RETRIEVAL DYNAMICS IN EPISODIC MEMORY – FROM COMPUTATIONS TO REPRESENTATIONS

By

JUAN LINDE DOMINGO

A thesis submitted to the

University of Birmingham

for the degree of DOCTOR OF PHILOSOPHY

School of Psychology
University of Birmingham
September 2018

Abstract

Understanding how our experiences are retrieved from long-term memory is fundamental in cognitive neuroscience. In this doctoral thesis I explore two essential questions regarding the temporal dynamics of episodic memory retrieval. First, I investigate how rapidly distinct components of a visual object representation (i.e., perceptual and conceptual aspects) are reactivated during retrieval, and how this temporal sequence evolves compared to visual encoding. Findings from a series of behavioural, scalp electroencephalography (EEG) and intracranial EEG experiments, using reaction times and time-resolved decoding analyses, suggest that retrieval is a hierarchical, multi-layered process that follows the reverse order compared to encoding, prioritizing semantic information over perceptual details. Second, I explore whether memories are reactivated following a specific oscillatory rhythm. Computational models, based on studies in rodents, suggest that encoding and retrieval processes occur at opposing phases of hippocampal theta oscillations. Evidence for such phase modulation in humans is still sparse. The present findings suggest that in humans, neural signatures of memory retrieval fluctuate with, and are time-locked to, the phase of theta oscillations. Altogether, this doctoral thesis supports the view that retrieval is an oscillatory process and the elements that form our memories are retrieved following a biased and sequential order.

Dedicated to my nephew Sergio

Acknowledgments

I would like to sincerely thank Dr Maria Wimber. Since I started this journey, Dr Wimber has been more than an excellent supervisor and always offered her overwhelming support, vast knowledge, trust and understanding. Apart from her very best qualities as a supervisor, I would like to thank Dr Wimber for being a reference in her love for science, her honesty and her motivation for always going a step beyond in her work. All skills and knowledge that I gained during these years would be meaningless without such important values.

I would also like to thank Dr Simon Hanslmayr for his support and his helpful feedback. I would like to thank Dr Hanslmayr not only for sharing his expertise and knowledge but especially for transferring his enthusiasm for research.

I am particularly grateful for the immense emotional and infinite support of Dr Rodika Sokoliuk. This doctoral thesis is, to a large extend, the result of your priceless help and encouragement. I also want to acknowledge Dr Catarina Ferreira and Casper Kerrén, not only for being excellent friends and colleagues, but also for their remarkable help to make this work possible.

I would like to offer my special thanks to every single member of the Memory and Attention Group who shared with me infinite and valuable discussions. Without your help, inspirational ideas and feedback, this work would not be the same. Also, I want to acknowledge every collaborator that offered their valuable

work during these years. Thanks to everyone who helped me collecting data and, of course, to every person who took part in these experiments.

I also wanted to offer my special thanks to my parents, my sister and Cristina Cano for your support during all these years and, especially, for helping me to take the first step that culminated in this thesis.

Finally, I would like to thank the Integrative Biosciences Training Partnership for their financial support of this thesis.

Thanks to all of you.

Juan Linde-Domingo

Publications and Presentations

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

1. Publications

Linde-Domingo, J., Treder, M., Kerren, C., & Wimber, M. (preprint 2018). Evidence for a reversal of the neural information flow between object perception and object reconstruction from memory. bioRxiv, 300913. Under review.

Kerren, C.*, Linde-Domingo, J.*, Hanslmayr, S., & Wimber, M. An optimal oscillatory phase for pattern reactivation during memory retrieval. Current Biology. (accepted)

* These authors contributed equally

2. Conference contributions

Linde-Domingo, J., Treder, M., Kerren, C., Ter Wal, M., Roux, F., Chelvarajah, R., Rollings, D., Sawlani, V., Staresina, B., Hanslmayr, S., Wimber, M. (2018). Reversal of the information processing hierarchy between perception and memory. Poster at Society for Neuroscience (SfN), San Diego, USA. (abstract accepted)

Linde-Domingo, J., Treder, M., Kerren, C., Ter Wal, M., Roux, F., Chelvarajah, R., Rollings, D., Sawlani, V., Staresina, B., Hanslmayr, S., Wimber, M. (2018). Tracking the reconstruction of episodic memories in behaviour and EEG time courses. Learning and Memory 2018, Huntington Beach, USA.

Linde-Domingo, J., Treder, M., Kerren, C. & Wimber, M. (2017). Tracking the reconstruction of episodic memories in behaviour and EEG time courses. International Conference for Cognitive Neuroscience (ICON), Amsterdam, Netherlands.

Linde-Domingo, J., Treder, C. & Wimber, M. (2017). Deconstructing episodic memories to track their reconstruction in EEG time courses. British Neuroscience Association 2017. Birmingham, UK.

Linde-Domingo, J. & Wimber, M. (2016). Deconstructing memories to track their reconstruction in EEG time courses. Poster at Society for Neuroscience (SfN), San Diego, USA.

Linde-Domingo, J. & Wimber, M. (2016). Sustained processing shift towards pattern separation versus completion in an associative memory task. Poster at International Congress on Memory (ICOM), Budapest, Hungary.

Table of Contents

Table of Figures

Chapter 1: General introduction

Chapter 1

This doctoral thesis is focused on understanding the temporal dynamics of memory retrieval. Throughout the different chapters that form this work, I will present a corpus of experimental findings that consistently suggests two important features of episodic memory reactivation. First, I will report a series of studies that indicate that our episodic memories are not unitary representations and that its different elements are reactivated following the reverse order found during encoding or visual perception. In particular, our findings revealed that when retrieving a complex past event, access to its high-level information (as semantic features) is prioritized over its low-level perceptual details. Secondly, based on our experimental results, we suggest that episodic memory reactivation fluctuates rhythmically and is time-locked to the phase of ongoing theta oscillations.

To give a general overview of the topic, I will briefly describe the term “episodic memory”, its relationship with other types of memories, and the neural network associated with this learning system. Secondly, I will review some findings and theories that are trying to explain how external information is encoded and consolidated by the hippocampus, forming the unique engrams of our episodic experiences. Then, I will summarise theories and evidence produced over the last decade regarding the question how previously encoded representations are being retrieved. Lastly, in order to understand the relevance of this work, I will point out some crucial questions about the temporal dynamics of episodic memory retrieval that still remain unanswered.

1. Episodic memory

We can distinguish three major memory systems that depend on different neural networks and that respond to separate functions and types of associations (for a review, see Eichenbaum, 2016): (i) a habit learning system that links stimuli and behavioural responses, (ii) an emotional learning system that allows us to remember the associations between stimuli and their consequences, and (iii) a declarative learning system. This last one, the declarative learning system, is the one that supports the organization of knowledge about general facts (semantic memory), but also the creation of single and unique experiences (episodic memory) through the binding of heterogeneous information. Although there are clear interactions between these different systems and their neural networks, (for a review about these interactions, see Poldrack & Packard, 2003; Yang & Wang, 2017) it is important to highlight that one of the most important brain areas supporting declarative memories is the medial temporal lobe (MTL) including the hippocampus and adjacent cortical areas. Notably, the MTL is a key structure for the neural processing of time (Howard & Eichenbaum, 2013), space (Moser, Kropff, & Moser, 2008) and abstract concepts (Quiroga, 2012): three crucial building blocks of episodic memory.

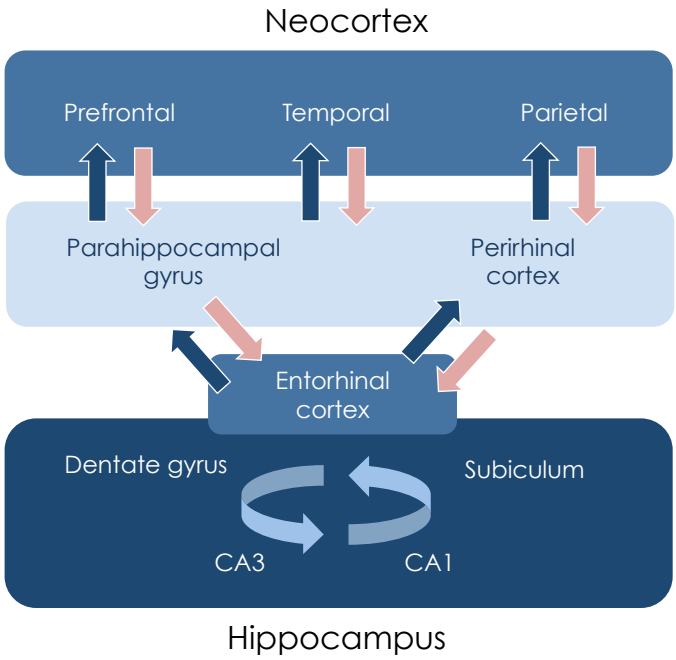
The term “episodic memory” was introduced by Tulving in 1972 (Tulving, 1972) and refers to the ability to retrieve personal events and to place our self in the past while a sense of time is maintained (Tulving, 1983, 2002). For instance,

this kind of memory system allows a song, a scent or a certain idea to reactivate a specific experience from our past as if we were living that moment again within a few seconds. That is, this memory system binds the space and time of our experiences and allows us to travel back in time in order to remember "what", "where" and "when" something happened. In contrast, semantic memories do not have an associated time-travelling feeling, and reflect our knowledge about regularities and general facts of the world (e.g. knowing that Burundi's capital is Bujumbura, or our date of birth).

When Tulving proposed both theoretical constructs, he also suggested that the episodic and the semantic memory system are two separate but interactive structures. This disconnection between types of memories received important critics from other memory theorists (like Baddeley or Roediger, see Tulving, 1984), and in response Tulving emphasized the similarities between both concepts in a later publication (Tulving, 1984). Specifically, in a response to these critics he stressed that "[episodic memory] is not a system parallel to the semantic system, standing as it were, side-by-side with it, but rather a sub-system, a system within a system". In this sense, a vast number of experimental findings have shown over the last decades the high degree of interactivity between both kinds of memories, but also between episodic memory and other cognitive domains as perception, language, decision-making or working memory (for a review, see Moscovitch, Cabeza, Winocur, & Nadel, 2016). In other words, despite the unique properties of episodic memory (i.e. the spatial-temporal properties and the sense of "time-travelling"), this system seems to

interact more with other cognitive processes than its original definition suggested, likely due to the role that the hippocampus and the rest of the brain network associated with episodic memory play in other domains. But what are the neural bases of episodic memory?

A large network of brain areas sustains the functioning of episodic memory (Fig 1). It includes neocortical areas related to control mechanisms (i.e. the prefrontal cortex) but also structures from the parietal and temporal cortex that are involved in perceptual processing. In this widespread network, the MTL, and more specifically the hippocampus and cortical areas surrounding the hippocampus, are a fundamental part of this neural system (for a review of this neurocircutry and associated disorders, see Dickerson & Eichenbaum, 2010) .



A. Chapter 1. Figure 1.

Figure 1. Block diagram of the episodic memory network (adapted from Rolls, 2010)**.** Pink arrows represent the forward connections to the hippocampus. Neocortical areas project to the hippocampus through the parahippocampal gyrus (scene information) and the perirhinal cortex (object processing). These projections arrive at the entorhinal cortex, which is the main connection between the hippocampus and the neocortex. From this structure, inputs are sent to inner hippocampal circuits starting at the dentate gyrus and continuing to CA3 and CA1, allowing episodic memory formation. During retrieval, outputs from the hippocampus are sent back to neocortical areas (blue arrows) via CA1 and the subiculum.

The hippocampus, which is composed of the dentate gyrus and the Ammon's horns (CAs), is the apex of the episodic memory system (Mishkin, Vargha-Khadem, & Gadian, 1998). Information from almost all neocortical association areas converges onto the cortical areas surrounding the hippocampus. The perirhinal cortex, sitting at the top of the hierarchy of object representation, and the parahippocampal cortex, involved in processing of spatial information, send

their projections to the entorhinal cortex. From the entorhinal cortex, these inputs reach the hippocampus through the perforant path, starting with the dentate gyrus and continuing to CA3 and CA1. The hippocampus integrates object and scene representations, capturing the spatial relation between the different parts of the perceived environment (Nadel & Peterson, 2013), and supports an initial learning of arbitrary associations of information (Kumaran, Hassabis, & Mcclelland, 2016; Randall C. O'Reilly, Bhattacharyya, Howard, & Ketz, 2014). The outputs of hippocampal processing are sent back from CA1 and the subiculum to parahippocampal regions, structures that possess feedback connections to neocortical association areas. As we will see, this complex network between the hippocampus, cortical areas surrounding the hippocampus and the neocortex is essential in organizing episodic memory formations and cortical representations (Dickerson & Eichenbaum, 2010).

In the next section, I will briefly review how external inputs that are processed by our senses are eventually encoded and integrated in our memory scheme. Later, I will cover theories and relevant findings that explain how these experiences are retrieved from the memory system.

2. From object representations to memory engrams

Information from our senses is one of the main inputs that is processed by the hippocampus (but not the only one). What we see or what we hear is encoded in order to create new memories, but external cues also help us to remember

our past. Although memory formation depends on various senses, due to the kind of stimuli that we used for our experiments, in the following I will focus on visual representations and how they are integrated in our memory system. Understanding the hierarchical process behind visual perception will be fundamental to comprehend one of our central hypotheses about memory reconstruction, since we propose that memory retrieval is also a hierarchical process that follows the reverse order of perception

2.1. The ventral and dorsal visual processing streams

Visual perception is classically described according to the influential “two-stream hypothesis” (Goodale & Milner, 1992; Milner & Goodale, 2008), suggesting the existence of two anatomically and functionally distinct processing streams: the ventral and the dorsal pathway. The ventral pathway (the “what” pathway) supports object recognition beyond changes in the environment, while the dorsal stream (the classical “where”) encodes motion, actions and the spatial relationship between visual elements. In both visual streams, the information is processed in a hierarchical manner, and each stage depends on the output of earlier steps, increasing the processing complexity along the stream. Although these visual pathways will be briefly described separately, visual perception depends on the interaction between both streams (for a review Milner, 2017).

The dorsal visual pathway has been traditionally described as the “where” pathway, however, the role of this visual stream remains still controversial. The conventional assumption that the dorsal stream supports the processing of object location has been questioned at several occasions, and single-unit studies have pointed out that processing of spatial information of objects depends on the same regions that support object recognition in the visual pathway (Aggelopoulos & Rolls, 2005; Rossi et al., 2006; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). For this reason, this widespread network that plays an important role in processes as motion detection, attention, actions or 3-D representations, should be understood as a “how” visual pathway (Rauschecker, 2018).

The dorsal visual stream starts in the thalamus’ lateral geniculate nucleus (LGN) that receives visual input from the retina, from the LGN, magnocelullar layers project to V1. Complex cells in V1 (the earliest area of the visual cortex) are sensitive to motion of moving edges, selective directions and speed of external stimuli (Hubel & Wiesel, 1968; Hubel, Wiesel, & Stryker, 1978; Orban, Kennedy, & Bullier, 1986). These outputs are sent to V2, an area that has been associated to directional maps; from there, information is projected to the middle temporal (MT) visual area (Born & Bradley, 2005) which also receives input from V1. MT neurons respond to 2D motion, speed and spatial frequency among others features (Brooks, Morris, & Thompson, 2011; Maunsell & Van Essen, 1983). The dorsal visual pathway continues to the parietal cortex which, apart from the projections from MT, also receives visual information through the

superior colliculus and the pulvinar nucleus (Gallivan & Goodale, 2018). The parietal cortex processes motion at a higher-level that involves the analysis of complex information, for instance, how does the object's motion change while the perceiver moves, but also the integration of optic flow with the head and eye position (Raffi, Persiani, Piras, & Squatrito, 2014). Importantly, the posterior parietal cortex is supposed to be involved in an important step within the dorsal stream where the sensorimotor information is transformed into actions, allowing object grasping and manipulation or online correction during movement (for a review Gallivan & Goodale, 2018).

With respect to the goals of this doctoral thesis, understanding the ventral visual pathway is key to comprehending how the original perception of a visual object (the main type of stimuli used in our experiments) is transformed into memory engrams (Fig 2). The ventral visual pathway culminates with object recognition (Cowey & Weiskrantz, 1967). This processing sequence starts with the analysis of low-level features as colours and shapes that are eventually combined into intermediate and more holistic object representations which include semantic information about the perceived object (Biederman & Cooper, 1991; Marr & Nishihara, 1978; Martin, Douglas, Newsome, Man, & Barense, 2018). This integration of low and high-level elements is one of the key features of the ventral visual processing hierarchy (for a review of some models, see Poggio & Ullman, 2013).

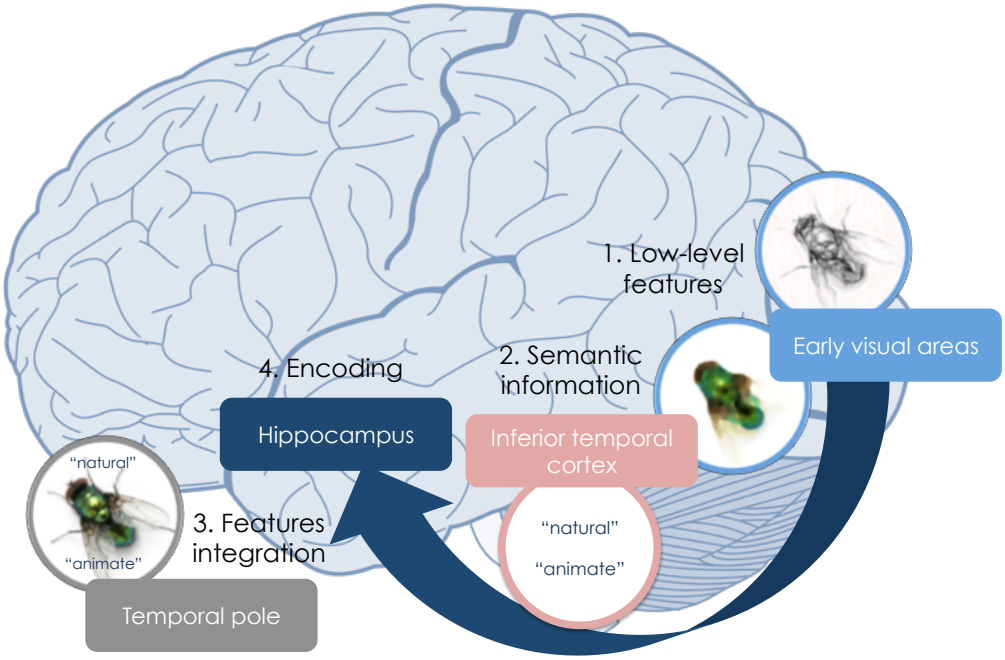
Most of the ganglion cells of the retina send projections to the lateral geniculate nucleus of the thalamus. Parvocelullar layers and some additional magnocellular layers of this thalamic structure project to V1, which is located in the visual cortex (Ferrera, Nealey, & Maunsell, 1994). V1 is subdivided into blobs and interblobs. On one hand, blobs are colour sensitive (Solomon, 2005) and, on the other hand, interblobs are selective to stimulus' orientation, including the processing of edges, bars and gratings (Hubel & Wiesel, 1968; Hubel, Wiesel, & Stryker, 1978). Although blobs and interblobs process these features independently of the type of object, there is evidence suggesting that feedback connections from higher-level processing areas can generate object-based modulations (Roelfsema, Lamme, & Spekreijse, 1998). A progression in shape processing is carried out in V2: the neurons of this visual area respond selectively to edge orientation (Peterhans & von der Heydt, 1989; von der Heydt, Peterhans, & Baumgartner, 1984) assign edges to objects, and encode to which object each border belongs (Zhou, Friedman, & von der Heydt, 2000). Processing outputs from V1 and V2 converge in V4, where a first stage of an intermediate object representation takes place. In V4, colour information is integrated in an object representation (Conway, Moeller, & Tsao, 2007; Schein & Desimone, 1990) but also shape processing progresses significantly here due to the combination of previous inputs that allow to place objects' curvatures with respect to the center of their shape (Pasupathy & Connor, 2001).

The next processing step of the ventral visual pathway takes place in the inferior temporal cortex (IT), where high-level object representations are

constructed despite changes in the retinal images or the surrounding environment. The IT processes a wide range of representational features that increase their complexity from posterior to anterior areas, including integration of shapes, textures, the real-word size of the perceived object (independently of the retinal size) and its semantic properties, for instance, whether the object represents an inanimate entity (T. Carlson, Tovar, Alink, & Kriegeskorte, 2013; Cichy, Pantazis, & Oliva, 2014; Clarke & Tyler, 2015; Konkle & Oliva, 2012; Tanaka, Saito, Fukada, & Moriya, 1991). Importantly, IT representations in the human brain predict the unique perceptual judgment of the individual, reflecting how each of us perceives the world (Charest, Kievit, Schmitz, Deca, & Kriegeskorte, 2014).

Finally, one of the latest stages of perception along the ventral stream is the combination of perceptual and semantic information. Although the multimodal integration of abstract conceptual features (but not perceptual) has been linked to the temporal pole (for a review, see Ralph, Jefferies, Patterson, & Rogers, 2016), recent findings indicate that the convergence of perceptual and conceptual (semantic) features is supported by the perirhinal cortex (Martin et al., 2018). As it was pointed out previously, the perirhinal cortex sends projections to the entorhinal cortex that finally arrive at the dentate gyrus of the hippocampus, the top of the representational hierarchy (Kent, Hvoslef-Eide, Saksida, & Bussey, 2016).

One of the main objectives of this thesis is to test whether when a visual representation is retrieved from memory, its components (i.e., low-level details as colours and semantic information) are also reactivated following a hierarchical stream. Specifically, we predicted that retrieval is also a sequential process that follows the reverse order of visual perception (i.e., we expected that semantic information would be reactivated before perceptual details). This alternative hypothesis will be explained in detail in the last section of this general introduction.



B. Chapter 1. Figure 2.

Figure 2. Schema of object processing in the visual ventral stream. (1) When a stimulus is perceived, its low-level features (like lines and colours) are first processed in early visual areas in the occipital cortex. (2) Then, higher-level information, including the semantic category of the perceived item (e.g. whether the stimulus is animate), is coded in the inferior temporal cortex. (3) At a later stage, low and high-level features are integrated. Although the temporal pole has been traditionally identified as the main area integrating these features, the perirhinal cortex has also been associated with this role (see main text). (4) Finally, together with temporal and spatial information, these inputs about the perceived object arrive in the hippocampus via the entorhinal cortex, where the different features of a complete episode are encoded.

2.2. The engram formation

Aiming to briefly describe the neural mechanisms behind the formation and access to episodic memory, I will follow the complementary learning system framework, one of the most influential computational models that describe the

putative mechanisms behind encoding, retrieval and consolidation processes (Kumaran et al., 2016; K. A. Norman, 2012; Randall C. O'Reilly et al., 2014). The main idea of this model is the requirement of two different learning systems to support declarative memories: one sparse system that allows quick episodic learning and depends on the hippocampal formation; and a second neocortical system that gradually integrates regularities across episodes to acquire semantic information. As we will see, this model describes the relationship between the episodic and the semantic system and the basic interaction between both systems required to consolidate our memories.

As we saw, the hippocampus sits in a privileged position to integrate distinct information about the external world and to create a rich episodic representation. The parahippocampal cortex, which is connected with both the dorsal and the ventral visual pathway, funnels information about both objects and scenes or spatial backgrounds (Harel, Kravitz, & Baker, 2013). On the other hand, the perirhinal cortex supports the processing of perceptual and semantic features of objects. Inputs from both structures, the parahippocampal and the perirhinal cortices, converge in the entorhinal cortex, which connects to the hippocampus proper. Specifically, it is connected to the dentate gyrus, CA3 and CA1. Moreover, the entorhinal cortex receives reward-related information from the orbitofrontal cortex and the amygdala, implementing inputs from the emotional learning system (W. A. Suzuki & Amaral, 1994; W. L. Suzuki & Amaral, 1994).

There is a key feature about how the hippocampus encodes this rich and heterogeneous ensemble of information. In order to reduce the degree of interference between unrelated events, the hippocampus is thought to keep different memory representations highly separated (or orthogonalized) from each other. This mechanism of transforming episodes into non-overlapping dissimilar events is known as *pattern separation* (for a review Rolls, 2013, 2015; Yassa & Stark, 2011). Various hippocampal theories suggest that the dentate gyrus (but not only this area; Kent et al., 2016) works as a competitive network that produces non-overlapping neural indices throughout its sparse projections to CA3 (i.e. mossy fibers) (Kesner & Rolls, 2015; R C O'Reilly & McClelland, 1994). This pattern separation mechanism allows CA3 neurons to bind information avoiding a likely interference with previously memorized episodes. Namely, different neurons in CA3 could take part in encoding memories of similar events, operating as an attractor network that forms differentiated objects, spatial and reward associations. However, as we will see in the next section, CA3 also supports access to past events based on partial information (Edmund T Rolls, 2013).

In summary, the hippocampal system allows a rapid incidental learning, forming a sparse neural representation of a rich, multiplexed experience. This way, the hippocampus encodes in a "one-shot" manner what we perceive, and creates an index pointing to the elements of each episode in the neocortex. Although this capacity is fundamental for episodic memory formation, it also implies a series of limitations. First, creating an engram of every association is

challenging for a finite capacity system (Treves & Rolls, 1994). Second, due to the unselective nature of these associations, a control system is needed that organises and gives meaning to our episodic memories. Because of the latter limitation (but also to avoid "catastrophic interference", see Randall C. O'Reilly et al., 2014), the declarative memory system is supported not only by the hippocampus, but also by the neocortical system.

The hippocampus and its adjacent structures are in continuous interaction with the neocortex. This communication between brain areas allows the slow formation of stable and long-term neocortical representations from the initial hippocampal traces, known as system-level consolidation (Dudai, Karni, & Born, 2015). One widely accepted idea is that this transfer of information from the hippocampus to the neocortex occurs mainly during "offline" periods of memory replay. Numerous evidence has shown that memories are reactivated during periods of sleep and rest (O'Neill, Pleydell-Bouverie, Dupret, & Csicsvari, 2010; Wikenheiser & Redish, 2015) and that CA3 plays a causal role in this system-level consolidation (Nakashiba, Buhl, McHugh, & Tonegawa, 2009). Moreover, it has been suggested that "online" reactivation during active retrieval could serve as a fast route for the formation of neocortical representations (Antony, Ferreira, Norman, & Wimber, 2017).

One of the principal features of this second neocortical system is that the neocortex is able to learn the statistical regularities from our daily events that serve to form semantic memories (i.e. knowledge about general facts that is

context independent: McClelland et al., 1995; Moscovitch, 2008). That is, while the hippocampal system supports the reactivation of spatiotemporal and detailed information of past events, the main function of neocortical memories is representing the “gist” of past episodes.

As we can experience analysing our own remote memories, specific details of past episodes are largely lost over time. Although this transformation into more gist-like memories has been associated with a representational trade-off from hippocampus to neocortex, recent work has pointed out that this transformation could depend on changes within the hippocampus (i.e. an increase of activity in the posterior over the anterior hippocampus, whose activity remains more stable over time; Dandolo & Schwabe, 2018). However, it is important to mention that the formation of semantic memories does not necessarily go along with a loss of specific details of episodic memories. Episodic and semantic memories can coexist and support each other, but the reactivation of specific spatiotemporal information about an episode seems to depend on the hippocampal system (Westmacott, Black, Freedman, & Moscovitch, 2004; Westmacott, Freedman, Black, Stokes, & Moscovitch, 2004).

Storing, modifying and consolidating our past experience would be meaningless if we could not accurately retrieve this information. In the next section I will therefore review theories and findings that try to address how the brain accesses these originally stored memory engrams and brings past representations back to our “mental eye”.

3. From memory engrams to mental representations

Sometimes, a simple cue, like a photograph of a house, is enough to awake remote but vivid memories associated with that place. Undoubtedly, it is fascinating that, despite the vast amount of information that the neural system retains, we can precisely access such a rich representation. However, before reviewing some relevant findings about the neural mechanisms behind memory retrieval, I would like to emphasize the reconstructive nature of memory recollection (Schacter, 2012; Schacter, Guerin, & St Jacques, 2011). As outlined previously, our memory engrams are in constant transformation and the continuous interaction between different memory systems allows us to preserve traces of relevant information (Horner & Doeller, 2017). Even active decisions seem to determine which representations will become consolidated (Murty, DuBrow, & Davachi, 2018). Forgetting and modifying parts of our memories is, unquestionably, a necessary mechanism for the correct functioning of our memory system (Kuhl, Bainbridge, & Chun, 2012; Williams, Hong, Kang, Carlisle, & Woodman, 2013; Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015). For this reason, it is important to understand that our memory representations are no “snapshots” of the past, but rather imperfect reconstructions whose errors, loss of details and distortions are part of the normal functioning of our memory system. Based on this premise, my thesis is concerned with the ways in which the memories we retrieve are systematically different from the original memories we encode.

How can a partial cue reactivate a past episode? As pointed out above, during the encoding (or learning) of a new experience, pattern separation mechanisms are essential. By contrast, retrieving information depends on a different computational process that has been termed pattern completion. Pattern completion is the process by which the hippocampal system is thought to be capable of retrieving multi-element representations from partial or noisy cortical activity that was present at the time of encoding (Horner & Burgess, 2014; R C O'Reilly & McClelland, 1994). In other words, an external input that partially overlaps with a stored memory trace could act as a reminder, re-awaken the hippocampal index, and in turn restore a relatively complete pattern of brain activity representing the original memory.

