (see also: contrast of Hits and Misses, below). This further emphasizes the link of power decreases to successful memory reinstatement. To identify the frequencies that showed the strongest decrease in oscillatory power, the power difference across all electrodes and time points in the retrieval interval, was averaged and subjected to a t-test. Power decreases peaked at 8 Hz (see figure 103c) when contrasting Hits and Correct Rejections in the visual ($t_{23} = -5.2696$, $P < 0.001$) and in the auditory condition ($t_{23} = -3.86$, $P < 0.001$). In the visual condition these 8 Hz power decreases displayed a broad topography that showed a parietal maximum over the left hemisphere and frontal maxima over both

hemispheres (figure 103d). In the auditory condition, power decreases at 8 Hz were equally broad. Maxima were located over left parietal and right frontal regions (figure 103e).

In order to identify brain regions where power decreases were maximal at 8 Hz, sources of the difference between Hits and Correct Rejections were reconstructed for that frequency (see methods). Statistical testing was run unrestricted on the whole brain level. After multiple comparison correction (see methods) a cluster of significant differences emerged in the visual ($P < 0.001$) and in the auditory condition ($P = 0.002$). Clusters of power decreases were broad and did not show statistical differences between the visual and the auditory condition. In the visual condition (figure 103f), the cluster of significant differences spanned parietal, temporal and frontal regions of the left hemisphere and mid-frontal and parietal regions of the right hemisphere. In the auditory condition (figure 103g), power decreases spanned left parietal, temporal, mid-frontal and right frontal regions.

Figure 103: Contrast of Hits and Correct Rejections



Legend to figure 103: Contrast of Hits and Correct Rejections

Successful memory reinstatement was associated with a cluster of strong power decreases in the lower frequencies (<30 Hz). (a-b). The sum of t-values across all electrodes in the cluster of significant differences is plotted in the visual condition (a) and in the auditory condition (b). c) The t-statistic of power decreases was averaged over electrodes and time showing a peak at 8 Hz. (d-e) Topography of power decreases in the visual condition (d) and in the auditory condition (e). Power decreases are plotted as t-values of average difference at 8 Hz between 0 and 4 seconds during retrieval. (f-g) Reconstruction of 8 Hz power difference in source space using an 'lcmv' beamforming-algorithm in the visual (f) and in the auditory condition (g).

Successful memory is associated with decreases in power (contrasting Hits and Misses)

To further explore the correlates of successful memory reinstatement, trials in which subjects successfully retrieved the associated content (i.e. High Confidence Hits) were contrasted with trials in which subjects could not remember the corresponding dynamic stimulus upon presentation of the retrieval-cue (i.e. Misses). The retrieval-cue itself could either be recognized as an old word or be mistaken for a new word (distractor) to be considered a Miss. In this contrast, only subjects were included for which at least 15 trials remained after preprocessing (visual session: N = 19, auditory session: N = 17).

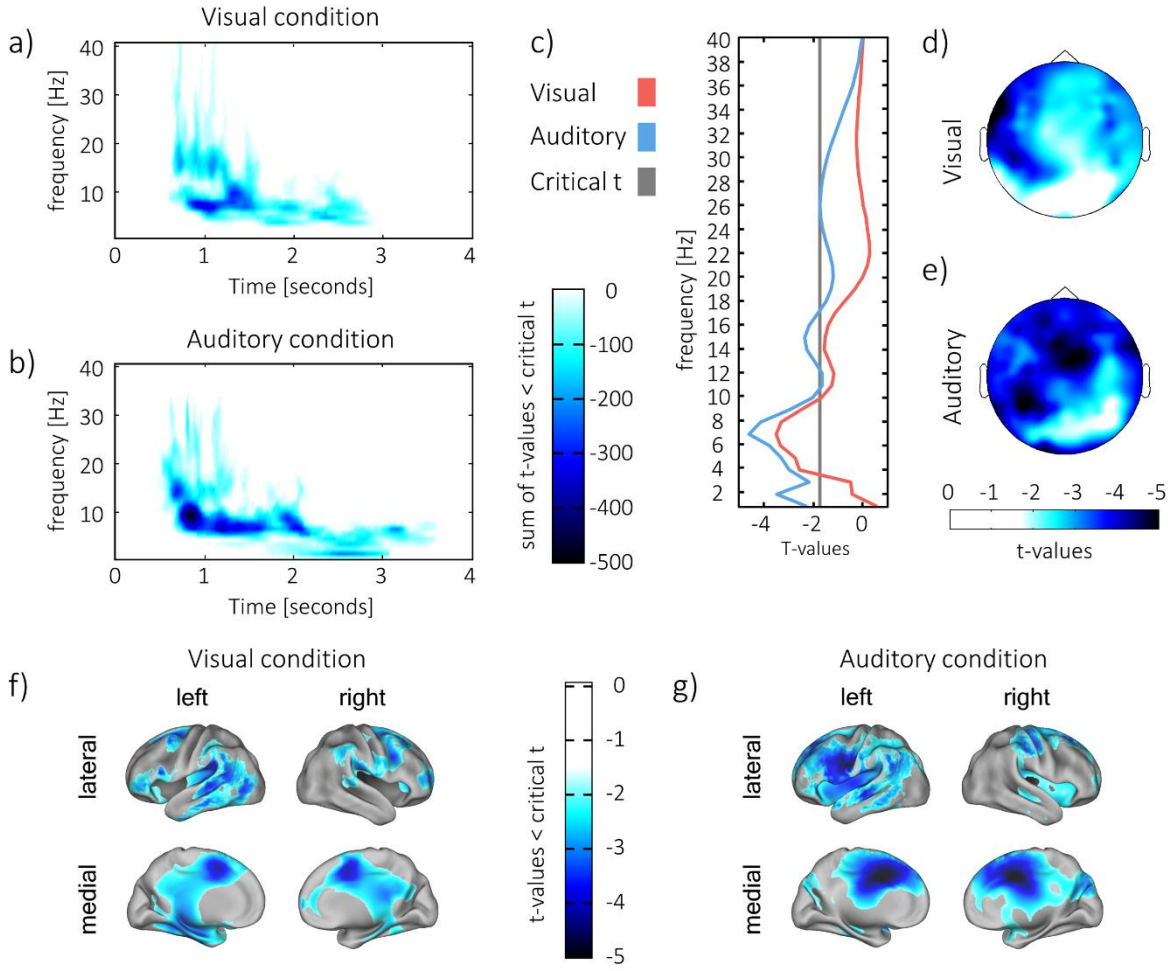
Consistent with the results from the contrast of hits and correct rejections, successful memory retrieval was associated with power decreases in the low frequencies (<30 Hz). The clusters, corrected for multiple comparisons, are displayed in figure 104. Stronger power decreases for hits were obtained when compared to misses in the visual (Figure 104a, P = 0.002) and in the auditory condition (figure 104b; $P < 0.001$). To confirm the frequencies that showed the strongest power decrease, differences were averaged across all electrodes and time points and subjected to a t-test. The strongest power decreases were observed at 7 Hz (see Figure 104c) in the visual (t_{19} = -3.51, P = 0.001) and in the auditory condition (t_{17} = -4.59, $P < 0.001$). Following results from the contrast of hits and correct rejections, the differences in power between hits and misses at 8 Hz are displayed in figure 104d-e.

In the visual condition, power decreases at 8 Hz were broad. The topography displayed a parietal and a frontal extreme over the left hemisphere (Figure 104d). In the auditory

condition the 8 Hz power decreases had a similarly broad topography. Maxima were situated over left parietal and central regions (Figure 104e).

To determine which regions expressed maximal power decreases at 8 Hz, the power of that frequency was reconstructed in source space and contrasted between Hits and Misses. Statistics were tested unrestricted on the whole brain level. Multiple comparison correction (see methods) revealed a cluster of significant differences in the visual ($P < 0.001$) and in the auditory condition ($P < 0.001$). Clusters of power decreases were again broad and did not differ statistically between the two conditions. In the visual condition (Figure 104f) the cluster of significant differences traversed parietal, temporal and mid-frontal regions of the left hemisphere and mid-frontal and parietal regions of the right hemisphere. In the auditory condition (Figure 104g) power decreases included left parietal, temporal and mid-frontal regions and central and frontal regions of the right hemisphere that bordered with the temporal lobe.

Figure 104: Contrast of Hits and Misses



Legend to figure 104: Contrast of Hits and Misses

Successful memory reinstatement was associated with a cluster of broad power decreases in the lower frequencies (<30 Hz). (a-b) Sum of t-values across the electrodes in the cluster of significant differences for the visual condition (a) and for the auditory condition (b). c) T-statistic of power decrease, averaged over electrodes and time. (d-e) Topography of power decreases in the visual condition (d) and in the auditory condition (e). Power decreases are plotted as t-values of average difference at 8 Hz between 0 and 4 seconds during retrieval. (f-g) Reconstruction of 8 Hz power difference in source space using an 'lcmv' beamforming-algorithm in the visual (f) and in the auditory condition (g).

Successful memory is associated with decreases in signal stationarity

In order to represent information in oscillatory phase, a signal cannot be stationary over time. If the time course of activity is relevant for the neural representation of content, one would rather expect frequent phase resets (i.e. a complex signal) going along with the representation of rich information. To test this requirement, the stationarity of the signal was estimated (see methods). Deviation from stationarity, i.e. the complexity of the signal was evaluated within each cycle of a frequency using a sliding window approach.

To measure the signal complexity, the distribution of phase-values within each cycle of a frequency was quantified (see methods). Low stationarity values reflect a more complex waveform, i.e. higher deviation from a uniform distribution. Average stationarity during the whole retrieval interval was then computed for every trial at every frequency.

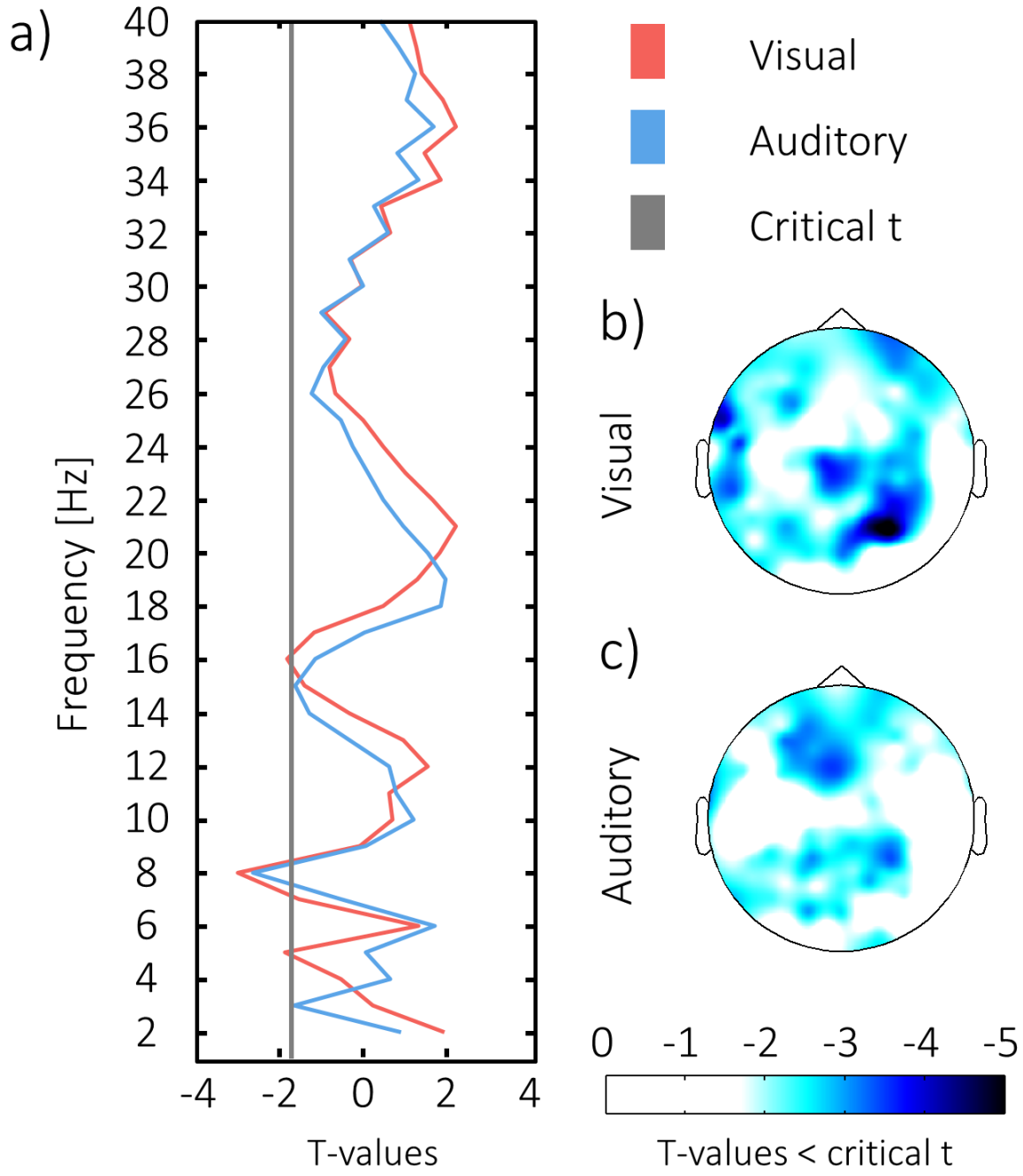
Comparing High Confidence Hits with Correct Rejections, a stationarity decrease at 8 Hz, i.e. an increase in signal complexity became evident in the visual condition (t_{23} = -3.01, P = 0.003, Figure 105a) and in the auditory condition (t_{23} = -2.63, P = 0.007, figure 105a), which was consistent with the power results.

The topography of 8 Hz stationarity decreases was broad, spanning most of the scalp in the visual condition (Figure 105b). Maxima were located over right parietal and left temporal scalp regions. In the auditory condition (Figure 105c) the topography spanned frontal, central, parietal and occipital regions, displaying maxima over fronto-central and parietal scalp areas.

To further examine the correlates of memory reinstatement, the analysis of signal stationarity was repeated comparing High Confidence Hits with Misses. Again only subjects were included

in the comparison, for which at least 15 trials remained after preprocessing (visual session: N = 19, auditory session: N = 17).

Figure 105: Stationarity, Hits minus Correct rejections



Legend to figure 105: Stationarity, Hits minus Correct rejections

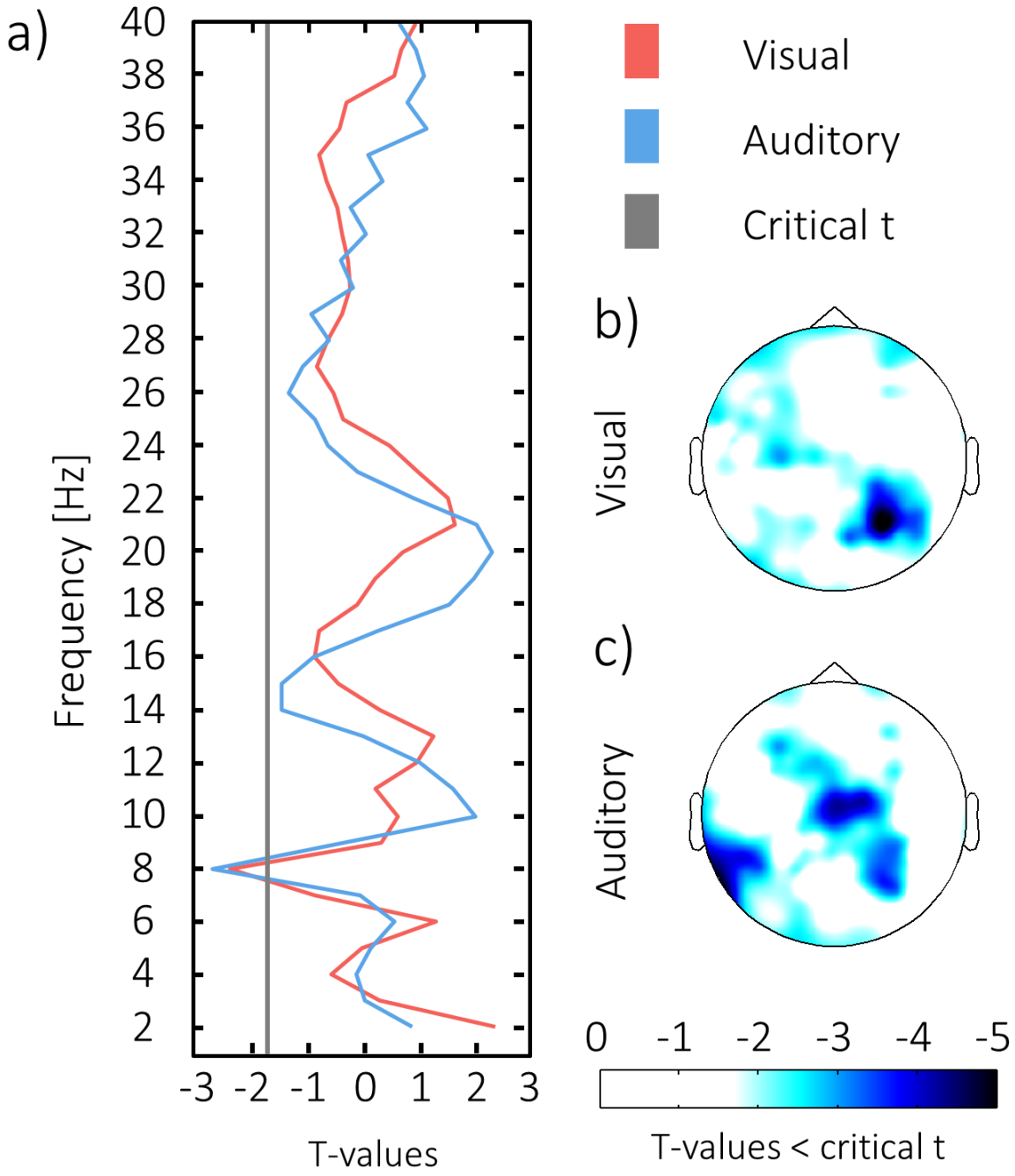
Successful memory reinstatement was associated with decreases in the stationarity of the signal. In the contrast of Hits and Correct Rejections, the decrease peaked at 8 Hz (a). Topographies of differences in 8 Hz stationarity are shown on the right in the visual (b) and in the auditory condition (c).

Results from the comparison of High Confidence Hits and Misses are displayed in figure 106. Statistical testing of the average difference revealed a decrease in signal stationarity at 8 Hz (an increase in signal complexity) when content was successfully reinstated. This effect was found in the visual (t_{19} = -2.4, P = 0.013, figure 106a) and in the auditory condition (t_{17} = -2.63, P = 0.009, figure 106a). The topography of 8 Hz stationarity decreases was again broad. In the visual condition it spanned occipital, parietal, left temporal and left frontal scalp regions. A maximum was located over right parietal regions (Figure 106b). In the auditory condition, the topography spanned frontal, central, parietal and occipital regions, displaying maxima over fronto-central, parietal and occipital scalp areas. Maxima were located over parieto-central regions (Figure 106c).

The result that power decreases go along with higher complexity of the signal, could reflect the hypothesized increase in information coding potential. However, the inherent relationship between power and signal-to-noise ratio of phase estimates needs to be considered.

As demonstrated by a simple simulation, under a constant level of noise, lower power leads to a less reliable estimate of oscillatory phase and therefore decreases stationarity, and increases signal complexity (see next section). Therefore, due to the possible influence of noise, higher signal complexity is a necessary, but not a sufficient condition to conclude that phase resets carry temporal information. For these reasons, only the temporal pattern similarity analysis can provide the crucial link between phase patterns and information coding, i.e. determine whether phase trajectories carry content-specific stimulus information during memory replay.

Figure 106: Stationarity, Hits minus Misses



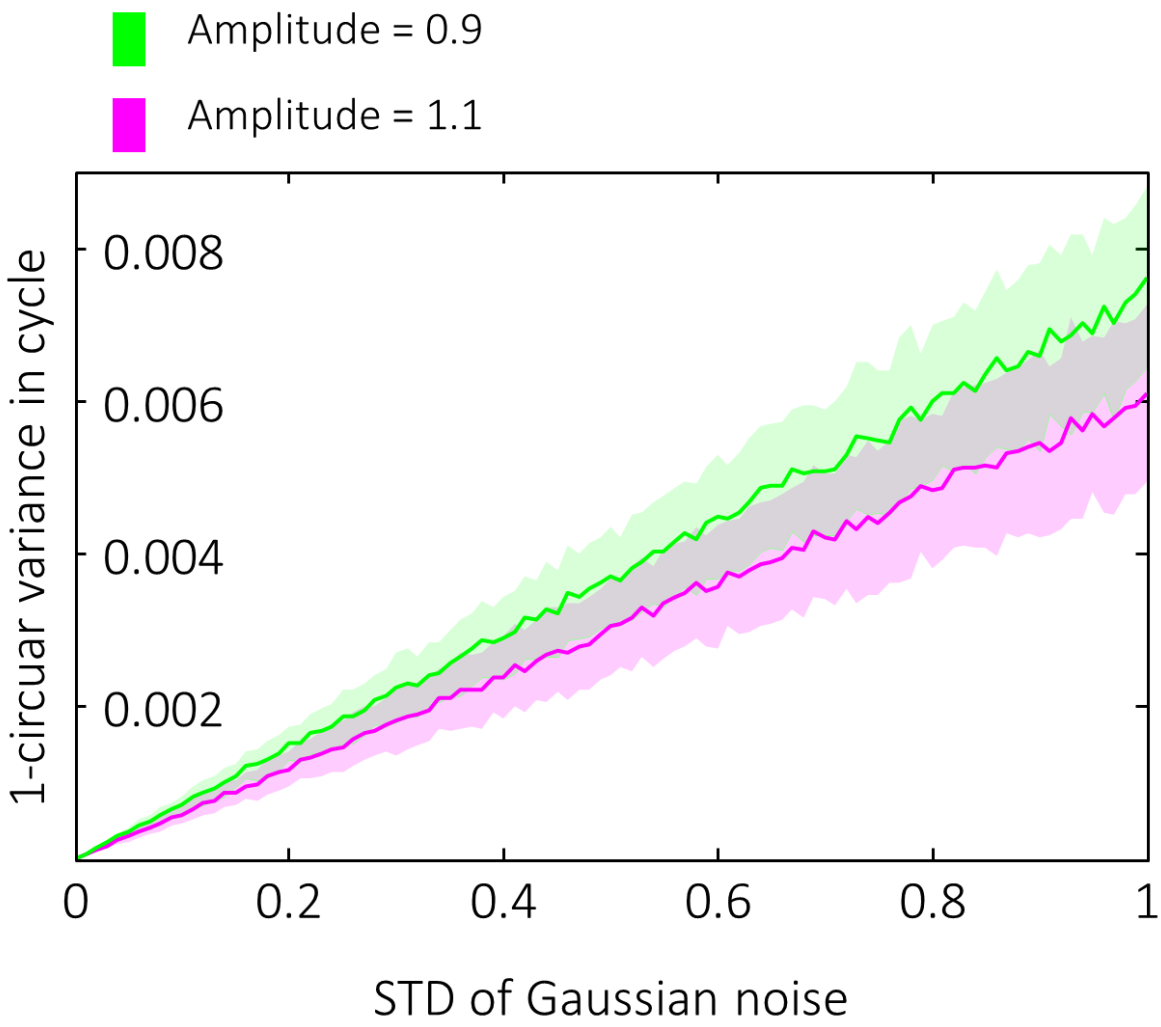
Legend figure 106: Stationarity, Hits minus Misses

Successful memory reinstatement was associated with decreases in the stationarity of the signal. In the contrast of Hits and Misses, the decrease peaked at 8 Hz (a). Topographies of differences in 8 Hz stationarity are shown on the right in the visual (b) and in the auditory condition (c).

Simulation of the relationship between Power and Phase stationarity

To illustrate the influence of signal to noise ratio (SNR) on the relationship between power decreases and phase stationarity, 2 sine waves of 8 Hz frequency, with amplitudes of 1.1 and 0.9 were created. Their phase at 8 Hz was extracted using a complex Morlet wavelet (see methods). As a measure of phase stationarity, the circular variance within each cycle of a sliding window was then computed over 4 seconds of the decomposed signal for 100 simulated trials. In subsequent steps, the data was masked with increasing levels of Gaussian noise. The stationarity of the signal was affected by the noise; the circular variance within the sliding window decreased with increasing noise, i.e. the signal appeared more complex. Importantly this effect was stronger for the signal with the lower amplitude (Figure 107). Therefore, differences in circular variance need to be interpreted with caution when two signals with different amplitude are contrasted.

Figure 107: Simulation of the relationship between signal amplitude, signal complexity (1-circular variance) and noise



Legend to figure 107: Simulation of the relationship between signal amplitude, signal complexity (1-circular variance) and noise

Two sine waves of different amplitude (0.9 green, 1.1 pink) were distorted by adding Gaussian noise of increasing strength. The standard deviation of that noise is plotted on the x-axis. The signal complexity of the distorted signal was subsequently assessed as 1 – circular variance and is plotted on the y-axis. The complexity of the signal with lower amplitude is more strongly affected by noise, demonstrating that signal complexity cannot be assessed independently of oscillatory power.

Temporal patterns differentiate between content during encoding

An important requirement for the detection of replay of temporal patterns during memory retrieval is that the stimulus content itself elicits a distinct time course of activity in the first place, i.e. while being perceived during encoding. In order to test this prerequisite, a modified version of the Pairwise Phase consistency (PPC) (Vinck et al., 2010) was contrasted between pairs of trials in which the same content was presented and pairs of trials that were of different content. This method assesses the degree of phase similarity that is specifically shared by trials that are instances of the same stimulus (i.e. content specificity of phase).

Content specificity of phase was assessed for every frequency band between 1 and 40 Hz. The time window for statistical testing was chosen between 500ms pre-stimulus and 3500ms after stimulus onset, to account for the temporal smearing of the wavelet decomposition. Importantly, the combination of trials was carefully balanced to avoid any possible bias (see methods). After correction for multiple comparisons, significant differences were obtained in both conditions in the form of two broad clusters in the visual ($P < 0.001$, $P = 0.003$, figure 108a) and one broad cluster in the auditory condition ($P < 0.001$, figure 108b). Importantly the clusters included 8Hz, which showed the strongest memory effects during replay (see above).

We hypothesized to later find reappearing temporal patterns in the frequency band of 8 Hz during retrieval, furthermore content specificity during encoding is a requirement for the detection of these patterns. For these reasons, temporospatial clusters in the data were now identified, in which the 8 Hz time course was maximally content specific. Hence, the statistical analysis was now restricted to 8 Hz only. After multiple comparison correction, the cluster at

encoding in which content could most reliably be differentiated (i.e. the cluster with the lowest p-value) was selected for further analysis.

In the visual condition, this cluster was identified between -152ms and 564ms ($P < 0.001$). Note that post-stimulus effects are temporally smeared into the pre-stimulus interval due to wavelet filtering. One further cluster was observed between 2650ms and 3300ms ($P = 0.016$).

In the auditory condition, the most reliable cluster of content specificity was identified in a time window between 22ms and 871ms ($P = 0.002$). Two further clusters were observed ranging from 1818ms to 2627ms ($P = 0.003$) and from 1203ms to 1504ms ($P = 0.047$). Therefore in both domains early and later time windows distinguished between different stimuli, reflecting the dynamic nature of the stimulus material.

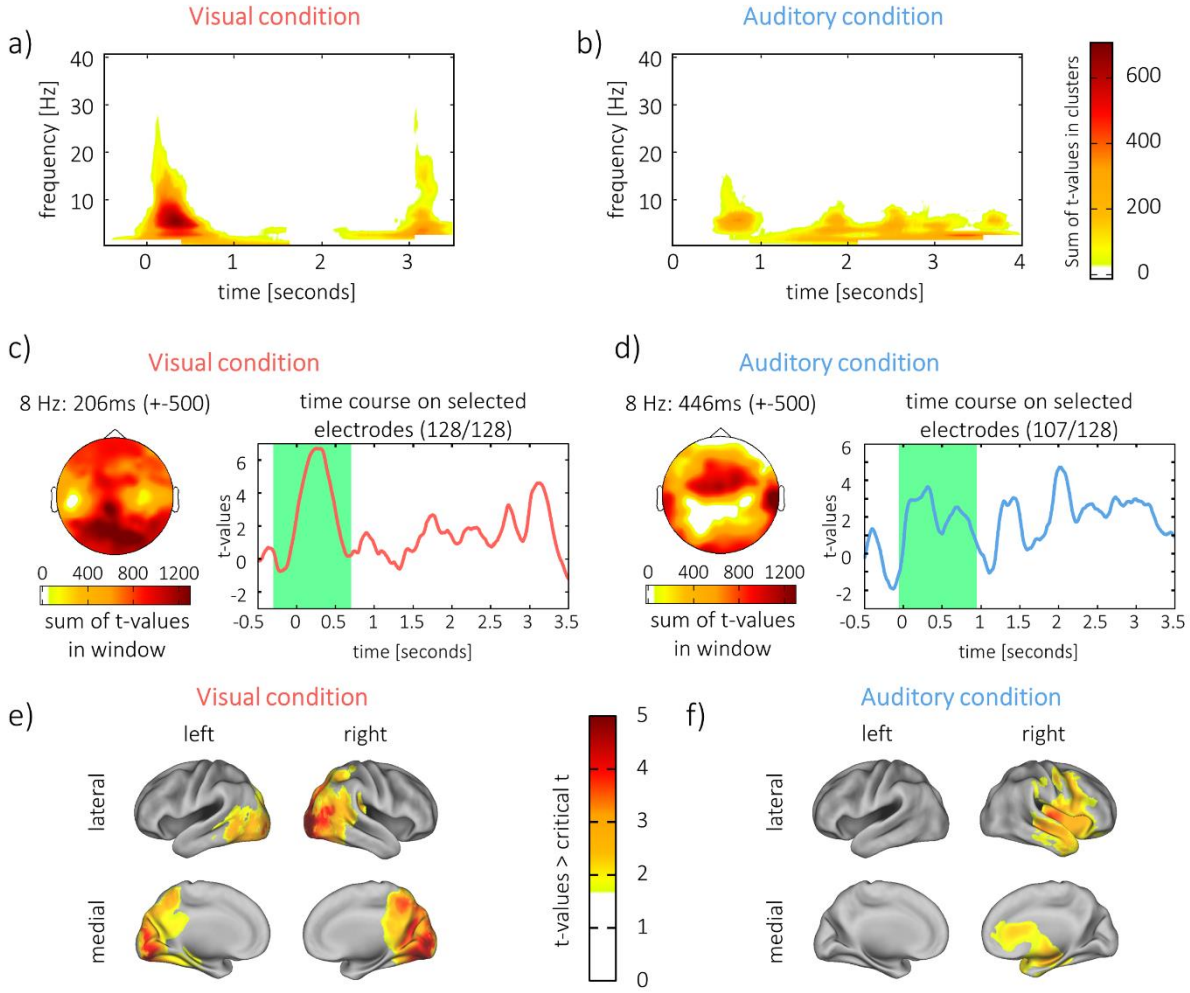
A 1 second time window was then defined around the centre of the most content specific cluster. In the visual condition, this centre was located at 206ms, thus the window ranged from -294ms to 706ms (figure 108c, right). Differences in phase similarity between trials of same and different content showed a clear visual topography within this window, i.e. the highest t-values were observed over posterior regions of the scalp (figure 108c, left).

In the auditory condition, the 1 second window was centred at 446ms (figure 108d, right), ranging from -54ms to 946ms. The topography of differences within that window showed a typical auditory distribution (figure 108d, left), i.e. high t-values were observed at fronto-central electrodes (Goff, Matsumiya, Allison, & Goff, 1977).