Computationally, area CA3 has been associated with pattern separation and pattern completion (for a review Deuker, Doeller, Fell, & Axmacher, 2014). Since both processes are computationally incompatible, it is thought that the hippocampus is able to shift between pattern completion and pattern separation depending on the received input. For instance, if the incoming information is highly consistent with previous memories (or expectations) the system will be biased towards pattern completion processes. However, if the incoming information is novel or unexpected, the hippocampus would change into an encoding mode, supported by pattern separation (Hasselmo & Schnell, 1994; Schapiro, Kustner, & Turk-Browne, 2012). Empirical support for this view comes from high resolution fMRI studies suggesting that CA1 works as a novelty detector that can shift the hippocampal system into a pattern separation

(encoding) or a pattern completion (retrieval) mode (Chen, Olsen, Preston, Glover, & Wagner, 2011; Duncan, Ketz, Inati, & Davachi, 2012). Here it is important to highlight that influential computational models have suggested that this shift between optimal states for pattern separation and pattern completion could be supported by the phase of hippocampal neural oscillations (i.e., theta oscillations; Hasselmo, Bodelón, & Wyble, 2002). In particular, rodent studies have shown that stimulating hippocampal neurons at opposite phases of a theta oscillation can be beneficial for pattern completion or pattern separation (Pavlides, Greenstein, Grudman, & Winson, 1988) and suggest that encoding and retrieval are oscillatory processes. This results have been replicated in additional rodent studies (Huerta & Lisman, 1993) and this role of theta oscillations has been implemented in several models of episodic memory (Buzsáki, 2002; Hasselmo & Eichenbaum, 2005; Kunec, Hasselmo, & Kopell, 2005; Parish, Hanslmayr, & Bowman, 2018). However, despite the influence of this model to explain how the hippocampus shifts between pattern completion and separation, there is no direct evidence in humans suggesting that encoding and retrieval processes are modulated by the phase of theta oscillations. As I will describe in a later section, one of the main objectives of this doctoral thesis is to test whether long-term memory reactivation is modulated by the phase of ongoing hippocampal theta oscillations.

Plenty of evidence suggests that once the memory system is in the optimal state for retrieval, the reactivation of an episode is supported by the reinstatement of the neural activity patterns from different cortical areas

produced during its encoding (Horner, Bisby, Bush, Lin, & Burgess, 2015; Edmund T. Rolls, 2017). Thus, in order to reactivate past experiences, the hippocampus needs to recruit neocortical networks (Treves & Rolls, 1994). How does the hippocampus send its outputs back to these areas? When the hippocampal system enters a retrieval mode, CA3 neurons send projections back to cortical regions (entorhinal cortex) throughout CA1. This is due to the capacity of CA1 to translate orthogonalized representations in CA3 into more overlapping representations in the entorhinal cortex. Specifically, these backward connections from CA1 project to the deep layer of the entorhinal cortex (the superficial layer projects forward connections into the hippocampus; for a review, see Norman, 2006). In other words, the entorhinal-hippocampal information flow changes its directionality from encoding to retrieval, integrating the engram held in different areas of the cortex (Fell et al., 2016; Staresina, Cooper, & Henson, 2013). In turn, entorhinal neurons send their signal back to neocortical areas that deliver input to the hippocampus during encoding (Lavenex & Amaral, 2000; Witter et al., 2000). This includes multimodal cortical areas like the superior temporal sulcus, but also unimodal association cortex like the inferior temporal visual cortex. This neocortical reactivation is also thought to be key in the formation of new long-term memory representations in both multimodal and unimodal areas, and could also support memory consolidation and the building of new semantic or gist-like memories (Antony et al., 2017; Edmund T. Rolls, 1991).

In the last decades, several human studies have shown that during retrieval, there is a reactivation of sensory and emotional areas that were active during encoding (for a review, see Danker & Anderson, 2010). This recruitment of neocortical areas has been associated to the mental reinstatement of previous events, including details that form these representations. In this respect, the episodic memory field has experienced a significant advance in the study of memory reactivation over the past years, using brain-imaging techniques. For instance, thanks to the implementation of multivariate pattern analyses (MVPA; for a review, see Norman, Polyn, Detre, & Haxby, 2006) in brain imaging studies, new approaches like representation similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) or decoding analyses (for a review, see Hebart & Baker, 2017) have been applied in order to track neural representations in the human brain (e.g., Kuhl et al., 2011; Richter, Chanales, & Kuhl, 2016; Staresina et al., 2012; Wimber et al., 2015). An increasing number of neuroimaging studies has used the neural activity patterns during retrieval to identify specific features of past events. For instance, analyses of the neural activity during retrieval have been applied to detect the context in which the event was encoded (Johnson, McDuff, Rugg, & Norman, 2009). But more important for the objectives of this thesis, it has been shown that low-level perceptual information of remembered stimuli can be identified in the neural signature (Bosch, Jehee, Fernandez, & Doeller, 2014; Waldhauser, Braun, & Hanslmayr, 2016; Wimber, Maaß, Staudigl, Richardson-Klavehn, & Hanslmayr, 2012). Furthermore, several studies have shown that the semantic category of past representations is reactivated during retrieval (Kuhl et al., 2011; Staresina

et al., 2012; Wimber et al., 2015). Altogether, these previous findings suggest that retrieval implicates the cortical reinstatement of a past neural state; but also that different features of previous encoded representations (e.g., perceptual or semantic information) can be identified in the neural signal using MVPA approaches.

These multivariate analyses used to investigate memory representations can be also applied following a time-resolved approach. That is, by using high-temporal resolution techniques as electroencephalography (EEG), it is possible to track how mental representations emerge on the millisecond level in order to understand their temporal dynamics (Cichy et al., 2014; Kurth-Nelson, Barnes, Sejdinovic, Dolan, & Dayan, 2015; Van de Nieuwenhuijzen et al., 2013). Throughout this doctoral thesis, I will present a set of findings that were obtained using a decoding time-resolved approach that allow us to investigate whether there is a systematic temporal pattern associated to memory reactivations and to examine the reactivation of perceptual and semantic details of memory representations over time.

4. Hypotheses about the temporal dynamics of retrieval

Investigating the temporal features of human episodic memory is the main objective of this doctoral thesis. In this last section, I will concisely introduce and contextualize the two central hypotheses that will be tested in later chapters: (i) the reverse reconstruction hypothesis and (ii) the role of theta oscillations in memory reactivation.

With respect to this thesis, there are four relevant conclusions that can be drawn from the literature reviewed above:

(a) Perception of a visual stimulus follows a hierarchical stream where the processing of low-level features (like colours and lines) precedes access to higher-level semantic information. Multiplexed visual representations with many distinct features are eventually encoded as episodic memories by the hippocampus (Fig 2).

(b) Retrieving a specific episodic memory is associated with the reactivation of neocortical areas that processed this information during encoding (including low-level visual and semantic processing areas)

(c) When a partial cue triggers a retrieval processing cascade, the information flow between the hippocampus and the entorhinal cortex reverses between encoding and retrieval, in terms of neural coupling.

(d) Memory engrams are in constant transformation. System-consolidation depends on the interaction between hippocampus and neocortex. Over

time, the neocortical system can extract regularities from one-shot episodic memories, creating more semantic-like representations.

Although important advances have been made regarding episodic memory reactivation, there is one fundamental question that remains unanswered. Specifically, it is still unknown whether the reactivation of past representations follows a hierarchical processing similar to visual perception. In other words, it is unclear whether distinct representational features of past events (i.e. low-level perceptual and high-level conceptual features) are reactivated in a systematic, hierarchical manner. Given how little is known about the temporal dynamics of the retrieval process, the work in this thesis is based on a novel working hypothesis that we call the reverse reconstruction hypothesis. Based on the above-mentioned conclusions, we hypothesise that retrieval is not an all-or-none process, but a hierarchical reconstruction process where, once triggered by a reminder, the various features that constitute a memory for a past event unfold in time in a specific order. Our main hypothesis is that during retrieval, semantic or high-level information of past events (i.e. whether a remembered entity is an animal or an object) will be retrieved before low-level features (i.e. the specific colour in which this entity was perceived). Namely, we expect that when a visual representation is reactivated from memory, there will be a processing stream of its components that will follow the reverse order of visual perception. In all later chapters, we will refer to this prediction as the “reverse reconstruction hypothesis”. This hypothesis will be tested throughout 6 different

studies that include behavioural and electrophysiological experiments. The main objective of each experiment will be explained in Chapter 2.

In this general introduction I presented computational models suggesting that the hippocampal system is able to shift between states that are optimal for encoding and retrieval processes (Hasselmo & Schnell, 1994; K. A. Norman & O'Reilly, 2003). Importantly, some findings obtained in rodent studies suggest that the shift between encoding and retrieval states is supported by the phase of ongoing hippocampal theta oscillations (Huerta & Lisman, 1993; Pavlides et al., 1988). Specifically, these results suggest that encoding and retrieval occur in a rhythmic manner in the hippocampus, where these processes are separated by a 180 degrees shift in theta phase. Based on the evidence obtained in these animal studies, different models about memory assume that retrieval is supported by hippocampal theta oscillations (Hasselmo et al., 2002; Kunec et al., 2005; Parish et al., 2018; Watrous & Ekstrom, 2014). However, despite the influence of these models, there is no direct evidence in human studies supporting that long-term memory reactivation is associated with a specific phase of the theta rhythm. Therefore, a second main objective in this doctoral thesis is to investigate in humans whether memory reactivation is also an oscillatory process modulated by the phase of the ongoing theta rhythm; but also whether this theta phase reverses between encoding and retrieval processes. This alternative hypothesis (i.e., whether memory reactivation in humans is an oscillatory process that is associated to a certain phase of the theta rhythm) was tested in an EEG experiment (Chapter 6) where we obtained

an index of memory reactivation over time through decoding analyses. The main objectives of this experiment will be explained in Chapter 2.

Chapter 2: Objectives

1. Evidence for a reversal of the neural information flow between object perception and object reconstruction from memory (Chapter 3)

The main objective of this first series of experiments was to test the reverse reconstruction hypothesis of episodic memory. We hypothesized that when a previously encoded visual representation is being retrieved from memory, its perceptual and semantic features are reactivated following the reverse temporal order compared to visual perception. We thus predicted that information about higher-level semantic contents is prioritized over low-level perceptual features when accessing a visual memory. To test this hypothesis, we used a simple cued recall paradigm where participants were asked to learn novel word-object associations, and were later asked to retrieve the object that had been associated. On each retrieval trial they were asked to respond to questions about perceptual and semantic features of these objects.

Experiment 1 and Experiment 2 tested the reverse reconstruction hypothesis and its replicability on the behavioural level, using reaction times (RTs) and accuracy profiles as dependent variables. Based on object recognition literature (T. Carlson et al., 2013; Cichy et al., 2014; Clarke & Tyler, 2015; Lehky & Tanaka, 2016; Martin et al., 2018), we predicted that, when objects are displayed on the screen, participants would be faster and more accurate to respond to questions about perceptual features of these objects (e.g., whether it was presented as a line drawing or as a photograph) than answering questions about semantic details (e.g., whether the object represented an animal or not).

Conversely, we expected that when participants remember these images in order to respond to perceptual and semantic questions, they would be faster and more accurate retrieving semantic information compared to low-level perceptual details.

This hypothesis was also tested in a third experiment (Experiment 3) using scalp electroencephalography (EEG) together with a time-resolved decoding analysis approach (T. Carlson et al., 2013; Cichy et al., 2014; Kurth-Nelson et al., 2015). This approach allowed us to identify at which specific moment in time the brain signal was more associated with processing the perceptual or the semantic features of a memory, both during encoding (i.e., visual perception) and retrieval. In this experiment we expected that during the time window in which objects were presented on the screen, we would find neural evidence of perceptual processing earlier than indications of semantic processing. During retrieval of the object represetation, we however predicted to find evidence suggesting that semantic information is reactivated before perceptual details.

2. Preliminary findings in an iEEG case study support the reverse reconstruction hypothesis (Chapter 4)

In this chapter, I will present some preliminary results of an intracranial electrophysiological case study (Experiment 4). In this work, the main objective was to test the reverse reconstruction hypothesis in an experiment where an

epileptic patient with implanted intracranial electrodes performed the same task that we used in Experiment 3 (Chapter 3).

Due to the electrode localisation that covered brain areas of special interest (i.e., semantic processing areas along the ventral stream and early visual processing areas), this case study allows us to validate the reverse reconstruction hypothesis with high spatial resolution, and to go a step further in understanding the role of different neural structures involved in retrieving semantic and perceptual features. Therefore, the main objective of this study was to gain first insights into how visual representations and their low and high-level features are represented in time within different brain areas of interest during encoding and retrieval using a similar decoding analysis approach than the one used in Experiment 3 (Chapter 3). Our main prediction was to find a reverse reconstruction effect restricting our decoding analysis to electrode contacts located along the ventral visual stream.

3. The reverse reconstruction effect across different perceptual and semantic manipulations (Chapter 5)

Two additional behavioural experiments (Experiment 5 and 6) were carried out to test if the results obtained in Experiment 1 and 2 could be generalized using different stimuli and other types of perceptual and semantic manipulations. Experiments 1-4 were all based on identical manipulations of perceptual and semantic content of a memory: items were perceptually manipulated by

presenting them either as colour photographs or as line drawings, and their semantic categories differed in whether they presented animate or inanimate objects.

Aiming to test the generalization of the reverse reconstruction hypothesis, the semantic manipulation in Experiment 5 and 6 was based on the object's artificialness: all items were either natural (e.g., fruits or plants) or artificial objects (e.g., electronic devices or music instruments). A new perceptual manipulation was used in each experiment, with objects being shown on the screen in two different retinal sizes in Experiment 5 (size manipulation); and objects with either a round or elongated shape in Experiment 6 (shape manipulation). We expected to replicate previous behavioural results (Experiment 1 and 2 in Chapter 3) in both experiments.

4. An optimal oscillatory phase for pattern reactivation during memory retrieval (Chapter 6)

The main objective of the last experiment presented in Chapter 6 was to investigate the role of brain oscillations in shifting the memory system between encoding and retrieval states. More specifically, based on evidence from rodent studies and computational models (Buzsáki, 2002; Hasselmo et al., 2002; Hyman, Wyble, Goyal, Rossi, & Hasselmo, 2003; Parish et al., 2018; Pavlides et al., 1988), we aimed to shed light onto the question how theta oscillations modulate memory reactivation in humans using scalp EEG.

For this purpose, we re-analysed data from Experiment 3 while focusing on semantic features only. We carried out decoding analyses on the EEG signal during retrieval (Carlson et al., 2013; Cichy et al., 2014; Kurth-Nelson et al., 2015) that provided us with a time-resolved, parametric index of memory reactivation (i.e., classifier fidelity for semantic features). This index was used to investigate whether neural memory reactivation oscillates following a theta rhythm during retrieval. Furthermore, in order to test whether maximum reactivation is linked to a specific phase of theta oscillations, we examined how these oscillations behaved around the time points when the classifier indicated maximum fidelity. We predicted that memory reactivation would fluctuate with the theta rhythm and that the peak of this reactivation index would be modulated by the phase of theta oscillations.

Chapter 3: Evidence for a reversal of the neural information flow between object perception and object reconstruction from memory

Juan Linde-Domingo[1], Matthias S. Treder[2], Casper Kerren[1] & Maria Wimber[1]

[1]School of Psychology & Centre for Human Brain Health (CHBH), University of Birmingham (UK), [2]Cardiff University Brain Research Imaging Centre (CUBRIC), Cardiff University (UK)

At the time of thesis submission, this chapter represents a near identical manuscript under revision in Nature Communication. An early pre-printed version of this manuscript has been published in BioRxiv (https://doi.org/10.1101/300913).

Abstract

Remembering is a reconstructive process. Surprisingly little is known about how the reconstruction of a memory unfolds in time in the human brain. We used reaction times and EEG time-series decoding to test the hypothesis that the information flow is reversed when an event is reconstructed from memory, compared to when the same event is initially being perceived. Across three experiments, we found highly consistent evidence supporting such a reversed stream. When seeing an object, low-level perceptual features were discriminated faster behaviourally, and could be decoded from brain activity earlier, than high-level conceptual features. This pattern reversed during associative memory recall, with reaction times and brain activity patterns now indicating that conceptual information was reconstructed more rapidly than perceptual details. Our findings support a neurobiologically plausible model of human memory, suggesting that memory retrieval is a hierarchical, multi-layered process that prioritizes semantically meaningful information over perceptual detail.

1. Introduction

When Rocky Balboa goes back to his old gym in the film Rocky V, the boxing ring and the feeling of the dusted gloves in his hands trigger a flood of vivid images from the past. Like in many other movies featuring such mnemonic flashbacks, the main character seems capable of remembering what the room looked like years ago, who was there at the time, and even an emotional conversation with his old friend and coach Mickey. Perceptual details like colours, however, are initially missing in the scene, like in a faded photograph, and only gradually saturate over time. This common way to depict memories in pop culture nicely illustrates that the memories we bring back to mind are not unitary constructs, and also not veridical copies of past events. Instead, it suggests that remembering is a reconstructive process that prioritizes more meaningful components of an event over other, more shallow aspects (Schacter, 2012; Schacter et al., 2011). We here report three experiments that shed light onto the temporal information flow during memory retrieval. Once a reminder has elicited a stored memory trace, are the different features of this memory reconstructed in a systematic, hierarchical way?

Considering our vast knowledge about the information processing hierarchy during visual perception, surprisingly little is known about the time course of memory recall. In the object recognition literature, it is generally agreed that the presentation of an external stimulus initiates a processing cascade that starts with low-level perceptual features in early visual areas, and progresses to

increasingly higher levels of semantic integration and abstraction along the inferior temporal cortex (Carlson, Tovar, Alink, & Kriegeskorte, 2013; Cichy, Pantazis, & Oliva, 2014; Clarke & Tyler, 2015; Lehky & Tanaka, 2016; Martin, Douglas, Newsome, Man, & Barense, 2018; Serre, Oliva, & Poggio, 2007). However, mental representations can also be re-created from memory, without much external stimulation: retrieving a scene from the movie Rocky V will elicit semantic knowledge about the film (e.g. that the actor is called Sylvester Stallone), but also mental images that can include fairly low-level details (e.g. whether the scene was in colour or in grey scale). How the brain manages to bring back each of these features when reconstructing an event from memory remains an open question. The present series of experiments tested our central working hypothesis that the stream of information processing is reversed during memory reconstruction compared with the perception of an external stimulus.

Over the last years, multivariate neuroimaging methods have made it possible to isolate brain activity patterns that carry information about externally presented stimuli, but also about internally generated mnemonic representations. Importantly, it has been shown that the neural trace that an event produces during its initial encoding is reinstated in brain activity during its later retrieval (Chen et al., 2017; Johnson, McDuff, Rugg, & Norman, 2009; Kuhl, Rissman, Chun, & Wagner, 2011; Michelmann, Bowman, & Hanslmayr, 2016; Staresina, Henson, Kriegeskorte, & Alink, 2012; Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015). Most of these studies focused on the reactivation of abstract information, including a picture’s category (Kuhl et al.,

2011; Staresina et al., 2012; Wimber et al., 2015) or the task context in which it was encoded (Johnson et al., 2009). Apart from these higher-level features, evidence also exists for the reactivation of low-level perceptual details in early visual areas (Bosch et al., 2014; Waldhauser et al., 2016). Moreover, a growing literature using electrophysiological methods has begun to shed light onto the timing of such reinstatement, typically demonstrating neural reactivation within the first second after a reminder is presented (Jafarpour, Fuentemilla, Horner, Penny, & Duzel, 2014; Michelmann et al., 2016; Sols, DuBrow, Davachi, & Fuentemilla, 2017; Staudigl et al., 2012), and sometimes very rapidly (Waldhauser et al., 2016; Wimber et al., 2012). However, because all existing studies focused on a single feature of a memory representation (e.g., its semantic category), the fundamental question whether memory reconstruction follows a hierarchical information processing stream, similar to perception, has not been investigated.

We hypothesize that such a processing hierarchy does exist, and that the information flow is reversed during memory retrieval compared with perception. That is, based on the widely accepted idea that memory reconstruction depends on back-projections from the hippocampus to neo-cortex (Marr, 1971; Moscovitch, 2008), we expect that those areas that are anatomically closer to the hippocampus (i.e. high-level conceptual processing areas along the inferior temporal cortex) should be involved in the reactivation cascade faster than areas that are relatively remote (i.e., low-level perceptual processing areas in earlier visual cortices). Therefore, we assume that once a reminder has initiated

the reactivation of an associated event, higher-level abstract information will be reconstructed before lower-level perceptual information, producing an inverse temporal order of processing compared with perception.

We tested this reverse reconstruction hypothesis in a series of two behavioural and one EEG experiment (see Fig. 1b, c, and Fig. 3a). All studies used a simple associative memory paradigm where participants learn a series of arbitrary associations between word cues and everyday objects, and are later cued with the word to recall the object. In order to test for a processing hierarchy, it was important to independently manipulate the perceptual and conceptual contents of these objects. Therefore, objects varied along two orthogonal dimensions: one perceptual dimension, where the object was either presented as a photograph or a line drawing; and a semantic dimension where the object represents an animate or inanimate entity (Fig. 1a). The two behavioural experiments measure reaction times while participants make perceptual or semantic category judgments for objects that are either visually presented on the screen, or reconstructed from memory. The EEG experiment uses a similar associative recall paradigm together with time-series decoding techniques (Carlson et al., 2013; Cichy et al., 2014; Kurth-Nelson, Barnes, Sejdinovic, Dolan, & Dayan, 2015), allowing us to track at which exact moment in time perceptual and semantic components of the same object are reactivated, and to create a temporal map of semantic and perceptual features during perception and memory reconstruction (Fig. 3b and c). Our behavioural and electrophysiological findings consistently support the idea that memory

reconstruction is not an all-or-none process, but rather progresses on each single trial from higher-level semantic features to lower-level perceptual details.

2. Results

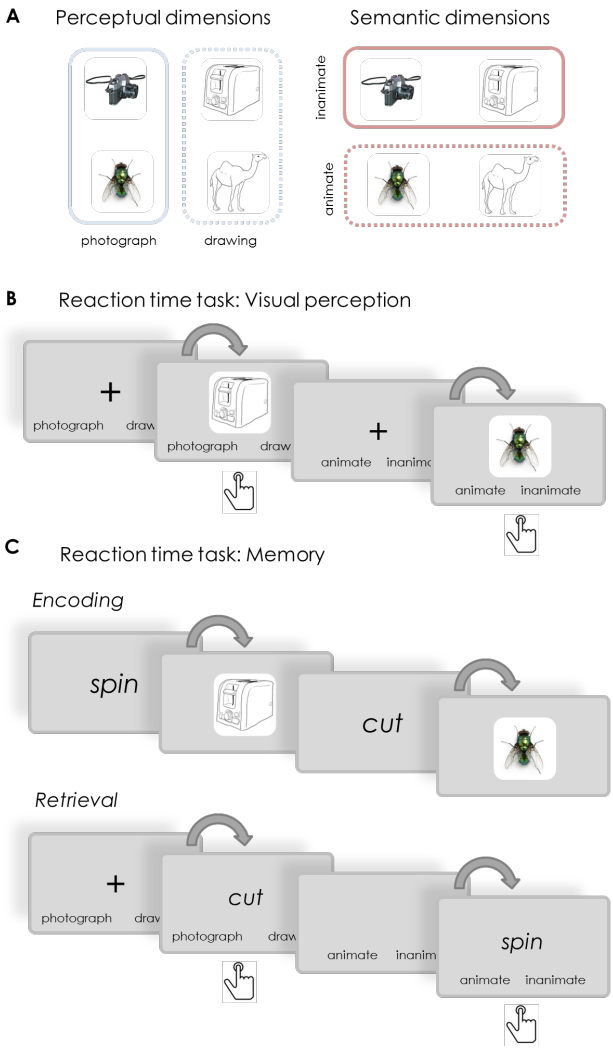
2.1. Behavioural experiments

Our two behavioural experiments used reaction times (RTs) to test our central hypothesis that the information processing hierarchy reverses between the visual perception of an object and its reconstruction from memory. We assumed that the time required to answer a question about low-level perceptual features (photograph vs. drawing) compared to high-level semantic features (animate vs. inanimate) of an item would reflect the speed at which these types of information become available in the brain. If so, we expected that reaction time patterns would reverse depending on whether the object is visually presented or reconstructed from memory: during visual perception, RTs should be faster for perceptual compared with semantic questions to mirror the forward processing hierarchy, while during retrieval RTs should be faster for semantic compared with perceptual questions if there is a reversal of that hierarchy.

Both experiments used a 2 x 2 mixed design (Fig. 1b and c), where all participants answered perceptual and semantic questions (factor question type, within-subjects) about the objects. Importantly, one group of participants was visually presented with the objects while answering these questions, whereas

the other group recalled the same objects from memory (factor task, between-subjects). The main difference between the two experiments was that in Experiment 1, both types of features were probed for a given object; and that in Experiment 2, objects were presented on background scenes (not of interest for the present purpose; see Methods section for details).

Overall accuracy in both experiments was near ceiling for the visual reaction time task (Experiment 1: M = 96.88%; SD = 2.40%; Experiment 2: M = 97.19%, SD = 2.99%), and high for the memory reaction time task (Experiment 1: 83.15%; SD = 0.92; Experiment 2: M = 66.23%, SD = 15.35). Note that Experiment 2 was more difficult because participants had to memorize background scenes in addition to the objects' semantic and perceptual features. In both experiments, only correct trials were used for all further RT analyses.



C. Chapter 3. Figure 1.

Figure 1. Stimuli and design of the behavioural experiments. (a) Illustration of the orthogonal design of the stimulus set. In all experiments, objects (a total of 128) varied along two dimensions: a perceptual dimension where objects could be presented as a photograph or as a line drawing; and a semantic dimension where objects could belong to the animate or inanimate category. (b) In the visual reaction time task, participants were prompted on each trial to categorize the upcoming object as fast as possible, either according to its perceptual category (photograph vs. line drawing) or its semantic category (animate vs. inanimate). (c) During the encoding phase of a memory reaction time task, participants were asked to create word-object associations (a total of 8 per block). Reaction times were then measured during the retrieval phase, where subjects were presented with a reminder word, and asked to recall and categorize the associated object according to its perceptual (photograph vs. line

drawing) or semantic (animate vs. inanimate) features. Button press symbols indicate at which moment in a trial RTs were collected.

2.1.2. Reaction times show the expected reversal in Experiments 1 and 2

To directly test for a reversal of the reaction time pattern between visual perception and memory reconstruction, we used generalized linear mixed-effect models (GLMM). GLMMs are ideally suited to model single trial data, like our RT data, without assumptions about the underlying distribution (e.g. normality), and they are able to capture variance explained both by fixed and random variables, including the experimental manipulations of interest (Lo & Andrews, 2015). For these analyses we used single trial RTs as target (dependent) variable. Our fixed effects were the kind of task (visual vs. memory task), question type (i.e. perceptual vs. semantic question) and the interaction between task and question type. Participant IDs and slopes were included as a random factor (including intercept).

Consistent with the reverse reconstruction hypothesis, we found that the interaction between task (visual vs. memory group) and question type (i.e. perceptual vs. semantic) significantly predicted RTs in both Experiment 1 ($F_{1, 9020}$ = 18.027, P < .001) and Experiment 2 ($F_{1, 3280}$ = 10.588, P = .001). In order to test whether the interaction was produced by differences in the expected direction (perceptual < semantic during encoding, and semantic < perceptual during retrieval), planned comparisons were then performed for the visual and memory task independently, with question type as the fixed effect. We found a

significant effect of question type in the visual task (Experiment 1: B = -.042, t = -3.973, P < .001; Experiment 2: B = -.048, t = -2.457, P = .014), where the negative sign of the coefficient indicates that the model indeed predicted lower RTs for perceptual compared to semantic questions. We also found a significant effect of question type in the memory task, but following the opposite pattern: positive coefficients now indicate significantly faster predicted RTs during semantic than perceptual questions (Experiment 1: B =.156, t = 2.551, P = .011; Experiment 2: B = .165, t = 2.523, P = .012).

For descriptive proposes, we also illustrate in Figure 2 the distribution of the participant-averaged RTs. During the visual task (Fig. 2A), participants on average were faster at answering perceptual (Experiment 1: M = 795ms; SD = 235ms; Experiment 2: M = 733ms; SD = 211ms) than semantic (Experiment 1: M = 842ms, SD = 185ms; Experiment 2: M = 797ms, SD = 235) questions. When performing the same task on objects reconstructed from memory, they were now on average slower responding to the perceptual (Experiment 1: M = 2502ms; SD = 561; Experiment 2: M = 3348ms, SD = 754) than the semantic (Experiment 1: 2334ms; SD = 534; Experiment 2: M = 3133ms, SD = 660ms) questions.

Reaction time analyses thus support our central hypothesis that the speed of information processing for different object features reverses between perception and memory, and this pattern fully replicated between Experiments 1 and 2.