Sources of the average difference in phase similarity between same and different content combinations were reconstructed for the 1 second windows, to identify the origin of content specificity in that window. T-tests were corrected for multiple comparisons on the whole brain level. In the visual condition, a cluster of significant difference ($P < 0.001$) emerged in

visual regions of the cortex, covering the occipital lobe as well as parts of the parietal lobe (figure 108e). The cluster exhibited a peak in the right middle occipital gyrus (MNI: 10; -100; 10; BA: 18). In the auditory condition, differences in similarity (P = 0.004) were lateralized to the right hemisphere which is in line with studies finding lateralization of musical processing to this hemisphere (Tervaniemi et al., 2000). Differences covered temporal and frontal areas, including primary and secondary auditory processing regions (figure 108f). The auditory cluster peaked in right sub-lobar insula (MNI: 40; -20; 0).

Figure 108: Content specificity of phase during encoding



Legend to figure 108: Content specificity of phase during encoding

Broad clusters of difference in phase similarity across time, frequency and electrodes were observed a) in the visual condition and b) in the auditory condition. T-values were summed across electrodes in the cluster to display the results. c-d) Topography of 8 Hz content specificity and time course of differentiation c) in the visual condition and d) in the auditory condition. The time course (c-d, right) is the t-statistic of averaged content specificity, across all electrodes that are included in the strongest cluster. The green window marks the time window of 1 second around the centre of the strongest cluster. e) Source reconstruction of content specificity in the visual condition and f) in the auditory condition. For consistency, sources were tested and plotted on the averaged similarity across the encoding time-window of 1 second around the peak of the strongest cluster (green window). Likewise topographies (c-d, left) show summed t-values on the scalp of content specificity across this time window.

Temporal patterns indicate replay of visual and auditory content

The crucial quest to identify a replay of temporal patterns from encoding during retrieval is challenged by the non-time-locked nature of retrieval. Indeed, replay of memory content during retrieval could happen at any point after presentation of the retrieval-cue, with the exact onset varying from trial to trial. Moreover we assumed that any temporal pattern from encoding could be replayed at any time during retrieval.

We therefore developed a procedure that is not affected by these time shifts; specifically we assessed the similarity between encoding and retrieval with a sliding window approach. To this end, phase similarity between combinations of encoding and retrieval time-windows was computed using a variation of the single-trial Phase Locking Value (S-PLV) (Lachaux et al., 2000; Mormann et al., 2000), namely the similarity of phase angle differences over time (see methods). This method is less susceptible to noise and allows for an estimation of similarity between two time windows in non-time-locked data. Again, phase-similarity of encoding-retrieval pairs that were of same content (e.g. perceiving A, remembering A), was contrasted with the similarity of pairs, that were of different content (e.g. perceiving B, remembering D).

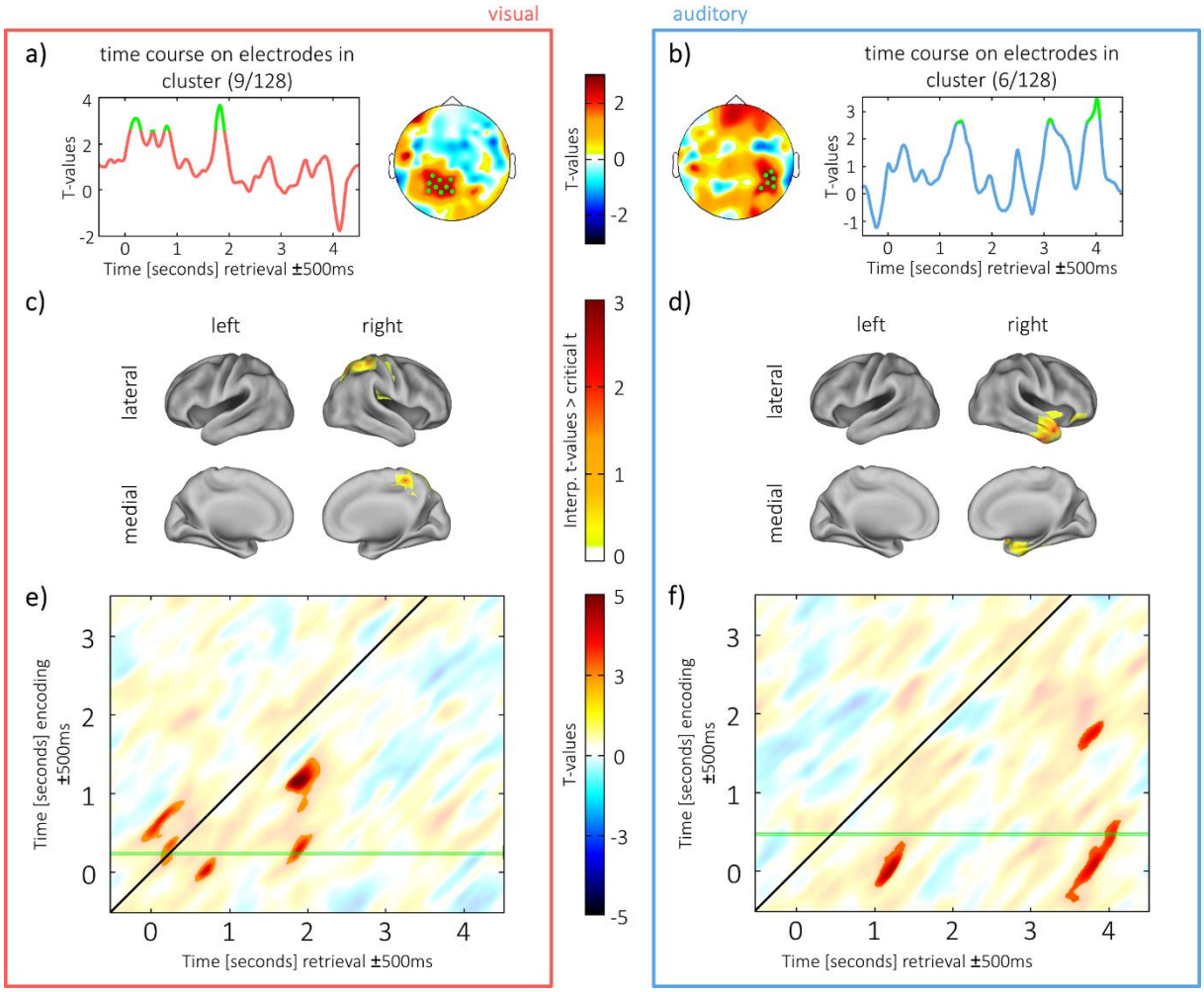
The time window that contained the temporal pattern from encoding was selected based on the highest content specificity of phase during encoding (see above). The width of the window amounted to 1 second (8 cycles) around the centre of the cluster that was located at 206ms (-294 to 706ms, see figure 108c, right) in the visual condition and at 446ms in the auditory condition (-54 to 946ms, see figure 108d, right). Since activity at encoding, under the null-hypothesis, would be independent from activity at retrieval, the specific selection of a

time window, based on results from encoding, can be used to increase the signal to noise ratio, without risking circular inference.

Phase-similarity to this pre-defined encoding window was now assessed by sliding the window over the whole retrieval episode. To slide the window into retrieval the prestimulus interval between -500ms and 0ms was used as padding. To slide it out at the end of the trial, the prestimulus interval between -1000ms and -500ms was used as padding (this was done because later time points were unsuitable for padding due to contamination with similar perception and responses). Note that the similarity at time point 0 is then assessed by comparing the encoding window to the retrieval window between -500ms and 500ms and similarity, at 4 seconds is assessed by comparing the encoding window to the concatenated retrieval window of 3500ms to 4000ms and -1000ms to -500ms.

The phase-similarity of the encoding window to episodes of replay of the same video/sound was now contrasted with the phase-similarity to episodes of replay where a different content was replayed from memory. The t-statistic was computed for every electrode on the averaged difference between same vs. different combinations between 0 and 4 seconds. A cluster-based permutation test indicated replay of encoding phase patterns during retrieval for both the visual (P = 0.002) and the auditory condition (P = 0.01). In the visual replay condition, the cluster of significant differences emerged over left parietal regions (figure 109a, right). In the auditory replay condition, a cluster of significant differences was observed over right posterior temporal areas (figure 109b, left). This signifies strong evidence for mnemonic replay of temporal patterns.

Figure 109: Encoding-Retrieval similarity



Legend to figure 109: Encoding-Retrieval similarity

a) Topography of visual cluster and time course on electrodes in the cluster. b) Topography of auditory cluster and time course on electrodes in the cluster. Electrodes in the cluster and time points that exceed threshold are highlighted in green. (c-d) Source reconstruction of encoding-retrieval similarity between 0 and 4 seconds of retrieval in the visual condition (c) and in the auditory condition (d). Statistical testing was run unrestricted on the whole brain level and for each condition the maximal cluster (i.e. with the highest summed t-values) was plotted. (e-f) Encoding-Retrieval similarity in the cluster (from a-b) between every time point of encoding and every time point of retrieval in the visual condition (e) and in the auditory condition (f). Clusters of significant differences are unmasked; remaining data is masked with transparency. The temporal imprecision of ±500ms is due to the width of the sliding window. The green lines in e and f highlight the original time window at encoding that is displayed in a-b (for e-f).

To test for frequency specificity, the same analysis was performed for 5 Hz and 13 Hz which are approximately in a golden ratio relationship (Pletzer, Kerschbaum, & Klimesch, 2010) to 8 Hz (i.e. maximally different in phase). Two further control frequencies were tested that showed peaks in power decreases in at least one of the conditions, namely 4 and 15 Hz. To this end, time windows from encoding were selected with the same criteria as for 8 Hz; electrodes for testing were again restricted to the electrodes in the significant cluster from encoding. Furthermore, the time window was likewise built from 8 cycles of the corresponding frequency. However, no effects were found in the visual or in the auditory condition for any of the control frequencies, suggesting that temporal reinstatement of phase patterns was specific to 8 Hz.

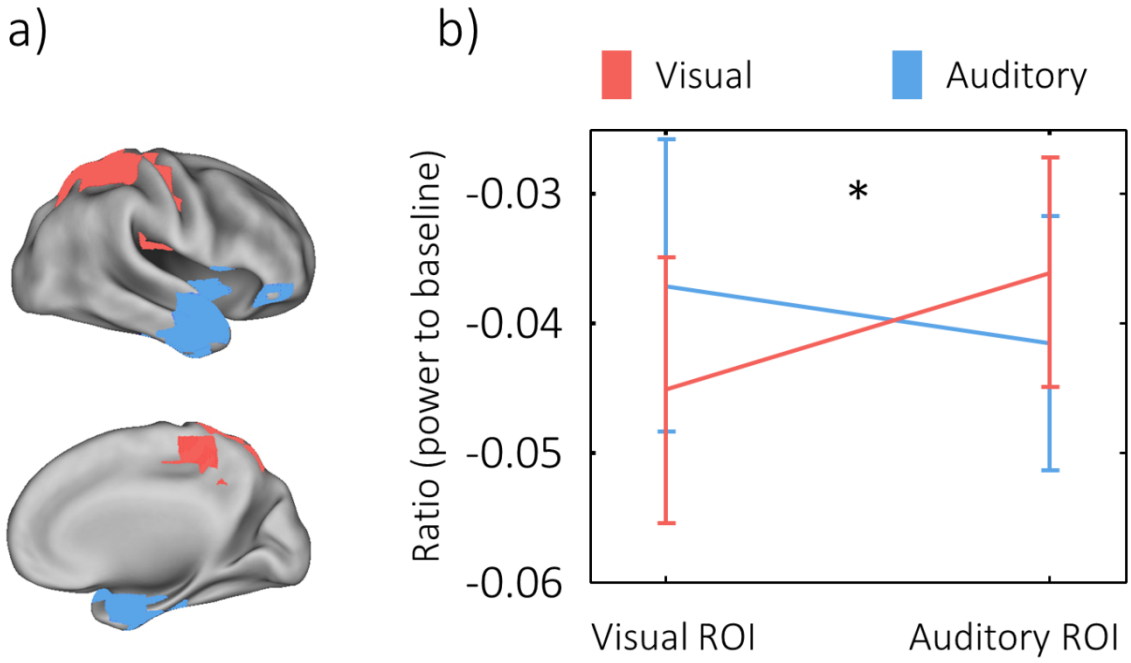
The temporal profile of the replay effect was then inspected by averaging phase similarity across electrodes within the cluster of significant differences. A t-test was computed at every time point, applying a probability of error below 0.01. For visual material, 4 episodes of replay could be identified (figure 109a, left), in which a one sided test exceeded the critical threshold (t_{23} = 2.5). These episodes peaked at 203ms (t_{23} = 3.09, P = 0.003), 547ms (t_{23} = 2.51, P = 0.01), 828ms (t_{23} = 2.75, P = 0.006) and 1844ms (t_{23} = 3.65, P < 0.001). For the auditory material, 3 episodes exceeded the critical t-value (figure 109b, right) peaking at 1406ms (t_{23} = 2.64, P = 0.007), 3125ms (t_{23} = 2.7, P =0.006) and 4016ms (t_{23} = 3.47, P = 0.001).

To reveal whether the encoding-retrieval similarity effects were maximal in material specific (i.e. visual/auditory) brain regions, encoding-retrieval similarity was assessed on the source level. Statistical testing was run unrestricted on the whole brain level and for each condition the maximal cluster (i.e. with the highest summed t-values) was plotted. For the visual material, the strongest cluster of encoding-retrieval similarity showed a peak in the superior

parietal lobule (MNI: 20; -50; 60, BA: 7, see figure 109c), overlapping with the similarity effects during encoding (compare to figure 108e) and in line with studies finding parietal lobe contributions to episodic memory retrieval (Rugg & Vilberg, 2013; Wagner, Shannon, Kahn, & Buckner, 2005). For the auditory material, similarity effects showed a peak in the right inferior temporal gyrus (MNI: 50; -10; -44, BA: 20, see figure 109d) also overlapping with the similarity effects during encoding (compare figure 108f) and in line with previously reported effects on memory for music (Groussard et al., 2010).

Since power decreases at 8 Hz spanned multiple brain regions in the visual and in the auditory condition (compare: figure 103 f-g), content specific decreases were still statistically unsubstantiated. In order to link phase-based similarity at 8 Hz with the power decreases during memory replay, we therefore compared the power-difference at 8 Hz between Hits and Correct Rejections (see above) within the regions of visual and auditory "replay". We computed a 2x2 ANOVA contrasting 8Hz power decreases on the source level, with the factors Region (visual/auditory) and Condition (visual/auditory). If power decreases are relevant for information coding, stronger power decreases should be observed in those sensory regions where replay occurred. This hypothesis was confirmed by a significant interaction ($F_{1,23}$ = 6.58, P = 0.017, see figure 110), showing that power decreases in the auditory region of interest were stronger during replay of auditory memories, whereas power decreases in the visual region of interest were stronger during visual memory replay.

Figure 110: The interaction of power decreases with memory replay



Legend to figure 110: The interaction of power decreases with memory replay

Regions of visual and auditory similarity (a) showed a significant interaction with conditions ($F_{1, 23}$ = 6.58, P = 0.017), such that power decreases in the auditory region of similarity were stronger in the auditory condition and power decreases in the visual region of similarity were stronger during visual retrieval (b).

To obtain a further understanding of the temporal dynamics of memory reinstatement, a follow up analysis within the electrode-clusters of significant differences was run for all combinations of retrieval and encoding time windows, resulting in retrieval time - encoding time diagrams (see figure 109e-f). It should be acknowledged that further analyses on this cluster will be biased towards being optimal for the time window, on which the electrodes were originally identified. Therefore the results are likely to show more reinstatement of phase patterns from early encoding. Primarily this analysis reveals which parts from the original sliding window (centred on the most content specific cluster from encoding), maximally contributed to the effect, when we tested for content specificity of reactivation (e.g. mostly activity from the beginning of the window). Moreover, on a descriptive level this analysis gives an idea about which phase patterns from encoding, in addition to the early ones, were also reactivated during retrieval.

It is noteworthy to keep two issues in mind when interpreting these plots. Firstly, similarity between two windows will always express temporal smoothing on the diagonal. The diagonal width can be seen as an indicator of the length of the episode that was replayed, but it is also affected by the length of the sliding window (i.e. longer windows will induce more smearing along the diagonal). Secondly the peak in these diagonals indicates which temporal pattern at encoding was actually replayed at which retrieval time point. When two time windows are aligned and they share a temporal pattern in their first quarter, this pattern would appear temporally delayed in a one-dimensional plot; however in two dimensions we can inspect the diagonal peak of similarity.

In the visual condition, a permutation test revealed significant differences in 5 clusters. The peaks of the clusters suggested that early, around 141ms, during retrieval, activity from

672ms during encoding was reinstated (P = 0.017). At 266ms of the retrieval interval, encoding patterns from around 359ms reappeared (P < 0.002), around 719ms, phase patterns from 31ms during encoding were reinstated (P = 0.004). Later, during retrieval at 1859ms, the phase patterns from 672ms during encoding were detected (P = 0.012) and at 1859ms, the activity from 1172ms during encoding showed a similarity effect (P = 0.022).

In the auditory condition, only 3 clusters could be identified. Peaks within the clusters suggested that 1203ms after the onset of the retrieval-cue, content from 15ms during encoding was replayed (P = 0.01). Later, at 3781ms, activity from 78ms at encoding reappeared (P = 0.017) and finally at 3797ms into the retrieval time, late encoding phase patterns from 1765ms could be detected (P = 0.014). Even though results are biased towards detecting replay from the early encoding window that served to identify the electrodes on which memory replay took place, this analysis could still give an idea of the temporal dynamics of reinstatement and show the potential of our method. It was observed that later encoding patterns did not appear until later in the retrieval episode, suggesting an ordered replay. Furthermore reactivation of visual patterns was observed very early, as was expected given recent evidence for early reactivation (Waldhauser, Braun, & Hanslmayr, 2016; Wimber et al., 2012), and notably earlier than reactivation of auditory patterns, which is in line with a worse memory performance of participants in the auditory condition of this study.

Discussion

In real life, most of our episodic memories are dynamic with an inherent temporal structure and are not bound to a single modality. We can re-experience information-rich memory traces with auditory and visual content and habitually reinstate these events with an abundance of subjective impressions in their correct temporal order. Although some of these temporal aspects of memory replay have been investigated in spatial navigation experiments in rodents (Ji & Wilson, 2007; Nádasdy, Hirase, Czurkó, Csicsvari, & Buzsáki, 1999; Skaggs & McNaughton, 1996), the temporal properties of episodic memory replay in humans were largely ignored in previous research. Consequently little is known about the neural mechanisms that orchestrate the replay of dynamic memories in humans.

In the present study, we identified content-specific temporal signatures of individual memories in the visual *and* in the auditory domain. These signatures were specific to a carrier frequency of ~8 Hz and could be localized to modality-specific regions, i.e. overlapping with those regions that carried the information of the stimuli during encoding. Strikingly, the 8 Hz frequency also showed the strongest power decrease during retrieval in both modalities. Likewise the power decrease in 8 Hz during retrieval was modulated in a sensory specific manner in those regions where memory replay took place, i.e. stronger power decreases in the parietal (visual) region during replay of videos and vice versa for replay of sounds in the temporal (auditory) region (see figure 110). These findings provide a link to other studies where a similar interaction between alpha power decreases and oscillatory phase has been proposed to temporally structure perceptual contents (Jensen et al., 2012). In line with these

findings our results suggest that similar oscillatory mechanisms which guide perception also guide the “re-perception” – that is memory replay – of these sensory events.

In order to detect the reinstatement of temporal neural patterns that indicate such replay of individual memories, we developed a novel dynamic phase based RSA method which is robust against variations in the onset of memory replay. This method can therefore be applied in conditions when the exact time point of the reinstatement of a neural pattern is unknown, like for example during offline replay in resting state or sleep.

RSA has been previously used to track episodic memories in EEG/MEG (Jafarpour et al., 2014; J. D. Johnson et al., 2015; Kurth-Nelson et al., 2015; Ng et al., 2013; Staudigl et al., 2015; Wimber et al., 2012) and iEEG (Yaffe et al., 2014; Zhang et al., 2015), however some important differences to these studies have to be considered. Firstly, we go beyond mere classification of memory content, since we use similarity measures to test a mechanistic hypothesis: that alpha power decreases are associated with the reinstatement of temporal patterns. Hence we can test whether temporal patterns reappear during retrieval and we can link this replay to a specific frequency band. Importantly the detection of temporal patterns was only made possible with our dynamic RSA approach.

Secondly, in our design we carefully avoided any sensory overlap between encoding and retrieval. We were therefore able to investigate mechanisms of purely memory driven reinstatement, as opposed to studies in which there was a high overlap in sensory stimulation between encoding and retrieval (Ritchey, Wing, LaBar, & Cabeza, 2012; Staudigl et al., 2015; Zhang et al., 2015). This aspect of our experimental design allows us to conclude that the brain actively reproduces a temporal pattern which is specific to a stimulus in order to re-experience this particular memory.

An important open question concerns how the hippocampus is involved in the replay of temporal patterns in the cortex as observed here. A critical involvement of the hippocampus, and the phase of theta oscillations therein, for memory replay is implicated by recent models and frameworks (Hanslmayr et al., 2016; Hasselmo, 2015; Ketz, Morkonda, & O'Reilly, 2013). Future studies are required that record simultaneously from both the hippocampus and the neocortex to investigate how the reinstatement of the temporal phase patterns described here interact with, or rely on, the hippocampus.

Studying the temporal aspects of memory replay has proven to be difficult because methods or stimulus material in previous studies did not allow investigating this question. Overcoming these previous limitations, we identified a potential domain general mechanism that orchestrates the replay of dynamic auditory and visual memories in humans. Specifically, our findings suggest an intimate relationship between power decreases in an 8 Hz frequency and a content-specific temporal code, carried by its phase. These results corroborate recent theories linking power decreases with the coding of neural information (Hanslmayr et al., 2016, 2012; Jensen & Mazaheri, 2010). Our findings open up new ways of investigating the temporal properties of memory replay in humans, which we only begin to understand.

Materials and Methods

Participants

24 healthy, right-handed subjects (18 female and 6 male) volunteered to participate. 7 further participants were tested, or partly tested, but could not be analysed due to poor memory performance (N=2), misunderstanding of instructions (N=2) and poor quality of EEG-recording and technical failure (N=3). All participants had normal or corrected-to normal vision. The average age of the sample was 23.38 (s.d. = 3.08) years. Participants were native English speakers (20), bilingual speakers (2) or had lived for more than 8 years in the UK (2). Ethical approval was granted by the University of Birmingham Research Ethics Committee, complying with the Declaration of Helsinki. Participants provided informed consent and were given a financial compensation of 24£ or course-credit for participating in the study.

Material and experimental set up

The cues amounted to 360 words that were downloaded from the MRC Psycholinguistic Database (Coltheart, 1981). Stimulus material consisted of 4 video clips and 4 sound clips in the visual and auditory session respectively. All clips were 3 seconds long; videos showed coloured neutral sceneries with an inherent temporal dynamic, sounds were short musical samples, each played by a distinct instrument. In both sessions a clip was associated with 30 different words. 60 words were reserved for the distractor trials and 12 additional words were used for instruction and practice of the task. For presentation, words were assigned to the clips or to distractors in a pseudorandom procedure, such that they were balanced for Kucera-Francis written frequency (mean = 23.41, s.d. = 11.21), concreteness (mean = 571, s.d.

= 36), imageability (mean = 563.7 s.d. = 43.86), number of syllables (mean = 1.55, s.d. = 0.61) and number of letters (mean = 5.39, s.d. = 1.24). Furthermore lists were balanced for word-frequencies taken from SUBTLEXus (Brysbaert & New, 2009). Specifically, "Subtlwf" was employed (mean = 20.67, s.d. = 27.16). The order of presentation was also randomized, assuring that neither the clips and their associates, nor distractor words were presented more than 3 times in a row or in temporal clusters. The presentation of visual content was realized on a 15.6 inch CRT-monitor (Taxan ergovision 735 TC0 99) at a distance of approximately 50 centimetres from the subjects eyes. The monitor refreshed at a rate of 75 Hz. On a screen size of 1280 x 1024 pixels, the video clips appeared in the dimension of 360 pixels in width and 288 pixels in height. 'Arial' was chosen as the general text-font, but font-size was larger during presentation of word-cues (48) than during instructions (26). In order to reduce the contrast, white text (rgb: 255, 255, 255) was presented against a grey background (rgb: 128, 128, 128). Auditory stimuli were presented using a speaker system (SONY SRS-SP1000). The 2 speakers were positioned at a distance of approximately 1.5 meters in front of the subject with 60 centimetres of distance between the speakers.

Procedure

Upon informed consent and after being set up with the EEG-system, participants were presented with the instructions on the screen. Half of the subjects started with the auditory session, the others were assigned to undertake the visual task first. Both sessions consisted of a learning block, a distractor block and a test block. The sessions were identical in terms of instructions and timing and differed only in the stimulus material that was used. During instruction, the stimulus material was first presented for familiarization and then used in

combination with the example words to practice the task. Instructions and practice rounds were completed in both sessions.

As a way to enhance memory performance, participants were encouraged to use memory strategies. The suggestion was to imagine the word in a vivid interaction with the material content, yet the choice of strategy remained with the subject. In the learning block, 120 clip-word sequences were presented. Each sequence started with a fixation cross that was presented in the centre of the screen for 1 second, then the video-clip played for 3 seconds. In the auditory condition, the fixation cross stayed on the screen and the sound-clip played for 3 seconds. Immediately after the clip, a word cue was presented in the centre for 4 seconds, giving the subject time to learn the association. After that, an instruction requested to subjectively rate on a 6 point scale how easy the association between the clip and the word was. After a press on the space bar, this scale was shown. Equidistant categories were anchored with the labels “very easy” and “very hard”; those labels were displayed at both ends above the scale. Participants used six response buttons to rate the current association (see figure 102).

In the distractor block, subjects engaged in a short unrelated working memory task, namely they counted down in steps of 13, beginning from 408 or 402 respectively. After 1 minute the distractor task ended. Following a short self-paced break, subjects refreshed the instructions on the retrieval block.

In this block, either a cue or a distractor was presented upon a button press on the space bar. Subjects were instructed to try to vividly replay the content of the corresponding video-clip or sound-clip in their mind upon presentation of the cue. The word stayed on the screen for 4 seconds, giving the subject the opportunity to replay the memory. Finally, a fixation cross was

presented for a varying time window between 250 and 750 milliseconds to account for movement and preparatory artefacts, before the response scale appeared on the screen.

The response-scale consisted of 6 options. 4 small screen shots of the videos or 4 black and white pictures of the featured instruments were presented in equidistant small squares of 30x30 pixels. Additionally, the options “new” and “old” were displayed in the form of text at the most left and most right position of the scale (see figure 102 c-d). Subjects could now either indicate the target (video/sound) they just replayed, by pressing the button corresponding to that clip. Instead, subjects could also indicate that the word was a distractor by pressing the button corresponding to the option “new”, or they would simply indicate that they remembered the word, but could not remember the clip it was associated with. In this last scenario subjects would press the button corresponding to “old”. The positions of “old” and “new” at the end of the scale, as well as the permutation of the 4 target positions in the middle of the scale were counterbalanced across participants. Finally, after making a decision, a further six point rating scale was presented on which subjects could rate the confidence in their response. Again a scale with equidistant categories was presented ranging from “guess” to “very sure”. An additional possibility was to press “F2” in case of an accidental wrong button press. In this case, the whole trial was discarded from analysis. Following the retrieval block, individual electrode positions were logged allowing for a break of approximately 30 minutes before beginning the second session. In addition to the 2 experimental sessions, all participants came to a separate session to record anatomical MRI-scans at the Birmingham University Imaging Centre (https://www.buic.bham.ac.uk/). This was later used to facilitate source localization (see below).

Data Collection

The recording of behavioural responses and the presentation of instructions and stimuli were realized using Psychophysics Toolbox Version 3 (Brainard, 1997) with MATLAB 2014b (MathWorks) running under Windows 7, 64 Bit version on a desktop computer. Response buttons were “s, d, f, j, k, l” on a standard “QWERTY” layout. Buttons were highlighted and corresponded spatially to the response options on the screen, so participants didn’t have to memorize the keys. To this end, the shape of corresponding fingers was also displayed under the scale. To proceed, participants used the space bar during the experiment. Physiological responses were measured with 128 sintered Ag/AgCl active electrodes, using a BioSemi Active-Two amplifier, the signal was recorded at 1024 Hz sampling rate on a second computer via ActiView recording software, provided by the manufacturer (BioSemi, Amsterdam, Netherlands). Anatomical data was acquired using magnetic resonance imaging (MRI) (3T Achieva scanner; Philips, Eindhoven, The Netherlands), electrode positions were logged with a Polhemus FASTRAK device (Colchester, VT, USA) in combination with Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011) implemented in MATLAB.

Preprocessing

The data was preprocessed using the Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data was cut into trial-segments from 2 seconds pre-stimulus to 4.5 seconds after stimulus onset (i.e. onset of the clip at encoding and onset of the word at retrieval). The linear trend was removed from each trial and a baseline correction was applied based on the whole trial. Trials were then downsampled to 512 Hz and a band-stop filter was applied at 48-52, 58-62, 98-102 and 118-122 Hz to reduce line noise at 50 Hz

and noise at 60 Hz; additionally a low-pass filter at 140 Hz was applied. After visual inspection for coarse artefacts, an independent component analysis was computed. Eye-blink artefacts and eventual heartbeat/pulse artefacts were removed, bad channels were interpolated and the data was referenced to average. Finally, the data was inspected visually and trials that still contained artefacts were removed manually. MRI scans of each participant were segmented into four layers (brain, cerebrospinal fluid, skull and scalp) using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) in combination with the Huang toolbox (Huang et al., 2013). On this basis, a volume conduction model was created with the Fieldtrip 'dipoli' method; individual electrode positions were aligned to the head model for every participant.

Behavioural analysis

For behavioural analysis, correct trials were defined as those of the retrieval phase in which the target was correctly identified and the confidence rating of the response was high (5 or 6). Trials were defined as correct rejections if a distractor-word was correctly identified as new; misses were defined as trials in which a cue-word was incorrectly identified as a new word or the response "old" was given to indicate that the subject recognized the word, but could not remember the target video or sound it was associated with. Hits of low confidence were not considered in subsequent analyses. Furthermore, selections of the wrong clip as well as accidental presses of the wrong button and distractor trials that were not recognized as distractors were discarded from further analysis.