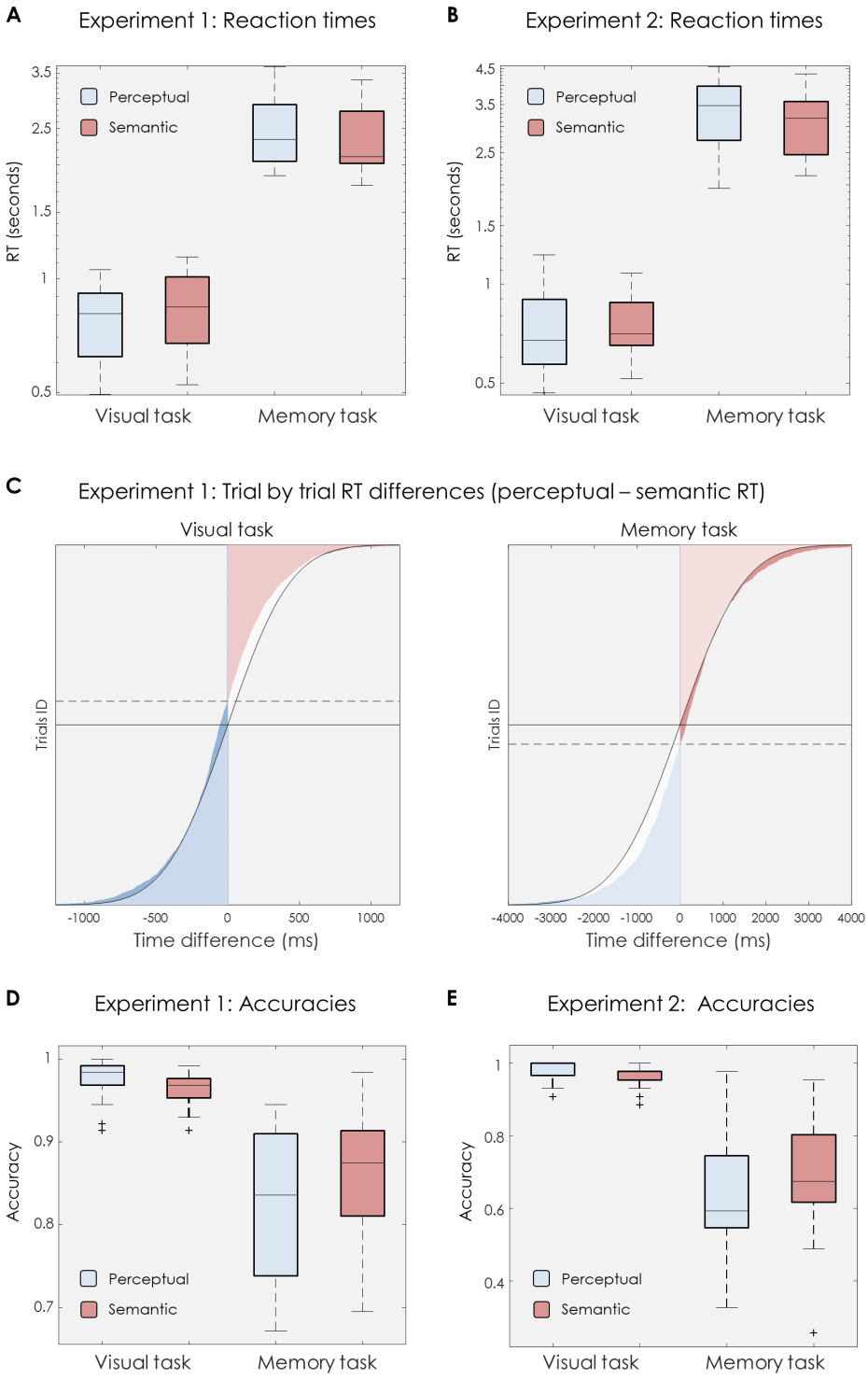
2.1.3. Accuracy results support a reversal between perception and memory

Next we investigated whether a similar pattern was, at least qualitatively, also present in terms of accuracy (Fig 2d and 2e). To perform these analyses we used a GLMM with a logistic link function and a binary probability distribution for our target variable (accuracy, which could be correct or incorrect on a given single trial). Fixed effects were the type of task (visual vs. memory task), question type (i.e. perceptual vs. semantic question), and the interaction between the two factors, the latter again being our effect of primary interest. Participant IDs and slopes were selected as a random factor, including intercept.

We found that in both experiments, the interaction between task (visual vs. memory group) and question type (perceptual vs. semantic question) significantly predicted participants' accuracy (Experiment 1: $F_{1, 11260} = 12.215$, $P < .001$; Experiment 2: $F_{1, 4124} = 8.383$, $P = .004$). When running planned comparisons separately for the visual and the memory task in Experiment 1, results for the visual task revealed that question type (perceptual or semantic) predicted participants' accuracy ($F_{1, 5886} = 5.066$, $P = .024$; $B = -.420$, $t = -2.251$, $P = .024$), suggesting that accuracy for perceptual questions (M = 97.42%; SD = 2.68%) was higher compared to semantic questions (M = 96.33%; SD = 1.99%;). In the memory task, question type also significantly predicted participants' accuracy ($F_{1, 5374} = 5.374$, $P = .001$; $B = .251$, $t = 3.222$, $P = .001$),

with negative model coefficients indicating that they were more likely to give a correct answer in response to semantic (M = 85.83%; SD = 7.57%) than perceptual (82.63%; SD = 8.79%) questions, in line with a reversed processing stream. Experiment 2 showed a similar trend in accuracy profiles. GLMM analyses for the visual task indicated that question type (perceptual or semantic) significantly predicted accuracy ($F_{1, 2062}$ = 4.371, P = .037; B = -.585, t = -2.091, P = .037), with participants showing better performance for perceptual (M = 97.97%; SD = 2.77%) than for semantic questions (M = 96.41%; SD = 3.07%). In contrast, for the memory task, we found evidence for the prioritization of higher-level information (semantic accuracy M = 69.57%; SD = 15.17%) over low-level details (perceptual accuracy M = 62.89%; SD = 15.09%). Here, question type also significantly predicted participants' accuracy in the expected direction ($F_{1, 2062}$ = 6.707, P = .010), with more accurate answers to semantic than perceptual questions (B = .319, t = 2.590, P = .010).

Altogether, the findings from our two behavioural experiments provide support for our main hypothesis that during retrieval of a complex visual representation, the temporal order in which perceptual and semantic features are processed reverses compared with the perception of the same object. The results suggest that reaction times can be used as a proxy to probe neural processing speed, as argued in previous studies (Ritchie, Tovar, & Carlson, 2015). In the next sections, we report the findings from an EEG study that more directly taps into the neural processes that we believe are producing the behavioural pattern.



D. Chapter 3. Figure 2.

Figure 2. Behavioural RT and accuracy results. (a) Box plots representing reaction times in Experiment 1 and Experiment 2 (b) for perceptual (blue) and semantic (pink) questions when an object was physically presented on the

screen (visual task, left) or cued by a reminder (memory task, right). We found that RTs were significantly predicted by an interaction between question type and kind of task ($P < .001$). For illustrative purposes the Y-axis in (a) and (b) is logarithmically scaled. (c) In Experiment 1, both types of questions were asked for each object representation. This allowed us to measure the difference in RTs between perceptual and semantic questions (X-axis) on a trial-by-trial level (Y-axis) during the visual task (left panel) and the memory task (right panel). Curved lines represent an expected normal distribution. The solid horizontal lines indicate the 50% point of the distribution (i.e., half of the trials), and dashed horizontal lines indicate the trial with a value closest to zero, where the perceptual-semantic difference is flipping from positive to negative. If differences were normally distributed, the solid and dashed lines would be on top of each other. (d) Accuracy results in Experiment 1 for perceptual (blue) and semantic questions (pink) when the object was presented on the screen (visual task) or had to be recalled (memory task). Behavioural analyses showed that an interaction between type of task (i.e. visual or memory) and question type (i.e. perceptual or semantic) significantly predicted accuracy. (e) Box plots representing accuracy in Experiment 2 during the visual and memory task, where the significant interaction effect between type of task and question type was replicated. In all box plots, the line in the middle of each box represents the median, and the tops and bottoms of the boxes the 25th and 75th percentiles of the samples, respectively. Whiskers are drawn from the interquartile ranges to the furthest minimum (bottom) and maximum (top) values. Crosses represent outliers.

2.2. EEG experiment

While it is reasonable to assume based on previous literature (Ritchie et al., 2015) that reaction times tap into the neural processing speed for a given feature, we also wanted to obtain a more direct signature of feature activation from human brain activity. We therefore used multivariate pattern analysis applied to electrophysiological (EEG) recordings, with the goal to pinpoint when in time, on an individual trial, the perceptual and semantic features of an object could be decoded from brain activity. We expected to find the maximum

decodability of perceptual information before semantic information when an object was visually presented on the screen, and expected the order of these peaks to reverse when the object was recalled from memory. The design closely followed the behavioural experiments, with the important difference that all factors were manipulated within subjects, such that each participant carried out a visual encoding phase that served to probe visual (forward) processing, and a subsequent recall phase used to probe mnemonic (backward) processing. The trial timing was optimised for obtaining a clean signal during object presentation and object recall, rather than for measuring reaction times (Fig. 3). We therefore presented the perceptual and semantic questions only during the recall phase in order to probe memory accuracy, and questions were presented at the end of each recall trial, such that they would not bias processing towards perceptual or semantic features of the object.

2.2.1 Accuracy in the EEG study replicates the response pattern found in the behavioural experiments

In the retrieval phase of the EEG experiment, subjects were again cued with a word and asked to retrieve the associated object. On average participants subjectively declared to retrieve the object on 93.6% of the trials (SD = 5.89%), with an average reaction time of 3046ms (SD = 830ms; minimum = 1369ms; maximum = 5124ms) to make this response. We then asked two objective questions at the end of each trial, one perceptual and one semantic, which participants answered with an overall mean accuracy of 86.37% (SD = 6.6).

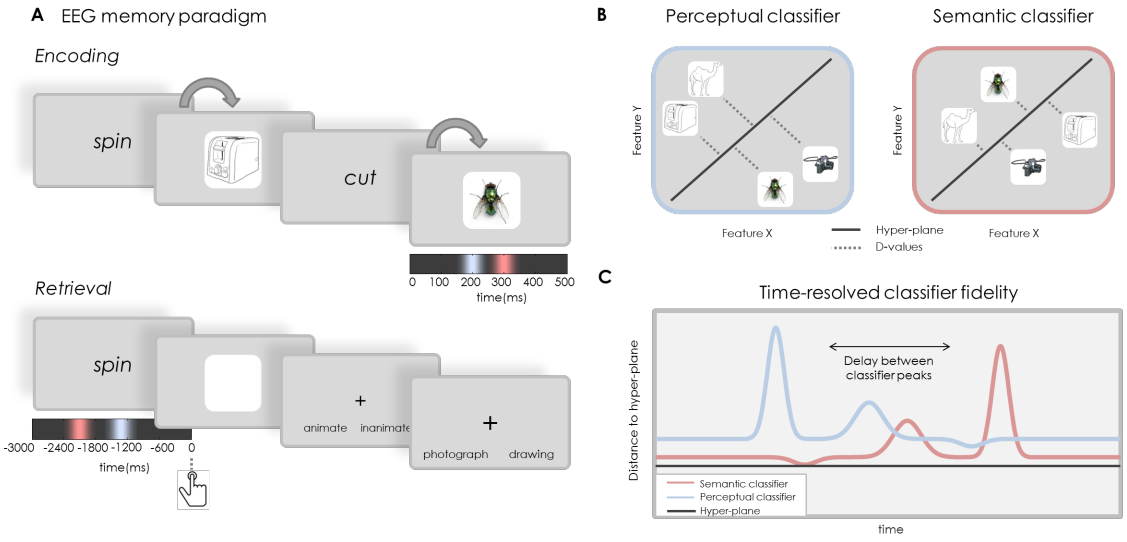
Mirroring our behavioural experiments, average hit rates were 87.65% (SD = 6.57%) when answering the semantic question, and 85.08% (SD = 6.53%) when answering the perceptual question. Within a GLMM, we found that the fixed factor question type significantly predicted accuracy ($F_{1, 5374}$= 7.706, P = .006), with perceptual questions showing a significantly lower hit rate than semantic questions (B = -.225, t = -2.776, P = .006). Note that the EEG task was not designed to measure reaction times, and participants were instructed to prioritize accuracy over speed.

2.2.2 Single-trial classifier fidelity suggests a reversal of the information processing cascade between perception and memory

In order to determine the temporal trajectory of feature processing on a single trial level, we carried out a series of time resolved decoding analyses. Linear discriminant analysis (LDA, see Method section) was used to classify perceptual (photograph vs. line drawing) and semantic (animate vs .inanimate) features of an object based on the EEG topography at a given time point, either during object presentation (encoding) or during object retrieval from memory (cued recall).

Our first aim was to confirm that there was a forward stream during perceptual object processing. Two separate classifiers were therefore trained and tested during encoding to classify the perceptual category (photograph vs. line drawing) and the semantic category (animate vs. inanimate) of the to-be-encoded object, respectively, in each trial and time point per participant (see

Fig. 3). For these analyses, decoding was performed in separate time windows starting 100ms before stimulus onset and up until 500ms post-stimulus. Our main interest was to determine the specific moment in each trial at which the two classifiers showed the highest fidelity in determining the correct perceptual and semantic categories (Fig. 3b and c). For the encoding data, we thus identified the highest *d* value peak per trial within 500ms of stimulus onset (see Methods section). This approach allowed us to compare, within each encoding trial, whether the classification peak for perceptual features occurred earlier than the classification peak for semantic features. Similarly, we used the cued recall time series to find the time points of maximum decoding performance of the perceptual and semantic classifiers during memory retrieval. All retrieval analyses are time-locked relative to the button press, i.e. the moment when participants declared that they had retrieved the associated object from memory. The time window used in this analysis covered 3sec prior to participants' responses, based on behavioural reaction times.



E. Chapter 3. Figure 3.

Figure 3. Design for EEG experiment and time resolved multivariate decoding. In the EEG experiment participants were asked to create word-object associations (panel A), and to later reconstruct the object as vividly as possible when cued with the word, and to indicate with a button press when they had a vivid image back in mind. EEG was recorded during learning and recall, with the aim to perform time-series decoding analyses that can detect at which moment, within a single trial, a classifier is most likely to categorise perceptual and semantic features correctly. Coloured time lines under object and cue time windows represent our reversal hypothesis regarding the temporal order of maximum semantic (pink) and perceptual (blue) classification during the perception (encoding) and retrieval of an object. All EEG analyses were aligned to the object onset during encoding, and to the button press during retrieval. (b) Decoding analyses were performed independently per participant at each time point. For each given time point during a trial, two linear discriminant analysis (LDA) based classifiers were trained on the EEG signal: one perceptual classifier discriminating photographs from line drawings, and one semantic classifier discriminating animate from inanimate objects. Classifiers were tested using a leave-one-out procedure, which allowed us to obtain a time series of confidence values (*d* values, reflecting the distance from the separation hyperplane) for each single trial. (c) Our main interest was to compare the time points of maximal fidelity of the perceptual (blue) and semantic classifiers (pink) on each trial, to test the hypothesis that the

perceptual maximum (blue) precedes the semantic one (pink) during perception, and importantly that this order is reversed during memory recall.

The first analysis of the single-trial peaks was very similar to the analysis conducted on reaction times in the behavioural studies. We again used a GLMM in order to test whether the relative timing of *d* value peaks from the perceptual and semantic classifiers reverses between encoding and retrieval. Like in the RT analyses, as fixed effects we included the type of classifier (perceptual or semantic), type of task (encoding or retrieval), and the interaction between both factors (type of classifier x type of task). Participant ID was included as random effect (with intercept) in our model. We found that the interaction between type of classifier and type of task significantly predicted the timing of the *d* value peaks ($F_{1,\ 5504} = 7.121$, $P = .003$). Planned comparisons between perceptual and semantic classifiers were then run separately for encoding and retrieval, with one fixed effect (type of classifier, perceptual or semantic) and including participant ID and slopes as random effects (with intercept). Type of classifier did not significantly predict the timing of *d* value peaks during encoding ($F_{1,\ 4326} = 0.328$, $P = .567$), but it did so during the retrieval task ($F_{1,\ 1180} = 3.879$, $P = .049$), with beta coefficients showing that the semantic peaks were predicted significantly earlier than the perceptual peaks ($B = 112.944$, $t = 1.969$, $P = .049$), as expected if there is a backward stream.

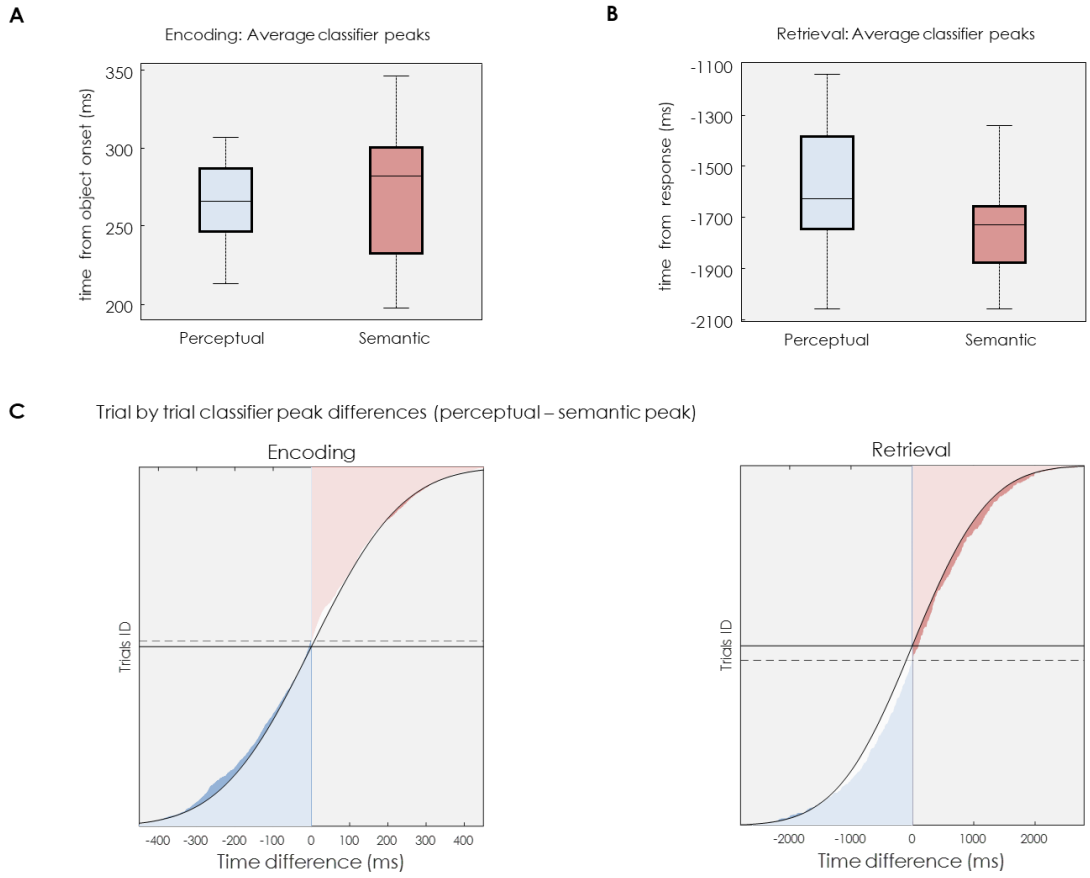
We followed up this GLMM result with an analysis specifically using the difference between each individual trial's semantic and perceptual classifier peak to test for their order relative to each other. At encoding, comparing the

pairwise difference of all single trial *d* value peaks against zero (Fig. 4c), we found a significant difference (T = -9.7642, P = .036) between the timing of perceptual and semantic peaks using a one-tailed clustered Wilcoxon signed rank test with random permutations (2000 repetitions; Jiang, Lee, & Rosner, 2017)). Fig. 4c shows that this difference was caused by a tendency of the single trial differences to be negative (learning towards the blue side), suggesting that confidence peaks for perceptual classification occurred before those for semantic classification. This result from the encoding phase of the experiment thus confirms previous studies showing that low-level features are processed before high-level features during visual perception (Carlson et al., 2013; Cichy et al., 2014; Clarke & Tyler, 2015; Lehky & Tanaka, 2016; Serre et al., 2007). The results also suggest that an analysis that takes into account the difference between the two paired classifier maxima from each single trial is more sensitive than our GLMM using the distributions of all single trials (which did not reveal a robust difference at encoding).

Importantly, following the same procedure, we next analysed the differences between the perceptual and semantic classifier peaks during memory reactivation, to test whether the order reversed during retrieval compared with encoding. The single-trial approach made sure that the relative temporal order of perceptual and semantic peaks within a trial would be preserved even if the retrieval process was set off with a varying delay across trials. Using a one-tailed clustered Wilcoxon signed rank test with random permutations (2000 repetitions; Jiang, Lee, & Rosner, 2017), a significant difference (T = 34.602, P

< .001) was found when we compared *d* value peak distributions of perceptual with those of semantic classification obtained from all single trials and participants (leaning towards the red side in Fig. 4c). Critically, the one-tailed test in this case confirms our central hypothesis that during memory retrieval, semantic information can be classified in brain activity significantly earlier than perceptual information, suggesting a reversal of information flow relative to perception.

Overall, the results again confirm our hypothesis that the information processing hierarchy reverses between perception (encoding) and recall, and that memory recall prioritizes semantic over perceptual information.



F. Chapter 3. Figure 4.

Figure 4. EEG multivariate analysis results. For illustrative proposes, box plots show group peak distribution of *d* values for perceptual and semantic categories during encoding (a; Perceptual peaks: M = 259, SD = 24; Semantic peaks: M = 267, SD = 43) and retrieval (b; Perceptual peaks: M = -1646, SD = 247; Semantic peaks: M = -1772, SD = 177) after averaging peaks within participants. All box plots elements represent the same metrics as in Figure 2. (c) Measuring classifier fidelity in terms of *d* value peaks on a single-trial level allowed us to measure the pairwise time distance between perceptual and semantic peaks during encoding (left panel) and retrieval (right panel). Y-axis represents each individual trial, with trials accumulated across participants. The time distance between classifier peaks (time of perceptual peak minus time of semantic peak on a given trial) is represented on the X-axis. The curved line represents an expected normal distribution. The solid horizontal line indicates the 50% point (half of the trials), and the dashed horizontal line indicates the point where the temporal distance values change sign from perceptual < semantic (blue) to semantic < perceptual (red).

2.2.3 Univariate ERP results are consistent with the reverse processing hypothesis

In a final step, we also sought to corroborate our classifier-based findings by more conventional event-related potential (ERP) analyses. If the differences in neural activity between perceptual (photograph vs. line drawing) and semantic (animate vs. inanimate) categories, as picked up by the LDA classifier, were produced by a signal that is relatively stable across trials and participants, these signal differences would also be visible in the average ERP time courses across participants. A comparison of the ERP peaks during encoding and retrieval would then reveal the same perception-to-memory reversal as found in our multivariate analyses.

Firstly, a series of cluster-based permutation tests (see Methods section) was performed during object presentation to test for ERP differences between perceptual and semantic categories. Contrasting objects from the two different perceptual categories (photographs and line drawings), we obtained a significant positive cluster (P_{corr} = .008) between 136ms and 232ms after stimulus onset, with a maximum difference based on the sum of T values at 188ms, and located over occipital and central electrodes (see Fig. 5a). Contrasting objects from the different semantic categories (animate and inanimate) revealed a later cluster over frontal and occipital electrodes (P_{corr} = .001) from 237ms until 357ms after stimulus presentation, with a maximum difference at 306ms (see Fig. 5a). The peak semantic ERP difference for

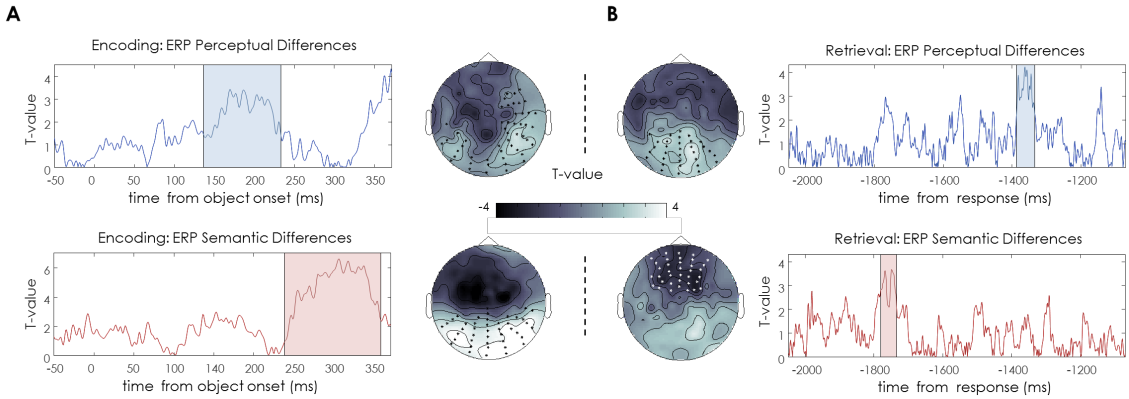
encoding thus occurred ~120ms after the peak perceptual difference, consistent with the existing ERP literature (Fabiani, M., Gratton, G., & Federmeier, 2007) . Similar contrasts between perceptual and semantic categories were then carried out during retrieval, again aligning trials to the time of the button press. We found a significant perceptual cluster distinguishing the recall of photographs and line drawings over occipital electrodes (P_{corr} = .046) between 1390ms and 1336ms before participants' responses, with a maximum difference based on the sum of T values at 1360ms prior to response time (see Fig. 5b). Comparing ERPs for the different semantic categories, we found a significant cluster distinguishing the recall of animate from inanimate objects over frontal electrodes (P_{corr} = .032) between 1781ms and 1735ms before object retrieval, with a maximum difference at -1770ms (see Fig. 5b). Therefore, during memory retrieval, the peak semantic ERP difference occurred ~400ms before the peak perceptual difference. Note that the timing of the effects also coincides with the timing of the classifier results in terms of the maximum differences between perceptual and semantic categories (see Fig. 4). Qualitatively, the ERP results thus mirror the results of our previous multivariate analyses in terms of the timing of the maximum signal difference between categories.

An additional analysis was carried out in order to statistically test for an interaction on the ERP level between type of task (encoding vs. retrieval) and representational features (perceptual vs. semantic). In each participant, we identified the time point of the maximum difference in each of our four

comparisons of interest (i.e. photographs vs. drawings during encoding and during retrieval; and animate vs. inanimate objects during encoding and during retrieval). These time-points of maximum difference were tested for an interaction in a 2x2 within-subjects ANOVA. We found a significant interaction between type of task (encoding vs. retrieval) and type of feature comparison (perceptual vs. semantic) ($F_{1, 42}$ = 7.798, P = .011).

Our final follow-up was aimed at probing if these ERP differences are driven by a specific combination of perceptual and semantic features. For example, it is possible that the cluster showing a semantic difference (animate vs. inanimate) is driven only by a difference in photographs but not line drawings. For each of the four clusters identified in the above ERP analysis, we therefore ran a 2x2 within-subjects ANOVA, averaging the signal separately for the four types of sub-categories (animate-photographs, animate-line drawings, inanimate-photographs, inanimate-line drawings). The ERPs for this analysis are illustrated in Supplementary Figure 1. During encoding, we did not find a significant interaction between semantic and perceptual categories in any cluster (perceptual cluster: $F_{1, 23}$ = 1.106, P = .304; semantic cluster: $F_{1, 23}$ = .640, P = .432). Similarly, interaction effects were absent during retrieval (perceptual cluster: $F_{1, 20}$ = 2.125, P = .160; semantic cluster: $F_{1, 20}$ = .403, P = .533). We therefore found no evidence indicating that our main ERP clusters showing perceptual and semantic differences at encoding and retrieval were produced by a selective difference in one of the sub-categories that constitute the orthogonal dimension.

Altogether, the ERP results confirm that perceptual aspects are coded in brain activity earlier than semantic aspects during visual processing, but semantic differences dominate the EEG signal earlier than perceptual ones during retrieval.



G. Chapter 3. Figure 5.

Figure 5. Univariate analysis results. (a) Left panels represent ERP group differences (T values) across time in those electrodes that formed a significant cluster during object presentation, locked to the onset of the stimulus. Top left panel shows the contrast of photographs vs. line drawings, and the bottom left panel differences between animate vs. inanimate objects. Scalp figures next to each contrast illustrate the maximum cluster's topography, averaged across the significant time-window, with all significant electrodes in a cluster being marked with an asterisk. (b) Right panels show ERP group differences (T values) over time in those electrodes that are contained in the maximum significant clusters during memory retrieval, time locked to participants' responses). The top right panel shows the perceptual contrast, and the bottom right panel the semantic contrast. Cluster topographies for each comparison are located next to each panel, and the temporal extent of significant clusters is shaded in colour.

3. Discussion

How does the neural fingerprint of a memory unfold in time when triggered by a reminder? While it is widely accepted that visual object recognition starts with low-level perceptual followed by high-level abstract processing (Carlson et al., 2013; Cichy et al., 2014; Lehky & Tanaka, 2016; Serre et al., 2007), much less is known about the mnemonic feature processing cascade. Here we demonstrate that the reconstruction of a visual memory does depend on a hierarchical stream too, but this mnemonic stream follows the reverse order relative to visual processing. Across three experiments, we found highly converging evidence in favour of such a reversal from behavioural reaction times and accuracy (Experiments 1 and 2), from multivariate classification analyses, and from univariate ERP analyses (Experiment 3).

The behavioural studies demonstrate that participants were significantly faster at detecting low-level perceptual differences than abstract, conceptual differences during a visual classification task, while the object was presented on the screen. Critically, however, when probing the perceptual and semantic components of objects recalled from memory, the reverse effect was found: subjects required significantly less time to correctly retrieve semantic information about the object compared to perceptual details (see Fig. 2a and 2b). This reversal was corroborated by a significant interaction between the kind of feature (perceptual or semantic) and the kind of task (visual perception or memory recall task). Based on signal-detection models (Ashby, 2000; O'Connell, Dockree, & Kelly, 2012), the RT findings suggest that during memory reconstruction, the decision threshold to identify abstract information of a

mnemonic representation is reached before a judgment about low-level information can be made. The response latency pattern therefore supports our central hypothesis that the temporal order in which features come online is reversed when retrieving a previously stored representation of an object, relative to its perception. In addition to reaction times, the same reversal pattern was present in accuracy profiles in both experiments, with significantly higher accuracy for perceptual than semantic questions during the visual task, but higher accuracy for semantic than perceptual questions during the memory task (see Fig. 2d). These findings suggest a prioritization of abstract semantic information over perceptual details of a mnemonic representation, a finding consistent with hierarchical memory system models (Henson & Gagnepain, 2010).

The results from our third, EEG experiment fully support the conclusions drawn from the behavioural studies. We used temporally resolved multivariate decoding analyses to observe when in time, during object perception and object retrieval, the perceptual and semantic features of an object would be maximally decodable from a participant's brain activity patterns. These analyses were carried out on a single trial level such that the fidelity peaks of the perceptual and semantic classifiers could be directly compared. When an object was visually presented during encoding, the maximum fidelity (*d* value) in classifying perceptual information (photograph vs. line drawings) occurred approximately 100ms earlier than the maximum for semantic information (animate vs. inanimate) (see Fig. 4a). This finding is consistent with a predominantly feed-

forward processing stream as described previously (Carlson et al., 2013; Cichy et al., 2014; Clarke & Tyler, 2015; Lehky & Tanaka, 2016; Serre et al., 2007). Note that perceptual and semantic peaks during visual perception only differed statistically when comparing their relative timing on a single trial level (i.e., the Wilcoxon signed rank test), suggesting that such an analysis is more sensitive to detecting relatively small timing differences in noisy data. When we asked participants to reactivate an object's representation from memory, peaks in classifying semantic information were found roughly 300ms before the peaks for perceptual categories (see Fig. 4b). Like in the behavioural experiments, a consistent reversal between perception and memory was supported by a significant interaction between the type of feature that was probed (perceptual or semantic), and the type of task participants were engaged in (encoding or retrieval). Finally, we also found the same reversal pattern in the ERP peaks when comparing the maximum ERP difference between perceptual and semantic object classes. During object perception, the largest perceptual ERP cluster occurred ~100ms before the semantic ERP cluster, whereas during retrieval the perceptual cluster followed the semantic one with a lag of about 400ms (see Fig. 5). In summary, our two behavioural experiments, together with the decoding results and the ERP analyses, provide robust evidence for our main prediction that semantic features are prioritized over perceptual features during memory recall, in the opposite direction of the well-known forward stream of visual-perceptual processing. Follow-up studies will need to test whether this reversed stream is robust under different conditions, for example in tasks that explicitly vary the encoding demands to emphasize perceptual over semantic

aspects of an event. If semantic information is always prioritized, this would suggest a hardwired characteristic of the output pathways from the hippocampus back to neocortex. Alternatively, and maybe more likely, the retrieved representation will to some degree also depend on what Marr (1971) called the “internal description” of a stimulus during encoding, including the rememberer’s goals and attentional state.

In our studies, the behavioural data were acquired separately from the EEG data, in a setting that was optimized for measuring reaction times. Previous studies simultaneously measuring RTs and neural activity suggest that a meaningful relationship exists, on a single trial level, between the *d* values resulting from EEG classification and human behaviour. In line with signal detection models (Ashby, 2000; O’Connell et al., 2012), it has been argued that the distance between two or more categories in a neural representational space can serve as a decision boundary that guides behavioural categorization (Ritchie et al., 2015). For example, Carlson et al. (Carlson, Ritchie, Kriegeskorte, Durvasula, & Ma, 2014) used fMRI-based activation patterns in late visual brain regions during an object recognition task, where participants had to make animacy judgements, similar to our semantic task. They found that the faster the reaction time on a given trial, the further away in neural space the object was represented relative to the boundary between semantic categories. Similarly, an MEG study (Ritchie et al., 2015) showed that the decision values during the time points of maximum decodability, derived in a way similar to our EEG study, were strongly correlated with reaction times for visual categorization. Both studies thus suggest that during object vision, single-trial

decoding measures reflect a distance between categories in a neural space that directly translates into behaviour. Even though we did not obtain reaction times during the same trials that were used for EEG decoding, our findings indicate that this meaningful brain-behaviour relationship extends to mental object representations during memory reconstruction.

How does the reverse reconstruction hypothesis fit with existing knowledge about the neural pathways involved in memory reconstruction? It is generally accepted that during memory formation, information flows from domain-specific sensory modules via perirhinal and entorhinal cortices into the hippocampus. Recent evidence suggests that during visual processing, the coding of perceptual object information is preserved up to relatively late perirhinal processing stages (Martin, Douglas, Newsome, Man & Barense, 2018). The hippocampus is considered a domain-general structure (Howard Eichenbaum, 2004; Moscovitch, 2008; Staresina & Davachi, 2008) whose major role is the associative binding of the various elements that constitute an episode (Davachi, 2006; H. Eichenbaum, Yonelinas, & Ranganath, 2007; Squire, Stark, & Clark, 2004). The hippocampal code later allows a partial cue to trigger the reconstruction of these different elements from memory. This memory reconstruction process is thought to depend on back-projections from the hippocampus to neocortical areas, causing the reactivation of memory patterns in at least a subset of the areas that were involved in perceiving the original event. Such reactivation has consistently been reported in higher-order sensory regions related to processing of complex stimulus and task information

(Johnson et al., 2009; B. A. Kuhl et al., 2011; Michelmann et al., 2016; Wimber et al., 2015), but also in relatively early sensory cortex (Bosch, Jehee, Fernandez, & Doeller, 2014; Waldhauser et al, 2016), suggesting that in principle higher- and lower-level information can be reconstructed from memory. Interestingly, however, recent evidence suggests that the structure of complex naturalistic events (movies) is transformed from low-level perceptual to memory codes (Chen et al., 2017). Our work suggests that higher-order meaningful information is prioritized over lower-level details during retrieval.

While the reverse reconstruction hypothesis is neurobiologically plausible and has strong intuitive appeal, direct empirical evidence so far has been lacking. Indirect evidence comes from an fMRI study showing that within the medial temporal lobe, regions that are involved in the processing of objects and scenes are also activated when retrieving objects and scenes from memory, but with a delay relative to the actual perception of objects and scenes, consistent with a reversed information flow (Staresina, Cooper, & Henson, 2013). Intracranial EEG recordings have shown that connectivity between the entorhinal cortex and the hippocampus changes directionality between encoding and retrieval (Fell et al., 2016), which could provide the functional basis for cortical reinstatement. Studies in rodents indicate that the hippocampus is in principle capable of replaying the neural code that represent a certain spatial memory in reverse order, in particular when the animal is awake and resting suggesting a potential role of reverse replay in active memory retrieval (Carr, Jadhav, & Frank, 2011). Finally, there is work using MEG decoding suggesting that it is

mainly the later processing stages of the perceptual stream that are reactivated during retrieval as well as during mental imagery, consistent with a prioritization of higher-level information (Dijkstra, Mostert, Lange, Bosch, & van Gerven, 2018; Kurth-Nelson et al., 2015). Our proposal of a reverse processing hierarchy is thus plausible based on functional anatomy and the existing literature, even though it has never been explicitly proposed or tested so far.

We regard our reverse reconstruction hypothesis as complementary to existing models that address the nature and timing of different retrieval processes, including the influential dual process model (for a review see Yonelinas, Aly, Wang, & Koen, 2010). Dual process models focus on recognition rather than recall tasks, and on the cognitive processes and operations required to access a stored memory rather than the reactivated features of a memory themselves. They assume that successful recognition of a previously stored stimulus can be based on a sense of familiarity, or on the additional recollection of contextual information associated with the stimulus during encoding, an influential idea in the memory field since the introspective analyses of William James (James, 1890). While the original model does not explicitly address the time course of these processes, there is evidence, based on the EEG literature, suggesting that familiarity signals occur earlier than recollection signals. Familiarity signals can be detected in the EEG as early as 300ms after the onset of a recognition probe, while recollection-related activity typically begins to emerge after 500-600ms (Bridson, Fraser, Herron, & Wilding, 2006; Klimesch et al., 2001; Mecklinger, 2006; Rugg & Curran, 2007). In contrast to the above-mentioned

studies, our studies probed memory via cued recall, where successful recall strongly depends on the recollection of associative information. Our results suggest that within this recollection process, the semantic “gist” of a memory is accessed before perceptual details. Assuming that familiarity signals reflect a more gist-like and less detailed stage of the retrieval process than recollection signals (an assumption that some find controversial, see Nyhus & Curran, 2009), the hierarchical progression from an early global semantic signal to more fine-grained recollection might thus be a fundamental principle of retrieval that is shared between recall and recognition memory.

Beyond specific models of declarative memory, there are also interesting parallels between our findings and visual learning phenomena like the Eureka effect (Ahissar & Hochstein, 1997). The general idea that perception is shaped by stored representations has been proposed over a century ago by von Helmholtz (Helmholtz, 1924). A wealth of findings now support the idea that previous exposures to a stimulus can exert a strong top-down influence on its subsequent perception (for a review; Aggelopoulos, 2015). Reminiscent of our present findings, Ahissar and Hochstein (2004) suggest that such visual learning is a top-down process that progresses from high-level to low-level visual areas with increasing practice. Specifically, they argue that improvements in visual discrimination tasks (e.g. identifying a tilted line among distractors) are guided by high-level information (e.g. “the gist of the scene”) during earlier stages of learning, and increasingly more by low-level information (e.g. line orientations or colours) at later stages. Our findings indicate that during the

reactivation of an object's stored representation, its high-level features are retrieved more rapidly than its low-level components. Abstract information might thus be reactivated more easily and during earlier stages of visual learning, and thus have a stronger driving influence on performance than more detailed information. Even though speculative at the moment, our reverse reconstruction framework might thus have explanatory value for findings in related fields of learning and memory.

How our brain brings back to mind past events, and enriches our mental life with vivid images or sounds or scents beyond the current external stimulation, is still a fascinating and poorly understood phenomenon. Our present results suggest that memories, once they are triggered by a reminder, unfold in a systematic and hierarchical way, and that the mnemonic processing hierarchy is reversed with respect to the major visual processing hierarchy. We hope that these findings can inspire more dynamic frameworks of memory retrieval that explicitly acknowledge the reconstructive nature of the process, rather than simply conceptualizing memories as reactivated snapshots of past events. Such models will help us understand the heuristics and systematic biases that are inherent in our memories and memory-guided behaviours.

4. Methods

4.1. Participants

A total of 49 volunteers (39 female; mean age 20.02 +/- 1.55 years old) took part in behavioural Experiment 1. Twenty-six of them (19 female; mean age 20.62 +/- 1.62 years old) participated in the memory reaction time task. Five out of these 26 participants were not included in the final analysis due to poor memory performance (<66% general accuracy) compared with the rest of the group ($t_{24} = 6.65$, $p < 0.01$). Another group of 23 participants (20 female; mean age 19.35 ± 1.11 years) volunteered to participate in the visual reaction time task. In a second behavioural experiment (Experiment 2), 48 participants were recruited (42 female; mean age 19.25 +/- 0.91 years). Twenty-four of them performed the memory reaction time task and another group of 24 took part in the visual reaction time task. For the electrophysiological experiment we recruited a total of 24 volunteers (20 female; mean age 21.91 ± 4.68 years). Since the first 3 subjects we recorded performed a slightly different task during retrieval blocks (i.e., they were not asked to mentally visualise the object for 3 seconds, and they had to answer only one of the perceptual and semantic questions per trial), we did not include these participants in any of the retrieval analyses. Since our paradigm was designed to test for a new effect, we did not have priors regarding the expected effect size. Behavioural piloting of the memory task showed a significant difference in reaction times in a sample of n = 14. We therefore felt confident that the effect would replicate in our larger

samples of n = 24 per group in each in the two behavioural experiments and the EEG experiment.

All participants reported being native or highly fluent English speakers, having normal (20/20) or corrected-to-normal vision, normal colour vision, and no history of neurological disorders. We received written informed consent from all participants before the beginning of the experiment. They were naïve as to the goals of the experiments, but were debriefed at the end. Participants were compensated for their time, receiving course credits or £6 per hour for participation in the behavioural task, or a total of £20 for participation in the electrophysiological experiment. The University of Birmingham's Science, Technology, Engineering and Mathematics Ethical Review Committee approved all experiments.

4.2. Stimuli

In total, 128 pictures of unique everyday objects and common animals were used in the main experiment, and a further 16 were used for practice purposes. Out of these, 96 were selected from the BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010), and the remaining images were obtained from online royalty-free databases. All original images were pictures in colour on a white background. To produce two different semantic object categories, half of the objects were chosen to be animate while the other half was inanimate. Within the category of inanimate objects, we selected the same

amount of electronic devices, clothes, fruits and vegetables (16 each). The animate category was composed of an equivalent number of mammals, birds, insects and marine animals (16 each). With the objective of creating two levels of perceptual manipulation, a freehand line drawing of each image was created using the free and open source GNU image manipulation software (www.gimp.org). Hence a total of 128 freehand drawings of the respective 128 pictures of everyday objects were created. Each drawing was composed of a white background and black lines to generate a schematic outline of each stimulus. For each subject, half of the objects were pseudo-randomly chose to be presented as photographs, and half of them as drawings, with the restriction that the two perceptual categories were equally distributed across (i.e. orthogonal with respect to) the animate and inanimate object categories. All photographs and line drawings were presented at the centre of the screen with a rescaled size of 500 x 500 pixels. For the memory reaction time task and the EEG experiment, 128 action verbs were selected that served as associative cues. Experiment 2 also used colour background scenes of indoor and outdoor spaces (900 x 1600 pixels) that were obtained from online royalty-free databases, which are irrelevant for the present purpose.

4.3. Procedure

4.3.1. Behavioural experiments

4.3.1.1. Experiment 1

Visual reaction time task

Before the start of the experiment, participants were given oral instructions and completed a training block of 4 trials to become familiar with the task. The main perceptual task consisted of 4 blocks of 32 trials each (Fig.1b). All trials started with a jittered fixation cross (500 to 1500ms) that was followed by a question screen. On each trial, the question could either be a perceptual question asking the participant to decide as quickly as possible whether the upcoming object is shown as a colour photograph or as a line drawing; or a semantic question asking whether the upcoming object represents an animate or inanimate object. Two possible response options were displayed at the two opposite sides of the screen (right or left). The options for “animate” and “photograph” were always located on the right side to keep the response mapping easy. The question screen was displayed for 3 seconds, and an object was then added at the centre of the screen. In Experiment 2, this object was overlaid onto a background that filled large parts of the screen. Participants were asked to categorize the object in line with the question as fast as they could as soon as the object appeared on the screen, by pressing the left or right arrow on the

keyboard. Reaction times (RTs) were measured to test if participants were faster at making perceptual compared to semantic decisions.

All pictures were presented until the participant made a response but for a maximum of 10 sec, after which the next trial started. Feedback about participants' performance was presented at the end of each experimental block. There were 256 trials overall, with each object being presented twice across the experiment, once together with a perceptual and once with a semantic question. Repetitions of the same object were separated by a minimum distance of 2 intervening trials. In each block, we asked the semantic question first for half of the objects, and the perceptual question first for the other half.

The final reaction time analyses only included trials with correct responses, and excluded all trials with an RT that exceeded the average over subjects by +- 2.5 standard deviations (SDs).

Memory reaction time tasks

The memory version was kept very similar to the visual reaction time task, but we now measured RTs for objects that were reconstructed from memory rather than being presented on the screen, and we thus had to introduce a learning phase first. At the beginning of the session, all participants received instructions and performed two short practice blocks. Each of the overall 16 experimental blocks consisted of an associative learning phase (8 word-object associations)

and a retrieval phase (16 trials, testing each object twice, once with a perceptual and once with a semantic question). The associative learning and the retrieval test were separated by a distractor task. During the learning phase (Fig. 1c), each trial started with a jittered fixation cross (between 500 and 1500ms) that was followed by a unique action verb displayed on the screen (1500ms). After presentation of another fixation cross (between 500 and 1500ms), a picture of an object was presented on the centre of the screen for a minimum of 2 and a maximum of 10 seconds. Participants were asked to come up with a vivid mental image that involved the object and the action verb presented in the current trial. They were instructed to press a key (up arrow on the keyboard) as soon as they had a clear association in mind; this button press initiated the onset of the next trial. Participants were made aware during the initial practice that they would later be asked about the object's perceptual properties as well as its meaning, and should thus pay attention to details including colour and shape. Within a participant, each semantic category and sub-category (electronic devices, clothes, fruits, vegetables, mammals, birds, insects, and marine animals) was presented equally often at each type of perceptual level (i.e. as a photograph or as a line drawing). The assignment of action verbs to objects for associative learning was random, and the occurrence of the semantic and perceptual object categories was equally distributed over the first and the second half of the experiment in order to avoid random sequences with overly strong clustering.

After each learning phase, participants performed a distractor task where they were asked to classify a random number (between 1 and 99) on the screen as odd or even. The task was self-paced and they were instructed to accomplish as many trials as they could in 45 seconds. At the end of the distractor task, they received feedback about their accuracy (i.e., how many trials they performed correctly in this block).

The retrieval phase (Fig. 1c) started following the distractor task. Each trial began with a jittered fixation cross (between 500 and 1500ms), followed by a question screen asking either about the semantic (animate vs. inanimate) or perceptual (photograph vs. line drawing) features for the upcoming trial, just like in the visual perception version of the task. The question screen was displayed for 3 seconds by itself, and then one of the verbs presented in the directly preceding learning phase appeared above the two responses. We asked participants to bring back to mind the object that had been associated with this word and to answer the question as fast as possible by selecting the correct response alternative (left or right keyboard press). If they were unable to retrieve the object, participants were asked to press the down arrow. The next trial began as soon as an answer was selected. At the end of each retrieval block, a feedback screen showing the percentage of accurate responses was displayed.

Throughout the retrieval test, we probed memory for all word-object associations learned in the immediately preceding encoding phase in

pseudorandom order. Each word-object association was tested twice, once together with a semantic and once with a perceptual question, with a minimum distance of 2 intervening trials. In addition, we controlled that the first question for half of the associations was semantic, and perceptual for the other half. Like in the visual RT task, the response options for “animate” and “photograph” responses were always located on the right side of the screen. In total, including instructions, a practice block and the 16 learning-distractor-retrieval blocks, the experiment took approximately 60 minutes.

For RT analyses we only used correct trials, and excluded all trials with an RT that exceeded the average over subjects by ± 2.5 SDs.

4.3.1.1. Experiment 2

Experiment 2 was very similar in design and procedures to Experiment 1, and we therefore only describe the differences between the two experiments in the following.

Visual reaction time task

The second experiment started with a familiarisation phase where all objects were presented sequentially. In each trial of this phase, a jittered fixation cross (between 500 and 1500ms) was followed by one screen that showed the photograph and line drawing version of one object simultaneously, next to each

other. During the presentation of this screen (2.5 sec) participants were asked to overtly name the object. After a jittered fixation cross (between 500 and 1500ms), the name of the object was presented.

After this familiarisation phase, the experiment followed the same procedures as the visual reaction time task in Experiment 1 except for the following changes. Objects were overlaid onto a coloured background scene (1600 x 900 pixels). Also, each object (286 x 286 pixels) was probed only once, either together with a perceptual question, a semantic question (like above), or a contextual question asking whether the background scene was indoor or outdoor. For the current purpose we only describe the RTs to object-related questions in the Results section. Another minor difference to Experiment 1 was that in this version of the task, the question screen was displayed for 4sec, and the two options to answer during stimulus presentation were removed from the screen as soon as the object/reminder appeared.

Memory reaction time task

The memory reaction time task in Experiment 2 also included, during the associative learning phase, a background scene (1600 x 900 pixels) that was shown on the screen behind each object (286 x 286 pixels), and participants were asked to remember the word-background-object combination. In this version of the task, each word-object association was tested only once, together with either a perceptual question about the object, a semantic question

about the object, or a contextual question regarding the background scene (indoor or outdoor). Therefore, one third of the objects were tested with a semantic question, one third with a perceptual question, and one third with a contextual question. Again, context was not further taken into account in the present analyses.

4.3.2. EEG experiment (Experiment 3)

Following the EEG set-up, instructions were given to participants and two blocks of practice were completed. The task procedure of the EEG experiment was similar to the memory task in Experiments 1 and 2 except for the retrieval phase (Fig. 3a). Each block started with a learning phase where participants created associations between overall 8 action verbs and objects. After a 40 sec distractor task, participants' memory for these associations was tested in a cued recall test. In total, the experiment was composed of 16 blocks of 8 associations each.

Each trial of the retrieval test started with a jittered fixation cross (500-1500ms), followed by the presentation of one of the action verbs presented during the learning phase as a reminder. Participants were asked to visualize the object associated with this action verb as vividly and in as much detail as possible while the cue was on the screen. To capture the moment of retrieval, participants were asked to press the up-arrow key as soon as they had the object back in mind; or the down-arrow if they could not remember the object.

This reminder was presented on the screen for a minimum of 2 sec and until a response was made (maximum 7 sec). Immediately afterwards, a blank square with the same size as the original image was displayed for 3 sec. During this time, participants were asked to “mentally visualize the originally associated object on the blank square space”. After a short interval where only the fixation cross was present (500-1500ms), a question screen was displayed for 10 seconds or until participant response asking about perceptual (photograph vs. line drawing) or semantic (animate vs. inanimate) features of the retrieved representation, like in the behavioural tasks. However, in this case both types of questions were always asked on the same trial, and they were asked at the end of the trial rather than before the appearance of the reminder. The first question was semantic in half of the trials, and perceptual in the other half. Therefore, each retrieval phase consisted of 8 trials where we tested all verb-object associations learned in the same block in random order.

4.4. Data Collection (behavioural and EEG)

Behavioural response recording and stimulus presentation were performed using Psychophysics Toolbox Version 3 (Brainard, 1997) running under MATLAB 2014b (MathWorks). For response inputs we used a computer keyboard where directional arrows were selected as response buttons.

Electroencephalography (EEG) data was acquired using a BioSemi Active-Two amplifier with 128 sintered Ag/AgCl active electrodes. Through a second

computer the signal was recorded at a 1024 Hz sampling rate by means of the ActiView recording software (BioSemi, Amsterdam, Netherlands). For all three experiments it was not possible for the experimenters to be blind to the conditions during data collection and analysis.

4.5. GLMM analyses

Generalized linear mixed models (GLMMs) were used to test our alternative hypotheses for accuracy (all experiments), reaction times (Experiments 1 and 2), and the relative timing of EEG classifier fidelity (*d* value) peaks (Experiment 3). We chose GLMMs instead of more commonly used GLM-based models (i.e., ANOVAs or t-tests) because they make fewer assumptions about the distribution of the data, are better suited to model RT-like data (REF) including our d-value peaks, and can accurately model proportional data that are bound between 0 and 1 (like memory accuracy). Our conditions of interest were modelled as fixed effects in the GLMM. Unless otherwise mentioned, these were the type of task (visual perception vs memory retrieval) and the type of feature probed (perceptual vs semantic). Our central reverse processing hypothesis was tested by an interaction contrast between the factors type of task and question type. Two further planned comparisons were then conducted to test if an interaction was driven by effects in the expected direction (e.g., reaction times perceptual < semantic during visual perception, and semantic < perceptual during memory retrieval). For all analyses, participant ID (including intercept) was modelled as a random factor. Wherever possible, we also

included slope as a random factor because GLMMs that do not take into account this factor tend to overestimate effects (that is, they are overly liberal; Barr, Levy, Scheepers, & Tily, 2013). In all cases, we used a compound symmetry structure based on theoretical assumptions and AIC and BIC values. We would like to emphasize that all of the effects reported as significant in the results section remain significant (with a tendency for even stronger effects) when excluding the random factor slope, but we chose to report the results from the more conservative analysis.

Due to the data structure (specifically, the Hessian matrix not being positive definite), slope as a random effect could not be modelled in 2 of the analyses in Experiment 3: (i) when analysing the interaction between type of task and type of classifier as predictive factor for EEG classifier peaks; and (ii) when testing behavioural accuracy. In these two cases, the results are reported for GLMMs that do not include slope as a random factor. For the interaction analysis in (i), we also had to apply a linear transformation to the data, because the d-values during encoding and retrieval (which are compared directly in the interaction contrast) differed too much in scale. Data was thus z-scored to avoid errors calculating the Hessian matrix, and a constant value of 1000ms was added to each value to avoid negative values in our target variable.

For all accuracy analyses we used a binomial distribution with a logistic link function. All models for analysing RTs and *d* value peaks used a gamma probability distribution and an identity link function. The choice of a gamma

distribution was justified because in all cases it fit our single trial distributions better than alternative models, for example inverse Gaussian or normal distributions (evidence from AIC and BIC available on request).

4.6. Clustered Wilcoxon signed rank test

To compare the pairwise differences between perceptual and semantic *d* value peaks in each encoding or retrieval trial (Experiment 3), and test whether the median of these differences deviates from zero in the expected direction (that is, perceptual < semantic during encoding, and semantic < perceptual during retrieval), we used a one-tailed Wilcoxon signed rank test that clustered the data per participant, using random permutations (2000 repetitions). This analysis was run using the R package “clusrank” (Jiang, Lee, & Rosner, 2017).

4.7. EEG Pre-processing

EEG data was pre-processed using the Fieldtrip toolbox (version from 3rd, August, 2017) for Matlab (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data recorded during the associative learning (encoding) phase was epoched into trials starting 500ms before stimulus onset and lasting until 1500ms after stimulus offset. The resulting signal was baseline corrected based on pre-stimulus signal (-500ms to onset). Retrieval epochs contained segments from 4000ms before until 500ms post-response. Since the post-response signal during retrieval will likely still contain task-relevant (i.e., object specific)

information, we baseline-corrected the signal based on the whole trial. Both datasets were filtered using a low-pass filter at 100 Hz and a high-pass filter at 0.1 Hz. To reduce line noise at 50 Hz we band-stop filtered the signal between 48 and 52 Hz. The signal was then visually inspected and all epochs that contained coarse artefacts were removed. As a result, a minimum of 92 and a maximum of 124 trials remained per participant for the encoding phase, and a range between 80 and 120 trials per subject remained for retrieval. Independent component analysis was then used to remove eye-blink and horizontal eye movement artefacts; this was followed by an interpolation of noisy channels. Finally, all data was referenced to a common-average-reference (CAR).

4.8. Time resolved multivariate decoding

First, to further increase the signal to noise ratio for multivariate decoding, we smoothed our pre-processed EEG time courses using a Gaussian kernel with a full-width at half-maximum of 24ms. Time resolved decoding via linear discriminant analysis (LDA) using shrinkage regularization (Lemm, Blankertz, Dickhaus, & Müller, 2011) was then carried out using custom-written code in MATLAB 2014b (MathWorks). Two independent classifiers were applied to each given time window and each trial (see Fig. 3b): one to classify the perceptual category (photograph or line drawing) and one to classify the semantic category (animate or inanimate). In both decoding analyses, we used undersampling after artefact rejection (i.e. for the category with more trials we randomly selected the same number of trials as available in the smallest

category). The pre-processed raw amplitudes on the 128 EEG channels, at a given time point, were used as features for the classifier. LDA classification was performed separately for each participant and time point using a leave-one-out cross-validation approach. This procedure resulted in a decision value (*d* value) for each trial and time point, where the sign indicates in which category the observation had been classified (e.g., - for photographs and + for line drawings in the perceptual classifier), and the value of d indicates the distance to the hyper-plane that divided the two categories (with the hyper-plane being 0). This distance to the hyper-plane provided us with a single trial time-resolved value that indicates how confident the classifier was at assigning a given object to a given category. In order to use the resulting *d* values for further analysis, the sign of the *d* values in in one category was inverted, resulting in *d* values that always reflected correct classification if they had a positive value, and increasingly confident classification with increasingly higher values.

Our main intention was to identify the specific moment within a given trial at which each of the two classifiers showed the highest fidelity, and to then compare the temporal order of the perceptual and semantic peaks. We thus found the maximum positive *d* value in each trial, separately for the semantic and perceptual classifiers. The time window used for *d* value peak selection covered 3sec prior to participants' response and, based on behavioural reaction times, only trials with an RT ≥ 3sec were included (rejecting a total of 1459 trials on a group level). For all further analyses we only used peaks with a value exceeding the 95th percentile of the classifier chance distribution (see section

on bootstrapping below), such as to minimize the risk of including meaningless noise peaks. The resulting output from this approach allowed us to track and compare the temporal "emergence" of perceptual and semantic classification within each single-trial. When a peak for a given condition does not exceed the 95th percentile threshold, we do not include the trial in further analyses. For encoding trials, including all participants, we excluded 1.77 per cent of the trials based on this restriction. In the case of retrieval trials, all maximum peaks found exceeded the value of the threshold. In addition to this single-trial analysis, we also calculated the average *d* value peak latency for perceptual and semantic classification in each participant to compare the two average temporal distributions. Note, however, that many factors could obscure differences between semantic and perceptual peaks when using this average approach, including variance in processing speed across trials, e.g. for more or less difficult recalls. We therefore believe that the single trial values are more sensitive to differences in timing between the reactivated features. We used these single trial classifier peaks as dependent variables in a GLMM (as described above) to test for an interaction between two fixed effect: the type of feature (perceptual vs. semantic) and the type of task (encoding vs. retrieval). Significant interaction results were followed up by planned comparisons to test for a significant effect of feature (perceptual vs. semantic) separately for encoding (expecting an earlier timing of perceptual than semantic peaks) and retrieval (expecting an earlier timing of semantic than perceptual peaks). Wilcoxon sign rank tests were then carried out to further corroborate the relative timing of the single-trial classifier peaks, as described in the next sections.