Power analysis

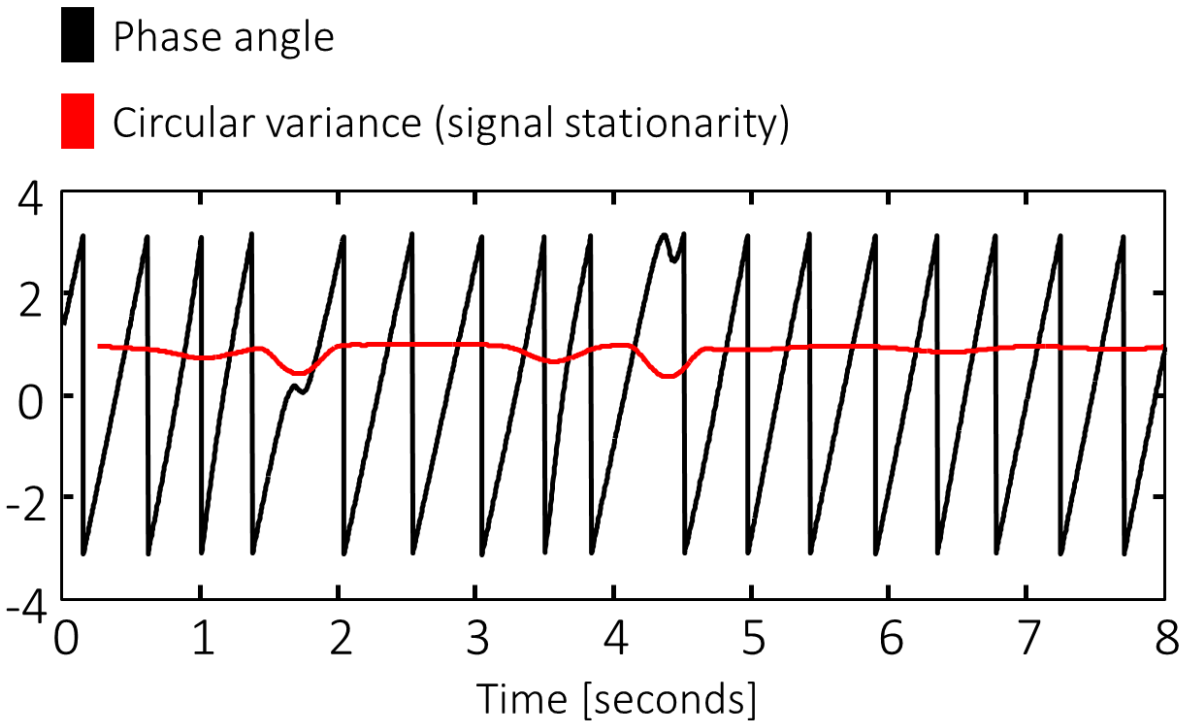
Power at retrieval was determined by multiplying the Fourier-transformed data with a complex Morlet wavelet of 6 cycles. Raw power was defined as the squared amplitude of the complex Fourier spectrum and estimated for every 4^{th} sampling point (i.e. sampling rate of 128 Hz). For each contrast (i.e. hits vs. misses, or hits vs. correct rejections), baseline normalization was performed separately. Therefore, a baseline was computed as the average power between -1 and 4 seconds of all trials within the contrast (Long et al., 2014). Every trial was then normalized by subtracting the baseline and subsequently dividing by the baseline $(activity_{tf} - baseline_f)/baseline_f$, where t indexes time and f indexes frequency. The relative power was calculated for all frequencies between 1 and 40 Hz.

Phase stationarity

For every frequency between 2 and 40 Hz, the stationarity of phase was defined within a sliding window of one cycle (see Supporting Information). Phase was estimated by multiplying the Fourier-transformed data with a complex Morlet wavelet of 6 cycles. The complex signal was then divided by its amplitude to standardize its power to 1. At every time point, the deviation from an even circular distribution within one cycle around this point was assessed, i.e. the circular variance (CV) of phase over time was computed. CV was interpreted as a measure of signal stationarity, since a perfectly stationary signal has an even distribution over one cycle and the circular variance within the cycle is maximal (i.e. reaches 1, see figure 111). Phase stationarity was baseline corrected in the same way as oscillatory power. A baseline was computed as the average stationarity between -1 and 4 seconds of all trials within the contrast. Every trial was then normalized again by subtracting the baseline and then dividing

by the baseline ($stationarity_{tf}$ – $baseline_f$)/$baseline_f$, where t indexes time and f indexes frequency.

Figure 111: Circular variance within a sliding window of one cycle as a measure of signal stationarity



Legend to figure 111: Circular variance within a sliding window of one cycle as a measure of signal stationarity.

The figure shows a time course of phase angles (black line) and its stationarity. The phase angles are taken from complex values of unit length, which approximate an oscillation at 2 Hz. These take values between π and –π on the y-axis. The circular variance within a sliding window of 1 cycle (i.e. 500ms for 2 Hz) describes the stationarity of this oscillation. When the signal is stationary and there are no phase resets, the circular variance reaches 1. A phase reset, on the other hand, causes a decrease in circular variance.

Content specificity of phase at encoding

While participants learned the associations in the encoding block, they repeatedly saw (heard) the same dynamic stimulus. Content specific properties could consequently be identified if they were shared by trials of the same content, but not by trials of a different content.

Hence, content specific phase was assessed by contrasting the phase-similarity between pairs of trials, in which the same content was presented, with the phase-similarity of an equal number of trial-pairs that were of different content. To achieve this, trials were grouped and combined in a random, but balanced way (see below). For each pair of trials, the cosine of the absolute angular distance was then computed and finally averaged across all (same or different) combinations (Vinck et al., 2010).

This resulted in an average similarity value at every time point, at every electrode and in every frequency of interest. This similarity was derived separately for the same pairs and for the different pairs and could consequently be subjected to statistical testing in order to define content specificity of phase.

Importantly, the way of combining the trials can result in bias. For this reason, the trial combinations were randomly selected in a carefully balanced way (figure 112). Firstly, the trials were grouped into four sets that were of the same content ($SE_{1\text{-}4}$), e.g. the same video. These sets were then recombined such that each set of content, say A, could be paired with a unique set of mixed content (say, B, C, and D) that was equal in size, i.e a contrast-set ($CE_{1\text{-}4}$). To make this possible, some trials were discarded from further analysis (figure 112a-b).

In order to form pairs of same content, all possible N*(N-1)/2 pairs within each of the four stimulus-sets (SE_{1-4}) were built. Then, to form pairs of different content, only N*(N-1)/2 pairs between the stimulus-set (SE_I) and its contrast-set (CE_I) were built. Importantly, wherever the second trial in the pairs of same content appeared in several combinations, it was replaced by instances of the same exclusive trial from the contrast set, while building the combinations of different content (figure 112c).

Figure 112: Trial combinations between same and different content during encoding

a)

$SE_1 = \{SE_{1T1}, SE_{1T2}, \dots, SE_{1TN}\}$

$CE_1 = \{SE_{2T11}, SE_{4T22}, \dots, SE_{3T3}\}$

$SE_2 = \{SE_{2T1}, SE_{2T2}, \dots, SE_{2TN}\}$

$CE_2 = \{SE_{1T12}, SE_{3T12}, \dots, SE_{4T17}\}$

…

$SE_4 = \{SE_{4T1}, SE_{4T2}, \dots, SE_{4TN}\}$

$CE_4 = \{SE_{3T29}, SE_{1T7}, \dots, SE_{2T4}\}$

b)

There were no common elements in stimulus groups and in contrast groups

$\forall_{I,J} \in \mathbb{N}\ (1 \le I,J \le 4 \wedge I \ne J \Rightarrow SE_I \cap SE_J = CE_I \cap CE_J = \emptyset)$

The stimulus group had the same size as its contrast group

$\forall_I \in \mathbb{N}\ (1 \le I \le 4 \Rightarrow |SE_I| = |CE_I|)$

c)

Same combinations Stimulus I

SE_I / SE_I	Trial 1	Trial 2	…	Trial N
Trial 1				
Trial 2				
…				
Trial n				

Different combinations Stimulus I

CE_I / SE_I	Trial r1	Trial r2	…	Trial rN
Trial 1				
Trial 2				
…				
Trial n				

Legend to figure 112: Trial combinations between same and different content during encoding

Each cell in the upper right regions (green and orange) indicates a pair of trials that are compared, i.e. cosine of angular distance. $SE_{k,Tm}$ denotes a trial, where k denotes one of the four videos (or sounds), and m a trial of that video (or sound). a) Trials at encoding were divided into sets of the same stimulus-content (SE_{1-4}). A contrast set, namely a set containing a random selection of trials of different content, was assigned to every stimulus set (CE_{1-4}). b) None of the same-content-sets or the contrast-sets shared any trial. The size of each stimulus-set was the same as the size of its contrast set. To ensure this, some trials were discarded from analysis. c) For N trials of content I (in the set SE_I), the N(N-1)/2 unique trial-combinations were built. This corresponds to the above diagonal region of a combination matrix (green cells). From all the possible combinations of different content, between a trial-set (SE_I) and its contrast set (CE_I), only the combinations above the diagonal were selected (orange cells) for contrast. This is equivalent to exchanging one side of the combinatory-pairs that were built within a stimulus set (i.e. left matrix), with trials from its contrast set (replacing all instances of one trial with instances of a trial from the contrast set). The same-content combinations and the different-content combinations of all 4 stimuli were then contrasted*

Content specific phase similarity between encoding and retrieval

Participants not only saw (heard) the same dynamic stimulus several times in the encoding block, they also repeatedly recalled the same memory content. This made it possible to detect content specific properties of memories if they were shared by trials in which the same content was learned and remembered (e.g. encoding A, remembering A) but not by trials in which different content was learned and remembered (e.g. encoding B, remembering C). Content specific phase was consequently assessed by contrasting the phase-similarity between encoding-retrieval pairs of same content, with the phase-similarity of encoding-retrieval pairs that were of different content.

Again trials were grouped and paired in a balanced randomization procedure to avoid potential bias. First the trials at encoding were grouped into four sets that were of same content (SE_{1-4}). Likewise, the trials at retrieval were grouped into four sets of same memory content (SR_{1-4}). These sets at retrieval were then recombined, such that each set of content A could be assigned a unique set of mixed content (B, C, and D) that was equal in size, i.e. a contrast-set (CR_{1-4}). To make this possible, some trials were discarded from further analysis (figure 113a-b).

Figure 113: Trial combinations of same and different content between encoding and retrieval

a) $SE_1 = \{SE_{1T1}, SE_{1T2}, \dots, SE_{1TN}\}$

$SE_2 = \{SE_{2T1}, SE_{2T2}, \dots, SE_{2TN}\}$

...

$SE_4 = \{SE_{4T1}, SE_{4T2}, \dots, SE_{4TN}\}$

$SR_1 = \{SR_{1T1}, SR_{1T2}, \dots, SR_{1TN}\}$

$CR_1 = \{SR_{2T1}, SR_{4T2}, \dots, SR_{3TN}\}$

$SR_2 = \{SR_{2T1}, SR_{2T2}, \dots, SR_{2TN}\}$

$CR_2 = \{SR_{1T1}, SR_{3T2}, \dots, SR_{4TN}\}$

...

$SR_4 = \{SR_{4T1}, SR_{4T2}, \dots, SR_{4TN}\}$

$CR_4 = \{SR_{3T1}, SR_{1T2}, \dots, SR_{2TN}\}$

b) There were no common elements in stimulus groups and in contrast groups

$\forall_{I,J} \in \mathbb{N}\ (1 \le I,J \le 4 \wedge I \ne J \Rightarrow SR_I \cap SR_J = CR_I \cap CR_J = \emptyset)$

The stimulus group had the same size as its contrast group

$\forall_I \in \mathbb{N}\ (1 \le I \le 4 \Rightarrow |SR_I| = |CR_I|)$

c)

Same combinations Stimulus I

SE_I \ SR_I	Trial 1	Trial 2	...	Trial N
Trial 1				
Trial 2				
...				
Trial n				

Different combinations Stimulus I

SE_I \ CR_I	Trial r1	Trial r2	...	Trial rN
Trial 1				
Trial 2				
...				
Trial n				

Legend to figure 113: Trial combinations of same and different content between encoding and retrieval

a) Trials were divided into sets of the same stimulus content at encoding (SE_{1-4}) and at retrieval (SR_{1-4}). Stimulus content at retrieval refers to the content held in memory. A contrast-set (CR_{1-4}), namely a set containing a random selection of trials of different content, was assigned to every stimulus set (SR_{1-4}) at retrieval. b) The different stimulus-sets at retrieval (SR_{1-4}) as well as the different contrast-sets (CR_{1-4}) had no common trials. Furthermore, every stimulus-set at retrieval had the same number of trials as its contrast-set. To ensure this, some trials were discarded before further analysis. c) The combinations of same content between encoding and retrieval consisted of all possible trial-pairs between a set of content I at encoding (SE_I) and the set of content I at retrieval (SR_I). However, combinations containing the same word cue were ignored (diagonal grey cells). Combinations of same content therefore correspond to the off-diagonal of a combinatory matrix (green cells). To build the combinations of different content for a stimulus, the trials from the very same set at encoding (SE_I) were then combined with all trials from the corresponding contrast-set to its content at retrieval (CR_I). Combinations on the diagonal were ignored accordingly; different content combinations correspond to the off-diagonal of the combinatory matrix (orange cells). The similarity of same-content pairs was finally contrasted with the similarity of different-content pairs for all 4 dynamic stimuli.

Encoding-retrieval pairs of same content were then formed by building all possible pairs of trials between each set of a content at encoding (SE_I) and the corresponding set of this memory content at retrieval (SR_I). In order to build the pairs of different content, the very same set of trials from encoding (SE_I) was combined with the corresponding contrast set (CR_I) at retrieval. Finally, pairs containing the same word-cue were ignored, this occurs, when the encoding-trial that was originally associated with a word-cue was combined with the retrieval-trial in which this cue was actually presented. Accordingly, in the combinations of different content, the pair between the discarded encoding trial and a random trial was ignored (figure 113c).

Between the pairs of same combinations, a similarity measure of phase was then computed (see below) and contrasted with the similarity between the pairs of different content. In order to maximize the signal to noise ratio in further analysis several restrictions were applied to define frequencies, time-windows, and electrodes of interest. The tested frequency was 8 Hertz, since both conditions expressed the strongest correlates of memory in this frequency band. Furthermore, the time-window at encoding was restricted to a 1 second episode, in which phase-patterns were maximally different between the stimuli. The window was defined around the centre of the cluster in which phase patterns were most reliably content specific during encoding (i.e. the cluster with the lowest p-value). Centring the encoding-window on the most content specific time course of activity should increase the sensitivity to detect differences from encoding at retrieval. Likewise, the electrodes for further analysis were restricted to the electrodes within that cluster (128/128 electrodes in the visual condition and 107/128 electrodes in the auditory condition). It needs to be emphasized that none of these restrictions leads to circular inference, since all of these prior restrictions are independent of

the similarity between encoding and retrieval trials. Most importantly phase similarity at encoding, under the null-hypothesis, is completely orthogonal to any neural activity at retrieval.

Phase similarity between two windows was then assessed with the Single-trial Phase Locking Value (S-PLV) (Lachaux et al., 2000; Mormann et al., 2000). This measure defines similarity between two windows (x and y) as the constancy of phase angle difference over time, where n denotes the width of the window and φ is the phase:

$$SPLV = \left| n^{-1} \sum_{t=1}^{n} e^{i(\varphi xt - \varphi yt)} \right|$$

If the two signals are very similar over time, the phase angle differences will not vary much (i.e. have low circular variance). In this way, the similarity of two windows can be quantified as 1 minus the circular variance of phase differences over time. S-PLV has the advantage of increased robustness for noisy data at the expense of temporal resolution. For the purposes of assessing similarity between two oscillatory patterns, this measure is convenient because it affords a high degree of temporal invariance and results in a value between 0 and 1 when two oscillatory patterns are compared. Therefore, despite the oscillatory nature of temporal patterns in the EEG, this makes it possible to assess the average similarity across time, trials and subjects. In their paper, the authors suggest to compute the S-PLV over 6-10 cycles of a frequency for a good signal to noise ratio (Lachaux, Rodriguez, Martinerie, & Varela, 1999; Mormann et al., 2000), for our purposes, S-PLV was applied to a time window of 8 cycles, which resulted in a 1 second window for 8 Hertz. Phase values were extracted by multiplying the fourier-transformed data with a complex Morlet Wavelet of 6 cycles. Phase-values were

then downsampled to 64 Hz. The similarity measure was computed for every pair of trials in the combinations of same content and in the combinations of different content. Importantly, a sliding window approach was used to account for the non-time-locked nature of the data (memory reactivation could happen at any time during retrieval).

For every combination of trials, this resulted in a single similarity value for every electrode and every time point at retrieval, i.e. the similarity to the 1 second encoding window (a similarity value at a single time point represented the similarity of the surrounding 1 second window at retrieval, to that window from encoding). Additionally, the retrieval window was truncated at 4 seconds in order to avoid potential confounds from post-stimulus images or responses; to assess similarity at 4 seconds, the time window was instead continued beginning from 1 second pre-stimulus (i.e. similarity at 4s reflects the similarity between the encoding window and the concatenated window from 3500ms to 4000ms and -1000ms to -500ms at retrieval).

In order to test for content specific phase patterns, the difference in similarity between same content combinations and different content combinations was averaged across the whole retrieval episode (between 0 and 4000ms), which resulted in a single value for every electrode for the same content combinations and for the different content combinations. Those values were then statistically tested across subjects, controlling for multiple comparisons with the fieldtrip permutation procedure (Maris & Oostenveld, 2007a). Additionally, 2 control frequencies were tested that were approximately in the golden mean ratio (i.e. maximally different in terms of phase) to 8Hz (Pletzer et al., 2010), namely 5 and 13 Hz, two further control frequencies were tested that showed the next strongest power decrease in one of the conditions, namely 4 and 15Hz. Encoding-time windows were defined

accordingly for these frequencies as 8 cycles around the centre of the most reliable cluster during encoding.

The electrode clusters of significant differences that resulted for 8 Hz, were subjected to further analysis in order to explore the temporal dynamics of reinstatement. In a first step, a series of post hoc t-tests was computed on the difference between same and different content combinations during every time point of retrieval. This resulted in a time series that is comparable to a cross-correlogram and can be interpreted as a time course of reinstatement (see figure 109a-b).

In a further step, the sliding window analysis was repeated with different time windows from encoding, however keeping with the electrodes in the cluster of significant differences. Thereby, similarity between any two time points could be estimated with a temporal uncertainty of +-500ms. The outcome of this analysis was a matrix of similarity between every time point at encoding and every time point at retrieval on each of the electrodes in the cluster (see figure 109e-f). The difference between combinations of same and different content was then averaged across electrodes and tested over subjects. The resulting clusters reveal the temporal relationship between presentation at encoding and reinstatement during retrieval, however it should be said that tests on this encoding-retrieval-matrix are not independent from the original identification of the electrodes.

Source reconstruction

To reconstruct the activity on the source level, a linearly constrained minimum variance (lcmv) beamforming approach was used as it is implemented in the Fieldtrip toolbox (http://www.fieldtriptoolbox.org/). Individual electrode positions were used together with

boundary element models that were constructed from individual MRI scans. With lcmv-beamforming, filters will be more accurate for the data that they were constructed on and will also be more accurate if constructed on a long time interval (Van Veen, Van Drongelen, Yuchtman, & Suzuki, 1997). This trade-off was addressed by computing each filter around the preprocessed data that contributed to the effect being localized. Power differences were localized with a filter based on -500ms to 4500ms at retrieval; for the phase similarity at encoding, the filters were estimated on the time window between -500ms and 3500ms of the encoding trials. Phase similarity between encoding and retrieval was reconstructed with a filter based on -500ms to 1000ms at encoding and -500ms and 4500ms at retrieval. Activity on 2020 virtual electrodes was thereby reconstructed and the analysis of the data was repeated in the same way on the virtual data.

Statistical analyses

Behavioural performance

Behavioural results were compared between the auditory and visual condition with a series of paired t-tests. P-values were compared against a Bonferroni-corrected threshold (Bland & Altman, 1995), however no specific hypothesis was tested.

Decreases in power

To test for differences in baseline corrected power, a paired t-test was first computed for every time point and frequency at every channel. For multiple-comparison-correction, a random permutation procedure was applied (Maris & Oostenveld, 2007a). This procedure sums up neighbouring t-values above a cluster forming threshold and compares the resulting clusters' sizes to the distribution of the maximal cluster sums that are derived, when

condition labels are randomly swapped with the Monte-Carlo method. The option for the minimum number of channels to be considered a cluster was specified with 3. This attenuates the confound coming from spatially high frequency noise only allowing clusters that contain at least 3 neighbouring channels above threshold; neighbouring electrodes were derived via the triangulation method of the Fieldtrip toolbox (http://www.fieldtriptoolbox.org/). The clusters were summed across time, frequency and channels, then labels were permuted 1000 times; computation of the clusters as well as the testing of the null hypothesis was addressed with a threshold for two-sided testing (alpha level of 0.025). Due to computational limitations, the power values were downsampled before the *unrestricted* test across time, electrodes and frequencies, such that values were included approximately every 16ms. To further identify frequencies with a reliable power difference, a paired samples t-test was computed for every frequency on the average power difference across channels and across the whole retrieval time window between 0 and 4 seconds (for frequencies below 6 Hz the time window was increasingly shorter, since the last point of data was at 4.5 seconds and the power of lower frequencies cannot be estimated towards the boundaries of the time window).

On the source level, the average power values between 0 and 4 seconds were compared for 8 Hz. Therefore, a t-test was computed for every virtual electrode and an unrestricted permutation procedure was run on the whole brain level in the same way as described above using 1000 permutations. Neighbouring t-values were now only spatially defined from neighbouring virtual electrodes.

Phase stationarity/signal complexity

In order to assess whether the frequency specific power decreases resulted in differences in phase stationarity, a paired t-test was computed for the average difference in stationarity

over time and electrodes (see Supporting Information and figure 105). These values were averaged over all data points in the time window between 0 and half a cycle before the end of the trial (4 seconds). Following the hypothesized dependency of phase stationarity on the power decreases, the frequency with the strongest power decrease was tested first. In subsequent tests, p-values were compared against a Bonferroni corrected p-value.

Phase similarity during encoding

Phase similarity at encoding was tested in the same way as power. A series of paired t-tests was computed to contrast the average similarity of combinations of same content with the average similarity of combinations of different content. T-values for every frequency band, electrode and time point were then corrected for multiple comparisons in an unrestricted cluster-based permutation approach. The cluster permutation compared again the sums of t-values across frequency, electrodes and time against the distribution of these clusters derived via the Monte-Carlo method. Due to computational limitations, the similarity values were downsampled only for the unrestricted test across time, electrodes and frequencies, such that approximately every 16ms a value was included in the test. A threshold for two-sided testing was applied, in order to test against the null-hypothesis. Later, the frequency 8 Hertz was tested separately with the same cluster permutation procedure against a one-sided threshold in order to identify a temporospatial cluster, in which 8 Hertz phase could differentiate content particularly well. On the source level, similarity was averaged over the defined one-second encoding window (see above) and contrasted between combinations of same and different content with a t-test on every virtual electrode. Multiple comparisons correction was performed again on the whole brain level, neighbouring t-values were summed and the distribution of resulting clusters was created using 1000 randomly drawn

permutations. The probability of the observed cluster was then assessed by comparing its size to this distribution, correcting for a threshold of two-sided testing.

Phase similarity between encoding and retrieval

The similarity between the encoding-window and the retrieval-episode was tested for differences between combinations of replay of the same versus replay of different content. In a first step, the average difference between 0 and 4 seconds was contrasted with a paired t-test on every electrode to test for a general effect. For multiple comparison correction again 1000 permutations were drawn and observed clusters of summed t-values were tested against the distribution of sums under random permutation of conditions; the threshold for significance was the p-value for two-sided testing. The test of the effect on the source level was small and did not survive multiple comparison correction. However, the maximal clusters of neighbouring t-values that exceeded a threshold for single-sided testing were assumed to reflect the effect that was significant on the electrode level. Therefore, we used only the maximal cluster of differences in source space as a region of interest for further analyses.

Interaction of power decreases and phase similarity in source space

These clusters of similarity were then tested for differences in power decreases by summing up power differences across all virtual electrodes within each region of interest (visual/auditory) in each condition (visual/auditory). Power in these Regions was then compared across conditions by subjecting the power decreases in the regions to a 2X2 repeated measurements ANOVA with the factors Region of Interest and Condition.

Exploration of encoding retrieval similarity in the cluster of memory-replay

Finally, on the electrode level, the similarity effect between encoding and retrieval was statistically explored. Within the electrode-clusters of significant differences, a series of post-

hoc t-tests was computed and thresholded with a probability level of 0.01 for a single-sided test in order to identify the time windows that caused the similarity-effect. Lastly, the Similarity matrices within the clusters of electrodes that indicated encoding-retrieval similarity were tested. Neighbouring t-values of difference between same and different content combinations were thresholded again at a p-value of 0.01 and summed up, however In order to allow for negative effects and rule them out, the threshold was adapted for two-sided testing. 1000 permutations were drawn and a distribution of the strongest clusters, second strongest clusters, etc. was built. The observed clusters were sorted and compared against the random distribution of clusters. A liberal approach was adopted, comparing the cluster with the highest sum of t-values, to the distribution of the maximal cluster and every following cluster to the distribution of next strongest clusters. Critical p-values for significance were divided by the number of the cluster (Bland & Altman, 1995).

Data Availability

Group statistical data and analysis scripts of this project are deposited in the Dryad repository: http://dx.doi:10.5061/dryad.ch110 (Michelmann, 2016)

Chapter 3 – Replay of Stimulus Specific Temporal Patterns during Associative Memory Formation

Preface

The previous chapter demonstrated that content specific temporal patterns can be detected during purely memory driven reinstatement. This reinstatement was linked to decreases in oscillatory power at a centre frequency of 8 Hz. Importantly, the experimental paradigm clearly separates perception of the dynamic stimuli from memory formation during encoding: When the association is formed, only the static word-cue is presented. Arguably, however, content specific patterns still need to be present in neural activity, in order to form the association between naturalistic stimuli and word-cues. In accordance with the information via desynchronization hypothesis, these patterns should again be marked by power decreases in the corresponding frequency bands. The following chapter will test this hypothesis on the same dataset that was presented in the last chapter. This research was submitted under the title: *Replay of Stimulus Specific Temporal Patterns during Associative Memory Formation* and is available in near identical form from biorxiv.org (Michelmann, Bowman, & Hanslmayr, 2017). At the time of this thesis, the paper was under review.

Contributions

The experiments were conceived and designed by SM and SH. SM performed the experiments.

All data analysis was performed by SM under supervision of SH, the manuscript was written by SM under supervision of SH and HB.

Abstract

Forming a memory often entails the association of recent experience with present events. This recent experience is usually an information rich and dynamic representation of the world around us. We here show that associating a static cue with a previously shown dynamic stimulus, yields a detectable, dynamic representation of this stimulus in working memory. We further implicate this representation in the decrease of low-frequency power (~4-30 Hz) in the ongoing electroencephalogram (EEG), which is a well-known correlate of successful memory formation. The maintenance of content specific patterns in desynchronizing brain oscillations was observed in two sensory domains, i.e. in a visual and in an auditory condition. Together with previous results, these data suggest a mechanism that generalizes across domains and processes, in which the decrease in oscillatory power allows for the dynamic representation of information in ongoing brain oscillations.

Introduction

Not everything we associate in our memory occurs at the same time. When our favourite football player is seeing the red card, for instance, we are able to bring this together with the events we just witnessed a few seconds before. Later, we are naturally able to recall all relevant information leading to the red card. In order to successfully make this association, our brain has to accomplish two things. First, it has to keep track of the past and maintain a representation of the events in the ongoing football match and second, form memories in which past events are connected to the red card. Processes during the encoding phase that will determine our ability to later remember events can be investigated with the so-called subsequent memory paradigm (Paller & Wagner, 2002; Wagner, Koutstaal, & Schacter, 1999). Subsequent memory effects refer to neural activity which distinguishes remembered from not remembered items at the time of encoding and are well documented in M/EEG and fMRI, showing involvement of cortical as well as medial temporal lobe regions (e.g. Long et al., 2014; Otten, Quayle, Akram, Ditewig, & Rugg, 2006). Concerning M/EEG power, decreases in low frequency (<40 Hz) brain dynamics have repeatedly and consistently been related to successful memory formation (Hanslmayr & Staudigl, 2014).

It has recently been proposed that cortical power decreases in the alpha/beta frequency range allow for a rich representation of memory content, since a desynchronized system has more flexibility to code information over a system of high synchrony. We call this view, the information via desynchronization framework (Hanslmayr et al., 2012). Confirming this idea, we have shown that sustained power decreases in the alpha band at approximately 8 Hz, contain item specific information about the remembered content, when subjects successfully

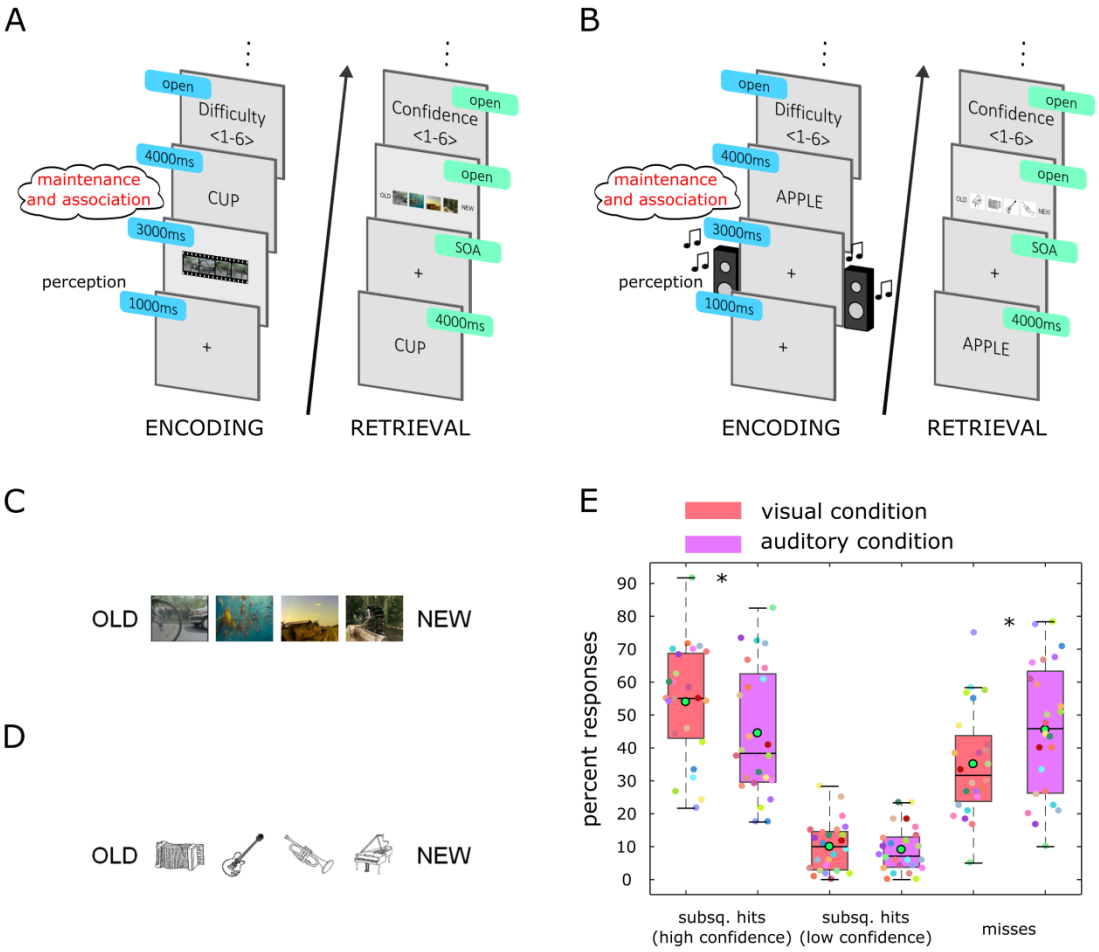
replay dynamic stimuli (i.e. video and sound clips) from memory (Michelmann et al., 2016). In this study, we provided direct evidence that power decreases are involved in the representation of stimulus specific information (Hanslmayr et al., 2012). Moreover these results are well in line with numerous studies showing that perception is not continuous but rather is rhythmically sampled at a frequency of ~7-8 Hz (Hanslmayr et al., 2013; Landau & Fries, 2012; VanRullen et al., 2007). These outcomes indicate that rhythmic patterns from the perception of dynamic stimuli can reappear during internally guided retrieval processes, in the absence of the stimuli themselves. Accordingly, these prior findings also suggest the possibility that the replay of temporal patterns can be observed in a situation where dynamic stimuli have to be maintained internally in working memory.

To address this question, we here analyse the data during the encoding phase from a previous dataset (Michelmann et al., 2016). The paradigm required subjects to associate a dynamic stimulus with a static word that was used as a cue in the later retrieval phase. Importantly, during encoding the perception of the dynamic stimulus and the presentation of the word-cue was temporally separated, i.e. in every trial, one out of four dynamic stimuli was followed by a unique word-cue (figure 201, a-b). In a visual condition, these dynamic stimuli consisted of four short video-clips, in an auditory session four short sound clips were used. In a later retrieval block, participants were presented with the word-cue and were tested whether they remembered the associated video/sound clip.

We hypothesize that, in order to associate the word-cue with the dynamic stimulus, subjects maintain (i.e. replay) a sensory representation of the dynamic stimulus in working memory, which is why we refer to this phase as the maintenance phase. Using temporal pattern similarity analysis, we should therefore be able to detect the replay of these patterns during

the maintenance phase, i.e. when the association between a word and the sound/movie is formed. In accordance with the information via desynchronization framework, we should observe stronger decreases for later remembered versus later not remembered items. This subsequent memory effect should be most evident in the frequency band that codes for the representation of the dynamic stimulus in working memory, i.e. 8 Hz as per our previous findings. Moreover, if power decreases enable a richer representation of the perceptual content, we should already observe stronger broad power decreases for later remembered compared to later not remembered stimuli during the perception of the dynamic stimulus.

Figure 201: Experimental design and behavioural results



Legend to figure 201: Experimental design and behavioural results

Experimental sequence in the visual (A) and in the auditory (B) session. During encoding (A, B left), participants perceived a dynamic stimulus that played for 3 s and was then followed by a word cue. The cue was presented for 4 s and subjects had to associate the word with the dynamic stimulus they just saw. Note that during encoding, the word cue was shown after the dynamic stimulus, separating the perception interval from the association interval, therefore participants had to maintain a representation of the dynamic stimulus in working memory. Participants learned 120 associations between four repeatedly shown dynamic stimuli and 120 different words. At the end of every encoding trial, they rated the perceived difficulty of the association on a scale from 1 to 6. In the retrieval block (A, B right) they recalled the dynamic stimulus upon presentation of the word-cue. Cues from encoding were mixed with 60 new words that served as distractors. After that, they indicated the stimulus they recalled. Response options (C, D) consisted of four small screenshots of the video clips in the visual session (C) and of four small instruments, representing the sounds, in the auditory session (D). The response option "NEW" represented the distractor, and was available to indicate that the word was not presented in the encoding block, the response option "OLD" was available to indicate that subjects remembered only the word, but not it's associate. At the end of every retrieval trial a confidence rating was collected on a scale from 1 to 6 (A, B right). (E) Behavioural performance for the associations from encoding. Hits are trials in which the correct associate was subsequently remembered (video or sound). A rating of high confidence was considered a rating > 4. Misses were defined as all trials in which the associate was later forgotten. Boxes are 25th and 75th percentiles around the median; whiskers represent

*minimum and maximum (disregarding outliers). Green points in black circles are arithmetic means small coloured points are single subject values. * $p < 0.05$*

Materials and Methods

Participants

24 healthy, right-handed subjects (18 female and 6 male) participated in this study. 7 further participants were tested, or partly tested, but could not be analysed due to poor memory performance (N=2), misunderstanding of instructions (N=2), and poor quality of EEG-recording and technical failure (N=3). All participants had normal or corrected-to normal vision. The average age of the sample was 23.38 (s.d. = 3.08) years. Participants were native English speakers (20), bilingual speakers (2) or had lived for more than 8 years in the UK (2). Ethical approval was granted by the University of Birmingham Research Ethics Committee, complying with the Declaration of Helsinki. Participants provided informed consent and were given a financial compensation of 24£ or course-credit for participating in the study.

Material and experimental set up

The cues amounted to 360 words that were downloaded from the MRC Psycholinguistic Database (Coltheart, 1981). Stimulus material consisted of 4 video clips and 4 sound clips in the visual and auditory session respectively. All clips were 3 seconds long; videos showed coloured neutral sceneries with an inherent temporal dynamic, sounds were short musical samples, each played by a distinct instrument. In both sessions, a clip was associated with 30 different words. 60 words were reserved for the distractor trials and 12 additional words were used for instruction and practice of the task. For presentation, words were assigned to the clips or to distractors in a pseudorandom procedure, such that they were balanced for Kucera-Francis written frequency (mean = 23.41, s.d. = 11.21), concreteness (mean = 571, s.d.

= 36), imageability (mean = 563.7 s.d. = 43.86), number of syllables (mean = 1.55, s.d. = 0.61) and number of letters (mean = 5.39, s.d. = 1.24). Furthermore lists were balanced for word-frequencies taken from SUBTLEXus (Brysbaert & New, 2009). Specifically, "Subtlwf" was employed (mean = 20.67, s.d. = 27.16). The order of presentation was also randomized, assuring that neither the clips and their associates, nor distractor words were presented more than 3 times in a row or in temporal clusters. The presentation of visual content was realized on a 15.6 inch CRT-monitor (Taxan ergovision 735 TC0 99) at a distance of approximately 50 centimetres from the subjects eyes. The monitor refreshed at a rate of 75 Hz. On a screen size of 1280 x 1024 pixels, the video clips appeared in the dimension of 360 pixels in width and 288 pixels in height. 'Arial' was chosen as the general text-font, but font-size was larger during presentation of word-cues (48) than during instructions (26). In order to reduce the contrast, white text (rgb: 255, 255, 255) was presented against a grey background (rgb: 128, 128, 128). Auditory stimuli were presented using a speaker system (SONY SRS-SP1000). The 2 speakers were positioned at a distance of approximately 1.5 meters in front of the subject with 60 centimetres of distance between the speakers.

Procedure

Upon informed consent and after being set up with the EEG-system, participants were presented with the instructions on the screen. Half of the subjects started with the auditory session, the others were assigned to undertake the visual task first. Both sessions consisted of a learning block, a distractor block and a test block. The sessions were identical in terms of instructions and timing and differed only in the stimulus material that was used. During instruction, the stimulus material was first presented for familiarization and then used in

combination with the example words to practice the task. Instructions and practice rounds were completed in both sessions.

As a way to enhance memory performance, participants were encouraged to use memory strategies. The suggestion was to imagine the word in a vivid interaction with the material content, yet the choice of strategy remained with the subject. In the learning block, 120 clip-word sequences were presented. Each sequence started with a fixation cross that was presented in the centre of the screen for 1 second, and then the video-clip played for 3 seconds. In the auditory condition, the fixation cross stayed on the screen and the sound-clip played for 3 seconds. Immediately after the clip, a word cue was presented in the centre for 4 seconds, giving the subject time to learn the association. After that, an instruction requested participants to subjectively rate on a 6 point scale how easy the association between the clip and the word was. After a press on the space bar, this scale was shown. Equidistant categories were anchored with the labels “very easy” and “very hard”; those labels were displayed at both ends above the scale. Participants used six response buttons to rate the current association (see figure 201).

In the distractor block, subjects engaged in a short unrelated working memory task, namely they counted down in steps of 13, beginning from 408 or 402 respectively. After 1 minute the distractor task ended. Following a short self-paced break, subjects refreshed the instructions on the retrieval block.

In this retrieval block, either a cue or a distractor was presented upon a button press on the space bar. Subjects were instructed to try to vividly replay the content of the corresponding video-clip or sound-clip in their mind upon presentation of the cue. The word stayed on the screen for 4 seconds, giving the subject the opportunity to replay the memory. Finally, a

fixation cross was presented for a varying time window between 250 and 750 milliseconds to account for movement and preparatory artefacts, before the response scale appeared on the screen.

The response-scale consisted of 6 options. 4 small screen shots of the videos or 4 black and white pictures of the featured instruments were presented in equidistant small squares of 30x30 pixels. Additionally, the options “new” and “old” were displayed in the form of text at the most left and most right position of the scale (see figure 201C-D). Subjects could now either indicate the target (video/sound) they just replayed, by pressing the button corresponding to that clip. Instead, subjects could also indicate that the word was a distractor by pressing the button corresponding to the option “new”, or they would simply indicate that they remembered the word, but could not remember the clip it was associated with. In this last scenario, subjects would press the button corresponding to “old”. The positions of “old” and “new” at the end of the scale, as well as the permutation of the 4 target positions in the middle of the scale, were counterbalanced across participants. Finally, after making a decision, a further six point rating scale was presented on which subjects could rate the confidence in their response. Again a scale with equidistant categories was presented ranging from “guess” to “very sure”. An additional possibility was to press “F2” in case of an accidental wrong button press. In this case, the whole trial was discarded from analysis. Following the retrieval block, individual electrode positions were logged allowing for a break of approximately 30 minutes before beginning the second session.

Data Collection

The recording of behavioural responses and the presentation of instructions and stimuli were realized using Psychophysics Toolbox Version 3 (Brainard, 1997) with MATLAB 2014b (MathWorks) running under Windows 7, 64 Bit version on a desktop computer. Response buttons were “s, d, f, j, k, l” on a standard “QWERTY” layout. Buttons were highlighted and corresponded spatially to the response options on the screen, so participants did not have to memorize the keys. To this end, the shape of corresponding fingers was also displayed under the scale. To proceed, participants used the space bar during the experiment. Physiological responses were measured with 128 sintered Ag/AgCl active electrodes, using a BioSemi Active-Two amplifier, the signal was recorded at 1024 Hz sampling rate on a second computer via ActiView recording software, provided by the manufacturer (BioSemi, Amsterdam, Netherlands). Electrode positions were logged with a Polhemus FASTRAK device (Colchester, VT, USA) in combination with Brainstorm (Tadel et al., 2011) implemented in MATLAB.

Preprocessing

The data was preprocessed using the Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld et al., 2011). Data was cut into trial-segments from 2.5 seconds pre-stimulus to 7 seconds after the onset of the dynamic stimulus. The linear trend was removed from each trial and a baseline correction was applied based on the whole trial. Trials were then downsampled to 512 Hz and a band-stop filter was applied at 48-52, 58-62, 98-102 and 118-122 Hz to reduce line noise at 50 Hz and noise at 60 Hz; additionally a low-pass filter at 140 Hz was applied. After visual inspection for coarse artefacts, an independent component analysis was computed. Eye-blink artefacts and eventual heartbeat/pulse artefacts were removed, bad

channels were interpolated and the data was referenced to average. Finally the data was inspected visually and trials that still contained artefacts were removed manually.

Behavioural analysis

For behavioural analysis, correct trials were defined as those in which the target was correctly identified. The confidence rating of the response was considered as high if a rating of 5 or 6 was selected. Misses were defined as trials in which a cue-word was incorrectly identified as a new word, the wrong clip was selected, or the response "old" was given to indicate recognition of the word without remembering the target video or sound it was associated with.

Power analysis

Oscillatory power was determined by multiplying the Fourier-transformed data with a complex Morlet wavelet of 6 cycles. Raw power was defined as the squared amplitude of the complex Fourier spectrum and estimated for every 4^{th} sampling point (i.e. sampling rate of 128 Hz). For the contrast of subsequent hits and subsequent misses, a baseline was computed as the average power between -1 and 7 seconds of all trials within the contrast (Long et al., 2014). Every trial was then normalized by subtracting the baseline and subsequently dividing by the baseline $(activity_{tf} – baseline_f)/baseline_f$, where t indexes time and f indexes frequency. The relative power was calculated for all frequencies between 2 and 30 Hz.

Phase pattern analysis during perception and maintenance

While participants learned the associations in the encoding block, they repeatedly perceived (saw/heard) the same dynamic stimulus. Content-specific properties could consequently be

identified if they were shared by trials of the same content but not by trials of a different content. Hence, content-specific phase during perception was assessed by contrasting the phase similarity between pairs of trials in which the same content was presented, with the phase similarity of an equal number of trial pairs that were of different content. For each pair of trials, the cosine of the absolute angular distance was then computed and finally averaged across all (same or different) combinations [29]. The average similarity value for same and different combinations was subjected to statistical testing across subjects at every time point, at every electrode and in every frequency of interest; this contrast embodies content specific phase patterns during perception.

Participants also repeatedly associated the same dynamic stimulus (one of four videos/sounds) with a different word cue. Therefore the temporal pattern during perception of the dynamic stimulus could be compared to the temporal pattern in different trials in which subjects maintained the same dynamic stimulus in working memory. Notably, excluding within trial combinations eliminates the potential confound of temporal autocorrelation. Likewise the temporal pattern during perception could be compared to trials in which subjects maintained a different dynamic stimulus in working memory.

In this way, the phase similarity between combinations of same content (e.g. perceiving content 1, maintaining content 1) was contrasted with the phase similarity between trials of different content (e.g. perceiving content 4, maintaining content 2). This contrast reveals phase patterns that are specific to the dynamic stimulus which subjects associated with the cue.

To maximize the signal to noise ratio, the following restrictions were applied: The tested frequency was 8 Hertz, following our previous results and hypotheses (Michelmann et al.,

2016); A time-window during perception was centred on the cluster in which phase patterns were most reliably content specific during encoding (i.e. the cluster with the lowest p-value) and subsequently used in a sliding window approach in order to detect content specific patterns.

Phase similarity between two windows was then assessed with the Single-trial Phase Locking Value (S-PLV) (Lachaux et al., 2000; Mormann et al., 2000). This measure defines similarity between two windows (x and y) as the constancy of phase angle difference over time, where n denotes the width of the window and φ is the phase:

$$SPLV = \left| n^{-1} \sum_{t=1}^{n} e^{i(\varphi xt - \varphi yt)} \right|$$

S-PLV assesses the phase coherence between two time windows and has the advantage of increased robustness for noisy data at the expense of temporal resolution.

(Lachaux et al., 2000) suggest to compute the S-PLV over 6-10 cycles of a frequency for a good signal to noise ratio, for our purposes, S-PLV was applied to a time window of 8 cycles, which resulted in a 1 second window for 8 Hertz. Phase values were extracted by multiplying the Fourier-transformed data with a complex Morlet Wavelet of 6 cycles. Phase-values were then downsampled to 64 Hz. The similarity measure was computed for every pair of trials in the combinations of same content and in the combinations of different content. Importantly, a sliding window approach was used to account for the non-time-locked nature of the data (temporal patterns could be present anywhere in the maintenance interval). This resulted in a time course of similarity for the combinations of same and of different content.

The difference in this similarity was first averaged across the whole maintenance episode (between 3500 and 7000ms) and then statistically tested across subjects with a random permutation procedure based on clusters of summed t-values across electrodes (Maris & Oostenveld, 2007a). In a second test, the time courses at every electrode were compared with a series of t-tests and subsequently tested with a cluster-based random permutation procedure, where clusters were summed across electrodes and time (see also: statistical analyses, below). Additionally, a control frequency was tested, namely 6 Hz, based on the results from the power analysis. Time windows were defined accordingly for this frequency as 8 cycles around the centre of the most reliable cluster during perception.

Statistical analyses

Behavioural performance

Behavioural results were compared between the auditory and visual condition with a series of paired t-tests. P-values were compared against a Bonferroni-corrected threshold (Bland & Altman, 1995), however no specific hypothesis was tested.

Decreases in power

To test for differences in baseline corrected power, a paired t-test was first computed for every time point and frequency at every channel. For multiple-comparison-correction, a random permutation procedure was applied (Maris & Oostenveld, 2007a). This procedure sums up neighbouring t-values above a cluster forming threshold and compares the resulting clusters' sizes to the distribution of the maximal cluster sums that are derived, when condition labels are randomly swapped with the Monte-Carlo method. The minimum number of neighbouring channels to be considered a cluster was specified with 3, which attenuates

the impact of spatially high frequency noise; neighbouring electrodes were derived via the triangulation method of the Fieldtrip toolbox (http://www.fieldtriptoolbox.org/). The clusters were summed across time, frequency and channels, then labels were permuted 1000 times; thresholding of the clusters as well as the testing of the null hypothesis was addressed with a threshold for single-sided testing (alpha level of 0.05). To identify frequencies with a reliable power difference, a paired samples t-test was computed for every frequency on the average power difference across all channels and across the whole time window of interest.

Phase similarity during perception of the dynamic stimulus

Phase similarity during perception was tested in the same way as power. A series of paired t-tests was computed to contrast the average similarity of combinations of same content with the average similarity of combinations of different content. T-values for every frequency band, electrode and time point were then corrected for multiple comparisons in an unrestricted cluster-based permutation approach. The cluster permutation compared again the sums of t-values across frequency, electrodes and time against the distribution of these clusters derived via the Monte-Carlo method. Later, the frequency 8 Hertz was tested separately with the same cluster permutation in order to identify a temporospatial cluster, in which 8 Hertz phase could differentiate content particularly well.

Phase similarity between perception and maintenance

The similarity between the time-window during perception and the maintenance-episode was tested for differences between combinations of same and combinations of different content. As mentioned above, in a first step, the average difference between 3.5 and 7 seconds was contrasted with a paired t-test on every electrode to test for a general effect. For multiple comparisons correction, again, 1000 permutations were drawn. Observed clusters of those t-

values that exceeded the critical threshold were summed across neighbouring electrodes and were tested against the distribution of sums under random permutation of conditions. In a second step, a paired t-test was computed for every electrode and time point during the maintenance interval and differences were again tested with a cluster based permutation approach. Now clusters were formed by summation of the thresholded t-values across electrodes and time and compared against the distribution of these clusters for 1000 random permutations.

Results

Behavioural performance

In the visual session, participants remembered on average 53.92% (standard deviation [s.d.] = 17.56%) of the video clips with high confidence (rating > 4), and they further remembered 9.97% (s.d. = 7.62%) of the clips with low confidence (figure 201E). In the auditory session, 44.44% (s.d. = 19.8%) of the audio clips were subsequently remembered with high confidence, which was significantly less than in the visual condition (t_{23} = -2.81, p < 0.01). An additional 9.06% (s.d. = 6.9%) of the audio clips were remembered with low confidence. In accordance, the number of subsequent misses was significantly lower in the visual session (mean 35.07%, s.d. = 16.43%) than in the auditory session (45.45%, s.d. = 20.27%, t_{23} = -3.33, p < 0.01).

Successful memory encoding is associated with low frequency power decreases in the visual and auditory condition

To find correlates of successful memory encoding, the oscillatory power between subsequently remembered (hits) and subsequently not remembered (misses) items was compared. Specifically, we contrasted trials for which associations were subsequently remembered with high confidence, with trials in which the associations were subsequently not remembered correctly. In this analysis, only those datasets were used, in which a minimum of 15 trials remained for hits or misses after preprocessing (N=18). Two crucial episodes for successful memory encoding were tested separately: (i) the time interval in which the dynamic stimulus was actually perceived (0 to 3 seconds) and (ii) the maintenance interval (3 to 7 seconds), in which the memory formation would be expected to have taken place. In the time interval from 0 to 3 seconds, a small cluster of power decreases was associated with successful memory in the visual condition; it displayed a trend towards significance ($p < 0.07$, figure 202A, left). Likewise, in the auditory condition a similar cluster of power decreases appeared ($p = 0.047$, figure 202B, left).

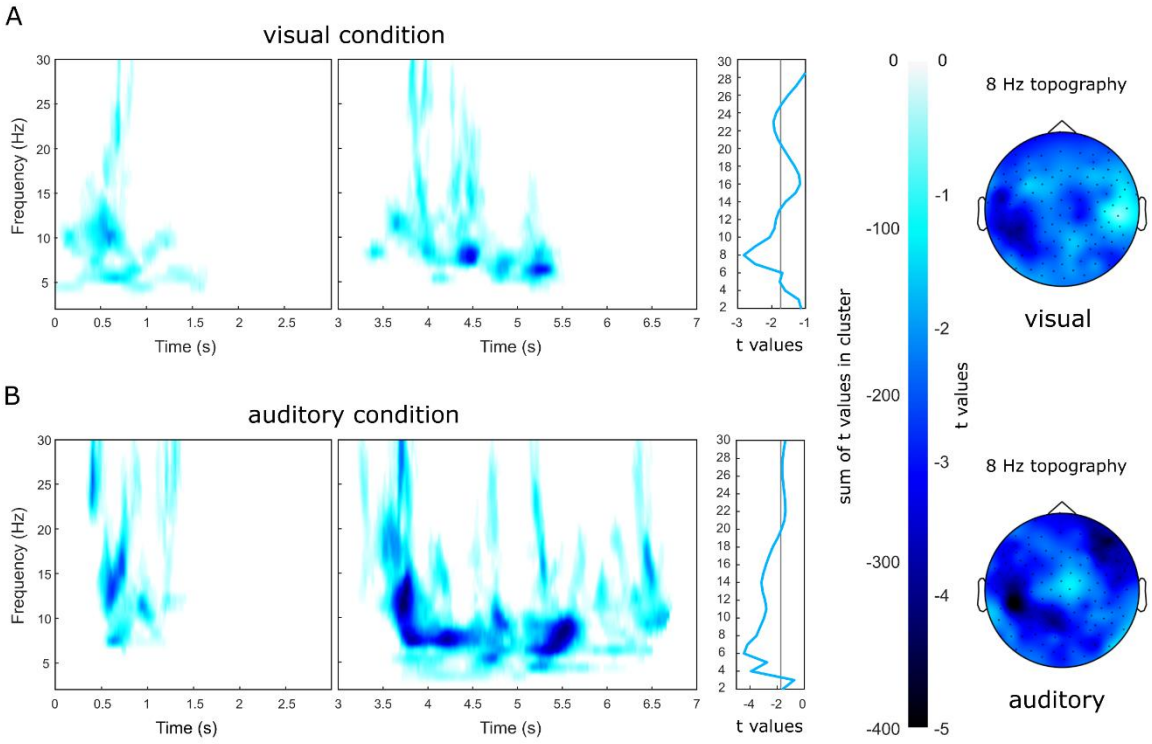
During the maintenance interval (3 to 7 seconds), substantially reduced power in the lower frequencies (<30 Hz) was observed for subsequent hits compared to subsequent misses (figure 202, middle) in both conditions. In the visual condition, a broad cluster emerged where power was significantly lower when tested against random permutations ($p = 0.031$, figure 202A, middle). Likewise a broad cluster of significant power decreases appeared in the maintenance interval of the auditory condition ($p < 0.003$, figure 202B, middle).

To identify frequencies that robustly exhibited lower oscillatory power for successful memory encoding, the power during the maintenance interval was averaged across all electrodes and time points and differences were subjected to a t-test. Following our previous results (Michelmann et al., 2016), we expected the strongest power decreases in both conditions to peak at 8 Hz. Indeed, a clear peak at 8 Hz was observed in the visual condition (t_{17} = -2.82, p < 0.01, figure 202A, middle). In the auditory condition, however, a peak was observed at 6 Hz (t_{17} = -4.45, p < 0.001, figure 202B, middle), yet power decreases also extended to 8 Hz (t_{17} = -3.53, p = 0.001).

For the visual condition, the power decreases at 8 Hz displayed a broad topography with a parietal maximum over the left hemisphere (figure 202A, right). Decreases in 8 Hz power were similarly broadly distributed in the auditory condition, with maxima over left parietal and right frontal regions (figure 202B, right).

Together, these results confirm the fundamental role of decreases in low frequency oscillatory power for the successful formation of memory.

Figure 202: Subsequent memory effects in oscillatory power



Legend to figure 202: Subsequent memory effects in oscillatory power

Successful memory encoding was associated with broad power decreases in the lower frequencies (<30 Hz) in the visual (A) and in the auditory (B) condition. In the first 3 seconds of a trial, when subjects perceived the dynamic stimuli, clusters of broad power decreases in the visual (A, left) and auditory (B, left) condition displayed a trend towards significance already during this interval. During the association with the word-cue (between 3 and 7 seconds within each trial) broad clusters of significant power decreases emerged in both conditions (A, B, middle). Time frequency plots show the sum of t-values across the clusters (A and B, left and middle panels). The t-value of average power difference across electrodes and time between 3 and 7 seconds is plotted to the right of the middle panels. Topographies of power decreases are plotted on the right as maps of t-values derived from the average power decreases between 3 and 7 seconds.

Temporal patterns are content specific during perception and can be detected during maintenance

The detection of content specific temporal patterns during the maintenance period necessitates that the dynamic stimuli themselves elicit temporally distinct neural responses. To address this, we first compared the pairwise phase consistency (PPC) (Vinck et al., 2010) between trials in which the same dynamic stimulus was perceived with the PPC between trials of different content. Oscillatory phase of the neural responses was specific to the dynamic stimuli in two broad clusters in the visual ($p < 0.001$, $p = 0.003$, figure 203C) and one broad cluster in the auditory condition ($p < 0.001$, figure 203G), confirming prior reports that the content of dynamic stimuli is tracked by the phase of low frequency oscillations (Ng et al., 2013). Vitally, both clusters included 8 Hz which was the oscillation for which we hypothesized to detect the reappearance of temporal patterns in the maintenance period.

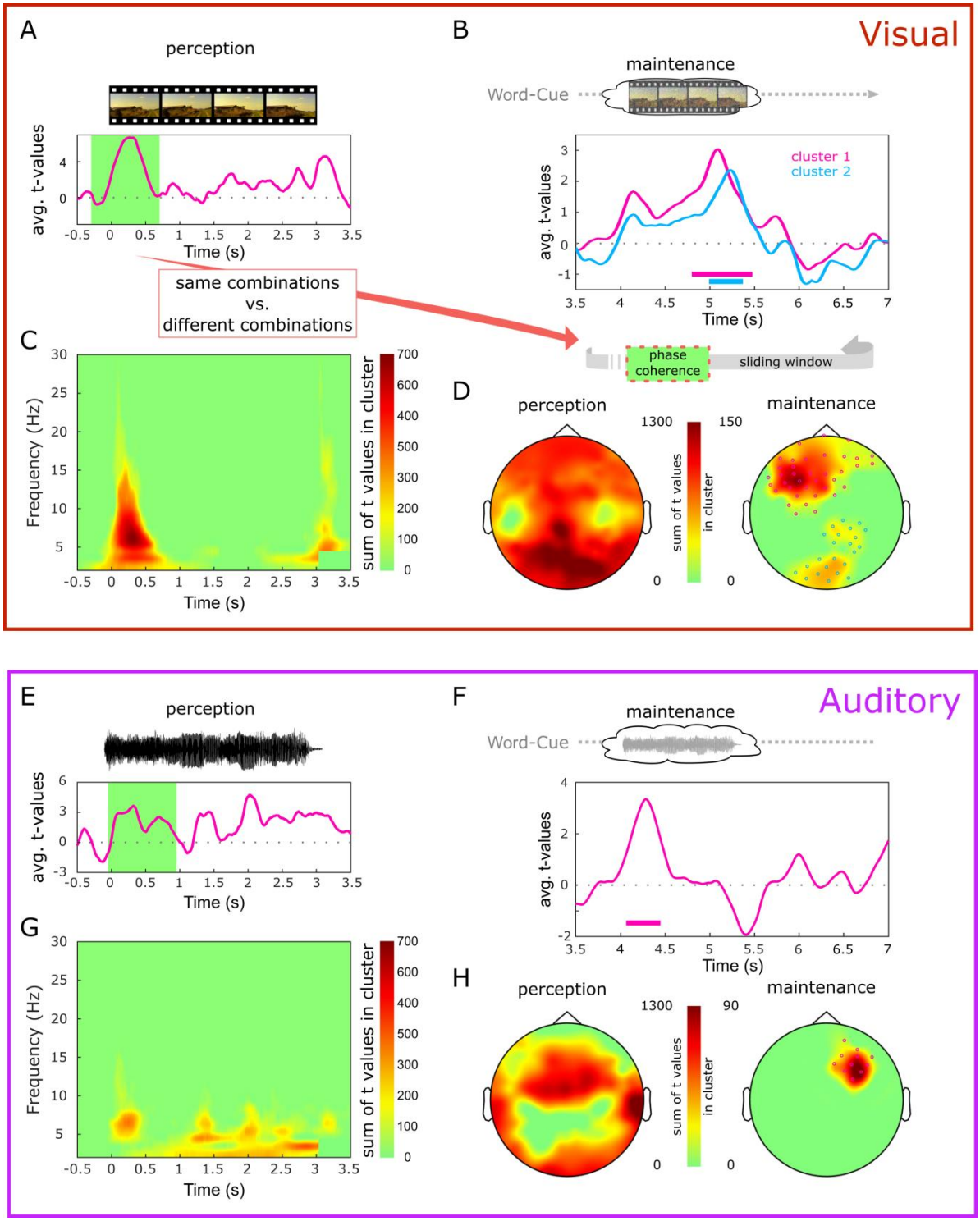
We now identified periods during perception in which the time courses at 8 Hz were maximally content specific by restricting the statistical test to 8 Hz only and selecting the cluster in which content could most reliably be differentiated during perception (i.e., the cluster with the lowest p-value). In the visual condition, this cluster extended from -152 ms to 564 ms ($p < 0.001$). Note that post-stimulus effects are smeared temporally into the pre-stimulus interval because of the wavelet decomposition. The most reliable cluster of content specificity in the auditory condition extended from 22 ms to 871 ms ($p = 0.002$). A further cluster in the visual condition was observed between 2,650 ms and 3,300 ms ($p = 0.016$). In the auditory condition further clusters emerged between 1,818 ms and 2,627 ms ($p = 0.003$)

and between 1,203 ms and 1,504 ms (p = 0.047) indicating that in both modalities early and later time windows showed content specific temporal patterns.

For the 8 Hz oscillation, a 1-second wide window was now centred on the cluster that most reliably distinguished content during perception (i.e. at 206 ms in the visual condition and 446 ms in the auditory condition, figure 203A, E). In a sliding window approach, a measure of phase coherence (S-PLV (Lachaux et al., 2000)) was then computed between this window and every 1-second-wide window between 3 and 7 seconds during the maintenance period (see figure 203A-B). For practical reasons, at the end of the trial the window was slid out back into the pre-stimulus interval (zero padding could be an alternative but more intricate approach). This time course of similarity (phase coherence) was now computed for trial-combinations comprising perception and maintenance of the same stimulus and for trial-combinations of perception and maintenance of different content. Importantly, the combinations of same content were never built within a trial, assuring a balancing of temporal autocorrelation between same and different combinations. In a first test, we subjected the average similarity across time to a t-test, contrasting same and different combinations at every electrode. A cluster-based permutation revealed a significant cluster in the visual condition (p < 0.001), but not in the auditory condition. In a follow-up test, we repeated the t-test for every time-point at every electrode and summed clusters across time and electrodes. A permutation test revealed 2 clusters of significant differences in the visual condition (p < 0.001, p = 0.035, figure 203B, D). The first cluster was located over left-frontal regions and extended from 4.8 to 5.41 seconds after stimulus onset (i.e. 1.8 to 2.41 seconds after the start of the maintenance phase). The second cluster was located over parietal and occipital areas, extending from 4.97 to 5.34 seconds (1.97 to 2.34 seconds of the maintenance phase, figure

203D right). We applied the same approach to the auditory condition; a cluster (p = 0.047) emerged over right-frontal regions extending from 4.11 to 4.44 seconds after stimulus onset (1.11 to 1.44 seconds of the maintenance phase), even though strictly interpreted, this cluster does not exceed a corrected alpha threshold (figure 203F, H right). Finally we also tested the frequency of 6 Hz, which showed the most reliable power decrease in the auditory condition, however no effects were found.