4.9. Generating an empirical null distribution for the classifier

Previous work has shown that the true level of chance performance of a classifier can differ substantially from its theoretical chance level that is usually assumed to be 1/number of categories (Combrisson & Jerbi, 2015; Jamalabadi, Alizadeh, Schönauer, Leibold, & Gais, 2016; Kowalczyk & Chapelle, 2005). A known empirical null distribution of *d* values would allow us to determine a threshold for considering only those *d* value peaks as significant whose values are higher than the 95th percentile of this null distribution. We generated such an empirical null distribution of *d* values by repeating our classifier analysis with randomly shuffled labels a number of times, and combined this with a bootstrapping approach, as detailed in the following.

As a first step, we generated a set of *d* value outputs that were derived from carrying out the same decoding procedure as for the real data (including the leave-one-out cross-validation), but using category labels that were randomly shuffled at each repetition. This procedure was carried out independently per participant. On each repetition, before starting the time-resolved LDA, all trials were randomly divided into two categories with the constraint that each group contained a similar number of photographs and line drawings, and approximately the same amount of animate and inanimate objects (the difference in trial numbers was smaller than 8%). The output of one such repetition per participant was one *d* value per trial and time-point, just as in the real analysis. This procedure was conducted 150 times per participant for object

perception (encoding) and retrieval, respectively, with a new random trial split and random label assignment on each repetition. For each participant we thus had a total of 151 classification outputs, one using the real labels, and 150 using the randomly shuffled labels.

Second, to estimate our classification chance distribution for the random-effects (i.e., trial-averaged) peak analyses, we used the 151 classification outputs from all participants in a bootstrapping procedure (Stelzer, Chen, & Turner, 2013). On each of the bootstrapped repetitions, we randomly selected one of the 151 classification outputs (150 from shuffled labels classifiers and one from a real labels classifier) per participant, and calculated the *d* value group average based on this random selection for each given time point. Real data was included to make our bootstrapping analyses more conservative, since under the null hypothesis, the real classifier output could have been obtained just by chance. This procedure was repeated with replacement 10000 times. To generate different distributions for the perceptual and semantic classifiers, we run this bootstrapping approach two times: once where the real labels output from each subject came from the semantic classifier, and once where the real *d* values came from the perceptual classifier.

4.10 Univariate event-related potential (ERP) analysis

A series of cluster-based permutation tests (Monte Carlo, 2000 repetitions, clusters with a minimum of 2 neighbouring channels within the FieldTrip software) was carried out in order to test for differences in ERPs between the two perceptual (photograph vs. line drawing) and the two semantic (animate vs. inanimate) categories, controlling for multiple comparisons across time and electrodes. First, we contrasted ERPs during object presentation in the encoding phase in the time interval from stimulus onset until 500ms post-stimulus. We then carried out the same type of perceptual and semantic ERP contrasts during retrieval, in this case aligning all trials to the time of the button press. We used the full time window from 3000ms before until 100ms after the button press, but we further subdivided this time window into smaller epochs of 300ms to run a series of T tests, again using cluster statistics to correct for multiple comparisons across time and electrodes. For all four contrasts, we reported the cluster with the lowest *p* value.

We were mainly interested in the temporal order of the ERP peaks that differentiated between perceptual and semantic classes during encoding and retrieval. The above procedure resulted in four statistically meaningful clusters across subjects: one each differentiating perceptual categories during encoding, semantic categories during encoding, perceptual categories during retrieval, and semantic categories during retrieval. To statistically test for an interaction in this timing of these clusters, we extracted the time point of the maximum ERP

difference for each individual participant, restricted to the electrodes showing an overall cluster effect but over the entire time window for encoding and retrieval. These time points were entered into a 2x2 within-subjects ANOVA with the factors type of feature (perceptual or semantic), and type of task (encoding or retrieval), with the only planned comparison in this analysis being the interaction contrast.

5. Acknowledgments

We thank Alexandru-Andrei Moise, Emma Sutton, Thomas Faherty, Laura De Herde and James Lloyd-Cox for helping with data collection, and Rodika Sokoliuk for her useful technical support. This work was supported by a European Research Council Starting Grant ERC-2016-STG-715714 awarded to M.W, and a scholarship from the *Midlands Integrative* Biosciences *Training* Partnership (MIBTP) awarded to J.L.D.

6. Author contributions

J.L.D. and M.W. designed the experiments. J.L.D. conducted the experiment. J.L.D., M.S.T. and C.K. analysed the data. All authors contributed to the analysis approach and to data interpretation. J.L.D. wrote the manuscript under the supervision of M.W. and all authors contributed to reviewing and editing.

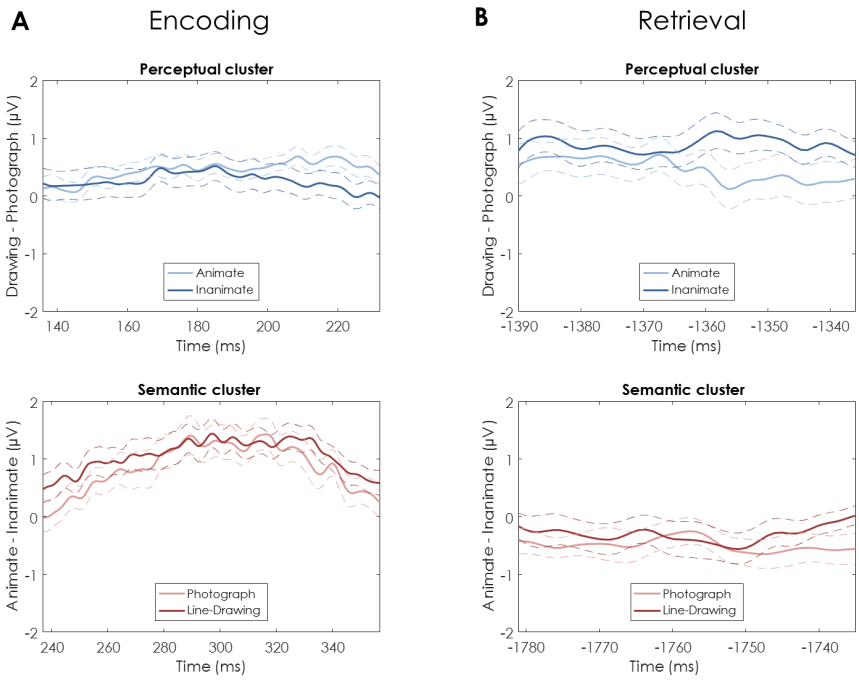
7. Declaration of interests

The authors declare no competing financial interests.

8. Data and code availability statement

The data and the custom code that support the findings of this study are available from the corresponding author upon reasonable request.

9. Supplementary figures



H. Chapter 3. Supplementary Figure 1.

Supplementary Figure 1. Additional ERP results. Within the four significant ERP clusters reported in the main results (see Fig. 4), we did not find evidence suggesting any cluster was driven by a particular combination of perceptual and semantic features. (a) Upper panel: ERP differences between line drawings and photographs shown separately for animate (light blue) and in inanimate (dark blue) objects, based on same electrodes and time window contained in the main perceptual cluster found at encoding. Lower panel: ERP differences between animate and inanimate objects plotted separately for photographs (pink) and line drawing (dark red), based on the main semantic cluster found at encoding. (b) Upper panel: ERP differences between line drawings and photographs shown separately for animate (light blue) and in inanimate (dark blue) objects, based on same electrodes and time window contained in the main perceptual cluster found at retrieval. Lower panel: ERP differences between animate and inanimate objects plotted separately for photographs (pink) and line drawing (dark red), based on the main semantic cluster found at retrieval. In all four plots, dashed lines represented standard error of the mean. The results of a statistical comparison of the average T values are reported in the main results.

Chapter 4: Preliminary findings in an iEEG case study support the reverse reconstruction hypothesis

J. Linde-Domingo[1], F. Roux[1], R. Chelvarajah[2], D. Rollings[2], V. Sawlani[2], B. Staresina[1], S. Hanslmayr [1] & M. Wimber[1]

[1]School of Psychology & Centre for Human Brain Health (CHBH), University of Birmingham (UK), [2]University Hospitals, Birmingham NHS Foundation Trust (UK)

This chapter represents a series of preliminary analyses from a larger ongoing project.

1. Introduction

In the previous chapter, we tested what we call the reverse reconstruction hypothesis: whether remembering a past visual representation is a hierarchical process where perceptual and semantic features unfold in time following the reverse order compared to visual perception. Consistent findings obtained from two behavioural experiments and one electroencephalography (EEG) study supported this alternative hypothesis. It is widely accepted that low-level perceptual details are processed before semantic information during encoding or visual processing of a complex image (T. Carlson et al., 2013; Cichy et al., 2014; Clarke & Tyler, 2015; Lehky & Tanaka, 2016). We demonstrated for the first time, however, that during the recall of the same images from memory, participants showed a consistent prioritization of semantic features over perceptual details, as seen behaviourally in reaction time and accuracy profiles. Similarly, decoding analyses of the EEG signal indicated that during retrieval the maximum peaks classifying semantic information significantly preceded perceptual classification peaks, showing the opposite order found during encoding.

Apart from the temporal dynamics previously described, knowing the neural pathway of this reconstruction-processing stream is essential to fully understand how memory representations and their details are retrieved over time. In the last decade, a growing body of evidence in the episodic memory field has suggested that retrieving features of past representations requires the activity of

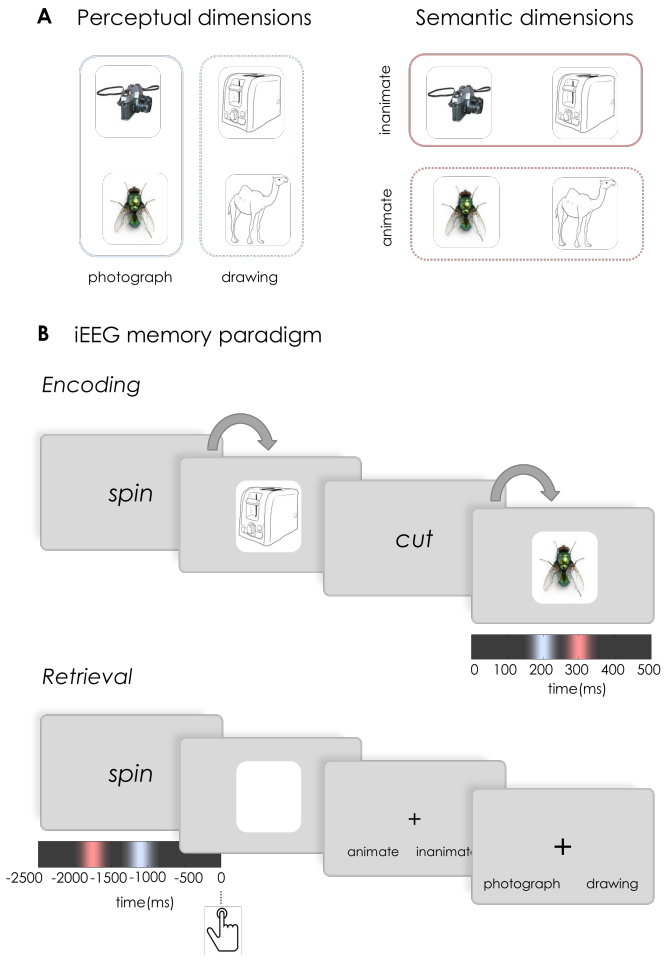
brain areas that process each type of information during encoding. For instance, retrieving semantic information implies activation of semantic processing areas in the temporal lobe (B. A. Kuhl et al., 2011; Staresina et al., 2012; Wimber et al., 2015), and access to low-level visual information elicits a reactivation of these features in occipital brain areas (Bosch et al., 2014; Waldhauser et al., 2016; Wimber et al., 2012). Since both semantic and perceptual processing areas are part of the perceptual pathway, it can be hypothesized that when a visual representation of an everyday object is retrieved, the reactivation of semantic details would be temporally prioritized compared to perceptual features especially along the ventral visual stream (VVS). In order to investigate the temporal dynamics of memory retrieval in brain structures or networks of interest, the use (or combination) of imaging techniques that allow measuring neural activity with a high temporal and spatial resolution is fundamental. In this sense, the use of intracranial EEG (iEEG) is an ideal approach to test the reverse reconstruction hypothesis along a spatial and a temporal dimension, allowing the study of the neural electrical activity on a millisecond scale from specific local field potentials. In humans, this valuable type of neural signal recording is usually carried out for medical reasons in patients that suffer from medication resistant epilepsy. Therefore, an important limitation of human iEEG studies is that electrode locations depend on clinical purposes.

In the present chapter, I will present preliminary analyses and results testing the reverse reconstruction hypothesis in an iEEG single case study. Due to the

unique location of intracranial electrodes that cover brain structures of special relevance for us (i.e. early visual areas and the temporal lobe) this study is an exceptional opportunity to introduce how the methodology used in Chapter 3 could be applied to test our alternative hypotheses in brain areas of interest. In particular, this single case allowed us to test whether previous results that suggest that the information processing flow reverses between encoding and retrieval can be replicated restricting our analyses to areas along the visual ventral pathway. Also, based on recent findings that indicated that memories are transformed from detailed (more perceptual) to gist-like (semantic) representations along the longitudinal axis of the hippocampus (Dandolo & Schwabe, 2018b), we tested whether the reverse reconstruction effect could be replicated using electrode contact located closed to the hippocampus.

To address these questions we used an adapted version of the associative memory paradigm used in Experiment 3 (Chapter 3), where the participant learned a series of random associations between word cues and everyday objects. Later, the participant was cued with these word cues and asked to mentally visualise the associated object (see Fig. 1). To identify in which moment perceptual and semantic features of object representations were processed, time-series decoding techniques were carried out on the iEEG signal when items were presented on the screen or when they were mentally retrieved after cue presentation.

Despite the early state of this work, and the fact that a bigger sample size and further analyses are needed to make a solid conclusion, these initial iEEG findings are in line with our previous behavioural and electrophysiological results and suggest that during memory retrieval semantic information is accessed before low-level perceptual features along the VVS stream and the hippocampus.



I. Chapter 4. Figure 1.

Figure 1. Stimuli and design of the iEEG experiment. (a) Illustration of the orthogonal design of the stimulus set. Objects that were presented in this experiment were the same (a total of 128) that were used in Experiment 3

(Chapter 3). As in the previous experiment, all objects varied along a perceptual dimension (i.e. objects were presented as a photograph or as a line drawing); and a semantic dimension (i.e. objects belonged to the animate or inanimate category). (b) In this iEEG experiment we used an adapted version of the EEG paradigm without a time limit for the participant's response (self-paced). The participant was asked to create word-object associations during encoding, and to remember the object as vividly as possible when cued with the word. Later, we asked questions about perceptual and semantic details of the retrieved episode. iEEG was recorded during encoding and retrieval, and time-series decoding analyses were performed on the recorded data in order to detect at which moment, within a single trial, a classifier is most likely to categorise perceptual and semantic features correctly. Button press symbols indicate at which moment the participant confirmed that the episode was remembered.

2. Results

2.1. Behavioural results.

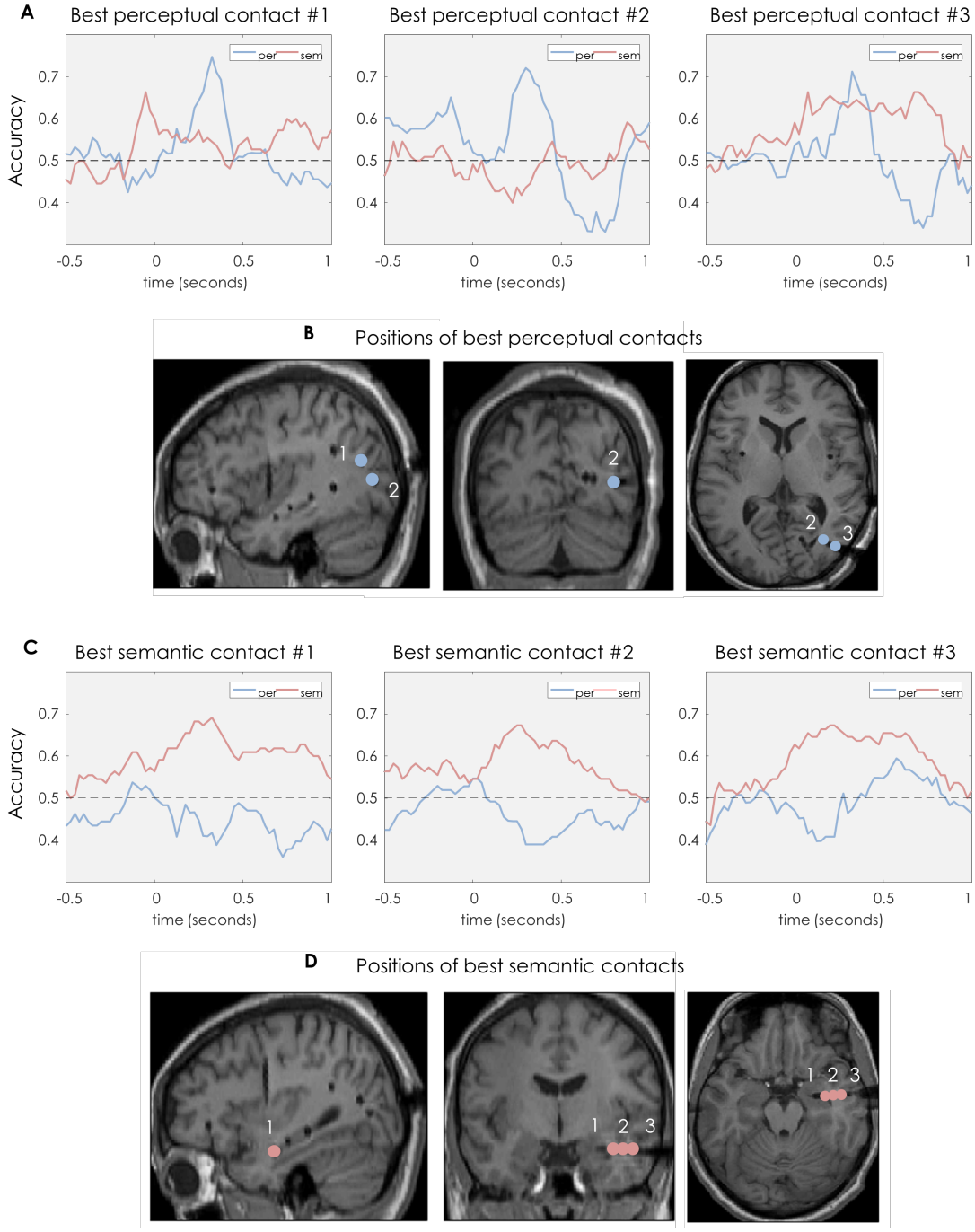
Behavioural analyses revealed that the participant performed well in the experiment. The general performance responding to questions about previous word-object associations was 78.91%. The participant's accuracy retrieving semantic details of these objects was 79.69% and 78.13% when responding to questions about perceptual details. Although this tendency to better remember semantic information than low-level details was in line with our previous behavioural results (i.e. Experiments 1, 2 and 3 in Chapter 3), we did not find that the kind of question predicted participant's correct responses when using generalised linear mixed-models analyses on this single case (GLMM; $F_{1, 254}$ = 0.069, P = .793).

2.2. Time resolved decoding results

Aiming to test the reverse reconstruction hypothesis in a concrete set of electrodes, we followed the rationale of Experiment 3 in Chapter 3. Specifically, we used a trial-by-trial decoding analysis that allowed us to find in which moment the brain signature associated with perceptual or semantic processing was peaking. However, in these series of decoding analyses we ran a separate classifier independently for each electrode contact. The main reason for using this approach was to investigate the temporal dynamics maximising the local information, allowing us to group these outputs into contacts of interest.

To achieve this we carried out a series of time resolved classifiers based on linear discriminant analyses (LDA) for each contact. Importantly, inspired by previous literature (Schönauer et al., 2017; Xiao & Ding, 2015) we used the power spectrum (between 2 Hz and 30 Hz) of each electrode contact and time point as the features for the classifier. In order to ascertain whether this decoding approach produces meaningful outputs in line with the object recognition field, we inspected the location of those electrode contacts that showed the highest accuracy decoding semantic and perceptual features. In particular, we expected that, during visual processing, electrodes located in early visual areas and in semantic areas along the ventral stream would present the best performance decoding perceptual and semantic information respectively. To test this prediction, for both classifiers (i.e. perceptual and semantic) we selected the three contacts with the highest general accuracy

peaks from 0ms to 1000ms relative to object presentation. Confirming our prediction based on vision literature (T. Carlson et al., 2013; Cichy et al., 2014), we found that the three electrode contacts that showed the highest accuracy in decoding perceptual information (75.73%, 72.09% and 71.18%; see Fig. 2a and b) were located in early visual processing areas in the occipital lobe, where the nearest grey matter was the middle occipital gyrus and the closest BA was BA 19 (according to the Talairach Atlas; Lancaster et al., 1997, 2000). On the other hand, and being consistent with previous findings in the field of object recognition (T. Carlson et al., 2013; Cichy et al., 2014), the three electrode contacts that obtained the highest accuracy (69.09%, 67.27% and 67.27%) in classifying the semantic category of objects were located in the anterior temporal lobe and the closest grey matter were the parahippocampal gyrus and the adjacent BA was BA 35 (see Fig. 2c and d). However, although these initial sanity checks are in line with widely replicated findings, a bigger participant number is fundamental to confirm statistically the effectiveness of this decoding approach with the material used.



J. Chapter 4. Figure 2.

Figure 2. Electrode contacts with highest accuracy during object presentation. In line with the previous findings, we observed that using the power spectrum as a feature for the classifier (A) electrode contacts with the highest accuracy in decoding perceptual details during encoding were located in early visual areas. Panel B represents the positions of the best perceptual contacts #1, #2 and #3 (numbered blue points) in the middle occipital gyrus. (C)

Decoding accuracy at three sematic electrodes. In accordance with the object processing literature, electrode contacts that reached the highest accuracy in classifying semantic information were located in the parahippocampal area (numbered pink points in panel D). In panels A and C, Y-axes represent general classifier accuracy for all trials decoding perceptual (blue line) and semantic information (pink line). X-axes indicate the time window in seconds relative to object presentation (time 0). Points in panel B and D, representing electrode contacts are not drawn to scale.

2.2.1. Semantic information is reactivated faster than low-level details along the ventral visual stream

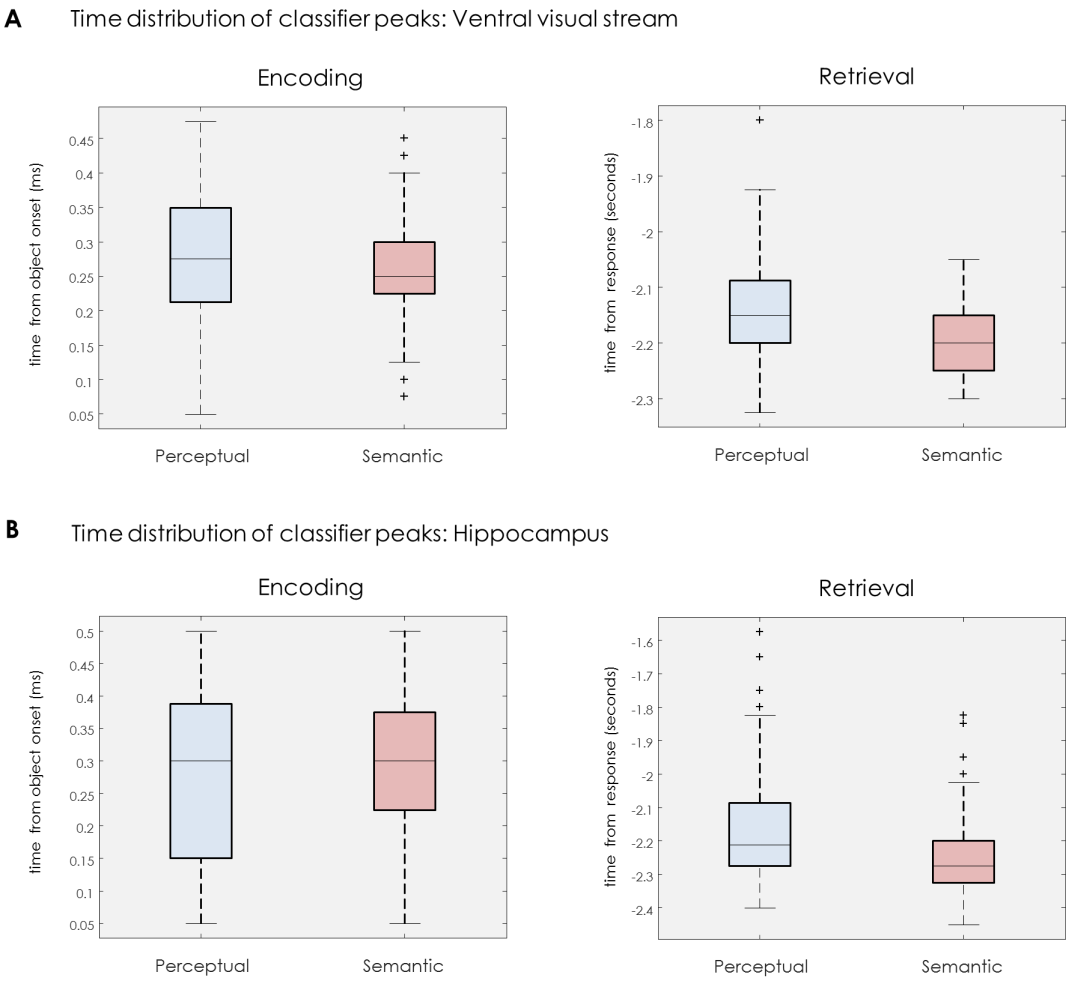
Confirming our alternative hypothesis, we found in an EEG study (i.e. Experiment 3, Chapter 3) that the interaction between the kind of task (encoding or retrieval) and the type of classifier (perceptual or semantic) predicted significantly at which time point the neural signature is more associated with perceptual or semantic processing. In line with the reverse reconstruction hypothesis, these results suggested that, when participants are remembering a past representation, they can process its semantic details earlier than its low-level perceptual features, following the reverse order found traditionally during object perception. In the present iEEG experiment, we tested whether this previous finding could be replicated limiting our analyses to those electrode contacts located in the VVS.

First, we selected the contacts of interest for further analyses based on purely anatomical criteria. In this case, we included all contacts placed along the right temporal lobe (a total of 16, while excluding the closest to the hippocampus)

and those situated in early visual areas (a total of 8). To select the latter, we took those contacts that were located closest to Brodmann area (BA) 17, 18 and 19 based on the Talairach Atlas (Lancaster et al., 1997, 2000). We ran two classifiers per contact (i.e. one perceptual and one semantic classifier) during the encoding and retrieval time windows. Per trial, we obtained a perceptual *d* value across time that reflected the classifier's fidelity selecting the correct perceptual category per observation (i.e. line drawing vs. photograph), and a semantic *d* value indicating the classification fidelity when deciding about semantic categories (i.e. animate vs. inanimate). We then calculated in which moment we obtained the maximum perceptual and semantic *d* values in a given time window for encoding (from object onset until 500ms post-onset) and retrieval (from -2500ms to time 0 relative to the participant's response indicating that they had mentally reinstated the object). To prevent the inclusion of non-significant *d* value peaks, we only used those peaks that exceeded a certain value. This threshold was calculated per electrode based on the classifier performance when using meaningless information (see details in Methods section). Then, per each individual trial, we averaged the perceptual and semantic peak time positions across the electrode contacts of interest.

Using the same analysis procedure presented in Chapter 3, peak distributions (Fig. 3a) were examined using generalized linear mixed-models (GLMMs). In these series of analyses, the time of *d* value peaks was selected as a target variable and three fixed factors were selected: the type of classifier (perceptual or semantic), type of task (encoding or retrieval), and the interaction between

both factors (type of classifier x type of task). We found that the interaction between type of task and type of classifier significantly predicted the time position of *d* value peaks ($F_{1,\ 388}$ = 12.323, *P* = .001). Then, planned comparisons were carried out for encoding and retrieval keeping the same parameters in the model. During encoding we did not find that the type of classifier significantly predicted the timing of *d* value peaks ($F_{1,\ 190}$ = 0.121, *P* = .729). However, during retrieval this factor (type of classifier) predicted significantly the time position of *d* value peaks ($F_{1,\ 198}$ = 25.209, *P* < .001) where beta coefficients suggested that the semantic peaks (M = -2195, SD = 68) appeared significantly earlier than perceptual peaks (M = -2139, SD = 104; *B* = 0.056, *t* = 5.021, *P* < .001), replicating previous results (i.e. Experiment 3, Chapter 3).