Figure 203: Content specificity of oscillatory phase during perception and maintenance of a dynamic stimulus



Legend to figure 203: Content specificity of oscillatory phase during perception and maintenance of a dynamic stimulus

Oscillatory phase distinguishes between content during perception (left) and can be detected again during maintenance (right). During perception, pairwise phase consistency between trial-combinations of same content and combinations of different content was contrasted with t-tests (A, C, E, and G). (C) shows these t-values for every time and frequency bin in the visual condition, (G) displays the auditory condition. A horizontal slice through these time frequency plots is represented in (A) and (E) for the frequency of 8 Hz. The green window denotes the time window that was selected in order to detect content specific maintenance in the subsequent time interval, where only a static word-cue was presented on the screen. A measure of phase coherence over time (S-PLV) was computed between the selected window and every time point during the maintenance period (A-B).Importantly this similarity was never computed within trials, in order to balance temporal autocorrelations. The result from this sliding window approach was tested with a series of t-tests contrasting same and different content-combinations (e.g. watching movie 1, maintaining movie 1 vs. watching movie 2, maintaining movie 4). (B) and (F) show the time-courses of t-values within the clusters of significant differences in the visual (B) and auditory (F) condition. Horizontal bars denote the time interval in which clusters emerged with significant differences. (D) and (H) display the sum of t-values across the selected time window during perception (D, H, left) and across time within clusters, during maintenance (D, H, right).

Discussion

For most of the memories that we form during the day, we rely on rich and dynamic ongoing representations of the world around us. At a later point, we then associate these representations with distinct events. Both of these properties of our natural experience are rarely captured in experiments that investigate episodic memory. First, most studies use non-information rich stimuli to study memory, like words or pictures, and second material for association is usually presented simultaneously.

In this study, we used a memory task that can mimic memory in a more naturalistic scenario: an ongoing representation of an information rich, dynamic stimulus is maintained in working memory, in order to be associated with a subsequent event. In one session, subjects repeatedly watched one out of four short video clips, which was immediately followed by a unique word-cue. In a second session, subjects listened to one out of four sound clips, which they subsequently associated with a cue (figure 201). In order to form an association, participants had to maintain a representation of the video/sound clip in working memory.

Investigating the correlates of subsequent memory, we found broad and sustained decreases in ongoing oscillatory power to be associated with successful memory formation. These power decreases were particularly strong while subjects maintained dynamic representations in working memory, namely while they formed the association. Importantly, we found that these power decreases carried stimulus specific information in their temporal pattern of activity. Specifically, the phase of an 8 Hz frequency, which we previously linked to content representation (Michelmann et al., 2016) and where power decreases were strongest in the visual condition, was modulated in a stimulus specific way.

These results form part of converging evidence for a general mechanism, in which desynchronization of brain oscillations in the cortex, indicated by power decreases, allows for the rich representation of information (Hanslmayr et al., 2012). Specifically, the decrease in oscillatory strength, which also signifies a release from inhibition (Haegens et al., 2011; Klimesch, Sauseng, & Hanslmayr, 2007), renders the oscillation less stationary, i.e. less predictable. In mathematical terms, this decrease of predictability means an increase in the amount of information that can be coded (Hanslmayr et al., 2012; Shannon & Weaver, 1949). When we previously observed this mechanism during episodic memory reinstatement, oscillatory patterns were localized in sensory-specific areas (Michelmann et al., 2016). In contrast, the pattern maintenance observed in this analysis displayed a different, i.e. more frontal topography, which is suggestive of working memory processes (e.g. Goldman-Rakic, 1995). The generalization of this desynchronization-mechanism across different processes is further complemented by its generalization across modalities; namely, in this study as well as in previous results, we observed oscillatory patterns in desynchronizing brain dynamics for visual and auditory stimuli.

Finally, the frequency band of 7-8 Hz has been previously implicated in the rhythmic sampling of perceptual content (Hanslmayr et al., 2013; Landau & Fries, 2012; VanRullen et al., 2007). These studies integrate well with our findings and suggest that the 8 Hz frequency temporally organizes the representations of stimulus specific information during perception, episodic memory reinstatement and working memory maintenance and that decreases in oscillatory power allow these temporal patterns to resurface.

Our results moreover inform current debates about the neural mechanisms underlying working memory. While some studies have previously shown that content specific activity

patterns can be decoded during working memory maintenance (Fuentemilla, Penny, Cashdollar, Bunzeck, & Düzel, 2010; Jafarpour, Penny, Barnes, Knight, & Duzel, 2017), other studies suggest that representations in working memory may not always be maintained online, but rather latently stored in synaptic weights or even via more complex mechanisms (Stokes, 2015). Those representations can then re-emerge when they become task relevant, or they can be evoked experimentally by either 'pinging' them with unspecific input (Wolff, Jochim, Akyürek, & Stokes, 2017) or by stimulating transcranially with a magnetic pulse (Rose et al., 2016). Hence, an important insight from this study is that a stimulus-representation is maintained online in working memory, when an association with this previously shown stimulus is formed.

The method that we used in order to observe these stimulus patterns was specifically tailored to the detection of patterns that are dynamic in nature. This is very relevant for studies that investigate working memory maintenance because patterns that are involved in the online maintenance of representations in Prefrontal Cortex and Parietal Cortex of nonhuman primates, have been found to be highly dynamic (Crowe, Averbeck, & Chafee, 2010; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008).

An interesting question that arises from our results is whether the online maintenance of temporal patterns is functionally relevant for the successful formation of memories. We could demonstrate subsequent memory effects for power decreases here, because a minimum of 15 trials per condition can yield stable power estimates (Hanslmayr, Spitzer, & Bäuml, 2009). We could further link power decreases to the presence of content specific temporal patterns; however because the trial count of forgotten associations for most of the subjects was too low for stable similarity estimates, it is not clear whether these patterns are functionally

involved in memory formation. Specifically, the present study was designed to produce a sufficient number of remembered trials and we consequently could not contrast stimulus-specific temporal patterns between remembered and forgotten associations. Repeating this study in a longer and more adaptive design, could therefore allow for the contrast of pattern maintenance during successful and unsuccessful memory formation.

Additionally, future studies should address whether content-specific temporal patterns are causally involved in memory formation, either by disrupting content specific temporal patterns and therefore tampering with memory formation or even by artificially introducing spurious patterns to cause forged associations.

Chapter 4 – Speed of time-compressed forward replay flexibly changes in human episodic memory

Preface

The previous chapters have established that both, episodic memory formation and retrieval, yield detectable representations in brain oscillations. These oscillations are marked by sustained power decreases. Some important questions however could not be addressed with these data.

Firstly, participants had variable memory performance in this experiment, resulting in very few forgotten associations in some subjects. Therefore, an open question is, whether the reinstatement of oscillatory patterns is relevant for successful memory, because pattern reinstatement could not be contrasted between successfully remembered and forgotten trials.

Secondly, we asked subjects to replay dynamic stimuli vividly in their mind. A new paradigm should elicit reinstatement in a natural way, i.e. without explicit instruction.

Finally, it was not possible to address the temporal dynamics of memory replay sufficiently. Since only patterns from a single time window were tracked, conclusions about temporal dynamics of reinstatement remained limited. Specifically, it was not possible to test whether overall reinstatement is compressed, i.e. if patterns from different time points during encoding reappear in closer temporal distance during retrieval. Additionally it was not possible to test whether distinct patterns from encoding remain in the correct temporal order when they are reinstated from memory. To this end, the following chapter will make use of distinct sub-events within a sequence of natural video-stimuli. These sub-events will then be tracked in episodic memory. In a behavioural experiment, this will be investigated via reaction

times; in an MEG experiment pattern-similarity will be leveraged. This will clarify the temporal dynamics of oscillatory pattern-reinstatement.

At the time of this thesis, the following chapter was about to be submitted under the title: *Speed of time-compressed forward replay flexibly changes in human episodic memory* in near identical form. Co-authors on this paper are Sebastian Michelmann[1], Bernhard P. Staresina[1], Howard Bowman[1, 2] and Simon Hanslmayr[1], affiliated to: 1. University of Birmingham, School of Psychology, Centre for Human Brain Health and 2. University of Kent, School of Computing.

Contributions

The experiments were conceived and designed by SM, SH and BPS. SM performed the experiments.

All data analysis was performed by SM under supervision of SH, the manuscript was written by SM under supervision of SH, BPS and HB.

Abstract

Remembering information from continuous episodes is a complex task. On the one hand, we must be able to recall events in a highly accurate way that often includes exact timing; on the other hand, we can somehow ignore irrelevant information and skip to events of interest. We here track continuous episodes that consist of different sub-events as they are recalled from memory. In behavioural and MEG data, we show that memory replay is compressed and forward. We detect neural replay by tracking temporally accurate patterns of activity, yet we statistically observe different compression levels in the neural data. Overall reinstatement of episodes is faster than their original perception; therein the replay of subevents occurs on a slower time-scale than the overall compression level permits. This renders memory replay as a flexible process in which participants replay fragments of fine-grained temporal patterns and are able to skip flexibly between sub-events.

Introduction

Episodic memory retrieval (Tulving, 1993) is a flexible process that operates at different timescales. In some instances, it is crucial for our behaviour to replay mentally events at the same speed of the initial experience: Re-enacting a classic movie scene relies on a temporally accurate representation of dialogue and events. In other instances it would be highly dysfunctional to recall our memories at the same speed they originally unfolded: We have to be able to reconstruct how we came to work today without zoning out at our desk for thirty minutes and must therefore be able to flexibly adjust the speed of our memory replay.

Previous studies that related the timescale between retrieval vs. perception of a particular event (Arnold, Iaria, & Ekstrom, 2016; Bonasia, Blommesteyn, & Moscovitch, 2016) showed that self-reported durations of memory replay are compressed. Findings of replay in rodents mirror this compression, but are mostly confined to the sequential reactivation of hippocampal place cells (Carr, Jadhav, & Frank, 2011; Foster & Wilson, 2006). One recent study observed the reactivation of static representations in oscillatory gamma power (Yaffe, Shaikhouni, Arai, Inati, & Zaghloul, 2017) on a faster timescale than during perception; however no time dimension was included in stimulus material or task, therefore limiting conclusions about replay trajectories. Notably, a recent fMRI study tracked episodic memory reinstatement over long episodes (50min) (Chen et al., 2017), finding reappearing spatial patterns on a compressed timescale. Importantly, however no study so far has leveraged electrophysiology to address directly the temporal dynamics of episodic memory reinstatement on a fine-grained temporal scale.

It is possible to investigate patterns from perception, as they reappear during memory retrieval with the use of multivariate similarity measures (Kriegeskorte, 2008; Nyberg, Habib, McIntosh, & Tulving, 2000; Ritchey et al., 2012; Staresina et al., 2012). Their extension to electrophysiological methods in humans (Jafarpour et al., 2014; Kurth-Nelson et al., 2015; Michelmann et al., 2016; Sols, DuBrow, Davachi, & Fuentemilla, 2017; Staudigl et al., 2015; Wimber et al., 2012; Yaffe et al., 2014) could now allow for the time resolved investigation of memory replay. Electrophysiological correlates that have been implicated in the representation of stimuli in perception and memory include the amplitude of high frequencies (Staresina et al., 2016; Yaffe et al., 2014, 2017; Zhang et al., 2015) and the phase and amplitude of low frequencies (Michelmann et al., 2016; Staresina et al., 2016; Staudigl et al., 2015). Importantly simultaneous EEG and multi-unit recordings in primates demonstrate an intimate relation between neural firing and the phase of slow oscillations in the EEG (Ng et al., 2013) during the perception of natural stimuli. This means that temporal properties of spiking neurons are reflected in the shape of waveforms (see also (Belluscio, Mizuseki, Schmidt, Kempter, & Buzsáki, 2012; Schyns et al., 2011)) and that similarity in oscillatory phase captures similarity of underlying neural activity (Michelmann et al., 2016; Ng et al., 2013; Schyns et al., 2011).

Importantly the notion of compression seems to be at odds with the observation of temporal similarity between perception and memory (Michelmann et al., 2016; Staudigl et al., 2015; Wimber et al., 2012; Zhang et al., 2015). A similar temporal pattern in memory implies that fragments of activity reappear at roughly the same speed. It is therefore unclear how these findings integrate into the temporal dynamics of mnemonic representations observed in behaviour. Investigating trajectories during memory replay requires a paradigm that prompts

participants to evoke continuous representations with distinct subevents from memory. This will make it possible to track fragments of these representations in episodic memory via multivariate analysis methods.

To this end, we asked subjects to associate static word-cues with 'video-episodes' consisting of a sequence of three distinct scenes. The three dynamic scenes thus formed a continuous six-second-long video. In encoding-trials, we presented a word-cue during one of the scenes. This allowed us to prompt memory replay in a natural way, i.e. we asked participants to recall in which of the three scene-positions they had learned an association during encoding. After completing this part of the task, we asked about the video-episode itself and confirmed correct memory. In a behavioural experiment, we investigated direction and speed of replay via measuring reaction times to the scene-position response. In a separate MEG study, we leveraged the content specific phase patterns that each scene elicited and used them as handles to track the speed of replay of the video-episodes. If memory replay were indeed compressed, we expected to find evidence for this compression in reaction times and in the reinstatement of neural patterns. This replay should be either forward or backward. In line with previous findings (Michelmann et al., 2016; Staudigl et al., 2015; Wimber et al., 2012; Zhang et al., 2015) we expected to find evidence for reactivation of temporal patterns, signifying replay at the same speed for fragments of neural activity. We further hypothesized that the rift between accurate representations and overall compression would be due to a flexible mechanism that allows subjects to skip between sub-events, as they replay episodes. Therefore, replay within sub-events should occur at a slower pace, whereas skipping between sub-events should occur fast.

Results

Reaction times in behavioural experiments suggest compressed and forward memory-replay

In the behavioural experiment, participants associated word-cues with one of three scenes within video-episodes (figure 301a). We used four continuous video-episodes, each consisting of three individual dynamic scenes. A trial-unique word-cue appeared in one scene during a video-episode. After a brief distractor task (figure 301b) subjects performed, in alternation, either a cued-recall (CR) retrieval task or an associative-recognition (AR) task (figure 301d, top). The AR task was included as a control condition, because active replay is arguably not required for recognition. In the CR blocks we presented participants with the word-cues (figure 301d, top-left). Their task was to recall the scene-position that was associated with the word-cue as quickly as possible. In AR blocks, subjects successively saw the word-cues superimposed on screenshots from encoding (figure 301d, top-right) and were asked to decide as quickly as possible, whether this association was intact or rearranged (figure 301d, top-right).

To address the direction and speed of memory replay, reaction times (RTs) at retrieval were compared between associations that were learned in the first, second and third scene-position of a video-episode. We only used RTs for correct hit trials (correct recall in CR and correctly recognized intact associations in AR blocks) and excluded trials in which the subjects were wrong or guessed (see Supplemental Information for the same analysis including correct guesses). This resulted in a 3x2 repeated measures ANOVA with the factors position and

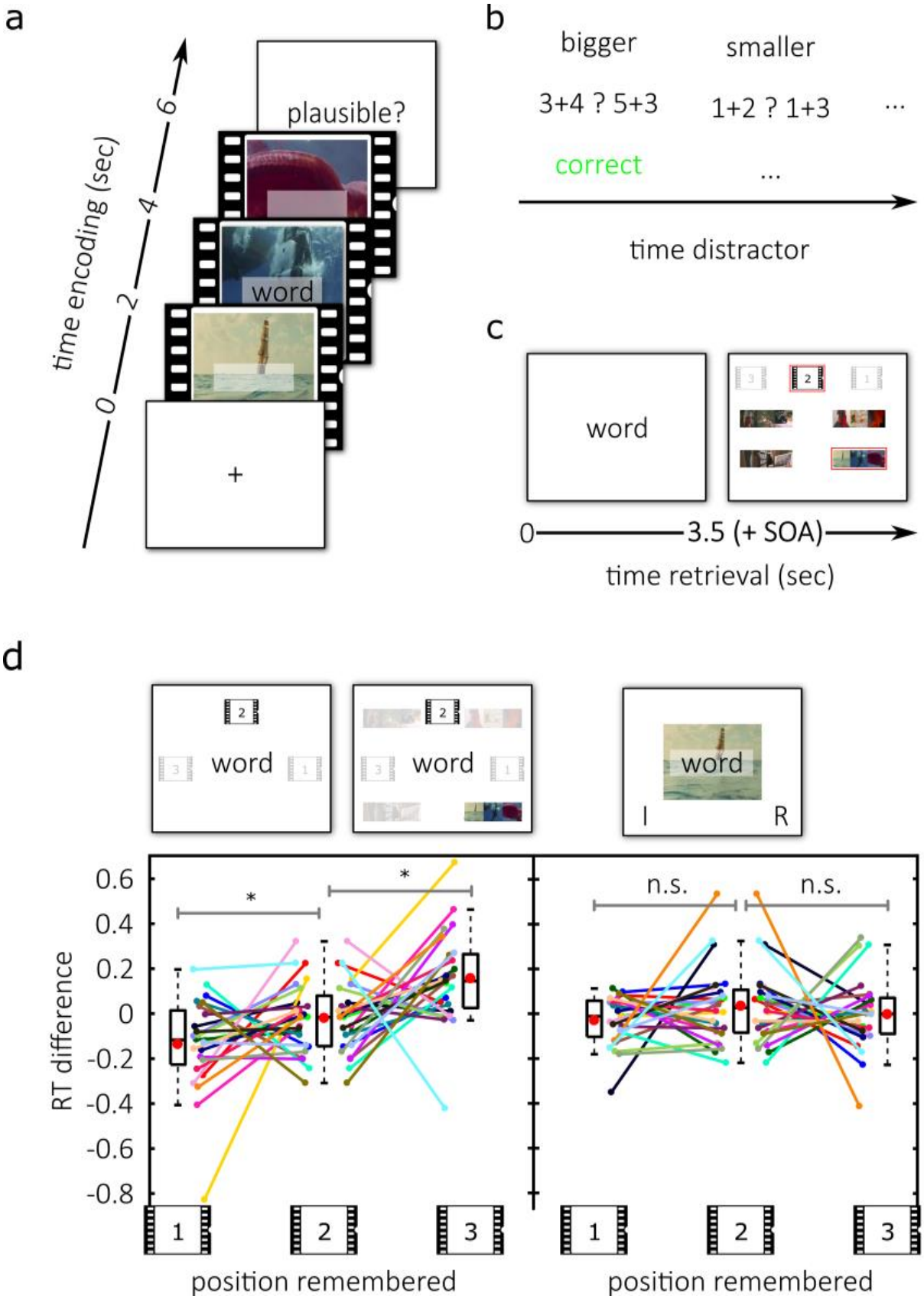
condition. Data from the CR task (figure 301d, left) indicated compressed replay in the forward direction. A significant main effect of position ($F_{1.85,\ 42.48}$ = 5.884, p = 0.007, log-RT: $F_{1.79,\ 41.26}$ = 3.375, p = 0.049) as well as a position by condition interaction ($F_{1.75,\ 40.34}$ = 5.9, p = 0.008, log-RT: $F_{1.76,\ 40.58}$ = 5.606, p = 0.009) were obtained. Both effects were driven by forward replay in the cued-recall condition (ANOVA: $F_{1.79,\ 41.19}$ = 9.082, p = 0.001, log-RT: $F_{1.60,\ 36.90}$ = 8.207, p = 0.002): During encoding, individual scenes of each video-episode lasted 2 seconds. During CR retrieval, however, associations that were learned in the first scene-position of a video-episode (mean RT = 2.5s) were recalled on average 116ms faster than associations that were learned in the second scene-position (t_{23} = -1.870, p = 0.037, log-RT: t_{23} = -2.4, p = 0.012). Associations that were learned in the second scene-position (mean RT = 2.617s in CR) were recalled on average 176ms faster than associations that were learned in the third scene-position (t_{23} = -2.767, p = 0.006, log-RT: t_{23} = -2.274, p = 0.016, (mean RT = 2.793s in CR)). The replay of the video-episodes was therefore compressed during CR, which replicated findings from a behavioural pilot experiment (see Supplemental Information). The average RT difference of 146ms per position corresponds to a compression factor of 13.7 during replay.

Might the effects be due to asymmetrical encoding of scene-positions? That is, one could argue that associations have a higher saliency when presented in the first scene-position, leading to higher confidence and shorter RTs during retrieval. Additionally, subjects can take more time to rehearse early associations during the remainder of the video-episode, perhaps resulting in the weakest memory trace for the last scene. Importantly, however, if the serial position merely affects the overall strength of the memory trace in our paradigm, we should observe comparable effects on cued recall (CR) and associative recognition (AR). Conversely,

if the effect is contingent on the need to mentally replay scene after scene, serial position at encoding should only exert an effect on the CR task.

Importantly, no differences in reaction times between scene-positions were evident in the AR task (figure 301d, right; ANOVA: $F_{1.64,\ 37.66} = 0.708$, $p = 0.472$, log-RT: $F_{1.61,\ 36.95} = 0.793$, $p = 0.435$, pairwise comparisons of positions: all ps > 0.199, all $BF_{01} > 2.158$). Together with the significant position by condition interaction, this confirms that the position effect on RTs is specific to the CR task and rules out a saliency-based explanation. Finally, we observed a significant main effect of condition with unscaled ($F_{1.00,\ 23.00} = 62.349$, $p < 0.001$) and log-transformed ($F_{1.00,\ 23.00} = 95.036$, $p < 0.001$) reaction times. This was due to faster RTs in associative-recognition blocks ($t_{23} = -7.896$, $p < 0.001$, log-RT: $t_{23} = -9.7487$, $p < 0.001$). Taken together these results are evidence that successful recall of elements from a continuous video-episode relies on compressed forward replay.

Figure 301: Experimental design and behavioural results



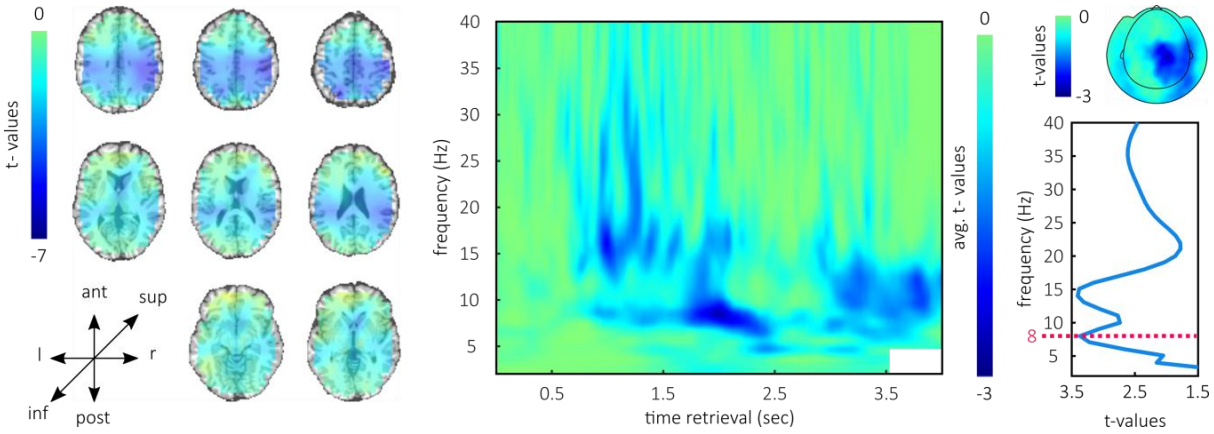
Legend to figure 301: Experimental design and behavioural results

(a) During encoding subjects repeatedly saw one out of four video-episodes. In one of three scenes, a word-cue appeared in the centre of the screen. (b) In the distractor block participants identified either the bigger or the smaller one of 2 simple sums. (c) In the MEG experiment participants saw the static word-cue during retrieval for 3.5 seconds, followed by a fixation cross for 250ms - 750ms. Subsequently they first picked the scene-position in which they learned the association and then confirmed the correct video-episode. (d) In the cued-recall (CR) condition of the behavioural experiment (left) participants selected the correct scene position as quickly as possible during retrieval. In an associative-recognition (AR) control condition (right) they decided whether the presented association (word superimposed on a screenshot) was intact or rearranged. In CR blocks, subjects were faster to recall an association that was learned in earlier scene-positions during encoding (bottom left). Importantly, in the control condition they performed the same encoding task and needed source memory for AR retrieval, however no modulation of reaction times was found. The y-axis denotes the difference to each participant's average reaction time in the respective condition. Spaghetti-plots show individual subjects. Boxplots are 25th and 75th percentile and the median; whiskers are maxima and minima, excluding outliers. Red dots within the boxplots depict the arithmetic mean. Significant differences are marked with a star, n.s. denotes non-significant in a post-hoc paired t-test comparison.

Broad decreases in oscillatory power accompany successful memory reinstatement

Successful memory reinstatement was associated with strong and sustained decreases in oscillatory power. Successfully remembered associations were those trials, in which subjects knew that they had identified the correct scene and the correct video-episode. Those trials were contrasted with the trials in which subjects either indicated a guess, or in which they selected the wrong scene-position and/or video-episode. A broad cluster emerged, in which oscillatory power was significantly lower when memory-retrieval was successful ($p_{cluster}$ < 0.001, figure 302, middle). This cluster included a sustained power-decrease in the lower alpha band. In a series of post-hoc t-tests, the same contrast was now tested on averaged oscillatory power across time and sensors. Inspection of t-values confirmed a local peak at 8 Hz (t_{22} = -3.367, p = 0.001, figure 302, right) which we previously linked to replay during episodic memory reinstatement (Michelmann et al., 2016). In order to derive the topography for the average power decrease at 8 Hz across time, a separate t-test was computed on every sensor. Maximal t-values were located over central sensors extending over right parietal sensors. The average power at 8Hz was next contrasted at every virtual sensor, resulting in an estimate of the spatial extent of power decreases in source space. Bilateral central and occipito-parietal areas as well as the medial temporal lobe displayed power decreases at this frequency (Figure 302, left). These findings replicate our previous findings of broad power decreases with a sustained decrease at 8Hz, in a paradigm that prompts subjects to replay a dynamic stimulus from memory.

Figure 302: Oscillatory correlates of successful memory



Legend to figure 302: Oscillatory correlates of successful memory

Successful memory was associated with broad decreases in oscillatory power. The middle panel shows the t-values averaged across time and sensors within the significant cluster. Low frequencies displayed a sustained effect over time. T-tests of the average power decrease across time and sensors expressed two local peaks in t-values, at 8 and 14 Hz (right panel). The topography of t-tests for the 8 Hz frequency at every sensor included central sensors and extended over right parietal sensors (right panel, top). Source reconstructions of the average power at 8Hz revealed power decreases on bilateral central and occipito-parietal areas as well as the medial temporal lobe.

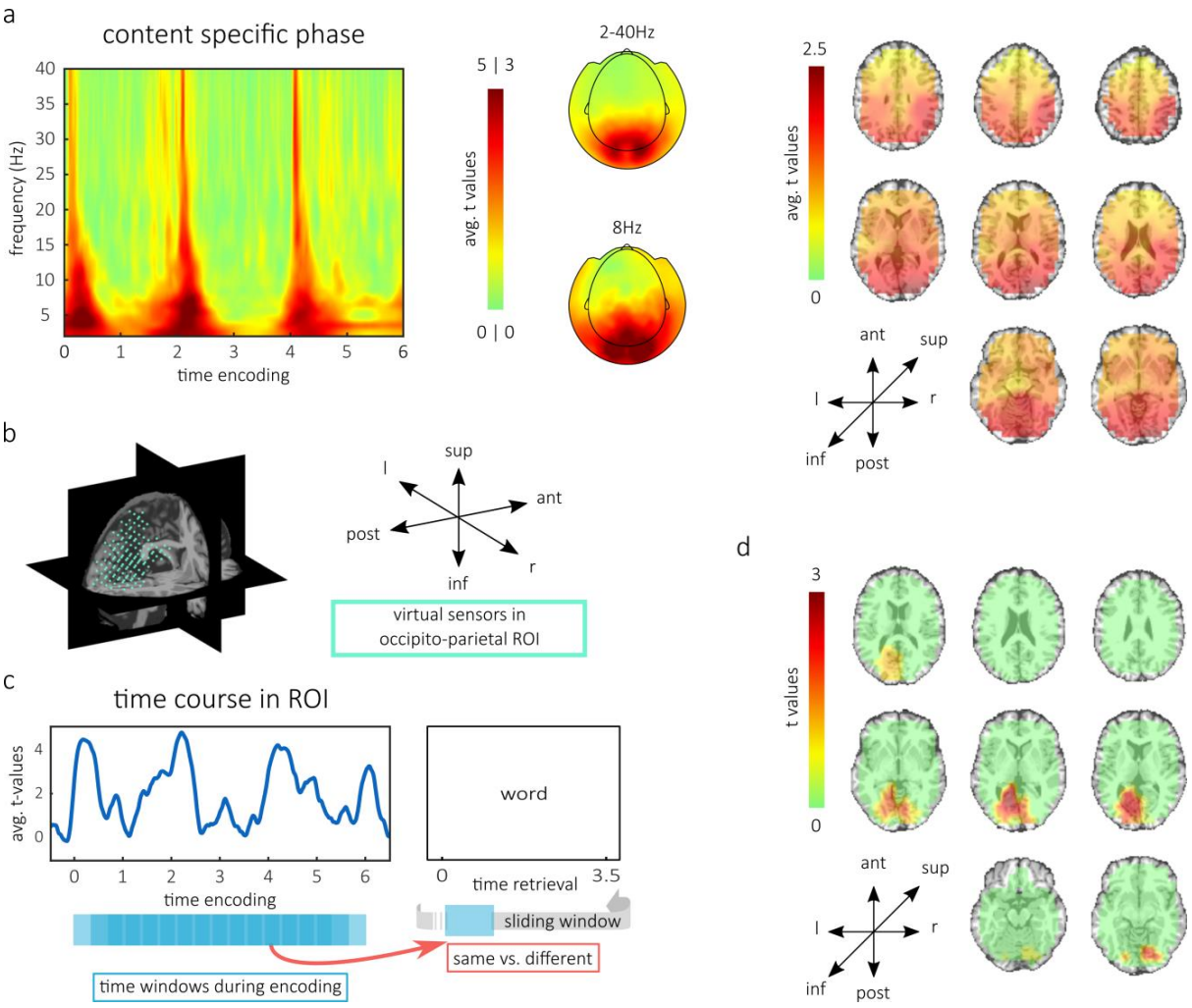
Low frequency phase patterns from encoding reappear during successful memory retrieval

In the MEG experiment, participants performed the same CR task as in the behavioural experiment, with the only difference being that they gave responses after the word-cue disappeared (figure 301c). In a first step, we asked whether perceptual content could indeed be distinguished based on oscillatory phase. To this end, we compared the inter-trial phase coherence (ITPC) between encoding-trials that we grouped according to their video-content, with the ITPC between trials that we grouped randomly. This has been used previously to reveal the content specific entrainment of cortical rhythms to naturalistic dynamic stimuli (Michelmann et al., 2016; Ng et al., 2013). The four video-episodes showed reliably distinguishable phase patterns during encoding ($p_{cluster} < 0.001$, figure 303a, left and middle). The significant cluster contained robust differences in the lower frequencies and showed a maximum over occipito-parietal sensors (figure 303a, middle). Consistent with our previous results (Michelmann et al., 2016), strongest differences were observed at the onset of each scene. Importantly, the 8 Hz frequency band was included in the cluster, which was previously linked to the reinstatement of phase patterns (Michelmann et al., 2016). Testing the 8 Hz phase differences on the source level revealed one broad cluster of content specificity during encoding ($p_{cluster} < 0.001$). Averaging t-values across this significant cluster over time revealed highest values in occipital and parietal locations (figure 303a right). Together, these results show that every sub-scene within the video-episodes was associated with a content specific fingerprint in oscillatory phase, which was maximal in a parieto-occipital region. In the

following, we used these sub-scene specific phase patterns at 8 Hz as handles to track replay in memory.

In a first step, we tested whether 8 Hz phase-patterns of the video-episodes were reactivated in memory. Therefore, we first contrasted phase-similarity between encoding-retrieval combinations of the same video-episodes (e.g. watching video A, recalling video A) with encoding-retrieval combinations of different video-episodes (e.g. watching video A, recalling video B). Similarity between encoding and retrieval phase patterns was analysed by a sliding-window approach (window size = 1 sec.), retaining a time resolved measure of memory replay (Lachaux et al., 2000; Michelmann et al., 2016; Mormann et al., 2000) (see figure 303c). On the source level, analysis was restricted to an anatomically defined occipito-parietal region of interest (ROI) following the results from the encoding phase and previous studies showing memory replay in these regions (Albers, Kok, Toni, Dijkerman, & De Lange, 2013; Ekman, Kok, & de Lange, 2017; Ji & Wilson, 2007; Michelmann et al., 2016) (figure 303b). Evidence for replay was found for hit trials (Hits; $p_{cluster}$ = 0.034; supplemental figure 306a, also see supplemental figure 306b for unmasked maps of t-values), suggesting that replay of video-episodes can be tracked in the phase of an 8Hz oscillation. Notably, we found no such replay effect for Misses, i.e. trials in which subjects either guessed, or did not remember the correct scene-position and/or video-episode. Furthermore, a direct contrast between Hits and Misses revealed significantly stronger replay for Hits compared to Misses ($p_{cluster}$ = 0.030, figure 303d), demonstrating the functional significance of this pattern-reinstatement for memory.

Figure 303: Reinstatement of oscillatory patterns from encoding



Legend to figure 303: Reinstatement of oscillatory patterns from encoding

(a) During encoding, the different video-episodes elicited content specific phase patterns. The left panel shows the averaged t-values across sensors in the cluster of significant content-specificity. Topographies in the middle are t-values within the same cluster, averaged across time and across all frequencies (top) or only for 8 Hz (bottom). Both topographies show maximal values over occipital and parietal sensors. The right panel shows the average t-values across time on virtual sensors, within the temporo-spatial cluster of significant differences at 8Hz. Occipital and parietal sensors expressed the maximal t-values. (b) Occipito-parietal region of interest (ROI) that we used for statistical testing of content-specific reactivation. (c) Time course of content specific phase during encoding, averaged across the ROI. Below, the sliding window approach is illustrated, in which all possible time windows from encoding were compared to each retrieval time window in phase coherence. Subsequently combinations of same and different content combinations were contrasted. (d) Cluster of significant differences between content-specific reactivation for successfully remembered and forgotten associations.

Compressed forward replay flexibly changes

Motivated by the above findings, we next addressed directly the question of direction and speed of replay, by statistically comparing at what time during retrieval, distinct phase patterns from encoding tend to reappear. We wanted to know if the phase-similarity to earlier encoding patterns was distributed more towards earlier times during retrieval than the phase-similarity to later encoding patterns. We therefore divided the encoding interval into 6 non-overlapping windows, centred at 0.5, 1.5, 2.5, 3.5, 4.5, and 5.5 seconds, then we derived and compared their distributions of phase-similarity across retrieval (figure 304a, left).

To test the direction of replay statistically across subjects, we used the following approach: We cumulated the similarity distributions across the whole retrieval time. This resulted in the cumulated similarity (CS) for every subject and every encoding-window. Similarity started at the beginning of the retrieval interval with a value of zero. It ended at the end of the retrieval interval, with a value of one (figure 304c). If phase-similarity to an encoding-window “A” cumulates earlier than phase-similarity to an encoding-window “B”, then the cumulated similarity for “A” is higher compared to “B” and consequently “A” is replayed earlier during retrieval than “B”. In other words, when the CS of one phase-pattern is higher than the CS of another, then the evidence for replay of that phase-pattern is leading over the other at that point. If, however replay of a phase-pattern is lagging behind the replay of another, the CS should be lower at that time point. We tested this relation statistically at every time point by comparing the cumulated similarity across all windows for each subject. The overall tendency is tested best by fitting a line across all six encoding windows. A negative slope indexes

forward replay, since earlier windows have higher values in the CS than later windows, a positive slope signifies backward replay.

Results revealed significant forward replay in two time windows (i.e. 135ms to 1919ms, and 3458ms to 3473ms after cue presentation, see Online Methods for some notes of precaution regarding the interpretation of the exact time-window). We can therefore conclude that there is a dominance of early encoding-patterns in early time points at retrieval relative to late encoding-patterns, which supports the notion of forward replay (see also Supplementary Information for further evidence supporting forward replay).

Notably the content specific reactivation that we found in temporal patterns signifies that subjects replay *fragments* of the video-episode at roughly the same speed as during encoding. We hypothesized that this rift between locally accurate replay and globally compressed replay was possible through the flexible skipping between salient elements (e.g. sub-events). We therefore wanted to test whether within sub-events (i.e. scenes) the compression level of neural replay was statistically weaker.

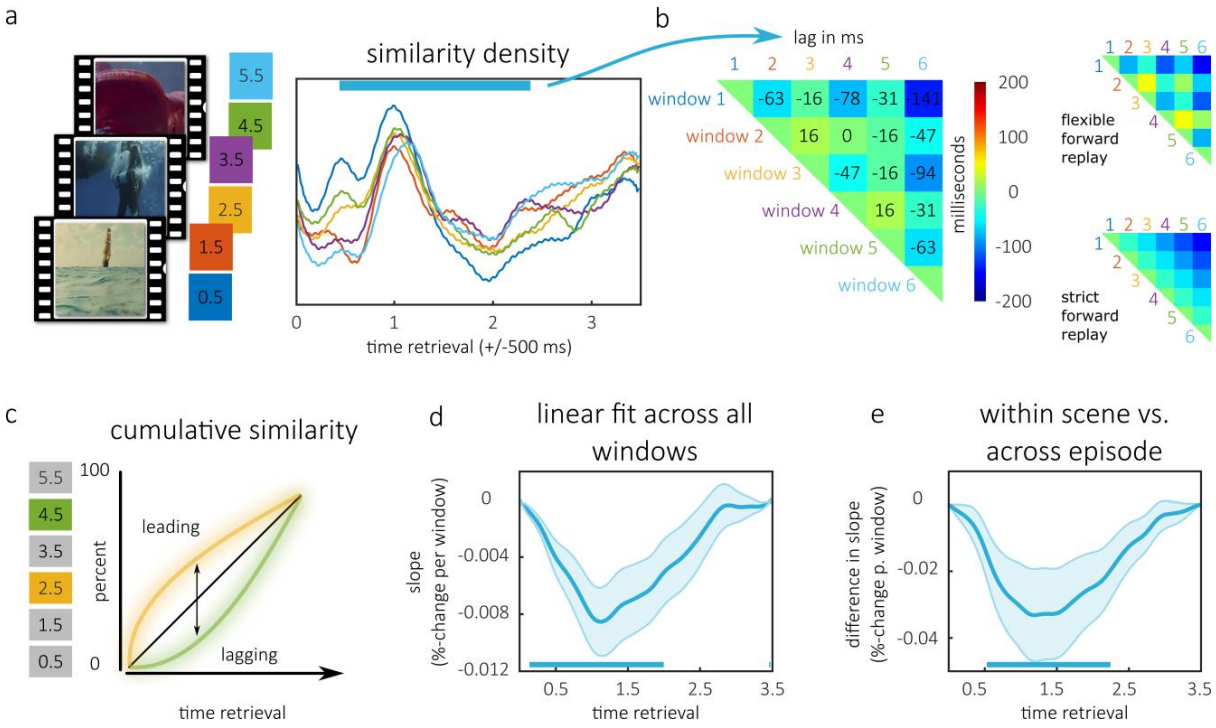
To this end, we extended the method of fitting a line across CSs to compare the compression of replay within individual scenes (i.e. within sub-events) to the overall compression level. Specifically, calculating the slope of the fitted line allows for an estimation of the speed of replay. This slope indicates the lag between replayed patterns in the retrieval interval, such that steep slopes indicate a long lag (i.e. slow replay). We fitted a separate line for each pair of encoding-windows that belonged to the same scene across their respective CSs and averaged the slopes across the three lines. The time interval between 442ms and 2350ms displayed slopes significantly below zero, confirming forward replay within scenes. More importantly, between 550ms and 2350ms at retrieval, slopes of windows within a scene were

significantly steeper (i.e. replay was slower) compared to the slope obtained across all encoding-windows (figure 304e). This means that, when participants replayed the first and second part of a scene, this replay was less compressed than we expected from the global compression level of the whole video-episode. Consequently, this also means that subjects did not replay every scene successively in every trial and not at the same speed. Taken together, these results show that memory replay does not occur at a constant speed; instead, the speed of replay seems to change flexibly depending on the replayed interval (figure 304b, right). We repeated these tests with those trials in which subjects did not remember the correct positional-scene or video-episode; however, we found no significant time-points for any of the contrasts, which demonstrates the implication of these replay effects in memory. In a further control analysis, we excluded the first 800ms of the retrieval interval for the similarity analysis in order to rule out that event related fields (ERFs) were driving similarities. Again, we found significant negative slopes between 812ms and 1212ms and slower replay within scenes in that window.

These results statistically support a flexible forward replay strategy. Via cross-correlations, we next derived a descriptive measure of the delay between the six sub-events during flexible memory replay (550ms-2350ms). The cross-correlation was computed on pairs of averaged and smoothed similarity distributions (figure 304b), which retained a time lag value for every combination of the six sub-events. The adaptive replay that we found is also visible in the pattern of time lags and can be illustrated with shorter lags between sub-events that belong to different scenes compared to sub-events that belong to the same scene (figure 304b, right). In contrast, to illustrate a strict and inflexible forward replay strategy, lags between the

sub-events should increase linearly according to their position at encoding (illustrated in figure 304b, right).

Figure 304: Chronometry of memory replay



Legend to figure 304: Chronometry of memory replay

a) *The 6 non-overlapping time windows from encoding illustrated next to a video-episode (left). The average similarity densities to these windows are on the right. The blue bar denotes where replay was significantly slower within scenes (see e).* ***(b)*** *Cross correlations of similarity densities within this window show the adaptive pattern. In this, lags between windows within scenes are bigger than lags between windows across scenes (right, top); with strict forward replay, all scenes would be replayed in order (right, bottom).* ***(c)*** *Illustration of the cumulative similarity (CS) approach used to test replay-dynamics. If evidence for a window statistically precedes evidence for another during retrieval, its cumulated similarity is higher.* ***(d)*** *Average slope of lines fit across all windows' CS, for each subject and time point. Negative slope indicates that earlier encoding-windows have higher CS values and signify forward replay.* ***(e)*** *Contrast of average slopes from the average fit across windows within scenes and a fit across all windows, supporting an adaptive replay framework. The blue bars in d and e denote significance.*

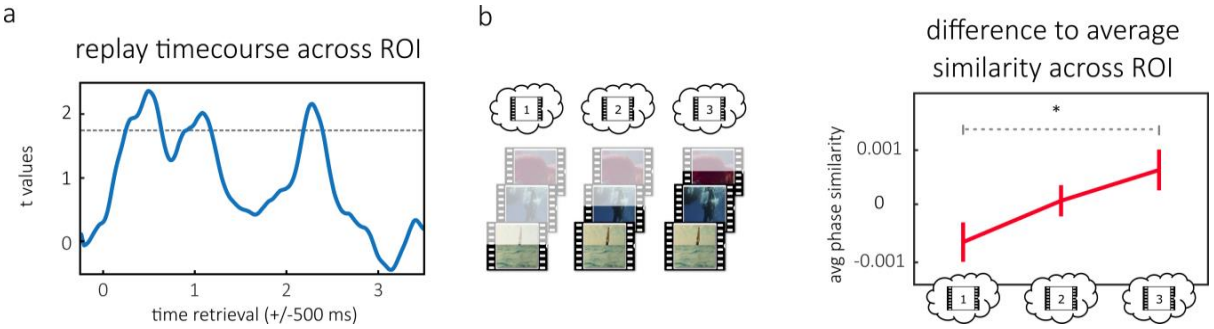
Time course of replay and further evidence for forward replay

We further investigated the time-course of reinstatement of the video-episodes. To this end, we computed a t-test of content-specificity at every time point during retrieval; precisely we assessed the average content-specificity across the whole ROI (figure 305a). Three peaks emerged at 442ms (t_{22} = 2.363, p = 0.014), 1042ms (t_{22} = 2.022, p = 0.028) and at 2163ms (t_{22} = 2.258, p = 0.017). Interestingly the last peak corresponds roughly to the period in which we observed average reaction times in the behavioural experiments.

Next, we further pursued forward replay. Following the behavioural results, we predicted that subjects replay overall more of the video-episodes when they have to recall later scene-positions (figure 305b, left). We reasoned that subjects accumulate evidence in a forward direction, until the correct association is identified. In the behavioural experiments, this would cause the increase in RT for associations that were learned later during encoding. Hence, in analogy to the analysis of the behavioural experiment, we split the retrieval trials according to the remembered scene-position. We assessed the average similarity to all sub-scenes in the corresponding video-episodes from encoding and compared it between trials: We contrasted trials in which an association from the first, second or third positional-scene of a video-episode was remembered. In this overall similarity to the corresponding video-episode should be higher, when participants recalled associations from later scene-positions. In a first ANOVA there was no significant difference in similarity depending on the remembered scene-position ($F_{1.83,\ 40.31}$ = 2.384, p = 0.109, linear contrast: $F_{1,\ 22}$ = 3.63, p = 0.07). There was, however, significantly less similarity to encoding, in trials in which subjects remembered an association from the first positional-scene compared to trials in which the third scene was recalled (t_{22} = -

1.909, p = 0.035, figure 305b). Note that this is equivalent to testing the hypothesis of a linear increase (i.e. directed linear contrast) with planned contrasts in the ANOVA.

Figure 305: Time course of reinstatement and forward replay in ROI



Legend to figure 305: Time course of reinstatement and forward replay in ROI

(a) *Time course of replay tested across the ROI. Three peaks are evident at 442ms, 1042ms and 2163ms.* ***(b)*** *Illustration of forward replay in which an association from the first, second or third scene was remembered (left) and difference to average similarity to encoding across the ROI (right, error bars are standard error of the mean). This shows higher similarity to encoding when associations from the third vs. first scene were remembered.*

Discussion

In this study, we tracked the replay of continuous episodes from memory. To this end, we used a novel paradigm in which participants associated unique word-cues with one out of three distinct scenes in seamless video-episodes. We prompted replay by asking volunteers, in which exact position (1, 2, or 3) they had learned each word-cue. Evidence in behavioural and neural data indicated that replay of memories takes place in a forward direction and at a compressed speed, i.e. memory replay was faster relative to perception. Notably, on a neural level we found indications for different speeds of replay: Fragments of temporal patterns reappeared at the same speed and the speed of replay within sub-events (i.e. scenes) of continuous video-episodes was slower than the overall compression level.

Importantly our finding of different compression levels implies that memory replay acts in a flexible way. The rift between the slower speed of replay within scenes and the overall compression is an aggregated observation that cannot hold on a single trial level. Specifically, it signifies that replay is not a simple concatenation of fragments because in a single trial, the sequential replay of three scenes would take longer than the overall compression permits. Consequently, participants must be able to skip between replayed fragments; importantly on average, the skipping between sub-events must take place on a faster temporal scale than the skipping within sub-events. A plausible interpretation of the observed pattern is therefore that replay of relevant information is initiated from the boundaries between scenes and that participants can flexibly skip between them. Event boundaries (Radvansky & Zacks, 2017) have been previously shown to trigger replay events during memory encoding (Sols et al., 2017). They could therefore also serve as starting points during memory retrieval, to initiate

the replay of information on a fine-grained temporal scale. Mechanistically the hippocampus has been suggested to preserve the temporal order of experiences (Davachi & DuBrow, 2015) and interactions between the hippocampus and visual cortex have been observed during memory replay in sleeping rodents (Ji & Wilson, 2007). In our data we consistently found reinstatement of fine-grained temporal patterns in sensory-specific regions (Michelmann et al., 2016). It is therefore possible that the hippocampus exerts control over sensory areas, when those regions realize the vivid reinstatement of sensory information. Specifically information-rich and temporally accurate representations could rely on sensory cortices whereas the hippocampus initiates replay, based on a sparse code (Hanslmayr et al., 2016). At first glance, the reinstatement of temporal patterns is also at odds with the observation of compression in general. An important implication from our findings is therefore that the temporally accurate reinstatement of patterns must be limited to fragments of the original perception. In other words, subjects probably omit non-informative (possibly redundant) parts of the video-episodes and therefore replay a shorter episode in memory, which contains less information. Previous work on mental simulation of paths supports this interpretation. The duration that participants take to mentally simulate a path increases, when this path includes more turns (Bonasia et al., 2016). In the same way, the duration of replay might depend on the overall number of relevant elements within a video-episode.

Another crucial result from our experiments is the forward direction of replay. This finding is in line with recent studies showing anticipatory activation of familiar paths in the visual cortex (Ekman et al., 2017) and evidence of forward replay of long narratives (Chen et al., 2017). Notably in the rodent literature, the task of spatial navigation appears to determine whether replay is backward or forward. At the end of a path awake rodents replay in a backward

fashion (Foster & Wilson, 2006), whereas animals that plan the path towards a goal display an anticipatory activation of place-cells in the forward direction (A. Johnson & Redish, 2007). Task requirements in our design could indeed have prompted participants to go mentally through the video-episodes in a forward manner. Speculatively, other designs (e.g. tasks requiring recency judgements) might therefore cause a backwards replay. This would be well in line with the flexibility in memory replay that we observed in the neural data, since a flexible mechanism could arguably guide replay in a forward and backward direction when skipping through events. An interesting additional question arising from this is, whether replay of fine-grained temporal patterns in the cortex can also be backwards.

Importantly our study also demonstrates how one can investigate these open questions. The design that we used to trigger the replay of distinct sub-events in a continuous episode can easily be adapted to a working memory context and our method to track oscillatory patterns allows for the investigation of replay in working memory, during rest and during sleep. We have repeatedly shown how to use the similarity in oscillatory phase to track content-specific reactivation, even when the exact onset of memory-reactivation is unknown. We here extended our previously developed method (Michelmann et al., 2016) to track distinct sub-events from continuous representations: In a statistically robust way we aggregated evidence across several repetitions and compared their distribution across time.

This investigation of temporal dynamics during human episodic memory replay has only recently become an option, when the tracking of multivariate patterns was extended to human electrophysiology (Jafarpour et al., 2014; Kurth-Nelson et al., 2015; Michelmann et al., 2016; Sols et al., 2017; Staudigl et al., 2015; Wimber et al., 2012; Yaffe et al., 2014). Leveraging a novel paradigm in combination with a method that can detect the individual

fingerprints in oscillatory patterns, we were now able, to observe for the first time to our knowledge, the fine-grained dynamics of memory replay on a behavioural and on a neural basis. Our data render memory replay as a flexible process, namely the compression level varies within replayed episodes: Some fragments reappear on a timescale that resembles the original perception and replay is less compressed within sub-events of continuous episodes, which suggests that participants were able to flexibly skip between sub-events during memory replay.

Methods

Participants

Balancing pilots

For each of the 2 pilot experiments that served to balance the video material, 18 subjects were tested (36 total). In the first balancing pilot, 16 female and 2 male, right handed subjects participated that were on average 18.67 years old (youngest: 18, oldest: 20). In the second balancing pilot, 15 female and 3 male right handed subjects were tested. Their average age was 21.39 years (youngest: 18, oldest: 47). 2 additional subjects were tested in balancing pilot 2, however their behavioural performance was at chance and they were excluded from the analysis.

Behavioural pilot and experiment

For the behavioural pilot 12 subjects (8 female, 4 male) participated that were on average 22.58 years old (youngest: 19, oldest: 29). 2 of the female participants were left handed, the rest were right handed. Data from 24 right handed volunteers (18 female, 6 male) was acquired for the behavioural experiment. The average age of this sample amounted to 22.79 years (youngest: 20, oldest 34).

MEG experiment

For the MEG experiment 24 volunteers (13 male, 11 female) participants were tested. Subjects were between 18 and 34 years old (mean: 23.92 years). 6 participants were left handed, 18 participants were right handed. 1 of the 24 subjects was excluded after pre-processing because of a persistent electrical artefact in the data that could not be removed with filtering.

6 additional subjects (4 female, 2 male) aged 19 to 28 years (mean: 22) were recorded but not analysed; they were discarded from analyses due to the following reasons: 2 subjects moved excessively throughout the recording session (maximal movement: 1.8 cm and 2.7 cm), 1 subject moved excessively throughout the session (maximal movement 1.4 cm) and fell asleep during the experiment. 1 subject felt unwell and aborted the experiment after approx. 10 % of the recording session, 1 subject only completed approx. 70 % of the recording session and moved more than 2 cm throughout the experiment. Finally 1 subject was lost due to technical failure during the recording. After preprocessing, the maximal movement of included participants across all trials (i.e. the range of all positions) was on average 5.89mm (s.d. = 2.62, min = 1.69, max = 9.09).

All included and excluded participants in the pilot studies, behavioural experiments and the MEG experiment, were native English speakers. Before participation they were screened for any neurological or psychiatric disorders. Their informed consent was obtained according to the ethical approval that was granted by the University of Birmingham Research Ethics Committee (ERN_15–0335A), complying with the Declaration of Helsinki.

Material and experimental set up

Videos

For each of the balancing pilots, a total of 12 short video-clips were used. Videos stemmed from a pool that was provided by Landesfilmdienst Baden-Württemberg, Germany, some of them were additionally edited. Each video-clip was a 2-second-long coloured, dynamic scene that featured a single action (i.e. a ship sailing or a diver jumping into the water). During the task, video-clips were always superimposed with a transparent text box (white box with alpha

value 0.9) in which the word-cue could appear. According to the behavioural results from balancing pilot 1, we edited or changed some of the scenes before the second balancing pilot. The final video-clips were 12 different scenes that belonged to four general topics. For the behavioural experiments and the MEG experiment the video-clips were then grouped into four seamless sequences of frames that formed a video-episode (i.e. a sequence of three scenes that belong to a general topic and form a short story). The 3 scenes of each video-episode were clearly distinguishable.

According to the second balancing pilot, scenes that were assigned to be in 1st, 2nd or 3rd position of video-episodes, did not differ significantly in difficulty (percent correct responses), when associated with a word-cue. Pairwise comparisons with t-test of positions 1 and 2 (t_{17} = 0.86, p = 0.4), 2 and 3 (t_{17} = 0.15, p = 0.88) and 1 and 3 (t_{17} = 1.4693, p = 0.16) and Bayes-Factor analysis supported the null Hypothesis of no difference between positions. This was supported either by substantial (BF_{01}>1.6) or strong (BF_{01}>3.3) evidence for the comparison of positions 1 and 2 (BF_{01} = 2.97) of positions 2 and 3 (BF_{01} = 4.07) and of positions 1 and 3 (BF_{01} = 1.65). Importantly reaction times in the second balancing pilot did not differ significantly between the video-clips that we finally assigned to be in position 1, 2 or 3. Pairwise comparisons with t-test of assigned positions 1 and 2 (t_{17} = -0.59, p = 0.56), 2 and 3 (t_{17} = -0.31, p = 0.76), and 1 and 3 (t_{17} = -1, p = 0.33) and Bayes-Factor analysis supported the null Hypothesis for the comparison of positions 1 and 2 (BF_{01} = 3.53) of positions 2 and 3 (BF_{01} = 3.95), and positions 1 and 3 (BF_{01} = 2.67).

Word-cues

Word-cues were downloaded from the MRC Psycholinguistic Database (Coltheart, 1981). For the balancing pilots, we divided 540 word-cues into 18 lists. Those lists did not differ in

Kucera-Francis written frequency (mean = 20.80, s.d. = 8.55), concreteness (mean = 506.50, s.d. = 90.07), imageability (mean = 521.04, s.d. = 69.51), number of syllables (mean = 1.63, s.d. = 0.68), number of letters (mean = 5.61, s.d. = 1.42) or word-frequencies taken from SUBTLEXus (mean = 15.22, s.d. = 14.07); specifically, "Subtlwf" was used (Brysbaert & New, 2009). In the balancing pilots, 12 of the lists were associated with a video-clip and 6 of the lists were assigned to become a distractor word. Across subjects each list was associated with every movie once and served as a distractor word six times. This was done to additionally control for list specific effects across subjects. An additional 9 words were randomly selected for practice.

For the behavioural pilot, the behavioural experiment and the MEG experiment, we divided 360 word-cues into 12 lists. Those lists were likewise balanced for Kucera-Francis written frequency (mean = 20.41, s.d. = 7.47), concreteness (mean = 518.72, s.d. = 78.39), imageability (mean = 530.78, s.d. = 60.17), number of syllables (mean = 1.56, s.d. = 0.62), number of letters (mean = 5.44, s.d. = 1.30) and word-frequencies taken from SUBTLEXus (mean = 15.07, s.d. = 13.04); again, "Subtlwf" was used (Brysbaert & New, 2009). Across participants each of the lists was associated with every video-clip twice. An additional 6 words were randomly selected for practice.

Response scales

To create the response scales (see figure 301c-d), we took Screenshots from the video-clips. In the balancing pilots, we adjusted brightness and contrast, so that no screenshot appeared more salient. For the behavioural pilot, the behavioural experiment and the MEG experiment the numbers 1, 2 and 3 were framed by a square which resembled a frame from an old film. Those represented the first response options, i.e. the choice between scene 1, scene 2 or

scene 3. For the second response, i.e. the response about the correct video-episode, the 3 screenshots from the concatenated video-clips were presented next to each other for each of the 4 choices. In the control condition of the behavioural experiment, the response option intact/rearranged was realized with a screenshot which was of the same size as the videos during presentations. This screenshot was superimposed by a transparent textbox containing a word-cue. The words intact and rearranged were displayed at the left and right of the textbox as response options. The left/right position of these options was balanced across participants.

Behavioural setup

Visual content was presented on an LED monitor (Samsung syncmaster 940n at a distance of approximately 60 cm from the subject's eyes. The monitor was set to a refresh rate of 60 Hz. On a screen size of 1280 x 1024 pixels, the video-clips had the dimension of 360 pixels in width and 288 pixels in height on the screen. "Helvetica" was chosen as the general text font, font size was set to 22 for instructions and to 28 for word-cues. Black text (rgb: 0, 0, 0) and movies were presented against a white background (rgb: 255, 255, 255).

MEG setup

MEG was recorded at the Sir Peter Mansfield Imaging Centre (SPMIC) in Nottingham, UK. Subjects performed the experiment in a seated position at a distance of approximately 60 cm from a white screen. The image was projected onto the screen using a PROPixx projector (VPixx Technologies, Saint-Bruno, Canada) that operated at a refresh rate of 60Hz and a resolution of 1920 x 1080 px. The projected image appeared at a size of approx. 40 x 22.5 cm on the screen. Accordingly, the video-clip appeared in a dimension of approx. 15 x 12 cm. An eye tracker (EyeLink 1000 plus, SR Research, Ontario, Canada) was placed in front of the

screen. The tracker was mounted in an upwards facing orientation, slightly below the visible display, on a small wooden board. In this setup it tracked the subject's left eye from below and from a distance of approximately 55 cm.

Procedure

Balancing pilots

The balancing pilots were realized to ensure that no material specific differences between the first, second or third position of a video-episode were to be expected in the following experiments. To this end, the considered video-clips were presented as single scenes during learning blocks, where they were superimposed by a transparent text box, containing a word-cue. Upon informed consent and completion of screening questionnaires, participants sat down in front of the screen and received a standardized instruction for the task. All subjects saw the video-clips for familiarization and completed a practice version of the task before starting. Participants performed 15 runs of an encoding block, in which their task was to vividly associate the word-cue with the corresponding video-clip, a short distractor block, in which they did some easy math and a retrieval block in which they retrieved the associated video-clips as quickly as possible, whilst presented with a word-cue.

During encoding a fixation cross was displayed for 2 seconds. Then the video-clip and word-cue played for 2 seconds. Finally a fixation cross appeared again for 2 seconds. In every encoding block each video-clip was presented twice amounting to a total of 24 trials in every block. The video-clips were presented in a balanced but randomized order such that no movie was presented more than 2 times in a row.

In the distractor block subjects solved simple math problems. For 45 seconds they were either presented with the word bigger or with the word smaller and two single digit sums (e.g. 4+5 and 3+2). Their task was to select the correct sum (i.e. either the bigger or the smaller sum). Feedback was given in the form of the words “correct” and “wrong” appearing in green and red respectively on the screen.

For each retrieval block the current 24 cues were mixed with 12 new distractor words in a randomized way, such that items corresponding to the same video-clip or targets corresponding to a distractor did not appear more than 2 times in a row.

In the retrieval block subjects were asked to select the video corresponding to the word-cue as fast as possible. The target (i.e. the screenshot from the correct clip) and two lures (two screenshots from a different clip) were presented on three positions around the word-cue. The positions formed a triangle with equal distance from the centre to the left and right and 2/3 of that distance above the word-cue. In addition to those three response options, a question mark was displayed below the centre of the screen. This response option was available in order to indicate that no video-clip-screenshot was identified as the correct target.

In order to control for effects from specific screen positions, the mapping of targets to positions was randomized but balanced, such that the target was presented on every position 8 times. To control for item specific effects, the two lures that were presented with the target were assigned, such that every video-clip-screenshot served four times as a lure to a target and the same screenshot was never on both lure-positions. The 12 additional distractor words were, by definition, only paired with lures, we balanced the random mapping, such that every video-clip-screenshot served once as a lure on every position.

In order to respond, participants placed the index finger of their dominant hand on the number 2 of the numeric keypad of the keyboard. The index finger rested there as long as no response was required. When presented with the cue-word subjects could either press one of the numbers 4, 6 and 8 which corresponded spatially to the response options (screenshots) on the screen and were in approximately the same distance from the starting position (number 2), or they could press 0 which corresponded to the question mark. Available buttons were highlighted with coloured stickers to facilitate orientation.

Whilst presented with the word-cue, subjects had maximally 4 seconds to select their answer. At the end of every retrieval block participants were reminded that associations from the previous block were now irrelevant and had the opportunity to take a self-paced break.

Behavioural pilot and behavioural experiment

In the behavioural pilot, subjects saw video-episodes that consisted of 3 distinct scenes. Those scenes comprised of the video-clips from balancing pilot 2, which ensured that no material specific differences were to be expected between position 1, 2 and 3 of the video-episodes; not in memory performance and most importantly not in reaction time. Participants first completed the screening questionnaire and gave informed consent. After instruction with the task, they saw the video-episodes twice for familiarization and were instructed to pay attention to their 3-scene-structure, such that they could confidently identify the first, second and third scene of each video-episode.

After a short practice version of the task, the experiment started. It was again a sequence of encoding, distractor and retrieval blocks. In each encoding block subjects learned a series of associations. They first saw a fixation cross on the screen for 2 seconds. After that one of the four video-episodes played for 6 seconds. During this video-episode a transparent textbox

was overlaid on the video. In one of the three scenes, a word-cue appeared in the textbox and disappeared again with the end of the scene. Subjects were instructed to form a vivid association between the word and the precise scene of the video-episode, such that they could later recall that exact scene and video-episode upon presentation with the word-cue. We randomized the presentation of the associations in a balanced way, such that no video-episode was presented more than twice in a row and a word-cue did not appear in the same position more than twice in a row. Additionally every position within every video-episode was associated with a word cue once within 12 subsequent associations.