K. Chapter 4. Figure 3.

Figure 3.* D *value peak distribution for perceptual and semantic features. (a) Box plots show peaks of *d* value distributions in VVS contacts for perceptual and semantic categories during encoding (perceptual peaks: M = 272, SD = 95; semantic peaks: M = 270, SD = 85) and retrieval (perceptual peaks: M = -2139, SD = 104; semantic peaks: M = -2195, SD = 68). GLMM analyses showed that an interaction between the type of task and type of classifier significantly predicted *d* value peak distributions ($F_{1,\ 388} = 12.323$, $P = .001$). (b) Peaks of *d* value distributions in hippocampal contacts for perceptual and semantic features for encoding (perceptual peaks: M = 274, SD = 134; semantic peaks: M = 296, SD = 111) and retrieval (perceptual peaks: M = -2168, SD = 170; semantic peaks: M = -2241, SD = 118). In this case, GLMMs analyses also suggested that an interaction between the type of task and type of classifier predicted significantly the distribution of classification peaks ($F_{1,\ 332} = 10.087$, $P = .002$). In general, although we did not find differences during encoding, hippocampal peaks occurred significantly earlier than VSS peaks during

retrieval ($z = -5.84$, $P < 0.001$). In all box plots, the line in the middle of each box represents the median, and the tops and bottoms of the boxes the 25th and 75th percentiles of the samples, respectively. Whiskers are drawn from the interquartile ranges to the furthest minimum (bottom) and maximum (top) values. Crosses represent outliers.

2.2.2. Semantic temporal prioritization during retrieval is also found in electrode contacts located close to the hippocampus

Based on recent findings that indicate that memory representations change from detailed memories into gist-like (semantic) representations along the anterior-posterior axis of the hippocampus (Dandolo & Schwabe, 2018b), we tested the reverse reconstruction hypothesis restraining our analysis to electrode contacts situated close to the hippocampal formation. More precisely, three electrodes were located along the longitudinal axis of the right hippocampus, and for each electrode we used the two contacts closest to the hippocampus (a total of 6 contacts). Apart from this, to compare the timing of perceptual and semantic *d* value peaks we kept the same procedure carried out previously.

Firstly, we compared the general time distribution of *d* value peaks obtained in hippocampal contacts relative to those peaks calculated along the VVS. Since the hippocampus is one of the final stages in the visual stream, we expected that during encoding VVS peaks would appear before hippocampal peaks. Conversely, the opposite pattern was expected during retrieval, where the hippocampus is thought to trigger the memory reconstruction cascade toward

neocortical areas (for a review, see Rolls, 2010). Although we found this trend during encoding where VVS peaks (M = 271ms; SD = 90ms) seemed to precede hippocampal peaks (M = 286ms; SD = 123ms), the difference was not significant when using a two-tailed Wilcoxon rank sum test (z = 1.43, P = 0.154). However, for retrieval outputs we found that hippocampal decoding peaks (M = -2204ms; SD = 151ms) were detected significantly before VVS peaks (M = -2167ms; SD = 92ms; z = -5.84, P < 0.001). These results confirm that during memory reconstruction, the retrieved content can generally be detected earlier in the hippocampus than in neocortical visual regions, consistent with the general notion of neocortical reinstatement.

GLMM analyses were then also performed to investigate the temporal distributions of perceptual and semantic classifier peaks in hippocampal contacts. Using the same parameters as in previous tests, the selected fixed factors were the type of classifier (perceptual or semantic), type of task (encoding or retrieval), and the interaction between both factors (type of classifier x type of task); the time of *d* value peaks was the target variable. Replicating previous results, these analyses resulted in a significant effect of the interaction between type of task and type of classifier predicting peak positions ($F_{1,\ 332} = 10.087$, $P = .002$). Further planned comparisons showed that during encoding the type of classifier did not significantly predict the timing of *d* value peaks ($F_{1,\ 134} = 1.113$, $P = .293$), although it did so during retrieval ($F_{1,\ 198} = 12.767$, $P < .001$), where beta coefficients indicated that semantic peaks (M = -2241, SD = 118) appeared significantly earlier than perceptual peaks (M = -

2168, SD = 170; B = 0.073, t = 3.573, P < .001). These preliminary results restricted to para-hippocampal contacts thus replicated the findings obtained along the VVS and our previous scalp EEG results (Experiment 3, Chapter 3), indicating that when retrieving a past representation its semantic information is reactivated before low-level details even at very early states of memory reinstatement.

3. Discussion

One of the main objectives of this doctoral thesis is to explore how representations of past events are retrieved in relation to their initial visual processing. In previous behavioural and EEG studies we found consistently that, although low-level perceptual information is processed faster than semantic details when objects are visually perceived, this hierarchical processing is reversed when these items are retrieved from memory. The rationale behind this brief chapter was to present how the trial-by-trial analysis procedure used in our previous work could be applied to iEEG recordings to further explore predictions related to the reverse reconstruction hypothesis on an anatomically more fine-grained spatial scale. Specifically, these findings can inform us about whether the reserve effect that we found in our scalp EEG study follows the predicted order along the VVS, and whether and when the hippocampus shows reinstatement of these representations.

Although the iEEG results presented here are still preliminary, and further analyses and a bigger sample size are required, we were lucky to be able to record from a patient who had electrodes implanted along the VVS as well as the hippocampus. The initial findings replicated earlier observations: in both regions, the reactivation of semantic information is prioritized over perceptual features when a visual representation is being retrieved from memory. Specifically, analyses based on GLMMs indicated that an interaction between the type of task (encoding or retrieval) and the type of classifier (perceptual or semantic) significantly predicted the moment in which classifiers showed the highest fidelity categorizing each observation. Further analyses also suggested that during retrieval (but not encoding) the factor “type of classifier” significantly predicted the timing of classification peaks. This pattern of results was the same when we analysed electrode contacts located along the VVS and the hippocampus, although hippocampal peaks in general (collapsed across perceptual and semantic) appeared significantly earlier than VVS peaks during retrieval. These pilot findings could build the basis of new hypotheses to be tested in future work. For instance, a relevant question is whether the time delay in reactivating low and high-level features as found in the cortical EEG signal is an effect caused directly by the time dynamics of the hippocampus, at the time the retrieval cascade is initiated. In other words, it could be predicted that this reverse reconstruction effect is generated in the hippocampus (i.e. triggering the reactivation of gist-like information before low-level details) and then propagates to the VVS.

One unexpected finding in these initial results was the lack of time differences between perceptual and semantic classification peaks during encoding. One potential explanation of these outcomes could be the need of a higher temporal resolution to detect an expected disparity between the two peaks (i.e. the temporal resolution was reduced to 40 Hz when time frequency analyses were performed). Additionally, some artefacts produced during time frequency analyses (as filter mirroring) could lead to a lower time precision when running classifiers on this type of signal. It is also likely that including a wider frequency power spectrum as features for the classifier (i.e., containing more information from the gamma band that is associated to visual object processing; Friese, Supp, Hipp, Engel, & Gruber, 2012; Jensen, Kaiser, & Lachaux, 2007; Tallon-Baudry & Bertrand, 1999) could shed light onto this question, and will be explored in future analyses.

In summary, this unique iEEG case study allowed us to test the reverse reconstruction hypothesis in brain areas of exceptional interest: the VVS and the hippocampus. This brain imaging technique together with decoding analyses could be crucial to identify the spatiotemporal map of memory retrieval in the human brain, understanding past events as multi-layered representations formed by diverse group features (e.g. perceptual and semantic information, but also contextual or emotional). Still, data from more participants and additional analyses are indispensable to corroborate these promising initial findings.

4. Methods

4.1. Participant

One female participant (age = 26) undergoing treatment for medication resistant epilepsy in the Queen Elizabeth Hospital in Birmingham (UK) took part in the experiment. Due to diagnostic purposes, the participant had implanted intracranial depth electrodes. The ethical approval was obtained from the NHS Health Research Authority (15/WM/0219) and informed consent was granted in accordance with the Declaration of Helsinki.

4.2. Stimuli

In this study we used the same stimuli described in the Chapter 3 for all experiments.

4.3. Procedure

This experiment used the same procedure followed in the EEG study presented in Chapter 3 (Experiment 3), except for the following three changes, all aimed at adapting the task for a clinical population. First, all stimuli (words and objects) were presented on the screen until the participant made a response to move on (i.e. pressing the up arrow) and there was no time limit to answer questions during retrieval. Second, in this study we did not include a performance

feedback screen for participants at the end of each block. Instead, information about performance was displayed at the end of the block in a masked manner, where only the experimenters were able to decode this information. This way, we tried to avoid discouraging participants in case of low accuracy. Finally, based on participant's performance, the experimenter was able to change the length of each experimental block (with a minimum of 2 word-object associations and a maximum of 8). In the case of the patient reported here, however, the length of all experimental blocks was 8 associations, and the patient thus performed at the same level as our healthy subjects in Experiment 3.

4.4. Electrode localisation

The native space coordinates of all contacts was determined by visual inspection of the participant's T1 scan after electrode implantation. Then, native space coordinates were transformed into MNI space via a transform matrix obtained by normalising T1 scans in SPM 12.

4.5. Signal pre-processing

We used the Fieldtrip Toolbox (version 3rd, August 2017) for MATLAB to pre-process the iEEG signal (Oostenveld et al., 2011). Data was recorded during the encoding phase of the task and epoched into trials from 4 seconds before stimuli onset until 4 seconds after stimuli onset. The recorded signal from the

retrieval phase of the task was segmented into trials starting 4 second before participant's response and lasting until 4 second after. We band-stop filtered the signal between 48 and 52 Hz in both datasets, aiming to reduce line noise at 50Hz. All trials were visually inspected and all epochs that contained coarse artefacts were rejected. In total, 111 encoding trials and 107 retrieval trials remained (out of a total of 128 in each phase). Finally, the signal from all individual contacts was re-referenced to a bi-polar reference to maximize the detection of a local signal.

4.6. Time-frequency analysis

Time-frequency power analyses from 1 Hz to 30 Hz were performed using Morlet wavelets on both encoding and retrieval pre-processed datasets using the Fieldtrip Toolbox (version 3rd, August 2017) for MATLAB. For these analyses we used a 0.5 Hz frequency-resolution, a cycle-length of 6 and 25ms temporal resolution. Encoding data were then segmented from -1000ms to 2000ms relative to stimuli onset; and retrieval data were epoched from -2500ms to 1000ms relative to participant's response, leading to an effective removing of all edge artefacts.

4.7. Time resolved multivariate decoding

Using custom-written MATLAB code, we carried out a time resolved decoding via linear discriminant analysis (LDA) with shrinkage regularization (Lemm et

al., 2011) for each electrode contact. Following the approach of Chapter 3, two independent classifiers were run individually per electrode contact, in each given time window and trial. That is, we used a “perceptual classifier” to decode whether a trial belonged to a photograph or line-drawing category; and a “semantic classifier” that decoded the higher-level category of each trial (animate or inanimate). Per time point, we used the power spectrum from 2 Hz to 30 Hz of each electrode as a feature for the classifier. In all cases a leave-one-out cross-validation approach was used.

This decoding procedure resulted in a decision value (*d* value) for each trial, time point and electrode contact, where the sign of this value indicated into which category (e.g., animate or inanimate for the semantic classifier) the observation was classified. In all cases, a value larger than 0 meant that the observation was classified in the correct category. The distance to 0 (being the value that divided the two categories) represents classifier fidelity in a given classification. *D* values were used on the single trial level in order to identify at which specific moment the classifier had the highest fidelity selecting the correct category for a single trial observation. To avoid the inclusion of meaningless *d* value peaks, we calculated the *d* value chance distribution of each electrode contact. Chance distributions were obtained using the same bootstrapping approach (10000 repetitions) as in Chapter 3 (for further details, see Methods section in Chapter 3), however, in this case we calculated how the classifier performed in a random-label scenario at each individual electrode contact. We

only selected peaks from a given electrode contact with a value that exceed the 95th percentile of its chance distribution.

4.8. GLMM analyses

To test our alternative hypothesis for participant's behavioural performance (accuracy) and for the relative timing of fidelity peaks (*d* value) we used generalized linear mixed models (GLMMs). Following the same procedure as in Experiment 1, 2 and 3 of Chapter 3, models for behavioural performance used a binomial distribution and a logistic link function; and all models for *d* value peak analyses used a gamma probability distribution and an identity link function.

5. Author contributions

J.L.D. and M.W. designed the experiments, conducted the experiment and contributed to the analysis approach and to data interpretation. J.L.D. analysed the data. F.R., R.C., D.R., V.S., B.S. and S.H. contributed to data acquisition within the hospital setup. J.L.D wrote the chapter under the supervision of M.W.

Chapter 5: The reverse reconstruction effect across different perceptual and semantic manipulations

J. Linde-Domingo & M. Wimber

School of Psychology & Centre for Human Brain Health (CHBH), University of Birmingham (UK)

At the time of thesis submission, this chapter represents a manuscript in preparation.

1. Introduction

Our previous findings from two behavioural experiments and two neuroimaging studies (i.e. scalp EEG and intracranial EEG) consistently support the reverse reconstruction hypothesis: although during encoding perceptual details of objects are processed before semantic features, semantic information is reactivated faster than perceptual features when participants retrieve these object representations from memory. The goal of the experiments reported in this chapter is to generalize these findings to different perceptual and semantic categories.

To shortly recapitulate, in the series of studies reported so far, we used a very simple associative memory task where, in an encoding phase, participants learnt a group of random associations between actions verbs and images of everyday objects. Later, subjects were cued with a verb and asked to mentally visualise the associated image. Aiming to test behaviourally how different representations' details are processed over time, we measured participants' reaction time and accuracy when they were asked about some features of objects while the objects were visually presented or recalled from memory (Chapter 3). These questions referred to perceptual details (i.e. whether the object was presented as a line drawing or as a photograph) or semantic information (i.e. whether the image represented an animate or an inanimate entity). Both behavioural experiments lead to the same results: participants responded significantly faster and more accurately when they had to retrieve

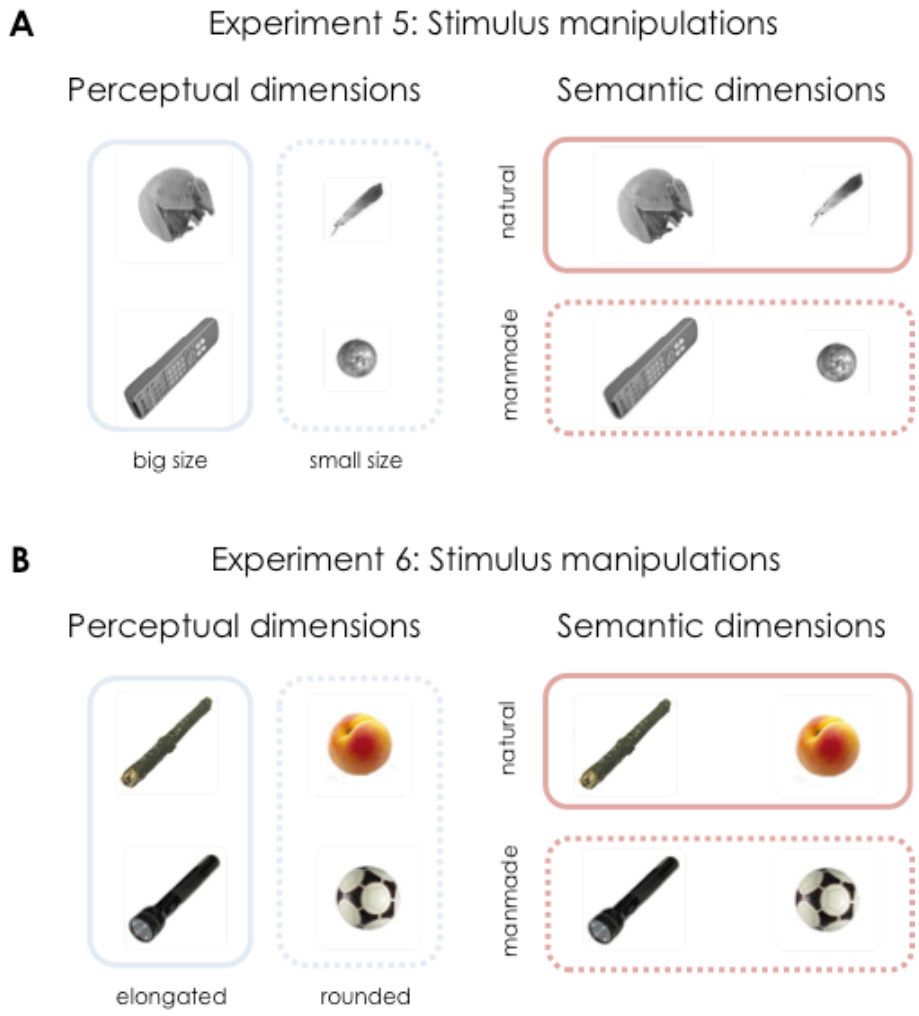
semantic information compared to perceptual details, following the reverse pattern found during encoding (or visual processing). To investigate the neural correlates of this temporal processing pattern, we ran this task while recording i) scalp EEG (Chapter 3) and ii) iEEG (Chapter 4) and performed temporally resolved decoding analyses (T. Carlson et al., 2013; Cichy et al., 2014; Kurth-Nelson et al., 2015) on the data. This decoding approach allowed us to identify in which specific moment the brain signal is more associated with perceptual or semantic processing. Electrophysiological results replicated a widely accepted processing hierarchy during visual recognition: perceptual details of images are prioritized over semantic information. Importantly, when participants were asked to bring back from memory these object representations, a reverse processing hierarchy was found. Results from both brain imaging experiments thus support the idea that semantic qualities of past representations (i.e. if the object was animate or inanimate) were recollected before their low-level details, as lines or colours.

Despite the robust replicability of these findings and their consistency with other findings in the literature (Ahissar & Hochstein, 2004; Hebart, Bankson, Harel, Baker, & Cichy, 2017), important questions about the reverse reconstruction effects are still unanswered. Among these interrogations is whether this prioritization of semantic information during retrieval depends on the specific type of perceptual and semantic categories used. To manipulate the perceptual details of our stimuli in these previous experiments, all items were presented either as colour photographs or as line drawings. Compared to colour

photographs, each drawing was composed of lines that formed a schematic outline of each stimulus. This experimental approach elicited the expected behavioural and neural responses during visual processing (see Chapter 3 and 4). In addition, to generate a high-level (semantic) distinction between items, images represented either inanimate objects (i.e. fruits, vegetables, clothes or electronic devices) or animate beings (i.e. insects, birds, mammals or sea animals). This semantic manipulation based on items' animacy has been widely used in the past, being a reliable approach to investigate visual processing (Carlson, Ritchie, Kriegeskorte, Durvasula, & Ma, 2014; Cichy et al., 2014; Cichy, Pantazis, & Oliva, 2016; Ritchie, Tovar, & Carlson, 2015). However, it is unclear whether the reverse stream reported in our former experiments is restricted to these particular categories, or if the same effect could also be found using different categories for manipulating low and high-level features.

To address this question, we carried out two additional behavioural experiments. In both studies, we maintained the same experimental procedure described for Experiment 1 in Chapter 3. However, here we employed different perceptual and semantic categories. Based on previous imagery studies (Konkle & Oliva, 2012), we used the retinal size of the object as a new perceptual manipulation for Experiment 5 (see Fig. 1a). That is, some of the images were displayed on the screen either in a small or a big size. Another perceptual manipulation was introduced in Experiment 6 (see Fig. 1b). In this second study, half of the items had a rounded shape (e.g. an orange or a ball), and the remaining stimuli had an elongated shape (e.g. a banana or a microphone). In both experiments, we kept the same, novel semantic

manipulation: instead of animacy, objects were grouped into two categories according to their naturalness: half of the objects were natural entities (e.g. fruits, plants or animals) and the other half represented manmade items (e.g. music instruments, tools or electronic devices). Importantly, since we used an orthogonal design (i.e. following the design of Experiment 1, 2 and 3 in Chapter 3), all perceptual and semantic categories were independent in Experiment 5 and 6.



L. Chapter 5. Figure 1.

Figure 1. Perceptual and semantic manipulations in Experiment 5 and 6. Orthogonal design of the stimulus set used in both experiments. Items (a total of 128 objects in each study) varied along two independent dimensions: a perceptual and a semantic one. (a) To manipulate the perceptual dimension in Experiment 5, objects were displayed on the screen using a small or bigger size (retinal size). At the same time, objects could belong to the natural or manmade category (semantic manipulation). (b) In Experiment 6, objects could have an elongated or rounded shape (perceptual manipulation). The different semantic categories were based on naturalness (natural vs. manmade) just as for Experiment 5.

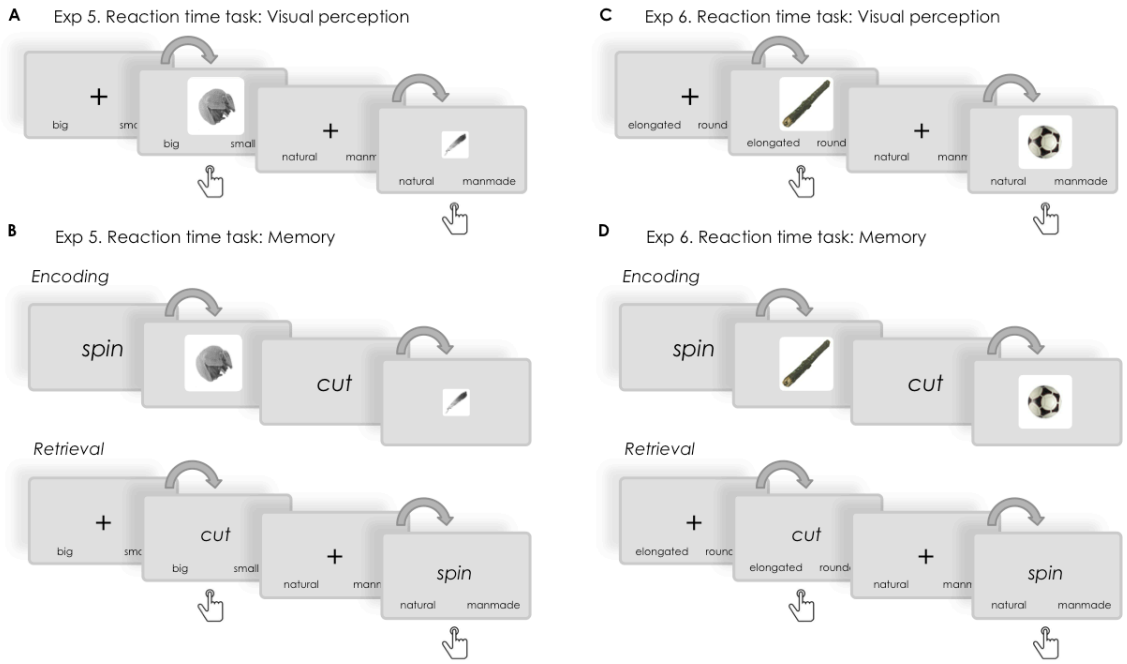
Consistent with our previous results, we found that RT and accuracy differences for perceptual and semantic questions changed depending on the task (visual

processing or memory task). Specifically, we found that in both experiments, when an object is on the screen (during visual perception) participants showed faster and more accurate responses when answering perceptual questions compared to semantic ones. During retrieval, we found a significant prioritization of semantic information in RT and accuracy in Experiment 5 (i.e. retinal size manipulation) but not in Experiment 6 (i.e. shape manipulation). A possible explanation for these results will be discussed below.

2. Results

To test whether the reversal of information processing between visual perception and retrieval can be generalized across different semantic and perceptual manipulations, we used the same paradigm as in Experiment 1 but with novel materials. As in our former behavioural studies, reaction times were our main dependent variable of interest in both experiments. Following the same assumption as in Chapter 3, we expected that the time required to answer a question about perceptual or semantic features of a given item would reflect the temporal dynamics of neural information processing. Based on our reverse reconstruction hypothesis and on our previous results, we predicted that, when an object is displayed on the screen, participants would be faster responding to questions about perceptual features of the item (i.e. size or shape) compared to semantic features (i.e. nature vs. manmade). Importantly, we expected the reverse pattern when the object is not visually presented but reconstructed from memory: RTs for semantic questions would be faster

compared to those of low-level perceptual questions. We expected to observe this pattern of results independently of the manipulation used.



M. Chapter 5. Figure 2.

Figure 2. Visual and Memory reaction time tasks. (a) In the visual RT task of Experiment 5, subjects were trained to categorize the upcoming object as fast as possible, according to its perceptual category (small vs. big size) or its semantic category (natural vs. manmade). (b) In the encoding phase of the memory RT task in Experiment 5, participants created word-object associations (a total of 8 per block). Then, during the retrieval phase, participants were presented with a reminder word, and asked to recall and categorize the associated object according to its perceptual (small vs. big) or semantic (natural vs. manmade) features as fast as possible. (c and d) The same visual and memory RT tasks were used in Experiment 6. However, although semantic questions were also based on items' naturalness, perceptual questions asked participants to categorise each object depending on its shape (elongated vs. rounded). Button press symbols indicate at which moment in a trial RTs were collected.

In both experiments we used a 2 x 2 mixed design (Fig. 2), mimicking the behavioural experiments reported in Chapter 3. All participants were asked to respond as quickly as possible to either perceptual or semantic questions (factor 'type of question', within-subject). While answering these questions, objects were displayed on the screen for one group of participants while another group of participants was cued to recall the objects from memory (factor 'task', between-subjects). The main difference between Experiment 5 and 6 was the kind of perceptual category used (Fig. 1a and b). Both experiments used the same semantic manipulation, with images representing either a natural or a manmade object, and participants answering semantic questions about the naturalness of the object. In order to manipulate the perceptual features of objects in Experiment 5, images could be presented at a small or a big size (see Methods section) and participants were asked to respond about the size of the images in the perceptual questions. In Experiment 6, all items appeared on the screen using the same size but they differed in their shape, which was either round or elongated (i.e. perceptual manipulation in Experiment 6, see Method). Here, perceptual questions were about the shape of each item. Accuracy analyses in both experiments revealed that, overall, participants performed well in the visual reaction time tasks (Experiment 5: M = 96.98%; SD = 1.86%; Experiment 6: M = 96.01%, SD = 3.80%) but also in the memory reaction time tasks (Experiment 5: M = 85.14%; SD = 5.70%; Experiment 6: M = 85.78%; SD = 7.22%).

2.1. Reaction times fully replicate previous results in Experiment 5 and a significant interaction between type of feature and task is found in Experiment 6

Aiming to test if the reverse reaction time pattern can be reproduced using different feature manipulations, we followed the same analysis procedure as in Chapter 3. Generalized linear mixed-effect model (GLMM) analyses were performed to model single trial data, including RT as the target variable and three fixed factors: kind of task (visual vs. memory), question type (perceptual vs. semantic) and the interaction between tasks and questions. Participant ID was included as a random effect in all GLMM analyses.