After each video-episode a fixation-cross showed for 1 second then subjects rated the plausibility of the association between word-cue and scene. Three response options were labelled with “not plausible”, “plausible” and “very plausible” and could be selected with the buttons 4, 5 and 6 on the numerical pad of the keyboard. The plausibility rating served to keep participants engaged in the task and support memory formation. In the distractor block, subjects were presented again for 45 seconds with simple math problems and had to decide which one of two single digit sums was either bigger or smaller. For the retrieval block the word-cues were now randomized again in a balanced way, such that word-cues corresponding to the same video-episode regardless of position, or to the same position regardless of video-episode, did not appear more than twice in a row.

Retrieval block started with a fixation cross, displayed for 2 seconds. Then a word-cue appeared in the centre of the screen and the three framed numbers appeared on a triangle around the word-cue. Participants were instructed to select, as quickly as possible, in which of the three scenes they learned the word. For this choice they only saw the numbers 1, 2 and 3; after they made this choice, screenshots forming the four video episodes appeared in the

four corners of the screen. Participants were asked to indicate now, to which of the four episodes the selected scene belonged. The position of the numbers 1, 2 and 3 as well as the mappings of the four screenshot-sequences to the screen positions were randomized in a balanced way, namely all possible permutations of 1, 2 and 3 were randomly mapped onto the three positions within 6 subsequent trials and all possible permutations of the four positions of the video-episode screenshots were used within 24 trials. This was done to control for any potential effects from specific screen positions on reaction times or position specific response preferences. In order to respond, volunteers were asked to place the index finger of their dominant hand on the number 5 of the numerical pad on the keyboard. The surrounding numbers 4, 6 and 8, which form a triangle around the number 5 were highlighted with red stickers and served as the response options for the scene-response (first response: 1, 2 or 3). Those buttons corresponded spatially to the position of the permuted numbers 1, 2 and 3 on the screen. Accordingly the buttons 1, 7, 9 and 3 which form a square on the numerical pad, were available for the second response which informed about the correct video-episode. Importantly subjects were instructed to make all responses with the index finger of the dominant hand and go back to the starting position after every response, i.e. leave the finger resting on the button 5. At the end of every retrieval trial, a scale appeared on which subjects rated the confidence in their response. Three options were labelled with “guess”, “sure” and “very sure” and corresponded to the buttons 4, 5 and 6 on the numerical pad.

Participants performed a variable amount of runs of encoding, distractor and retrieval blocks that varied in length according to their individual memory performance. The first block comprised of 24 items, subsequently its length was adjusted. If more than 70% of items were

recalled correctly in the last block (i.e. correct scene and movie were selected), 12 items were added to the next block, if less than 50% were recalled correctly, 12 items were removed from the following block. All blocks comprised at least of 12 associations that had to be learned all participants completed 360 trials in total.

In the final behavioural experiment subjects performed exactly the same task as in the behavioural pilot experiment, however, every other block was performed with a different retrieval task. Specifically subjects performed the same learning paradigm, yet they did alternating retrieval blocks of cued-recall (CR, see above) and associative recognition (AR). In the AR blocks subjects were presented with a screenshot of a single video-clip, representing one of three scenes within a video-episode. The centre of the screenshot was again superimposed with the transparent textbox containing one of the previously learned word-cues. The association between word-cue and video-clip could either be intact, i.e. the word was learned in this exact position within the video-episode, or it could be rearranged. In the latter scenario, a different video-clip from the same video-episode was superimposed by the word-cue. This means that word-cues were either presented in the correct position or in the wrong position within the video-clip. Participants were again instructed to decide as quickly as they could, whether the association was intact or rearranged. Block-size was adjusted in the same way with percent of correct responses measured as 200*(Hits - False Alarms)/N, with Hits being the number of correctly identified intact associations and False Alarms referring to the number of rearranged associations that were declared intact and N referring to the number of trials in the last block. Response buttons for the intact/rearranged choice were 4 and 6 on the numerical pad, which are in equal distance from the number 5, where the index finger of participants' dominant hand rested comfortably at the beginning of each trial. After

the experiment participants answered a few interview questions regarding eventual strategies and their subjective experience of the task.

MEG experiment

In the MEG experiment volunteers learned associations between video-episodes and word-cues in the same way as in the behavioural experiment. Memory retrieval was similar to the behavioural pilot experiment (i.e. a cued-recall task); however a fast response was not required (see below). Upon informed consent and screening questionnaires, participants received the instructions for the task on a laptop outside the scanner. They familiarized themselves with the video-episodes twice, paying close attention to their structure. It was ensured that every participant was able to identify the three different scenes of a video-episode. In a short practice, they performed a block of encoding, distractor and retrieval with the six example words. The head-localization coils of the MEG system were attached to the participants' head and their positions were logged along with the shape of participant's head (see Data Collection). Subsequently volunteers were seated in a comfortable position under the MEG helmet. Subjects used a single button on each of two response pads with their left and right index finger. After the eye tracker was mounted and calibrated, the experiment started.

The MEG experiment was again a sequence of encoding, distractor and retrieval blocks. In each encoding block subjects learned a series of associations between scenes in video-episodes and unique word-cues. Participants first saw a fixation-cross on the screen for 1 second. After that one of the four video-episodes played for 6 seconds overlaid with a transparent textbox. In one of the three scenes of the video-episode, the unique word-cue appeared in the textbox and disappeared again with the end of the scene. The task was again

to form a vivid association between the word and the precise scene of the video-episode, to recall later the exact scene and video-episode, when only presented with the word-cue.

After the video-episode, the fixation-cross appeared again for 500ms. Finally, the two response options 'plausible' and 'not plausible' appeared on the left and right of the screen. Subjects used the left or right button to indicate whether the association between video-scene and word-cue was plausible to them. This task kept participants engaged and supported their memory performance.

The order of presentation was randomized in a balanced way: no video-episode was presented more than twice in a row and a word-cue did not appear in the same position more than twice. Additionally every position within every video-episode was associated with a word cue once within 12 subsequent associations. In the distractor block, subjects solved simple math problems for 45 seconds: They had to decide which one of two single digit sums was either bigger or smaller, using a left or right button press. For the retrieval block the word-cues were now randomized again in a balanced way, such that word-cues corresponding to the same video-episode regardless of position, or to the same position regardless of video-episode, did not appear more than twice in a row.

Trials of the retrieval block started with a fixation cross that was displayed for 1 second. Then a word-cue appeared in the centre of the screen for 3.5 seconds. In this time interval subjects remembered in which exact scene they had seen this word. After a random time interval between 250ms and 750ms the response scale appeared. The time interval for retrieval was chosen based on reaction-time data from the behavioural experiments, such that participants could comfortably remember the correct association. The first response option required the selection of the correct scene. To this end pictograms featuring the numbers 1, 2 and 3 were

displayed on the top of the screen. The mapping of the numbers 1, 2 and 3 to the three screen-positions was randomized in a balanced way such that all possible permutations appeared within 6 subsequent trials. Participants could now move a red square, which framed the current selection. By pressing the left button they changed their selection by moving the frame clockwise. This selection was confirmed by pressing the right button. Note that this button assignment ensured that subjects would always prepare the same response during the retrieval trial, regardless of the memory content. This is important to control for trivial but systematic differences that correlate with memory content in the retrieval interval.

After the position was selected, the two other position pictograms were overlaid with transparency (alpha = 0.9), such that the selected option remained highlighted on the screen. The concatenated screenshots from the video-episodes appeared below the position-pictograms and the red selection frame could be moved clockwise with the left button. Again the selection was confirmed with the right button. To ensure that subjects tried to recall the correct position as soon as they were presented with the word-cue (and did not wait until the response scale was presented), there was a time limit of 4 seconds to select the correct position and again to select the correct movie. To allow for flexibility due to hasty or imprecise selections, 200ms were added to this time limit, whenever the selection-frame was moved. Participants did not know about this increment; all participants selected their responses quickly but not hastily. If the time limit was exceeded, the message 'too slow' appeared at the centre of the screen for 5 seconds. Altogether the time limits were designed, such that subjects could comfortably remember the correct association during the presentation of the word-cue, and were eager to select the two responses straight away. After the associated video-episode was selected, unselected response options were overlaid

with transparency for 300ms, then the two options 'guess' and 'know' were presented on the screen to give the participant the opportunity to communicate whether the selected answers were based on a guess.

Participants performed a variable amount of runs of encoding, distractor and retrieval blocks. The blocks varied in length according to their individual memory performance. The first block comprised of 24 items, subsequently its length was adjusted. If more than 90% of items were recalled correctly in the last block (i.e. correct scene and movie were selected), 24 items were added to the next block, If more than 70% of items were recalled correctly, 12 items were added to the next block, if less than 50% were recalled correctly, 12 items were removed from the following block, if less than 40% were recalled correctly, 24 items were removed. All blocks comprised at least of 12 associations that had to be learned; all participants learned and recalled a total of 360 associations.

Data Collection

Stimulus presentation and the collection of behavioural data was realized on a standard desktop computer running MATLAB 2014b (MathWorks) under Windows 7, 64 Bit version. Stimuli were presented through the Psychophysics Toolbox Version 3 (Brainard, 1997). In the behavioural experiments, responses were collected from button presses on the numerical pad of a wired keyboard (Model 1576, Microsoft Corporation, Redmond, US). In the MEG experiment, fibre optic response pads were used.

Neurophysiological data were collected with 275-channel CTF MEG (CTF, Coquitlam, BC, Canada) at the Sir Peter Mansfield Imaging Centre (SPMIC) in Nottingham, UK. The system was used in third-order gradiometer configuration, recording at a sampling frequency of 600

Hz over the whole duration of the experiment. Three localization coils that were attached to the participants' left preauricular point (LPA), right preauricular point (RPA) and to a point slightly above the nasion (NAS) were energized during the recording session. This was done to localize the head position relative to the sensors.

Head digitization was collected with a Polhemus ISOTRAK device (Colchester, Vermont, USA). A minimum of 500 points on the scalp were logged relative to the positions of the three fiducial points (LPA, RPA, NAS). Individual anatomical data was acquired via magnetic resonance imaging (MRI) (3T Achieva scanner; Philips, Eindhoven, the Netherlands) with an MPRAGE sequence covering the whole head at $1mm^3$ resolution. MRIs were either measured at the SPMIC or at the Birmingham University Imaging Centre (BUIC).

For 17 of the included subjects (23), eye tracking (Eyelink 1000 Plus, SR Research, Ontario, Canada) was recorded on a separate Computer provided by the manufacturer at a sampling rate of 2000 Hz. The data was additionally written into 3 analogue input channels of the MEG system via the EyeLink Analog Card. The eye tracker was used in remote mode tracking the pupil and corneal reflection with a 16mm lens. It was calibrated and validated using 13 points on 80% of the screen, which contained all of the task relevant information.

Analysis of Reaction Times

We defined reaction time (RT) as the time to the first response after onset of the word-cue. All RTs faster than 200ms were considered implausible and discarded from further analysis. Additionally RTs that were 2.5 standard deviations above the mean RT were discarded. The means of remaining RTs were then tested statistically. To account for the non-normal

distribution of RTs (Ratcliff, 1979), all statistical tests are also reported for log-transformed RTs.

Preprocessing of Neural Data

The data was preprocessed in MATLAB 2015a (MathWorks) with a combination of functions from the Fieldtrip toolbox for EEG/MEG analysis (Oostenveld et al., 2011) and custom written scripts.

For the sensor level analysis the 3rd order gradiometer correction was first applied, then the continuous recording was filtered with a Butterworth IIR filter of 4th order with a stopband of 49.5 to 50.5 and its harmonics (99.5 - 100.5, 149.5 - 150.5, 199.5 - 200.5, and 249.5 - 250.5) to reduce the line noise artefact. Additionally the data was filtered with a stopband of 59 – 60 to attenuate noise with a centre frequency of 59.5 Hz.

Subsequently the data was segmented into trials that started 1.5 seconds prior to video-onset and ended 7.5 seconds after video-onset at encoding. Trials at retrieval started 1.5 seconds prior to the onset of the word-cue and ended 5 seconds after onset of the word-cue. The dataset was combined with the downsampled and segmented trials from the eye tracking.

To remove activity from eye blinks and noise, and to detect heartbeats, Independent Component Analysis (ICA) was used (Delorme & Makeig, 2004). For the computation of the ICA unmixing matrix, trials containing coarse artefacts or showing strong muscle activity were heuristically excluded. Additionally the data was downsampled to 250 Hz and cut to 1 second-long segments; the obtained unmixing matrix was then applied to the original trials.

When possible, we compared independent components with the eye tracking data; we removed those components that picked up eye-blinks or eye-movement related activity.

Additional components that picked up channel-noise or electrical noise were likewise removed from the data. Components which contained a clear R-wave of the QRS complex in a heartbeat were stored for later peak-detection and regression; the remaining components were projected back to a channel representation.

Finally, all data was inspected visually and trials containing artefacts were removed from later analysis. After visual inspection 84.26 % (S.D. = 8.29 %) of trials remained.

Heartbeats were removed with a regression based approach: An iterative peak detection algorithm was applied to the ICA-component showing the clearest R-wave. It served as a proxy for ECG. This was done only for the remaining trials after visual inspection. Before peak-detection the heartbeat-component was highpass-filtered (4Hz, 4^{th} order Butterworth). The peak detection algorithm first calculated a plausible maximum of heartbeats that were not to be exceeded. The signal was z-scored and thresholded. Local peaks were detected by finding local maxima in clusters of z-scores that were above threshold. Subsequently the threshold was lowered up to a z-score of 2. With lowering threshold, increasingly bigger areas around the peaks were excluded from further peak detection. If the maximum number of plausible peaks was exceeded the threshold was no longer lowered. A heartbeat template was now created by averaging 500ms long segments around the peaks. Gaps in the continuous recording were subsequently zero-padded in order to convolve the component with the template. Peak detection was then repeated on the convolved time course and a new template was built from these peaks for subsequent convolution (Tal & Abeles, 2013). After a few repetitions the template converged and the resulting peaks were controlled manually, even though errors rarely needed to be corrected.

Instead of simply subtracting the mean template from the data, the trials were now split into four big segments and a general linear model (GLM) was built around the peaks in each segment. A high pass filter (1Hz, 4^{th} order Butterworth) was applied to the data, only for the purpose of fitting the model. The GLM consisted of a separate repeated measure factor for each time point in the heartbeat, beginning 280ms before the peak and ending 720ms after the peak. Additionally a separate factor was included for every heartbeat, which modelled the offset between 280ms pre-peak and 720ms post-peak. Furthermore an offset factor for the overall segment was included. The solved model was then applied to every channel. The data model $\hat{y}$ was built by using only the repeated measure factors which modelled each time point within the heartbeat (i.e. the beta weights for offsets were set to 0). After visual inspection, this resulting model of the heartbeat was subtracted from each original channel.

For the source level analysis the anatomical data was first aligned to the digitized head positions. This was done by extracting the surface of the head from the anatomical MRI; in a first step a rough alignment was done manually, then the Iterative Closest Point (ICP) algorithm implemented in fieldtrip was used to match the surface to the point-cloud of the head digitization, finally this solution was controlled and eventually corrected again manually. The transformation to the aligned space was subsequently applied to the segmentation of the brain, which was likewise extracted from the anatomical images. To correct for head movements, the average head positions within the trials were first clustered, such that one positional-cluster was built for every 10 trials. Subsequently a separate lead field was computed for every cluster and then averaged to obtain an average lead field across all trials for each participant (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). Importantly 'all trials' refers to the trials that were included in a given contrast (e.g. for the contrast of Hits and

Misses at retrieval, encoding trials were not included in the computation of the lead field). Before the source level analysis, the 3rd order gradiometer correction was applied to the cut raw-data, lead fields were adjusted accordingly. Finally the data was demeaned and bandpass filtered between 4 and 15 Hz. The position of virtual sensors in individual brains was derived from a 1 cm spaced grid, which was placed 6mm below the surface of the cortex into the MNI brain and then spatially warped into individual brains. This was done via the inverse of the transformation describing their normalization and resulted in 1407 individual virtual sensor positions which were anatomically equivalent. Finally, to reconstruct activity on virtual sensors a linearly constrained minimum variance (lcmv) beamforming approach, implemented in the Fieldtrip toolbox (Oostenveld et al., 2011), was used. Filter coefficients were again computed on all data in a given contrast.

Analysis of oscillatory power

To estimate oscillatory power at retrieval, the Fourier-transformed data was multiplied with a complex Morlet wavelet of six cycles. This was done in steps of 10ms for every full frequency between 2 and 40Hz. The raw power was then obtained from the squared amplitude of the Fourier spectrum. Across all trials within the contrast (i.e. Hits and Misses), a baseline was computed as the average power between 1 second pre-stimulus and 4 second after stimulus onset (Long et al., 2014). Trials were then normalized by subtracting the baseline and dividing by it $(activity_{tf} - baseline_{f})/baseline_{f}$, with t indexing time and f indexing frequency.

Region of Interest (ROI)

An occipito-parietal region of interest (ROI) was derived from the AAL atlas (Tzourio-Mazoyer et al., 2002). To obtain the ROI in form of a group of virtual sensors, the sensor-positions in MNI-space were assigned to the nearest described AAL-region, based on their Euclidean distance. The occipito-parietal ROI comprised of bilateral AAL-regions: angular gyrus, calcarine sulcus, cuneus, inferior occipital cortex, inferior parietal lobule, lingual gyrus, middle occipital gyrus, precuneus, superior occipital gyrus, superior parietal lobule, supramarginal gyrus.

Content specific oscillatory phase at encoding

During encoding participants repeatedly watched the same video-episodes. Hence, it was possible to assess content specific properties if they were more similar between trials of same content than between trials of different content. In order to determine whether the ongoing oscillatory phase was specific to individual perceptual content, trials were grouped into 4 sets according to the video-episode that was perceived. The complex Fourier spectrum was again derived by multiplying the Fourier-transformed data with a complex Morlet wavelet of six cycles. Then, inter-trial phase coherence(Tallon-Baudry et al., 1996) (ITPC) was computed across the trials of same content (i.e. for each of the four trial-groups). This was done at every full frequency between 2 and 40 Hz in steps of 10ms starting 1 second before the onset of the video-episodes and ending 7 seconds after the offset of the video-episodes. Following that, the trials were shuffled and grouped randomly into 4 sets of mixed-content-trials. Sets were of equal size to the 4 sets of same-content-trials. Again ITPC was computed separately for each of the 4 sets. To balance the contribution of the 4 sets, a Rayleigh Z-correction was applied with $N*ITPC^2$, where N refers to the number of trials in a set. Finally the corrected

ITPC was averaged across the 4 sets in the ordered and in the shuffled condition. Their difference indicated content specificity of phase which could be statistically tested (Busch, Dubois, & VanRullen, 2009; Ng et al., 2013). The analysis in source-space was done in the same way using the virtual sensors; however the frequency was restricted to 8 Hz.

Content specific phase similarity between encoding and retrieval

The reactivation of temporal patterns was estimated on virtual sensors for the frequency of 8 Hz. To this end, the oscillatory phase coherence between encoding and retrieval was contrasted between trial-combinations of same content (e.g. watching video-episode A, recalling video-episode A) and random trial-combinations of different content (e.g. watching video-episode A, recalling video-episode B). The combinations were balanced, such that in both conditions (same vs. different combinations) exactly the same trials were used in the same amount of combinations. We only changed the pairing between encoding and retrieval trials. For each trial-combination, 1-second long windows from the encoding trial were now compared to every time point at retrieval starting at the onset of the word-cue and ending at its offset after 3.5 seconds. This comparison was done with a sliding window approach. As a metric of phase-similarity, the phase coherence across time (Lachaux et al., 2000; Michelmann et al., 2016; Mormann et al., 2000) (i.e. across the 1 second window) was computed. All possible windows from encoding were used in this sliding window approach, with the first window ranging from 0 to 1 seconds and the last window ranging from 5 to 6 seconds during the video-episode (compare figure 303 d). Note that the response scale set on between 250ms and 750ms after the word-offset, additionally the first response-scale did not contain content-information (only the numbers 1, 2, and 3) and all responses required a

button-press on the left button. Therefore no confounds from the response interval were expected to bleed into the tested interval. Oscillatory phase was estimated by multiplying the Fourier-transformed data with a complex Morlet wavelet of six cycles in steps of 15.6ms consistent with our previous analyses (Michelmann et al., 2016). The average similarity between all time-windows and combinations was subsequently averaged to derive a single value of similarity for combinations of same content and a single value for combinations of different content at each virtual sensor. Note that this method (Michelmann et al., 2016) enables the investigation of highly dynamic patterns in a robust way, because a measure that captures dynamic changes in ongoing oscillations is accumulated across encoding time, retrieval time and ten thousands of trial-combinations.

Time courses of Replay

To observe the temporal scale of reactivation, the distribution of similarity to the remembered stimulus content (i.e. phase coherence) across retrieval was compared between different sliding windows from encoding. By definition a distribution is normalized to an area under curve of 1 and therefore accounts for differences in total similarity between windows. To robustly compare the distribution of similarity between 6 non-overlapping windows, phase-coherence was cumulated across time, such that at the beginning of the retrieval time zero similarity to all windows was present and at the end of retrieval (i.e. at 3.5 seconds after word onset) 100 % of similarity was reached. This made it possible to compare at each time point, whether the similarity to a window had come up earlier than to another window. In other words: If patterns from window “A” tend to appear earlier than patterns from window

“B” across subjects, then the cumulated similarity to window A should be statistically higher than the cumulated similarity to window “B”, at several time points.

In order to test for a general tendency for forward replay, a line was fitted across all 6 windows and tested against a slope of 0. Hence a negative slope of this line means that earlier windows from encoding appear earlier during retrieval. In order to test the hypothesis that the replay of individual scenes takes places on a slower timescale, 3 lines were fitted across the 2 non-overlapping windows within each scene, and their slope was averaged. If the average slope of these 3 lines is more negative than the slope of the line across all windows, then replaying individual scenes takes place on a slower temporal scale.

Importantly this way of cumulating the similarity distributions allows for robust testing across subjects, at the expense of introducing temporal dependencies between time points. Specifically, if more similarity to a window is present at an early point this can propagate to later points, if similarity thereafter increases at the same speed for all windows. In another scenario, similarity to a window could only appear late during retrieval. This means that other windows would lead during the whole retrieval interval, which is correct, however one should not interpret the fact that in this scenario some windows are already leading over others at very early retrieval. The extent of significant time intervals should therefore be interpreted with caution. Another disadvantage of this method is that the slope is interval scaled and its absolute value is not interpretable.

In order to quantify the actual lag between time windows from encoding descriptively, the distributions of similarity were averaged across subjects and smoothed with a moving average kernel of 250ms, to attenuate noise. The cross-correlation between distributions was then computed to estimate the lag between them: The shape of one similarity distribution is

matched to another. This was done within the time interval in which the slowing down of replay was observed; specifically in which the slope for lines fitted within a scene was significantly more negative than the slope across all windows (i.e. between 550ms and 2350ms at retrieval).

Statistical analyses

Behavioural performance and Reaction times

Behavioural performance was tested with a repeated-measures-ANOVA, on the percent of correct responses. Post-hoc tests were then performed with 2 separate ANOVAs for the final behavioural experiment and with a series of one-sample t-test (see Supplemental Information).

RTs in the balancing pilots were first contrasted with one-sample t-tests. In order to statistically test the null hypothesis the Scaled JZS Bayes Factor (Rouder, Speckman, Sun, Morey, & Iverson, 2009) to the one-sample t-tests was computed. RTs in behavioural pilot experiment were compared with a repeated-measures-ANOVA with the factor position (1, 2 and 3). In the final behavioural experiment, a 2x3-repeated-measures-ANOVA was computed with the factors retrieval task (cued-recall vs. associative recognition) and position (1, 2, and 3). Post-hoc tests were then performed with 2 separate ANNOVAs. Reaction times for the 3 different positions were subsequently compared with a series of post-hoc one-sample t-tests. Greenhouse-Geisser correction was used with all ANOVAs, null-effects of interest were tested with Bayesian t-tests (Rouder et al., 2009).

Content specific oscillatory phase at encoding

Content specific phase at encoding was statistically tested by contrasting average ITPC across arranged groups with the average ITPC across shuffled groups. This was done with a series of t-test at every time point between 0 and 6 seconds after onset of the video-episode, at every frequency between 2 and 40 Hz and at every sensor. Multiple comparison correction was done via Monte-Carlo permutation of contrast labels as implemented in the fieldtrip toolbox (Maris & Oostenveld, 2007b; Oostenveld et al., 2011). 3-dimensional clusters and cluster-sums were formed across time, frequency and sensors. The cluster-forming threshold corresponded to the critical t-value (alpha < 0.05) of a single-sided one-sample t-test, 1000 random permutations were drawn. On the source level content specific phase was assessed for the frequency of 8Hz. Again the ITPC of arranged groups and the ITPC of shuffled groups were contrasted with a one sample t-test that was computed at every time point and every virtual sensor. Clusters were summed across neighbouring sensors and time points in 1000 random permutations. To obtain time courses within the parieto-occipital ROI, t-values were averaged across all virtual sensors within the ROI.

Content specific phase similarity between encoding and retrieval

Based on previous results (Michelmann et al., 2016), statistical testing for content specific reactivation was done for the frequency of 8 Hz, restricted to an occipito-parietal region of interest (ROI) derived from the AAL atlas (Tzourio-Mazoyer et al., 2002). Averaged similarity values of encoding-retrieval combinations were contrasted between combinations of same content and combinations of different content. This was done with a one-sample t-test on every virtual sensor within the ROI. Subsequently t-values were thresholded with a t-value corresponding to a one-sided alpha value of 0.05; clusters were built across neighbouring

virtual sensors. Statistical testing was done again via 1000 random permutations. A series of post-hoc t-tests was done on every time-point at retrieval in order to estimate the contribution to the effect from encoding windows (see supplemental figure 306a)

Time courses were obtained by averaging across the ROI, which allows for an unbiased investigation of the time-courses of reactivation (see figure 305a). Importantly, the cluster correction approach results in a biased noise-distribution within the cluster of significant reactivation. This renders the interpretation of its shape and any post-hoc analysis on sensors within the cluster problematic (Maris & Oostenveld, 2007b), see also (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). Since 86.46% of the t-values in the ROI were positive, we therefore decided to average across all virtual sensors within the anatomical ROI for the analyses of all time courses that were statistically tested.

Likewise, similarity densities were computed on the averaged similarity values across all virtual sensors within the ROI. The cumulated similarity density distributions for 6 non-overlapping encoding-windows were obtained for every subject. Consequently at every retrieval time-point a line could be fitted across 6 values for every subject. The slope of that line was subsequently subjected to a t-test against 0 across all subjects. The resulting time-course of t-values across the whole retrieval time was finally subjected to a multiple comparisons correction by controlling the false discovery rate (Benjamini & Hochberg, 1995). The average slope fitted across two windows each (windows within scenes) was statistically tested against the slope across all encoding windows with a series of one-sample t-tests. T-values were obtained again at every time point during retrieval and the false discovery rate was controlled in order to correct for multiple comparisons. To estimate at which time-points reinstatement could be detected best (figure 305a), a series of one-sample t-tests was

computed at every retrieval time point, between encoding-retrieval similarity of same content combinations and encoding-retrieval combinations of different content combinations. Finally, the average similarity to all encoding time points was compared within the ROI, between trials in which an association from the first, second or third scene was recalled (figure 305a b). This was done with a repeated-measures-ANOVA with the factor position and pairwise post-hoc t-tests.

Oscillatory power

Baseline corrected oscillatory power was contrasted on the sensor level with a series of one-sample t-tests. Multiple-comparison correction was realized with a cluster-based Monte-Carlo permutation as implemented in the fieldtrip toolbox (Oostenveld et al., 2011). 1000 permutations of contrast-labels were used; the clusters were formed from neighbouring values below a threshold (see below). Neighbouring values were derived across time from 0 to 4 seconds after the onset of the word-cue, across frequency from 2 to 40 Hz and spatially across sensors The threshold was the t-value which corresponds to a threshold of alpha = 0.05 for a single sided test. The maximal cluster-sum of real data was then compared to the distribution of maximal cluster-sums under random permutations. In order to find the most robust frequencies that showed oscillatory power decreases, a t-test was computed for the average power difference across time (0 – 4s), sensors and frequencies. On the source level, baseline-corrected power at 8 Hz was averaged over time between 0 and 4 seconds and subjected to a one-sample t-test. Multiple comparison correction was addressed with the same cluster-based permutation approach; however, clusters were formed across neighbouring virtual sensors.

Supplemental Information

Behavioural performance

In the behavioural pilot experiment, participants recalled the correct position and video-episode in 72.29% (*SD* = 11.69%) of the trials. A main effect of position indicated decreasing performance when an association had been learned in a later scene of a video-episode (ANOVA: $F_{1.27,\ 13.95}$ = 4.988, p = 0.036, means: 74.86%, 72.29%, 69.72%); post-hoc tests indicated that only associations from the first position were recalled more often than associations from the third position (t_{11} = 6.27, p < 0.001).

In the alternating blocks of the behavioural experiment, participants recalled on average 69.47% (*SD* = 23.21%) of the correct word-scene associations in cued-recall (CR) blocks. They further recognized 90.27% (*SD* = 10.74%) of intact associations (Hits) and erroneously named 12.40% (*SD* = 14.38%) of rearranged associations intact (False Alarms) in an associative-recognition (AR) blocks. Performance in CR (i.e. percent correct responses) and in AR (i.e. percent Hits minus percent False Alarms) was compared with a 2x3 ANOVA. This revealed a significant main effect of condition ($F_{1,\ 23}$ = 38.30, p < 0.001), driven by a better performance in the associative-recognition blocks (t_{23} = 6.189, p < 0.001) and a significant factor position ($F_{1.84,\ 42.24}$ = 1.145, p = 0.002, interaction condition with position *n.s.*). This was driven by a slightly better performance in the cued-recall condition, for associations that were learned in the second position of a video-episode (ANOVA: $F_{1.58,36.24}$ = 2.794, p = 0.086, position 1 vs. 2: t_{23} = -2.804, p = 0.02, position 2 vs. 3: t_{23} = 1.961, p = 0.062) and a worse performance in associative-recognition for associations that were learned in the third position (ANOVA: $F_{1.86,42.68}$ = 5.552, p = 0.008, position 2 vs. 3: t_{23} = 3.879, p < 0.001, all other *ps* > 0.14).

In the MEG experiment subjects remembered on average 63.54% (*SD* = 11.768%) of associations, excluding guesses. After preprocessing on average 200.348 trials (*SD* = 38.645) remained for known correct associations and an additional 116 trials (*SD* = 39.425) were guessed or incorrect responses.

Reaction time in the behavioural experiment (including correct guesses)

The analyses of reaction times were repeated including those trials in which participants indicated that they had guessed the response. The 2x3 ANOVA of RTs revealed a significant main effect of condition ($F_{1.00,\ 23.00}$ = 66.254, p < 0.001, log-RT: $F_{1.00,\ 23.00}$ = 98.52, p < 0.001) driven by overall faster reactions in the associative-recognition condition (t_{23} = -8.14, p < 0.001, log-RT: t_{23} = -9.619, p < 0.001). A significant main effect of scene-position ($F_{1.90,\ 43.67}$ = 5.304, p = 0.010, log-RT: $F_{1.87,\ 43.09}$ = 2.823, p = 0.074) and the interaction of scene-position with retrieval-condition ($F_{1.96,\ 45.11}$ = 5.041, p = 0.011, log-RT: $F_{1.89,\ 43.39}$ = 5.771, p = 0.007) were both due to a strong forward replay effect in the cued-recall condition (ANOVA: $F_{1.80,\ 41.36}$ = 8.796, p = 0.001, log-RT: $F_{1.64,\ 37.82}$ = 8.304, p = 0.002). Specifically, associations that were learned in the first scene-position of a video-episode (mean RT = 2.5 sec) were recalled on average 132ms faster than associations that were learned in the second scene-position (t_{23} = -1.752, p = 0.047, log-RT: t_{23} = -2.127, p = 0.022). Associations that were learned in the second scene-position (mean RT = 2.617 sec) were recalled on average 170ms faster than associations that were learned in the third scene-position (t_{23} = -2.864, p = 0.004, log-RT: t_{23} = -2.539, p = 0.009).

In the AR condition, subjects performed the exact same encoding task, which also required source-memory. Importantly, no differences in reaction times were evident between

associations that were learned in the first, second or third position during encoding (ANOVA: $F_{1.44,\ 33.09}$ = 0.185, p = 0.759, log-RT: $F_{1.52,\ 35.05}$ = 0.591, p = 0.515, pairwise comparisons of positions: all ps > 0.5, Bayes-Factor supporting the null Hypothesis: position 1 vs. 2, BF_{01} = 3.771, position 2 vs. 3, BF_{01} = 4.466, position 1 vs. 3, BF_{01} = 4.504, log-RT: all ps > 0.39, position 1 vs. 2, BF_{01} = 3.688, position 2 vs. 3, BF_{01} = 3.317, position 1 vs. 3, BF_{01} = 4.048).

Reaction times in the behavioural pilot

In the behavioural pilot experiment, participants associated word-cues with one of three scenes within video-episodes (figure 301a). Four continuous video-episodes each comprised of three individual scenes. A trial unique word-cue appeared in one scene during a video-episode. After a brief distractor task (figure 301b) a cued retrieval task (figure 301d, top-left) was conducted where participants were presented with the word cues. Their task was to recall the scene-position that was associated with the word-cue as quickly as possible. After that, participants indicated which video-episode out of four was associated with the word.

Faster reaction times to associations that were associated with early position compared to later positions were observed (ANOVA: $F_{1.40,\ 15.41}$ = 4.257, p = 0.045, ANOVA of log-transformed RTs: $F_{1.58,\ 17.38}$ = 4.903, p = 0.027). On average, reaction times (RT) to first scene-positions were faster than RTs to second scene-positions (2.044 vs. 2.212 sec., t_{11} = -3.558, p = 0.005, log-RT: t_{11} = -3.626, p = 0.004), and trended to be faster than for third scene-positions (2.221 sec, t_{11} = -2.05, p = 0.065; log-RT: t_{11} = -2.227, p = 0.048). RTs for second scene-positions were only numerically, but not significantly, faster compared to third scene-positions. These results suggest that memory replay is forward and compressed. During encoding, individual scenes of each video-episode lasted 2 seconds. During retrieval,

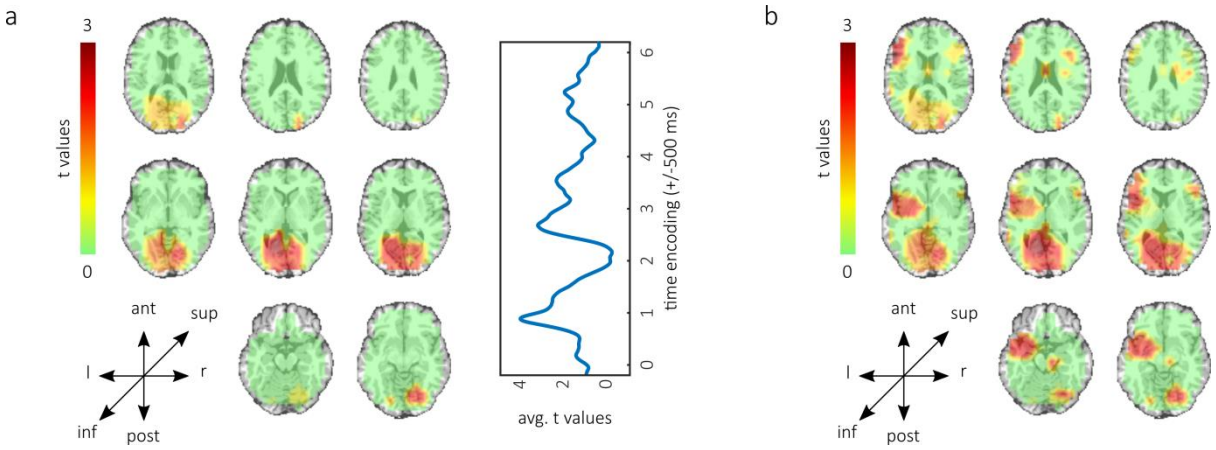
however, subjects took on average 167.8ms longer to recall an association from the second scene-position and an additional 10ms longer to recall an association from the third scene-position. Importantly, these effects cannot be explained by material specific differences between positions, because pilot experiments ensured that there were no differences in RTs when the scenes from position 1, 2 and 3 were associated with a word-cue in isolation (see Online Methods).

Reinstatement of encoding patterns and further evidence for forward replay

Overall we found a cluster of significant evidence for the reactivation of phase patterns from encoding during retrieval for hit trials (Hits; $p_{cluster}$ = 0.034; supplemental figure 306a, supplemental figure 306b for unmasked maps of t-values). In this, we wanted to assess how much each sub-part from the video-episodes contributed to this effect. To this end, we computed a series of post-hoc t-tests for every encoding time-window. We obtained the highest t-values for the reinstatement of earlier time-windows during encoding (Supplemental figure 306a, right).

If participants start to replay from the beginning of a video-episode and typically progress until they have the correct word-scene association in memory, then early time windows from encoding should be reactivated more often and more thoroughly (see above), therefore this pattern is consistent with forward replay.

Supplemental figure 306: Content specific pattern reinstatement



Legend to supplemental figure 306: Content specific pattern reinstatement

a) *Cluster of significant reactivation of phase-patterns from encoding for successfully remembered associations (left) and contribution to effect (right). Early encoding windows express the highest t-values and contribute more to the effect than later ones.* ***(b)*** *Unmasked map of t-values for reactivation of phase-patterns from encoding in successfully remembered trials.*

Chapter 5 – General Discussion

The presented studies have consistently demonstrated the tracking of content specific patterns in brain oscillations. Reinstatement of oscillatory patterns was localized in sensory specific cortices; mechanistically, it was linked to memory related decreases in oscillatory power at a centre frequency of 8Hz. These findings were possible with a new methodological approach: Representational Similarity Analysis (RSA) is combined with measures of phase coherence to track content specific oscillatory patterns. These patterns were detected during episodic memory formation and retrieval. Eventually, these methods were extended and combined with a new paradigm: Memory replay was elicited in a natural way that allowed for the tracking of sub-events within continuous episodes. Behaviourally, reaction times showed that memory replay is forward and faster than the original perception (i.e. memory replay was compressed). On a neural level, the distribution of phase similarity to different sub-events showed that memory replay is forward and that the speed of memory replay is flexible. Specifically, participants replay fragments of activity at the same speed and can skip between replayed elements. This skipping was faster between distinct sub-events than within these events.

Methodological advances

The method that was introduced in this thesis tracks oscillatory patterns with a combination of phase coherence and RSA. It is unique in leveraging fine-grained temporal patterns of activity that are confined to an oscillation. Primarily, it is therefore useful to test hypotheses regarding the role of specific oscillations for distinct mental states. In theory, all measures that are used to assess connectivity between channels are suited as a similarity metric for RSA. They can be chosen based on considerations that typically go into the choice of a

connectivity metric (Greenblatt, Pflieger, & Ossadtchi, 2012). Their properties will determine the interpretation of results. A very similar method, for instance, is the correlation of power spectra in combination with RSA (Staresina et al., 2016). This method is rather directed at overall spectral changes, whereas phase coherence may be the preferred similarity metric, if fine-grained temporal patterns are tracked.

A crucial strength of the presented method is the ability to track oscillations when the onset of their reappearance is unknown. In the described studies, the retrieval of an associate in a cued-recall paradigm happened upon presentation of a word-cue. The exact time point of memory retrieval however was variable. For this reason, a metric of phase coherence was chosen that assesses connectivity over a time window in a way that is robust to time shifts (Lachaux et al., 2000; Mormann et al., 2000). This consideration of onset is critical, when similarity between oscillating time series is assessed. The cross-correlation between two oscillations, for instance, will itself be an oscillation. Across time, it can therefore average to zero, even when patterns are similar. On the other hand, if two patterns are thought to occur at the same time, metrics that factor in time shifts, like pairwise phase consistency (PPC) (Vinck et al., 2010), may arguably be more sensitive.

Several other methods have recently been introduced to track content specific patterns in electrophysiology. An important distinction therein, is which patterns of activity are leveraged to distinguish mental states. Temporal patterns (Michelmann et al., 2016; Staresina et al., 2016; Staudigl et al., 2015), spatial patterns (Fuentemilla et al., 2010; Jafarpour et al., 2014; King & Dehaene, 2014; Wolff et al., 2017) or spatiotemporal patterns (Lu, Wang, Chen, & Xue, 2015; Sols et al., 2017) of activity can be differentiated. The pattern that is used for analysis will in the end determine the interpretation of significant findings: If content specific spatial

patterns are tracked, the conclusions derived from differentiable states is that distinct topographies underlie these mental states at a given time point (Stokes, Wolff, & Spaak, 2015). Crucially, the method presented in this thesis, specifically leverages information that can only be measured with electrophysiological methods, namely fine-grained temporal patterns. For this reason, it can provide qualitatively new information, i.e. more than just a time-resolved extension of spatial similarity.

Importantly, the way in which content can be differentiated based on neural activity will partly depend on the stimulus material. The Parahippocampal Place Area (PPA) (Epstein & Kanwisher, 1998) and the Fusiform Face Area (Kanwisher, McDermott, & Chun, 1997) have a functional preference to process places and faces. It is therefore likely that a condition that requires the processing of places and a condition that requires the processing of faces are marked by distinct topographies. The distinction of temporal patterns on the other hand implies that at a given location (i.e. sensor or electrode), unique time courses of activity are present. Arguably, temporal similarity measures will therefore have a higher sensitivity to distinguish dynamic stimuli. Importantly, temporal and spatial similarity measures will partially capture the same signal (Staudigl et al., 2015), i.e. spatial and temporal patterns in the EEG and MEG are entangled (Cohen, 2011). In general, it is important to consider stimulus material and the method to detect content specific patterns carefully, based on the hypothesis that is being investigated.

Support for the information via desynchronization hypothesis

The most robust finding throughout this thesis is support for the information via desynchronization hypothesis (Hanslmayr et al., 2012): A relation between decreases in oscillatory power and the representation of information was consistently demonstrated.

Successful memory of information-rich naturalistic stimulus material elicited particularly strong desynchronization in the presented studies. Crucially, these power decreases were linked to the reappearance of content specific oscillations that were localised in sensory specific cortices. In this thesis, the first demonstration of this principle in chapter 2 relied on strong memory traces. These were achieved via explicit instructions to replay content vividly. Furthermore the analysis was confined to associations that were remembered with high confidence. Eventually, in chapter 4 a new paradigm elicited memory reinstatement in a natural way; a homogeneous memory performance was achieved with a design that adapted to individual performance. Consequently, it was possible to contrast the pattern reinstatement of successfully remembered and forgotten associations and demonstrate the relevance of content specific oscillations for successful memory. Overall, content specific oscillations were linked to power decreases in two studies that used different recording methods, i.e. EEG and MEG. Additionally, power decreases in a visual and auditory modality contained content specific oscillations during two processes, namely episodic memory formation and retrieval.

From the data presented in this thesis, it can therefore be concluded that decreases in oscillatory power are not just a by-product of memory but are rather involved in the reinstatement of content representations from memory.

In order to specifically test the hypothesis that desynchronizing frequencies contain information, the analysis was confined to oscillatory patterns. Clearly, however, oscillations reappearing in sensory cortices are not the only patterns that are reinstated in episodic memory. Information about reinstated content can be detected with other measures (Kurth-Nelson et al., 2015; Lu et al., 2015; Staresina et al., 2016; Yaffe et al., 2014, 2017; Zhang et al.,

2015) and in other regions (Ritchey et al., 2012; Staresina et al., 2012). An important open question is consequently, how the cortical reinstatement of oscillatory patterns, that is linked to power decreases, integrates with other patterns that are reinstated. Specifically, the overall mechanism that functionally enables the brain to learn and retrieve episodic memories may represent information in different ways and at different levels of abstraction.

The Hippocampus, for instance, has been proposed to code information sparsely (Lisman & Jensen, 2013; McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003). It is therefore possible that representations are spectrally and functionally transformed in this region, but they still express similarities of recorded brain activity between encoding and retrieval (Staresina et al., 2016, 2012). In line with this, a recent extension of the information via desynchronization hypothesis (Hanslmayr et al., 2016) states that information rich content is coded in desynchronized patterns in the alpha beta frequency band; those desynchronized representations rely on the Neocortex whereas the Hippocampus provides a sparse code that relies on synchronized activity in the theta and gamma band (Lisman & Jensen, 2013), and serves to bind representations together.

A centre frequency of 8 Hz

Interestingly, the frequency band that was consistently linked to reinstatement of content specific phase patterns was centred at 8 Hz. Consistent with the information via desynchronization hypothesis, the 8 Hz frequency band displayed strong power decreases throughout different studies and contrasts. Studies that find memory related power decreases however often report stronger effects in the upper alpha and beta band (Hanslmayr & Staudigl, 2014; Hanslmayr et al., 2012), which is at odds with the finding of 8 Hz

in this thesis. A prime candidate to explain this difference to other memory experiments is the use of information-rich dynamic stimuli in this thesis.

While the 7-8 Hz frequency band has been previously implicated in the rhythmical sampling of continuous perception (Hanslmayr et al., 2013; Landau & Fries, 2012), we implicated this band for the first time in the reinstatement and maintenance of episodic memories of naturalistic stimuli (Staudigl & Hanslmayr, 2018). An 8Hz oscillation could therefore be required to temporally structure and coordinate information-rich continuous reinstatement. Alternatively, the reinstatement of oscillatory patterns could extend to other frequencies but was simply not detected in the presented studies: Even though control analyses were performed in other frequency bands, it is still possible that factors like the signal to noise ratio moderate the sensitivity of the presented method. Currently, however, it remains an open question, why specifically the 8 Hz frequency band was consistently implicated in the representation of information.

Flexible forward replay

Finally, it was possible to leverage similarity in oscillatory patterns to track the reinstatement of distinct sub-events, when continuous representations were replayed from memory. This replay was forward in behavioural data; in the neural data the statistical relation between phase-patterns during memory replay supported a forward direction at a compressed level. These different speeds that were observed suggest a flexible mechanism that allows us to skip between accurately represented fragments of activity in a forward direction.

Interestingly, data from animal studies suggests that even the direction of neural replay is not fixed but rather task dependent (Foster & Wilson, 2006; A. Johnson & Redish, 2007).

Speculatively, flexibility in replay could therefore also change the direction of replay, depending on the task demands.

Mechanistically, previous studies and theories have implicated the Hippocampus as a structure that is in control of timing in memory (Davachi & DuBrow, 2015; Heusser, Poeppel, Ezzyat, & Davachi, 2016; Lisman & Jensen, 2013; Sols et al., 2017), yet, the observed flexible patterns were located in sensory cortices. A possible explanation for these discrepancies is that the hippocampus initiates reinstatement in sensory areas, when the vivid reinstatement of sensory information is required. This explanation is well in line with a framework, in which the Neocortex represents information-rich content that is bound together in the Hippocampus (Hanslmayr et al., 2016). An important future question is therefore, how the Hippocampus and the Neocortex interact during episodic memory reinstatement, specifically whether the Hippocampus exerts flexible control over sensory regions, when the vivid reinstatement of patterns is required.

Possible clinical relevance

An important open question is how these findings integrate in a clinical context, specifically how patients might benefit from this research in the future. Several parallels can be observed between these studies and the memory related post-traumatic stress disorder (PTSD). PTSD patients often report vivid unwanted replay of previous experiences that resemble their original experience, i.e. intrusions (Friedman, Resick, Bryant, & Brewin, 2011). The naturalistic stimulus material that was used in this thesis and the vivid reinstatement thereof can be considered an approximation of such vivid intrusions in an experimental setting.

Interestingly, the finding of replay in the cortex fits nicely into the framework of a dominant theory that addresses intrusions in PTDS patients: The dual representations theory of PTSD

(Brewin, Gregory, Lipton, & Burgess, 2010) proposes that a verbally accessible memory system relies on the hippocampus and supports the controlled access to memory in the context of narratives, whereas the situationally accessible memory system (SAM) relies on sensory cortical areas. Intrusions are suggested to arise within the SAM. This mirrors the presented finding of temporally precise pattern replay in sensory specific cortices.

Another interesting link regards eye movement desensitization and reprocessing (EMDR) (Shapiro, 2001). EMDR is an effective treatment approach for PTSD and usually relies on eye movements that are carried out while the patient recalls traumatic memories. Eye movements are known to elicit saccade-evoked potentials (SEPs) (Burdette, Walrath, Gross, James, & Stern, 1986). Speculatively, these SEP could therefore interfere with traumatic intrusions: A mechanistic hypothesis would be that SEPs reset the phase of ongoing oscillations in visual regions which interferes with the vivid representation of traumatic memories.

Conclusion

Decreases in low frequency (<30 Hz) power are a well-known correlate of successful memory encoding and retrieval (Hanslmayr et al., 2012; Long et al., 2014; Zion-Golumbic et al., 2010). This thesis demonstrates their role in the representation of information: A new method made it possible to show that memory related power decreases harbour information about the content of memory during episodic memory encoding and retrieval. This information in oscillations, marked by power decreases, was finally tracked in memory. Replay of sub-events in continuous episodes was forward and its speed was faster in memory than during perception. This result was further supported by behavioural data. Finally, the tracking of replayed patterns in memory demonstrated that subjects can skip flexibly between sub-

events of continuous episodes and slow down when they replay uninterrupted segments. This renders episodic memory as a dynamic process in which power decreases harbour content specific patterns that are reinstated in a flexible way.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431.

Arnold, A. E. G. F., Iaria, G., & Ekstrom, A. D. (2016). Mental simulation of routes during navigation involves adaptive temporal compression. *Cognition*, *157*, 14–23.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839.

Belluscio, M. A., Mizuseki, K., Schmidt, R., Kempter, R., & Buzsáki, G. (2012). Cross-frequency phase-phase coupling between θ and γ oscillations in the hippocampus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(2), 423–435.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statisitical Society, Series B*, *57*(1), 289–300.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ : British Medical Journal*, *310*(6973), 170.

Bonasia, K., Blommesteyn, J., & Moscovitch, M. (2016). Memory and navigation: Compression of space varies with route length and turns. *Hippocampus*, *26*(1), 9–12.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Brewin, C. R., Gregory, J. D., Lipton, M., & Burgess, N. (2010). Intrusive Images in Psychological Disorders: Characteristics, Neural Mechanisms, and Treatment Implications. *Psychological Review*, *117*(1), 210–232.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Burdette, L. J., Walrath, L. C., Gross, J., James, B., & Stern, J. A. (1986). A comparison of saccade evoked potentials recorded during reading and tracking tasks. *Physiology and Behavior*, *37*(4), 527–532.

Burgess, A. P., & Gruzelier, J. H. (2000). Short duration power changes in the EEG during recognition memory for words and faces. *Psychophysiology*, *37*(5), 596–606.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The Phase of Ongoing EEG Oscillations Predicts Visual Perception. *Journal of Neuroscience*, *29*(24), 7869–7876.

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Canavier, C. C. (2015). Phase-resetting as a tool of information transmission. *Current Opinion in Neurobiology*, *31*, 206–213.

Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. In *Nature Neuroscience* (Vol. 14, pp. 147–153).

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

Cohen, M. X. (2011). It's about Time. *Frontiers in Human Neuroscience*, *5*(January), 2.

Collin, S. H. P., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, *18*(11), 1562–1564.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505.

Crowe, D. A., Averbeck, B. B., & Chafee, M. V. (2010). Rapid Sequences of Population Activity Patterns Dynamically Encode Task-Critical Spatial Information in Parietal Cortex. *Journal of Neuroscience*, *30*(35), 11640–11653.

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: the role of prediction and context. *Trends in Cognitive Sciences*, *19*(2), 92–99.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, *8*, 15276.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(6), 701–711.

Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*(7084), 680–683.

Friedman, M. J., Resick, P. A., Bryant, R. A., & Brewin, C. R. (2011). Considering PTSD for DSM-5. *Depression and Anxiety*.

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Düzel, E. (2010). Theta-Coupled Periodic Replay in Working Memory. *Current Biology*, *20*(7), 606–612.

Goff, G. D., Matsumiya, Y., Allison, T., & Goff, W. R. (1977). The scalp topography of human somatosensory and auditory evoked potentials. *Electroencephalography and Clinical Neurophysiology*, *42*(1), 57–76.

Goldman-Rakic, P. . (1995). Cellular basis of working memory. *Neuron*, *14*(3), 477–485.

Greenblatt, R. E., Pflieger, M. E., & Ossadtchi, A. E. (2012). Connectivity measures applied to human brain electrophysiological data. *Journal of Neuroscience Methods*, *207*(1), 1–16.

Groussard, M., Viader, F., Hubert, V., Landeau, B., Abbas, A., Desgranges, B., … Platel, H. (2010). Musical and verbal semantic memory: Two distinct neural networks? *NeuroImage*, *49*(3), 2764–2773.

Haegens, S., Nacher, V., Luna, R., Romo, R., & Jensen, O. (2011). -Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, *108*(48), 19377–19382.

Hanslmayr, S., Spitzer, B., & Bäuml, K. H. (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral Cortex*, *19*(7), 1631–1640.

Hanslmayr, S., Staresina, B. P., & Bowman, H. (2016). Oscillations and Episodic Memory: Addressing the Synchronization/Desynchronization Conundrum. *Trends in Neurosciences*.

Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories — A processing based perspective on oscillatory subsequent memory effects. *NeuroImage*, *85*, 648–655.

Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience*, *6*(April), 1–12.

Hanslmayr, S., Volberg, G., Wimber, M., Dalal, S. S., & Greenlee, M. W. (2013). Prestimulus oscillatory phase at 7 Hz gates cortical information flow and visual perception. *Current Biology*, *23*(22), 2273–2278.

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, *12*(9), 509–523.

Hasselmo, M. E. (2015). Current questions on space and time encoding. *Hippocampus*, *752*, 744–752.

Heusser, A. C., Poeppel, D., Ezzyat, Y., & Davachi, L. (2016). Episodic sequence memory is supported by a theta-gamma phase code. *Nature Neuroscience*, *19*(10), 1374–1380.

Huang, Y., Dmochowski, J. P., Su, Y., Datta, A., Rorden, C., & Parra, L. C. (2013). Automated MRI segmentation for individualized modeling of current flow in the human head. *Journal of Neural Engineering*, *10*, 66004.

Jafarpour, A., Fuentemilla, L., Horner, A. J., Penny, W. D., & Duzel, E. (2014). Replay of very early encoding representations during recollection. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(1), 242–248.

Jafarpour, A., Penny, W. D., Barnes, G. R., Knight, R. T., & Duzel, E. (2017). Working Memory Replay Prioritizes Weakly Attended Events. *Eneuro*, *4*(4), ENEURO.0171-17.2017.

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, *16*(4), 200–206.

Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, *4*(November), 1–8.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, *10*(1), 100–107.

Johnson, A., & Redish, A. D. (2007). Neural Ensembles in CA3 Transiently Encode Paths Forward of the Animal at a Decision Point. *Journal of Neuroscience*, *27*(45), 12176–12189.

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.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

Ketz, N., Morkonda, S. G., & O'Reilly, R. C. (2013). Theta Coordinated Error-Driven Learning in the Hippocampus. *PLoS Computational Biology*, *9*(6), e1003067.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.

Klimesch, W. (1997). EEG-alpha rhythms and memory processes. *International Journal of Psychophysiology*, *26*(1–3), 319–340.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88.

Klimesch, W., Schimke, H., Doppelmayr, M., Ripper, B., Schwaiger, J., & Pfurtscheller, G. (1996). Event-related desynchronization (ERD) and the Dm effect: Does alpha desynchronization during encoding predict later recall performance? *International Journal of Psychophysiology*, *24*(1–2), 47–60.

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 1–28.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.

Kurth-Nelson, Z., Barnes, G. R., Sejdinovic, D., Dolan, R., & Dayan, P. (2015). Temporal structure in associative retrieval. *eLife*, *4*, e04919.

Lachaux, J. P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J., & Varela, F. J. (2000). Studying single-trials of phase-synchronous activity in the brain. *International Journal of Bifurcation and Chaos*, *10*(10), 2429–2439.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*, 194–208.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*(11), 1000–1004.

Lisman, J. E., & Jensen, O. (2013). The Theta-Gamma Neural Code. *Neuron*, *77*(6), 1002–1016.

Long, N. M., Burke, J. F., & Kahana, M. J. (2014). Subsequent memory effect in intracranial and scalp EEG. *NeuroImage*, *84*, 488–494.

Lu, Y., Wang, C., Chen, C., & Xue, G. (2015). Spatiotemporal neural pattern similarity supports episodic memory. *Current Biology*, *25*(6), 780–785.

Maris, E., & Oostenveld, R. (2007a). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Maris, E., & Oostenveld, R. (2007b). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Marshall, T. R., Bergmann, T. O., & Jensen, O. (2015). Frontoparietal Structural Connectivity Mediates the Top-Down Control of Neuronal Synchronization Associated with Selective Attention. *PLoS Biology*, *13*(10), 1–17.

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 connessionist models of learning and memory. *Psychological Rewiew*, *102*(3), 419–457.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *Journal of Neurophysiology*, *100*(3), 1407–1419.

Michelmann, S. (2016). Data from: The temporal signature of memories – Identification of a general mechanism for dynamic memory replay in humans.

Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. *PLoS Biology*, *14*(8).

Michelmann, S., Bowman, H., & Hanslmayr, S. (2017). Replay of Stimulus Specific Temporal Patterns during Associative Memory Formation. *bioRxiv*.

Mormann, F., Lehnertz, K., David, P., & E. Elger, C. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D: Nonlinear Phenomena*, *144*(3–4), 358–369.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., & Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *19*(21), 9497–9507.

Ng, B. S. W., Logothetis, N. K., & Kayser, C. (2013). EEG Phase Patterns Reflect the Selectivity of Neural Firing. *Cerebral Cortex*, *23*(2), 389–398.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646.

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences*, *97*, 11120–11124.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9.

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. *Nature Neuroscience*, *9*(4), 489–491.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102.

Pletzer, B., Kerschbaum, H., & Klimesch, W. (2010). When frequencies never synchronize: The golden mean and the resting EEG. *Brain Research*, *1335*, 91–102.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133–140.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*(3), 446–461.

Ritchey, M., Wing, E. a., LaBar, K. S., & Cabeza, R. (2012). Neural Similarity Between Encoding and Retrieval is Related to Memory Via Hippocampal Interactions. *Cerebral Cortex*, *23*(12), 2818–2828.

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136–1139.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*(2), 255–260.

Schyns, P. G., Thut, G., & Gross, J. (2011). Cracking the Code of Oscillatory Activity. *PLoS Biology*, *9*(5), e1001064.

Shannon, C. E. C., & Weaver, W. (1949). *The mathematical theory of information*. *Urbana University of Illinois Press* (Vol. 97).

Shapiro, F. (2001). *Eye movement desensitization and reprocessing: basic principles, protocols, and procedures (2nd ed.)*. *New York: Guilford Press, 2001. xxiv, 472 pp.*

Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science (New York, N.Y.)*, *271*(5257), 1870–1873.

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Current Biology*, *27*(22), 3499–3504.e4.

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic reinstatement in the medial temporal lobe. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(50), 18150–18156.

Staresina, B. P., Michelmann, S., Bonnefond, M., Jensen, O., Axmacher, N., & Fell, J. (2016). Hippocampal pattern completion is linked to gamma power increases and alpha power decreases during recollection. *eLife*, *5*(AUGUST).

Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology*, *23*(12), 1101–1106.

Staudigl, T., & Hanslmayr, S. (2018). Reactivation of neural patterns during memory reinstatement supports encoding specificity. *bioRxiv*, 255166.

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. *Journal of Neuroscience*, *35*(13), 5373–5384.

Stokes, M. G. (2015). "Activity-silent" working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405.

Stokes, M. G., Wolff, M. J., & Spaak, E. (2015). Decoding Rich Spatial Information with High Temporal Resolution. *Trends in Cognitive Sciences*, *19*(11), 636–638.

Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *NeuroImage*, *68*, 39–48.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, *2011*, 1–13.

Tal, I., & Abeles, M. (2013). Cleaning MEG artifacts using external cues. *Journal of Neuroscience Methods*, *217*(1–2), 31–38.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *16*(13), 4240–4249.

Tervaniemi, M., Medvedev, S. V, Alho, K., Pakhomov, S. V, Roudas, M. S., Van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Human Brain Mapping*, *10*(2), 74–79.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization memory* (pp. 381–402). New York: Academic Press.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science*, *2*(3), 67–70.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289.

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 Transactions on Biomedical Engineering*, *44*(9), 867–880.

VanRullen, R., Busch, N. A., Drewes, J., & Dubois, J. (2011). Ongoing EEG Phase as a Trial-by-Trial Predictor of Perceptual and Attentional Variability. *Frontiers in Psychology*, *2*(April), 1–9.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19204–19209.

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., & Pennartz, C. M. a. (2010). The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. *NeuroImage*, *51*, 112–122.

Wagner, A. D., Koutstaal, W., & Schacter, D. L. (1999). When encoding yields remembering: Insights from event-related neuroimaging. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *354*(1387), 1307–1324.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453.

Waldhauser, G. T., Braun, V., & Hanslmayr, S. (2016). Episodic Memory Retrieval Functionally Relies on Very Rapid Reactivation of Sensory Information. *The Journal of Neuroscience*, *36*(1), 251–260.

Watrous, A. J., Deuker, L., Fell, J., & Axmacher, N. (2015). Phase-amplitude coupling supports phase coding in human ECoG. *eLife*, *4*, e07886.

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*(4), 582–589.

Wimber, M., Maaß, A., Staudigl, T., Richardson-Klavehn, A., & Hanslmayr, S. (2012). Rapid memory reactivation revealed by oscillatory entrainment. *Current Biology : CB*, *22*(16), 1482–1486.

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871.

Yaffe, R. B., Kerr, M. S. D., Damera, S., Sarma, S. V, Inati, S. K., & Zaghloul, K. A. (2014). Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proceedings of the National Academy of Sciences*, *111*(52), 18727–18732.

Yaffe, R. B., Shaikhouni, A., Arai, J., Inati, S. K., & Zaghloul, K. A. (2017). Cued Memory Retrieval Exhibits Reinstatement of High Gamma Power on a Faster Timescale in the Left Temporal Lobe and Prefrontal Cortex. *The Journal of Neuroscience*, *37*(17), 4472–4480.

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research* 1. *Journal of Memory and Language*, *46*(3), 441–517.

Zhang, H., Fell, J., Staresina, B. P., Weber, B., Elger, C. E., & Axmacher, N. (2015). Gamma Power Reductions Accompany Stimulus-Specific Representations of Dynamic Events. *Current Biology*, *25*(5), 635–640.

Zion-Golumbic, E., Kutas, M., & Bentin, S. (2010). Neural Dynamics Associated with Semantic and Episodic Memory for Faces: Evidence from Multiple Frequency Bands. *Journal of Cognitive Neuroscience*, *22*(2), 263–277.

Zumer, J. M., Scheeringa, R., Schoffelen, J.-M., Norris, D. G., & Jensen, O. (2014). Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. *PLoS Biology*, *12*(10), e1001965.