In line with the reverse reconstruction hypothesis and replicating our previous results, these analyses showed that the interaction between task and question type predicted RTs in both experiments. That is, this interaction was replicated using a new semantic (natural vs. manmade) and perceptual (small vs. big) manipulation in Experiment 5 ($F_{1,\ 8810}$ = 62.296, P < .001). Moreover, this interaction effect was found again in Experiment 6 where the semantic category for objects was based on objects' naturalness, and the perceptual manipulation depended on objects' shape (Experiment 6: $F_{1,\ 9804}$ = 294.308, P < .001). Planned comparisons were performed to confirm whether the significant interactions found were due to RT differences in the predicted direction (i.e. perceptual < semantic during visual perception, and semantic < perceptual during memory reactivation). Confirming the anticipated direction and previous

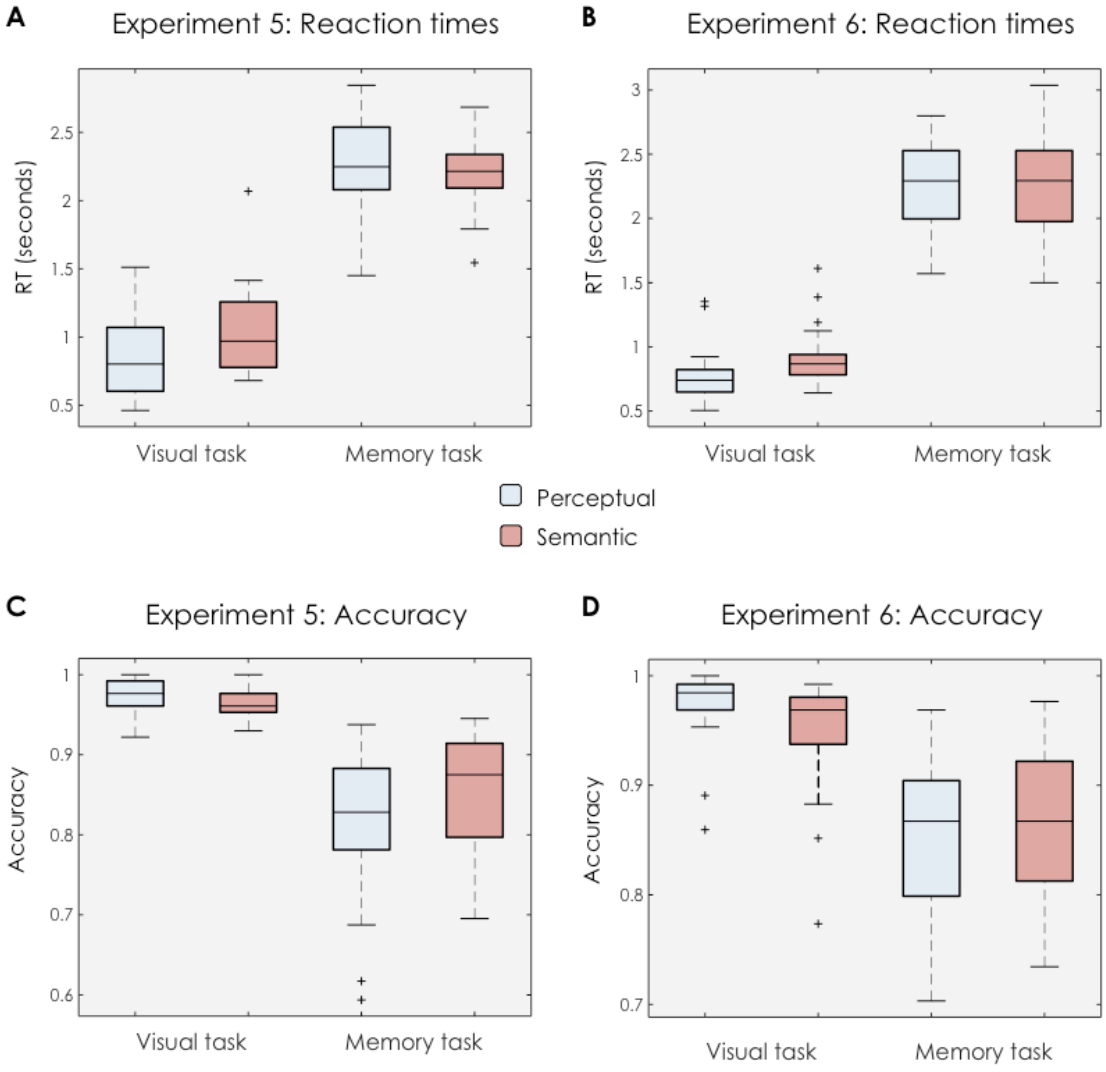
results, these analyses showed that in the visual reaction time task, the type of question significantly predicted reaction times in Experiment 5 ($F_{1, 5054}$ = 429, P < .001) and in Experiment 6 ($F_{1, 5356}$ = 380.794, P < .001). In both experiments, the sign of the model coefficient indicated that RTs for perceptual questions were predicted to be faster than RTs for semantic questions (Experiment 5: B = -.155, t = -20.712, P < .001; Experiment 6: B = -.111, t = -19.514, P < .001). However, in the memory reaction time task RTs were not significantly predicted by the kind of question in any of the experiments (Experiment 5: $F_{1, 3756}$ = 2.559, P = .110; Experiment 6: $F_{1, 4448}$ = .230, P = .631), although in both cases the RTs were estimated to be longer for perceptual questions (Experiment 5: B = .041, t = 1.6, P < .110; Experiment 6: B = .011, t =. 480, P = .631).

One obvious confound in Experiment 5, where our perceptual manipulation was based on the retinal size of the objects (small vs. big), is that we did not control the real-world size of our stimuli. Although objects were presented in different retinal size, the real-world size of some objects (i.e. a palm tree or a surf board) was bigger compared to other items (i.e. an apricot or a CD-ROM). Previous studies have shown that late processing areas in the ventral stream (i.e. parahippocampal gyrus) respond to stimuli's real-world size in a way that interacts with the retinal size (Konkle & Oliva, 2012). Motivated by this work, we carried out the same GLMM analysis as above, but now excluding those objects with a real-world size bigger than a shoebox (see Methods section) from further analysis. After controlling this potential confound, new analyses showed that our previous RT results (Experiment 5 and 6, Chapter 3) were fully replicated.

First, we found that the interaction between task (visual or memory task) and type of question significantly predicted participants' reaction time ($F_{1,\ 6032}$ = 51.170, P < .001). Second, the factor "type of question" predicted RTs in the visual task ($F_{1,\ 3460}$ = 277.312, P = < .001), where participants were significantly faster responding to perceptual questions compared to semantic ones (B = -.149, t =-16.653, P = .001). Importantly, we found that "type of question" also predicted response latencies in the memory task ($F_{1,\ 2572}$ = 4.688, P = .03). But in this case, RTs followed the inverse pattern compared to the visual task and participants responded faster to semantic questions than to perceptual ones (B = .068, t = 2.165, P = .03).

Figure 3a and 3b shows the distribution profile of participant-averaged RTs in each task and type of question for both experiments (controlling real-world size in Experiment 5). When the object was presented on the screen, participants showed a faster accurate response for perceptual questions (Experiment 5: M = 856ms; SD = 295ms; Experiment 6: M = 773ms; SD = 202ms) than for semantic questions (Experiment 5: M = 1.045ms; SD = 343ms; Experiment 6: M = 912ms; SD = 228ms). However, when participants needed to retrieve this object from memory, we did not find clear differences in RTs responding to perceptual (Experiment 5: M = 2254ms; SD = 355ms; Experiment 6: M = 2452ms; SD = 347ms) and semantic (Experiment 5: M = 2178ms; SD = 273ms; Experiment 6: M = 2485ms; SD = 374ms) questions.

In summary, reaction time results confirmed that compared to visual perception, temporal processing of perceptual and semantic features changes during retrieval. The existence of a significant interaction in both experiments suggests that temporal processing for high and low-level information varies depending on whether the object is externally presented or retrieved from memory. Independently of the manipulation used, all RT results of the visual tasks showed lower response latencies when responding to perceptual details. Supporting the reverse reconstruction hypothesis, we replicated our previous behavioural findings (see Chapter 3) in Experiment 5 when the real-world size of items was kept relatively constant. However, in Experiment 6 (i.e. shape manipulation), RT analyses indicated a lack of significant differences between question types.



N. Chapter 5. Figure 3.

Figure 3. Reaction time and accuracy results. Box plots represent reaction times in (a) Experiment 5 (controlling real-world size) and (c) Experiment 6 (b) for perceptual (blue) and semantic (pink) questions when an object was physically presented on the screen (visual task, left) or cued by a reminder (memory task, right). We found that RTs were significantly predicted by an interaction between question type and kind of task in both experiments (P < .001). The type of question significantly predicted RTs for the visual task in both experiments (P < .001), where participants responded faster to perceptual questions. However only in Experiment 5 we found a significant difference in RTs during the memory task; suggesting faster responses when subjects retrieved semantic information. Accuracy results in Experiment 5 (c) and Experiment 6 (d) for perceptual (blue) and semantic questions (pink) when the object was presented on the screen (visual task) or recalled (memory task).

Behavioural analyses showed that an interaction between type of task (i.e. visual or memory) and question type (i.e. perceptual or semantic) significantly predicted accuracy. Planned comparison analyses showed that the kind of question only predicted participants' accuracy for the visual task in Experiment 6 (i.e. indicating a better performance for perceptual questions) and during the memory task in Experiment 5 (i.e. suggesting higher accuracy for semantic questions). In all box plots, the line in the middle of each box represents the median, and the tops and bottoms of the boxes the 25th and 75th percentiles of the samples, respectively. Whiskers are drawn from the interquartile ranges to the furthest minimum (bottom) and maximum (top) values and crosses represent outliers.

2.2. Accuracy results support a reversal between perception and memory

The behavioural findings reported in Chapter 3 also supported the existence of a reversed accuracy profile between the visual and the memory task. When an object was displayed on the screen, participants performed better answering questions about low-level perceptual details compared to semantic questions. However, the opposite profile was found in the memory task: when we asked them to retrieve semantic information from mentally reinstated representations they performed significantly better than when they had to retrieve perceptual details. To examine whether a similar accuracy pattern was also present using our novel feature manipulations (Fig. 1), a series of GLMM analyses were performed. Aiming to confirm whether our fixed effects (i.e. type of task, question type and the interaction between task and question type) predicted participants' accuracy, we used a binary probability distribution and a logistic link function. Again, in all analyses participant IDs were selected as a random factor, including intercept.

In both experiments, we replicated a significant interaction between task (visual vs. memory) and question type (i.e. perceptual vs. semantic question). That is, independent of the type of perceptual and semantic manipulation, results revealed that this interaction significantly predicted participant's accuracy (Experiment 5: $F_{1, 10748}$ = 12.508, P < .001; Experiment 6: $F_{1, 12028}$ = 22.854, P < .001). Next, planned comparisons were run independently for the visual and the memory task in both experiments. In the visual task, we found that question type significantly predicted accuracy in both experiments (Experiment 5: $F_{1, 5630}$ = 4.092, P = .043; Experiment 6: $F_{1, 6142}$ = 27.665, P < .001), with a positive coefficient (Experiment 5: B = -.320, t = -2.023, P = .043; Experiment 6: B = -.735, t = -5.260, P < .001) that suggested a higher accuracy for perceptual questions (Experiment 5: M = 97.44%; SD = 2.17%; Experiment 6: M = 97.33%; SD = 3.31%;) compared to semantic ones (Experiment 5: M = 96.52%; SD = 2.07%; Experiment 6: M = 94.69%; SD = 5.19%;). We performed the same accuracy analyses for the memory reaction time tasks. Results for Experiment 5 revealed that the question type (perceptual vs. semantic) significantly predicted participants' performance during retrieval ($F_{1, 5118}$ = 12.508, P < .001; B = .306, t = 3.836, < .001), where a higher accuracy was found for semantic questions (M = 87.03%; SD = 5.8%) compared to perceptual ones (M = 83.24%; SD = 8.35%). However, analyses for the memory task in Experiment 6 showed that the type of question could not predict accuracy ($F_{1, 5886}$ = .117, P = .732; B = .026, t = .342, P = .732), since participants showed a similar performance retrieving perceptual (M = 83.24%; SD = 8.35%) and semantic information of past representations (M = 85.63%; SD = 7.34%).

The same analyses on subjects' performance were carried out for Experiment 5 when including only objects of a comparable real-world size (i.e. excluding items with a real-world size bigger than a shoebox). Results revealed that the interaction between task and type of question significantly predicted participants' accuracy ($F_{1, 7388}$ = 4.113, P < .043). Further planned comparisons indicated that the type of question (e.g. perceptual or semantic) did not predict accuracy patterns in the visual task ($F_{1, 3870}$ = 0.063, B = .038, t = .251, P = .802). However, when participants were asked to retrieve object representations in the memory task, we found that the type of question was a significant factor predicting accuracy ($F_{1, 3518}$ = 12.480, P < .001) and that subjects were better at retrieving semantic information over low-level details (B = .337, t = 3.533, P < .001).

Overall, accuracy results in both experiments replicated some of our previous findings. First, these outcomes revealed that an interaction between task and the type of question significantly predicted participants' accuracy. However, planned comparisons suggested different profiles per task in each experiment. In Experiment 6, the interaction was driven by a significant difference in the expected direction (perceptual > semantic) in the visual task, but no difference in the memory task. Conversely, in Experiment 5 we found differences in the expected direction in both tasks (perceptual > semantic in the visual task, and semantic > perceptual in the memory task). However, here the visual processing differences disappeared when we took into account the real-world

size of the stimuli, possibly due to the high performance and related low variance in the response patterns in this task.

3. Discussion

In the object recognition literature it is widely assumed that semantic processing of an image is preceded by the integration of its low-level perceptual details. Our previous findings suggested that this neural perceptual processing hierarchy can be measured behaviourally using reaction times, and importantly, that it is reversed when an object representation is reactivated from memory. We followed up these initial findings with two additional behavioural experiments exploring if the reverse reconstruction effect can be replicated when using different low- and high-level features. Instead of animacy, we used naturalness (natural vs manmade) as semantic categories, and retinal size (Experiment 5) or shape (Experiment 6) as perceptual categories. RT and accuracy analyses showed that in general, the interaction pattern indicating a processing reversal between perception and memory remains robust to the various feature manipulations. Some effects were shared by both of the new experiments, while others varied depending on the kind of perceptual manipulation used.

Starting with those effects common to both experiments, we found that independently of the perceptual manipulation (retinal size of objects in

Experiment 5; shape of the objects in Experiment 6), subjects were significantly faster answering perceptual questions compared to semantic ones (naturalness) when the object was physically presented on the screen (visual reaction time task). In terms of accuracy, when objects were visually perceived, participants exhibited better performance for perceptual than for semantic questions in Experiment 6 (shape manipulation), and in Experiment 5 (but only if real-world size of objects was not controlled; when we controlled this factor, participants showed a ceiling effect for perceptual and semantic questions). Importantly, in both experiments GLMM analyses indicated an interaction between task (visual or memory reaction time task) and the type of question (perceptual or semantic) that significantly predicted RTs and accuracy. That is, given that the visual task results suggest a clear prioritization of perceptual details over semantic information, the significant interactions indicate that in the memory task the processing hierarchy changed in both experiments. The most important difference between the two perceptual manipulations (i.e. size in Experiment 5 and shape in Experiment 6) was the way in which the processing hierarchy was modified when retrieving objects from memory. When controlling real-world size of objects in Experiment 5, subjects were faster and more accurate remembering semantic categories of past representations compared to perceptual details. This pattern of results fully replicates our previous behavioural findings (Chapter 3). Contrary to our predictions, however, behavioural analyses of Experiment 6 indicated a lack of a difference between perceptual and semantic questions during memory retrieval, both in terms of RT and accuracy.

In the following, potential explanations will be discussed for why the pattern of results diverges in the memory task when using retinal size compared with shape. Note that the expected forward stream was found during visual object processing with both manipulations. Why, however, was a reverse reconstruction effect during memory retrieval found when we used retinal size as the low-level feature manipulation, but not when using object shape instead?

Human perception of objects is always the result of a dynamic interplay between low and high-level feature processing along the visual ventral stream. It is worth noting that retinal size and shape represent visual information that is processed at different stages of the neural visual pathway. Compared to the retinal size of objects that is processed in early visual areas which are spatially organized (Felleman & Van Essen, 1991; Fox et al., 1986), shape processing is associated with higher-order visual areas (e.g. V4) that are located closer to semantic processing areas (for a review, see Connor, Brincat, & Pasupathy, 2007). In fact, findings in monkeys indicate that shape similarity is more strongly represented than semantic categories in inferotemporal neurons that are typically regarded as late, high-level ventral visual stream areas (Baldassi et al., 2013). It should be noted that these putative differences in the processing stage along the visual stream were not obvious during the visual task in terms of reaction times or accuracy, because in fact in both experiments we found the same pattern consistent with a forward stream. However, the differences in neuronal processing stage might indicate different degrees of dependency

between the low-level perceptual and the semantic categories used (natural vs. manmade). In particular, we suspect that there will be a more intimate link between late perceptual stages (like shape) and semantic processing than between more early perceptual stages (like retinal size) and semantic processing. During the last decade, a growing body of evidence has pointed out how prior knowledge (as semantic information) can guide low-level processing by resolving ambiguities under certain circumstances (e.g. degraded visual input) (for a review, see Panichello, Cheung, & Bar, 2013). Importantly, it has been shown that higher visual processing stages (i.e. shape processing) depend on long-term memory, and that the tuning of neurons at this processing stage is strongly dependent on experience with object categories (Connor et al., 2007). That is, having access to semantic categories helps to predict perceptual features that are processed at a later stage (i.e. shape) compared to those that depend on early visual areas (i.e. retinal size). Taking this evidence into account when interpreting our findings, we can hypothesise that if semantic information is available early on during memory reconstruction, it will be easier to reactivate (or predict) those higher order perceptual details (e.g. shape) than the lower order ones (e.g. retinal size). In other words, when retrieving the semantic category of a past representation (e.g. whether an object is a fruit or not), it will be more likely to gain fast and accurate access to perceptual details about shape than details about retinal size. This is particularly true if the perceptual information is highly predictable from the semantic category (e.g. round shapes are more associated to fruits than square shapes). In an extreme case, it can be argued that as soon as the participant remembers that the

associated object is a banana, the shape information will come “for free”. On the other hand, remembering that the object was a fruit or a banana has no predictive value with respect to our other perceptual category, retinal size. That is, in order to correctly answer the perceptual question about retinal size, participants will need to go beyond the semantic level and reconstruct the associated memory down to a very low level of perceptual detail. This rationale can explain why in our memory task, there were no RT differences between semantic and perceptual questions when using shape as perceptual feature, but there was evidence for a reverse stream when using retinal size as perceptual feature. Within this line of reasoning, our results suggest that a robust reverse processing stream can be found during memory retrieval when object semantics are not predictive of perceptual details, like in the case of Experiment 5 (retinal size) and all previous experiments reported in Chapters 3 (using photographs vs line drawings as perceptual manipulation). We are aware that it is difficult to draw a clear conclusion based on this post-hoc explanation, and that new experiments are necessary to explicitly probe this “dependency hypothesis”.

In our life, what we perceive remains available for just a small fraction of time. Thinking about the past or imagining our future means to bring back to our mind some representations that are no longer in front of our eyes. Along a series of experiments, we added more evidence to a widely accepted and important idea in the memory field: retrieval is a dynamic process that emphasises some types

of information over other details (Schacter, 2012; Schacter et al., 2011). Although sometimes our memories seem to re-appear vividly in front of our mental eye, the neural processes and time dynamics behind these reconstructions seems to be different compared to visual perception. Our present findings shed light onto how different perceptual and semantic features of past representations are reactivated from memory, compared to the processing order observed during memory formation (or encoding).

4. Methods

4.1. Participants

Forty-seven volunteers (43 female; mean age 19.09 +/- 0.9 years old) participated in Experiment 5. Twenty-four of them completed the visual reaction time task. In this visual task, two participants were excluded since both showed a low general accuracy (< 65%) compared to the rest of the group (t_{21} = 12.99, P < .001). The other 23 participants performed the memory reaction time task. In this experiment, 3 participants were excluded due to poor memory performance (< 70%) compared with the rest of the group (t_{21} = 6.0190, P < .001).

A second group of 49 volunteers (43 female; mean age 19.16 +/- 0.83 years old) were recruited for Experiment 6. In total, 24 of them took part in the visual reaction time task; the remaining 25 participants completed the memory

reaction time task. Two participants that completed the memory task were not included in the final analysis due to poor performance (< 70% memory accuracy) compared to the rest of the group (t_{23} = 3.90, P < .001).

All subjects confirmed having normal (20/20) or corrected-to-normal vision, normal colour vision, no history of neurological disorders and being native or highly fluent English speakers. All of them gave us written informed consent before the beginning of the experiment. Participants were blind with respect to the aim of the experiments, although they were debriefed at the end of the experimental session. We compensated participants for their time, receiving course credits or £6 per hour of participation. All experiments were approved by the University of Birmingham's Science, Technology, Engineering and Mathematics Ethical Review Committee.

4.2. Stimuli

Experiment 5

A dataset of 128 pictures of unique everyday objects and animals was used in this experiment. In addition, 16 extra images were used only during the training block. These images were selected from online royalty-free databases and from the BOSS database (Brodeur et al., 2010). To create a semantic manipulation, half of these images represented manmade objects (i.e. electronic devices, tools, music instruments and sport equipment) and the other half depicted natural elements from flora and fauna. To generate a perceptual manipulation,

half of these 128 images were pseudo-randomly selected for each participant to be displayed at the centre of the screen with a rescaled size of 200x 200 (4.17 degrees of visual angle), and the other half with a size of 533 x 533 pixels (19.30 degrees of visual angle), with the restriction that both sizes were equally distributed across the natural and manmade object categories. For a better control of low-level perceptual details, all images were transformed into grey scale images and luminance was controlled by means of the SHINE toolbox (Willenbockel et al., 2010). In this experiment, object shape (i.e. rounded and elongated shape), even though not constituting one of the relevant response categories, was counterbalanced across all semantic and size categories. There was thus no systematic confound between shape and any of the relevant object categories. We initially did not control the real-world size of these items, but we retrospectively decided that it is important to account for this factor, given the existing literature (Konkle & Oliva, 2012). Therefore, only objects that fit in a shoebox were included in our principal analyses (i.e. a total of 88 objects), excluding a total of 14 natural and 26 artificial objects. We used a chinrest to control that the distance to the screen was the same (70 cm) for all subjects.

Experiment 6

In this experiment, the same 128 images and 16 training images as for Experiment 1 were used. All these images were colour photographs of objects

and animals on a white background. Compared to Experiment 5, the semantic manipulation was the same: 64 objects illustrated natural elements and 64 represented manmade objects. In order to produce two different perceptual categories based on shape, we used the open source GNU image manipulation software (www.gimp.org). Half of the 128 images (32 natural and 32 manmade objects) were selected based on their round shape to fit with the round shape template as much as possible. The other half of the images were chosen based on their long shape to fit the long shape template. All images were displayed at the centre of the screen with a rescaled size of 500 x 500 pixels.

In both experiments, a total of 128 action verbs were selected as associative cues for the memory reaction time task. All verbs used were the same in Experiment 1, 2 and 3,

4.3. Data Collection

Stimulus presentation and behavioural response recording were performed using Psychophysics Toolbox Version 3 (Brainard, 1997) running under MATLAB 2014b (MathWorks). Directional arrows of a computer keyboard were used as buttons for response input. For the experimenter it was not possible to be blind to experimental conditions during data collection and analysis of both experiments.

4.4. Procedure

Visual reaction time task (Experiment 5 and 6)

Visual reaction time tasks followed the same procedure as reported in Experiment 1 of Chapter 3, except for the following changes (Fig. 2): On each trial, participants were asked to respond as quickly as possible either to a perceptual or a semantic question about the presented object. In Experiment 5, perceptual questions asked whether the object was displayed on the screen in a small or big size. In Experiment 6, perceptual questions were about the shape of the object (rounded or elongated shape). In both experiments, semantic questions were about whether the object represents a natural or a manmade object. For all questions, the two possible answers were displayed at the two opposite sides of the screen (right or left), mirroring the directional arrows of a computer keyboard which was used as response input. To keep the response mapping easy, the options for “natural”, “small” and “long” were always located on the right side of the screen.

Memory reaction time task (Experiment 5 and 6)

Both memory reaction time tasks used in Experiment 5 and 6 followed the same procedure described in Chapter 3 for Experiment 1. As reported for the visual reaction time task, the main differences were the perceptual and semantic questions that participants answered. Perceptual questions in Experiment 5 asked about the size of objects (small vs. big), and about object’s shape in

Experiment 6 (rounded vs. elongated). Semantic questions were about whether the object was natural or manmade.

4.5. GLMM analyses

Aiming to test our memory reconstruction hypothesis, we used generalized linear mixed models (GLMMs) to analyse participants' accuracy and RTs on the single trial level in both experiments. These models were set up in the exact same was as for the behavioural analyses in Chapter 3. For accuracy analyses, we used a binomial distribution with a logistic link function. For RTs, we selected a gamma distribution and an identity link function based on previous literature (Lo & Andrews, 2015) and evidence from AIC and BIC compared to other models (e.g. inverse Gaussian or normal distributions). Participant ID was selected as a random factor (including intercept) in all analyses. Only correct trials were used for RT analyses, and all trials that exceeded a RT of average RT over subjects +- 2.5 SDs were rejected.

5. Acknowledgments

We thank Sophie Watson, Wing Tse, and Jonathan Burton-Barr for helping with data collection.

6. Author contributions

J.L.D. and M.W. designed the experiments. J.L.D. conducted the experiment, analysed the data and wrote the chapter. M.W. contributed in reviewing and editing. All authors contributed to the analysis approach and to data interpretation.

Chapter 6: An optimal oscillatory phase for pattern reactivation during memory retrieval

Casper Kerrén[1]*, Juan Linde-Domingo[1]*, Simon Hanslmayr & Maria Wimber

[1]School of Psychology & Centre for Human Brain Health (CHBH), University of Birmingham (UK)

* These authors contributed equally

This chapter represents a near identical manuscript that has been accepted for publication in Current Biology. An earlier pre-printed version of this manuscript has also been published in Sneak Peek.

Abstract

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

1. Introduction

Our episodic memory defines us by storing a record of our past experiences and allowing us to consciously access these records. It is widely agreed that the hippocampus and neocortical areas work in conjunction during the formation and later retrieval of a memory (McClelland et al., 1995; E T Rolls, 1996; T J Teyler & DiScenna, 1986; E Tulving & Markowitsch, 1998). At encoding, the hippocampus is thought to continuously store a sparse and non-overlapping index that points to ongoing activity patterns in cortical space. This hippocampal index can later be reactivated by a reminder, and lead to the reconstruction of a previously stored memory pattern in neocortex (Alvarez & Squire, 1994; Marr, 1971; McClelland et al., 1995; K. A. Norman & O'Reilly, 2003; T J Teyler & DiScenna, 1986). Many recent studies have tested these computational assumptions by tracking the reinstatement of memory-related brain activity patterns during retrieval. The basic premise that content-specific neural patterns are reactivated during retrieval has been confirmed using fMRI (for reviews, see Danker & Anderson, 2010; Rissman & Wagner, 2012) and more recently also EEG and MEG (Jafarpour et al., 2014; Johnson, Price, & Leiker, 2015; Kurth-Nelson et al., 2015; Michelmann et al., 2016; Staudigl, Vollmar, Noachtar, & Hanslmayr, 2015; Waldhauser et al., 2016; Wimber et al., 2012). However, no study has so far investigated the temporal fluctuations of memory-related patterns in human long-term memory, and whether they are systematically linked to brain oscillations.

A major computational challenge for our memory system is to effectively separate the information arriving from external sensory sources from the information generated in internal circuits. In other words, if the brain constantly pattern completes, how does it make sure that the neural coding of this internally (and possibly incorrectly) generated information does not interfere with the coding of new, incoming information? One promising explanation suggests that this is accomplished by means of neural oscillations. In particular, it has been argued that the phase of the hippocampal theta oscillation supports the chunking of mnemonic information such that the neural assemblies involved in encoding and retrieval are temporally segregrated (Hasselmo et al., 2002; Nyhus & Curran, 2010). In a seminal paper, (Pavlides et al., 1988) showed that stimulating a hippocampal assembly at one phase of the theta rhythm induced long-term potentiation (LTP), whereas stimulating at the opposite phase induced long-term depression (LTD). This finding has since been replicated many times in rodents (Huerta & Lisman, 1993; Hyman et al., 2003), and implemented in computational models of episodic memory and the hippocampus (Buzsáki, 2002; Hasselmo et al., 2002; Hasselmo & Eichenbaum, 2005; Kunec et al., 2005; Parish et al., 2018). These models share the assumption that successful retrieval is most likely at one specific phase of the hippocampal theta rhythm, opposing the optimal encoding phase (Hasselmo, 2005; Hasselmo et al., 2002). Memory retrieval should be a continuously oscillating process that is locked to the hippocampal theta phase .

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

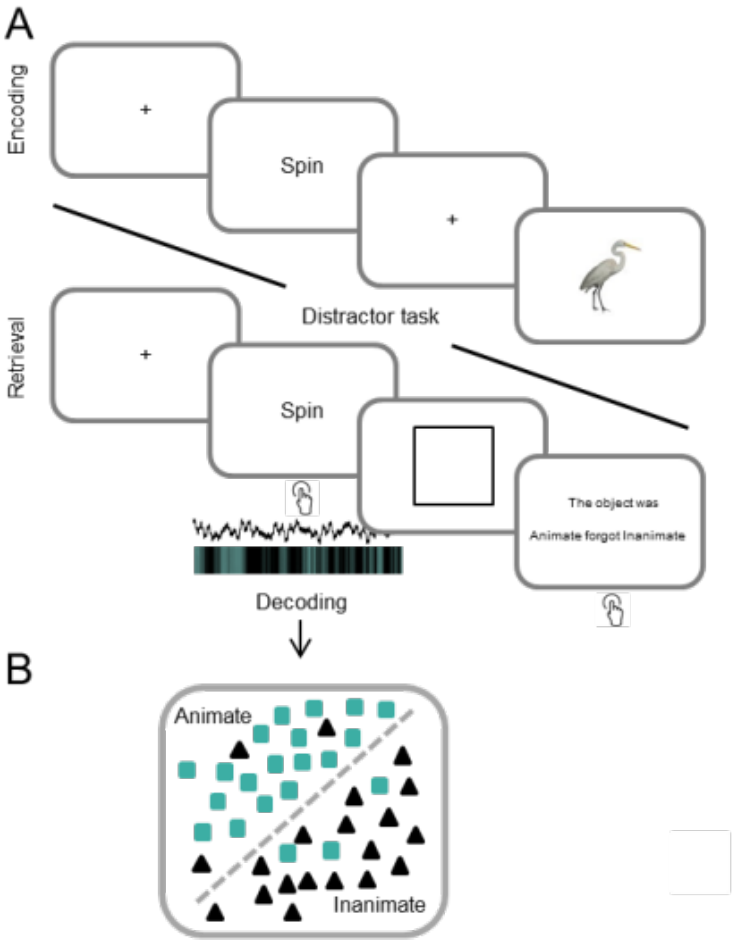
2. Results

2.1. Participants retrieve the episodic memories with high accuracy

The paradigm was a simple word-object associative memory task designed to yield a high number of correct trials (Figure 1A). Participants studied

associations between action verbs and objects in random pairings, and were later cued with the word to retrieve the object. Two measures of memory accuracy confirmed that participants performed the task well. The first was a subjective measure where participants indicated, via a button press after cue onset, whether and when they recalled the associated object. Participants on average indicated that they remembered the object on 94.21% (SD = 5.75%) of the trials. A second, more objective measure was accuracy in response to a question about the object's semantic category (animate vs inanimate), which appeared at the end of each retrieval trial, and which participants answered correctly on 88.20% (SD = 6.57%) of the trials. These two measures were highly correlated ($r_{Spearman}$ = 0.60, $p < .05$). Average accuracy for perceptual detail (photograph vs line drawing) was 85.31% (SD = 6.45%).

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



O. Chapter 6. Figure 1.

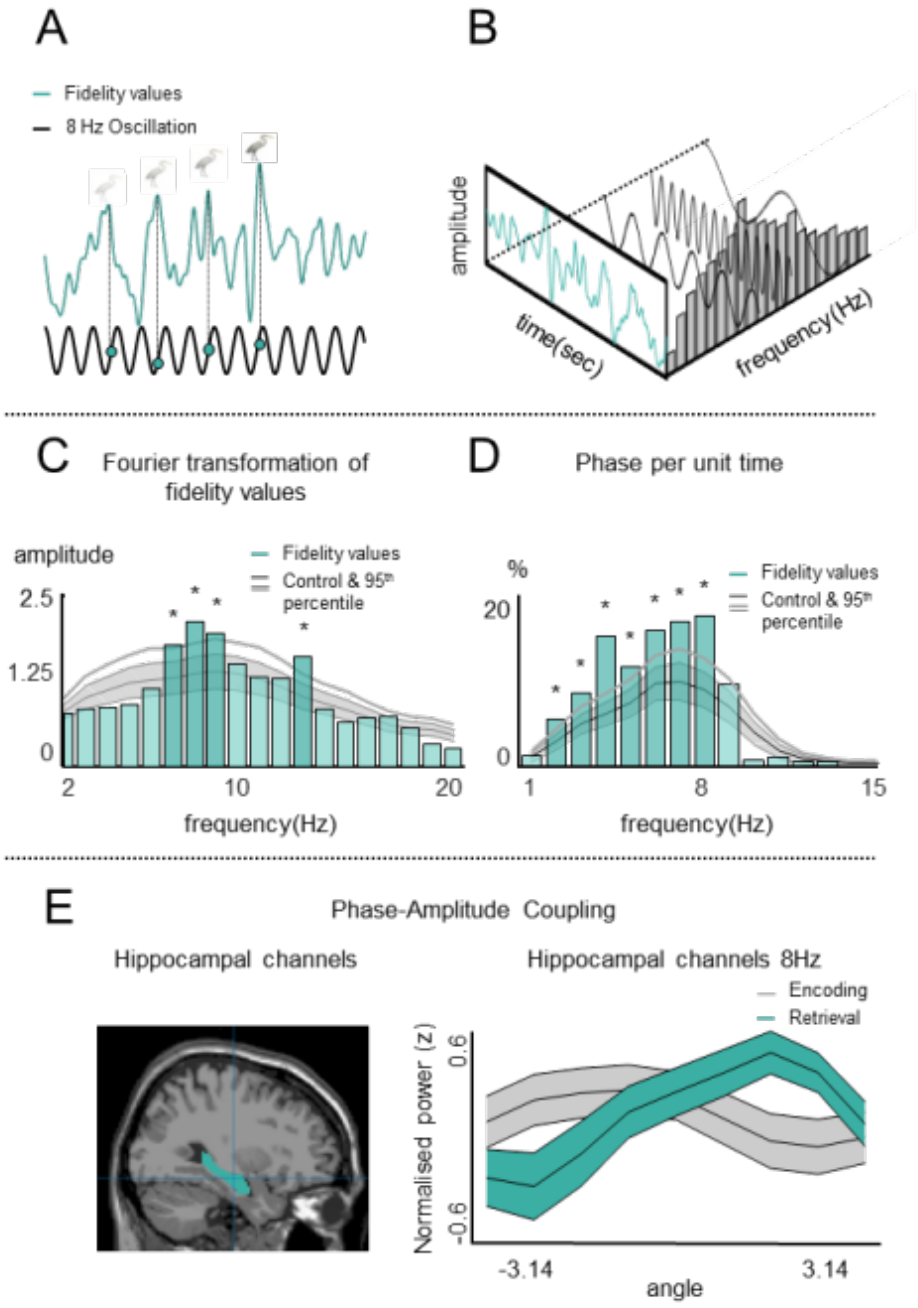
Figure 1. Trial structure and Multivariate Pattern Analysis. (**A)** At encoding, participants associated action verbs with images depicting either an animate or inanimate object. After a short distractor task, participants were tested on the previously learned associations. The action verb was shown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

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

Our primary goal was to test whether the neural signatures of memory retrieval wax and wane in a theta oscillatory rhythm. Our neural index of memory retrieval was obtained from a linear discriminant analysis (LDA) trained to detect evidence for the reactivation of the correct object category (animate vs. inanimate) during retrieval (Figure 1B, see methods for details). The LDA was trained and tested independently per participant at each retrieval time point starting with the onset of the word cue, using a leave-one-out procedure. The input into the LDA was a feature vector containing the signal amplitudes from all 128 EEG channels at a given time point. The major output of interest was the fidelity (distance, or d-) values available for each trial and time point. These values represent the distance from the hyperplane that optimally separates the two classes of retrieved objects (animate vs inanimate), and their timecourses served as our time-resolved, parametric index of memory reactivation. For the purpose of this study, the LDA was trained and tested during cued recall in order to isolate a purely retrieval-based signature or memory retrieval, which could then (below) be compared with a purely encoding-based index of memory classification. Additional analyses using classifiers trained on encoding and tested at retrieval are reported in the supplementary materials (Figure S1 and S4).

We first asked whether evidence could be found for an oscillation in these time-resolved indices of memory reactivation (Figure 2A-B). Fidelity timecourses from the recall task were averaged across trials per participant and subjected to a Fourier Transformation. If memory reactivation fluctuates in a theta rhythm, the resulting power spectra will show a selective increase in a band-limited lower (theta) frequency band. We compared the power spectra obtained from the real classifier outputs with a bootstrapped baseline (Stelzer et al., 2013), the latter using the d-value outputs from classifiers that were trained and tested on the same EEG trials but with randomly shuffled category labels (see Method section). This procedure controls for spurious power peaks that are driven by the frequency characteristics of the raw data (e.g. a dominant oscillation in the single trials). Significant power differences between the real and shuffled data were found in frequency bins at 7-9Hz and 13Hz, all exceeding the 95th percentile of the empirical null distribution (Figure 2C). Power at 7-9Hz was significantly higher ($t(1,23) = 1.9425$, $p = .03$) when including only correctly retrieved trials that when including all trials, suggesting a relationship of the classifier fluctuation to memory success (Siegel, Warden, & Miller, 2009). An alternative method with more stringent criteria to determine the presence of oscillations (Watrous, Miller, Qasim, Fried, & Jacobs, 2018) confirmed that oscillatory power in the classifier time series was increased above baseline in the 7-9Hz frequency range (Figure 2D). Moreover, a similar power spectrum was found when the classifier was trained on encoding and tested on retrieval (Figure S4.) The frequency characteristics of the classifier fidelity time courses

thus suggest a rhythmic fluctuation in memory reactivation that was most consistent in the 7-9Hz frequency range.



P. Chapter 6. Figure 2.

Figure 2. Analysis rationale and results of the time-frequency analyses relating classifier fidelity to theta oscillations and phase-modulation. (A) Example of a single-trial output from the LDA, reflecting the fidelity of the classifier in detecting the retrieved object's correct category at each time point during a retrieval trial. The black line represents a theta oscillation to illustrate

our assumption that neural indices of memory reinstatement (i.e., the d-value time series) rhythmically fluctuate, and that the time points of maximal classifier fidelity should be consistently related to a particular phase of the underlying oscillation. **(B)** D-values were subjected to a Fourier transformation which reveals the power in each frequency band. **(C)** The resulting power spectrum shows significant deviations from an empirical null distribution at 7 to 9Hz and 13Hz. The baseline power spectrum was obtained from a combination of random label classifiers and bootstrapping, and is shown in grey (mean and SD). Values of the real classifier outputs exceeding the 95th percentile of the baseline distribution are marked as significant. **(D)** Frequency decomposition of the classifier time series using an alternative approach to detect frequencies (Oostenveld et al., 2011), again showing above baseline power at slow frequencies including 7-8Hz. Figure showing mean ± SEM for baseline (grey lines) and 95th percentile (thick grey line). **(E)** Phase-amplitude coupling between EEG phase and classifier fidelity at source level revealed a significant modulation index averaged over hippocampal virtual channels (mask shown on the left) for 8Hz. Figure showing mean ± SD. See also Figure S4.

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

Our next two analyses were aimed at specifically testing for coupling between neural reactivation (i.e., classifier timeseries) and the phase of hippocampal theta-band oscillations. For this purpose, the raw EEG trials were projected into source space using an LCMV beamforming algorithm (Gross et al., 2001; Van Veen, van Drongelen, Yuchtman, & Suzuki, 1997), and a hippocampal mask was used to extract the 8Hz phase of the hippocampal virtual channels for each trial and time point. We computed a phase modulation index (MI) (Bialek, de Ruyter vab Steveninck, Rieke, & Warland, 1997) reflecting the strength of coupling between the hippocampal 8Hz phase and the amplitude of the classifier output. Classifier fidelity as a function of hippocampal theta phase is

plotted in Figure 2E (green line). This analysis revealed a significant modulation index (M = .0071, SD = .0042; baseline: M = .0056, SD = .0006), t(1,23) = 1.8191, p < .05, one-sided t-test, indicating that fidelity of the retrieval classifier was modulated by the phase of the hippocampal 8Hz oscillation (Figure 2E).

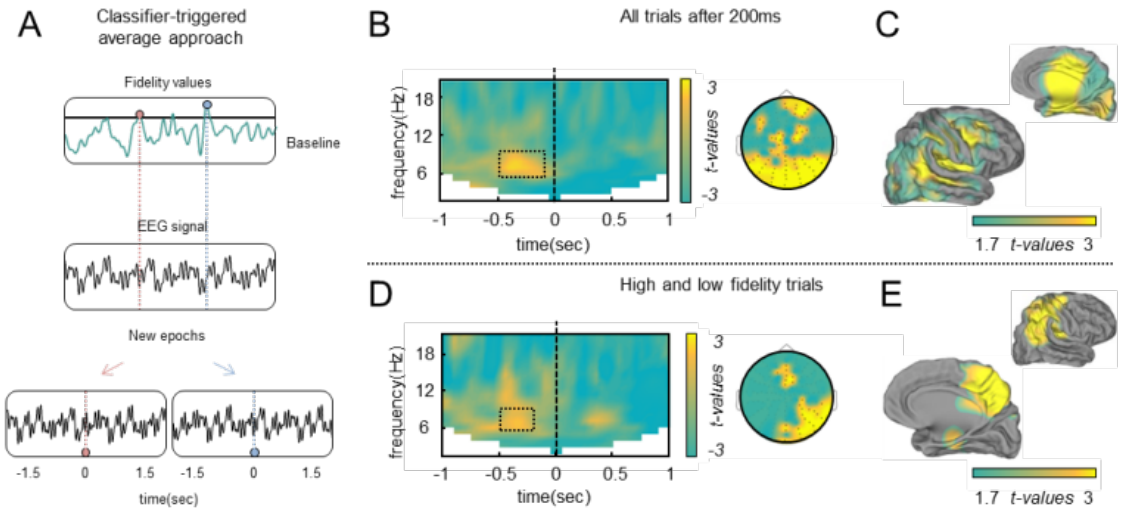
We next directly compared the theta phase at which classifier fidelity was maximal during encoding and retrieval. All basic analysis steps were repeated for the encoding EEG data, where an LDA discriminating animate from inanimate objects was trained and tested at each time point from 200ms before until 1500ms after object onset. The full time generalization matrices showing classifier performance for encoding and retrieval can be found in Figure S1. The 8Hz phase at encoding was then extracted from hippocampal virtual channels to calculate the phase modulation index. Classifier fidelity as a function of hippocampal theta phase during encoding is shown in Figure 2E (grey line). A significant phase modulation was found also for encoding (M = .0068, SD = .0029; baseline: M = .0052, SD = .0007), t(1,23) = 2.7494, p < .05, one-sided t-test). In order to directly compare the encoding and retrieval phases, we identified the phase at which encoding or retrieval classification was optimal in each subject. A Rayleigh circular statistic comparing the absolute phase angles at which encoding and retrieval classification was maximal revealed that these angles significantly differed from each other, z(1,23) = 5.5342, *p* = .001. Similar statistics were obtained by fitting a sine wave to the data and identifying and extracting the phase at which classification was optimal. Together, the results of

the phase modulation analyses show that retrieval fluctuates as a function of hippocampal theta (8Hz) phase, and that the optimal retrieval phase is on average 188 degrees phase shifted compared with the optimal phase during encoding.

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

Having established that the neural retrieval patterns oscillate and are coupled to an 8Hz oscillation, we next investigated the temporal relationship between theta phase and memory reinstatement. The analysis was inspired by the use of spike-triggered averages in animal intracranial work (Bialek et al., 1997; Douchamps, Jeewajee, Blundell, Burgess, & Lever, 2013). We here adopted a similar approach computing *classifier-locked averages* around the time points of maximal memory reactivation (see Methods for details). On each single trial, those time points of maximal classifier fidelity that exceeded the 95th percentile of a bootstrapped baseline were marked as new events of interest, the corresponding time stamps were located in the raw EEG epochs, and the ongoing EEG signal surrounding these maxima was then analysed for phase consistency across all electrodes (Figure 3A). We used a non-parametric cluster-based permutation test to compare the real data with a temporally shuffled baseline that keeps the EEG trial structure intact but produces a random temporal alignment between the classifier maxima and the ongoing phase (see Method section). Comparing the “real” times of maximum classifier

fidelity with the temporally shuffled baseline revealed a cluster of significant (p_{corr} < .05) phase consistency from 500ms to 50ms before the classifier maxima, centred at 7Hz (Figure 3B). Note that in this analysis, several classifier peaks per trial can exceed the 95th percentile criterium, and many of the classifier-locked EEG epochs will thus overlap, resulting in temporal smearing of the phase-locked activity. When running the same analysis extracting only one maximum per trial (Figure S2C), we found a similar cluster of phase locking but with a more narrow temporal extent from 500ms to 150ms pre-maxima, suggesting that the stongest phase-consistency effect was present roughly two theta cycles (corresponding to 2*143ms = 286ms) before mnemonic information could be confidently decoded. This finding supports our primary hypothesis that memory reinstatement shows a consistent oscillatory timing across trials and participants, in the same 7-9Hz frequency band at which the classifier fluctuates (Figure 2C).



Q. Chapter 6. Figure 3.

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

It might seem counterintuitive that the strongest phase consistency was observed prior to the time points of maximum classification fidelity, rather than at the maxima themselves. However, this temporal relationship is to be

expected if the phase-locked signal originates from a different, upstream region in the processing hierarchy compared to the signal that the classifier's decision is based on. Our findings are consistent with a model where the re-instantiation of a memory trace is triggered at a consistent phase of a hippocampal/MTL theta oscillation, followed by memory reinstatement in a broader range of neocortical regions representing the stored memory (Buzsáki, 2002; McClelland et al., 1995; Randall C. O'Reilly et al., 2014; Randall C. O'Reilly & Norman, 2002). The aim of the next analysis was to identify the brain regions involved in producing the observed clusters of theta phase consistency, with the hypothesis that the effect should be present in MTL areas (McClelland et al., 1995; T J Teyler & DiScenna, 1986).

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

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

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

Comparing classifier maxima of higher and lower fidelity revealed a significant (p_{corr} < .05*)* cluster at 7Hz preceding the maxima by 500ms to 200ms (Figure 3D). This cluster highly overlapped in timing, frequency and topography with our previous classifier-triggered average analyses. When conducting the same analysis in source space, we found sources that spanned the parietal and the right medial temporal lobes (maximum MNI coordinates xyz = 50 -30 30, inferior parietal lobule; Figure 3E), strongly reminiscent of the core recollection or memory success network typically found in fMRI studies (Rugg & Vilberg, 2014). Our data thus suggest that the neural signatures of memory retrieval are linked to a specific phase of a theta oscillation, and this phase relationship becomes stronger with more confident neural reactivation. The source level analysis additionally confirms our a priori assumption that the phase-locked signal that precedes memory reactivation involves the MTL and other core recollection areas.

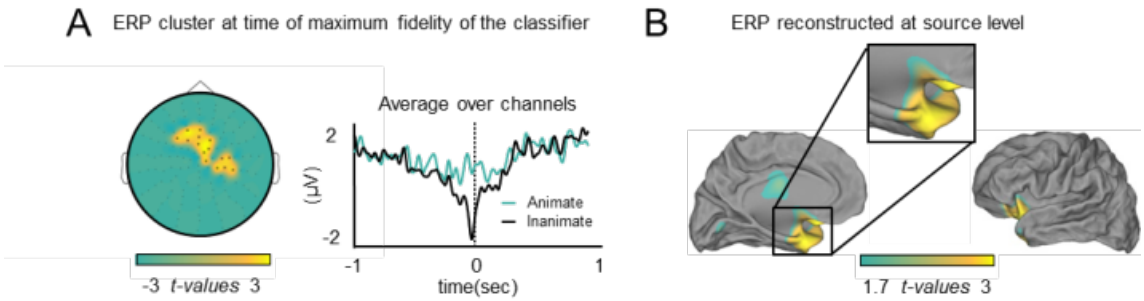
2.6. Theta phase-locking is unlikely to be produced by early cue-related effects

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

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

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

The second ERP analysis again contrasted animate and inanimate trials, but this time locked to the time points of maximal classifier fidelity (as used in previous analyses). A cluster-based permutation test revealed a significant cluster ($p_{corr} < .05$) over frontal electrodes, spanning from 90ms before to 120ms after the classifier maxima (Figure 4A). Reconstructed at source level (Figure 4B), this effect showed a maximum in left anterior temporal lobe (maximum MNI coordinates xyz = -30 10 -40, superior temporal gyrus; and -40 0 -40, inferior temporal gyrus). The results confirm that the single-trial classifier maxima indeed reflect a meaningful difference in the neural patterns elicited by retrieving different types of objects, rather than reflecting random fluctuations in classifier performance. The most likely source of the effect was found in anterior temporal lobe, an area strongly linked to semantic memory processing (Duvernoy, Cattin, & Risold, 2013; Patterson, Nestor, & Rogers, 2007), where previous studies found tight links between classifier fidelity and the speed at which participants behaviourally categorize objects as animate or inanimate (T. A. Carlson et al., 2014). Together, the two ERP analyses validate our LDA approach and provide converging evidence that retrieval-related differences between animate and inanimate objects fluctuate in the theta range and are most pronounced over neocortical regions involved in high-level semantic processing (Tyler et al., 2013).



R. Chapter 6. Figure 4.

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

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

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

frequency characteristics are similar to our main, purely retrieval-based metric. The results are summarized in Figure S4.

Encoding analyses were conducted on epochs time-locked to the onset of the animate and inanimate objects (-200ms to 1500ms). As a first step, an LDA was trained on encoding and also tested during encoding (Figure S4A). In line with the existing literature on object perception (T. Carlson et al., 2013), animate vs inanimate category membership could be best decoded in a time window around 300ms after object onset, with an accuracy peak at 305ms. The classifier fidelity timecourses were then averaged within participants and subjected to a Fourier Transformation, following the same procedure as for the retrieval data. The resulting spectra (Figure S4B) showed the strongest power in lower frequencies with peaks at 3, 5, and 6 Hz exceeding the 95^{th} percentile of the random label chance distribution.

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

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

3. Discussion

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

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

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

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

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

neural memory reinstatement as measured by EEG classifiers, in line with the computational models that motivated our hypotheses (Hasselmo et al., 2002; Ketz, Morkonda, & O'Reilly, 2013; Kunec et al., 2005).

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

represent the various components of an episode (McClelland et al., 1995; T J Teyler & DiScenna, 1986).

The exact time course of the interaction between hippocampus and neocortex during retrieval is still not fully understood. Electrophysiological studies using time-resolved multivariate methods have detected memory reactivation in the typical recollection time window (Jafarpour et al., 2014; Johnson et al., 2015; Michelmann et al., 2016). Consistent with this timing, our classifier maxima had a tendency to cluster in the recollection window around 400-800ms post-cue (Figure S2A). Our main interest in this study, however, was whether neural reactivation was linked to a consistent oscillatory phase in the theta band irrespective of when exactly it is triggered within a trial. Our findings provide strong evidence for such phasic modulation within a recall trial, in line with models suggesting that memory retrieval is initiated at an optimal phase of a hippocampal theta oscillation (Hasselmo et al., 2002).

At the exact time of the classifier maxima, we observed a significant difference in the ERPs distinguishing between the different types of retrieved memories (i.e., animate vs inanimate, Figure 4). The main source of this difference was localized to the anterior temporal lobe, consistent with this region’s role in representing abstract object information (Patterson et al., 2007). Note that it is not surprising that we observed such an ERP effect, since the classifier requires a reliable signal difference in order to detect differences in reactivated content.

The source of this signal is interesting, however, indicating that the classifier's decisions are based on information that originates from neocortical sources that are likely to represent the reactivated memory's content, and have little overlap with the sources of the theta phase-locked signal. Overall, our findings suggest that a few hundred milliseconds before the brain reinstates a memory in neocortex, an oscillating process in the MTL initiates retrieval, leading to a memory signal that oscillates and is modulated by the hippocampal theta phase (T J Teyler & DiScenna, 1986; Timothy J. Teyler & Rudy, 2007).

To our knowledge, our study is the first that directly links memory reinstatement to theta phase in human long-term memory. Previous studies have investigated the role of theta phase in working memory, and have provided first evidence for a phase shift between encoding and retrieval (Rizzuto, Madsen, Bromfield, Schulze-Bonhage, & Kahana, 2006). They also suggest that theta phase plays a role in orchestrating gamma (30-80Hz) oscillations during periods of working memory maintenance (Fell & Axmacher, 2011; O Jensen, 2006). High frequency activity in the gamma range is thought to represent the firing of cell assemblies that code for the content of mental representations, and lower frequencies presumably provide the time windows for the firing of these assemblies (Fell & Axmacher, 2011; Fuentemilla et al., 2010; O Jensen, 2006; Nyhus & Curran, 2010; Tallon-Baudry & Bertrand, 1999). Following this logic, Fuentemilla et al. (Fuentemilla et al., 2010) used a delayed match-to-sample working memory task to investigate how gamma patterns representing the encoded material re-emerged during maintenance. Reactivation took place several times over a 5-sec delay, and these reactivations were phase-locked to

a theta oscillation. Rodent work also suggests a link between gamma oscillations and theta phase. Different hippocampal subfields produce faster or slower gamma oscillations depending on whether the animal is encoding novel information or retrieving familiar information, and these two gamma rhythms are coupled to distinct phases of the hippocampal theta rhythm (Colgin et al., 2009). Our results provide the first evidence for a similar relationship in human long-term memory, using a classifier-based metric rather than gamma oscillations as a proxy for memory reinstatement and its relationship to the ongoing EEG.

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

encoding, during offline periods following encoding, and when reactivating memories during retrieval.

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

4. Methods

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

4.1. Participants

Twenty-four healthy participants (19 female) aged 18-32 years (mean = 22.1, SD = 4.7 years) received credits or monetary payment for participation.

Participants had normal or corrected-to-normal vision and reported no history of neurological disorders.

4.2. Material and Setup

The material consisted of 64 images depicting animate objects (equal number of mammals, birds, insects, and marine animals) and 64 images depicting inanimate objects (equal number of electronic devices, clothes, fruits, and vegetables), taken from BOSS database (Brodeur et al., 2010) and from online royalty-free databases, and was used due to previous success at distinguishing these categories using multi-variate pattern analysis (T. Carlson et al., 2013). All images were scaled to 500 x 500 pixels. A black-and-white drawing version of each image was manually created using GNU imaging manipulation software (www.gimp.org). The photographs vs. drawings served as an additional perceptual category (not of interest for the purpose of our current analyses). In addition to the material used for the experiment, 16 images were used for demonstrative purpose. Images from both semantic classes were randomly split into 16 sets, so that each set consisted of 8 images, 4 animate and 4 inanimate. Each set constituted one learning block. In addition, a list of 128 action verbs was generated for the experiment, serving as cue words in the cued recall task.

The experiment was set up via custom written MATLAB 2016a (©The Mathworks, Munich, Germany) code using functions from the Psychophysics

Toolbox Version 3 (Brainard, 1997). The presentation was done on a 15-inch computer screen with Windows 64 bit.

4.3. Paradigm

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

A distractor task followed each learning phase. Here participants had to respond if a given random number (between 1 and 99) presented on the screen was odd or even. They were instructed to accomplish as many trials as they

could in 45 seconds, and received feedback about their accuracy at the end of each distractor block.

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

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

4.4. EEG Data Analysis

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

4.4.1. Preprocessing

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

Data were divided into trials from 700ms pre-stimulus to 2000ms post-stimulus onset (before implementing MVPA), or 2500ms before the classification maxima to 2500ms after the classification maxima (epochs created based on points of maximum fidelity). A high-pass filter of 0.1 Hz, a low-pass filter of 195 Hz, and a

band-stop filter (48 to 52 Hz; 99 to 101 Hz, and 149 to 151 Hz), were applied to the data. At the edges of each trial, 500ms was then cut out to remove edge artifacts from filtering the epoched data. Trials were visually inspected before an ICA was computed to remove components related to eye-blink artifacts and muscle tension. After components were removed, all trials were again visually inspected, and trials still containing artifacts were manually removed. On average 112 out of 128 trials were kept (min = 100, max = 124, SD = 7). Bad channels were interpolated using the triangulation method. Data were then re-referenced to average.

4.4.2. Multivariate Pattern Analysis

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

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

4.4.3. Power spectrum of the classifier fidelity time series

The first analysis investigated the frequency characteristics of the classifier timeseries using fast fourier transformation (FFT). This and all subsequent

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

A baseline for the LDA outputs was created using a classifier with randomly shuffled labels. The labels of the two classes that the classifier later used for training and testing were shuffled pseudo randomly (to keep the same number of photographs and line drawings in each class), and fed into the LDA 25 times for each participant, such that the newly created groups had approximately the same number of trials from both classes. The parameters for running the classifier were the same as previously described for the real labels. In line with the procedure outlined in (Stelzer et al., 2013), and identical as for the real data, for each participant we drew (with replacement) 100 random accuracy maps (i.e., either a baseline that was created using shuffled labels, or the real

classification of the data), which were then averaged within participants. These accuracy maps were tapered with a Hann window, frequency transformed, and averaged into a group accuracy map. The background 1/f signal was subtracted using a least-squares linear regression, as described above. This procedure was repeated 1000 times, and resulted in an empirical chance distribution, which allowed us to investigate whether the results from the real-labels classification had low probability of being obtained due to chance ($p<.05$) (i.e., exceeding the 95^{th} percentile).

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

To investigate the relationship between the continuous classifier outputs and the EEG data, the Modulation Index (MI) was computed in accordance with (Tort, Komorowski, Eichenbaum, & Kopell, 2010). Following the same procedure as outlined under *Source Analysis* below, we projected the data from scalp level to source level, where each filter was computed using baseline corrected pre-processed data (-.2 – 0 sec), and frequencies below 15Hz (i.e., -200 before to 1500ms after cue onset). Epochs were then reconstructed for 2015 virtual electrodes, rather than the original 128 electrodes. The phase of the EEG signal was estimated by convolving the data with a complex Morlet wavelet of 6 cycles. Each complex value data point was then point-wise divided by its magnitude (absolute value or complex modulus), which gave us a 4D-matrix of phase values, containing trials*channels*frequencies*time. We then

binned the phase values at a given electrode (e.g. a virtual hippocampal electrode), and at a given frequency of interest (e.g. 8Hz), into 10 adjacent bins, ranging from – π to π. The z-scored amplitudes (d-values) of the classifier output from corresponding time points were then sorted into their corresponding phase bins, and the mean amplitude of each phase bin was calculated. Following this sorting procedure at a given frequency, the modulation index was calculated. The MI was computed by comparing the distribution of classifier fidelity values across the 10 phase bins against a uniform distribution (using the mean across bins to construct the uniform distribution). The Kullback-Leibler (KL) distance was then calculated using the equation in (Tort et al., 2010):

$$\mathrm{D}_{KL}(P,Q) = \sum_{j=1}^{N} P(j)\log\left[\frac{P(j)}{Q(j)}\right]$$

A statistical control analysis was then performed to infer whether the MI was significantly different from a distribution that could be obtained by chance. The baseline was computed by running the same analysis as described above, but by cutting the classifier outputs into two segments at a random time point, and inserting the second data segment at the beginning of the trial. This procedure is recommended in (Mike X Cohen, 2014), because it keeps the temporal structure of the classifier outputs largely intact while randomizing their relationship to the EEG phase at any given time point. The newly created random classifier outputs were then paired with the real EEG phase time series from their corresponding trial, and were binned in the same way as the real data. This procedure was repeated 500 times, and the MI was calculated for

each iteration. The 95th percentile across iterations was determined, and the real modulation index for each subject was compared against this subject's 95th percentile using a paired samples t-test. Note that this is a very conservative analysis, resulting only in statistically significant phase modulation, if across participants real phase modulation values significantly exceed the 95th percentile of the time-permuted baseline.

Based on our initial FFT findings, all phase modulation analyses were focused on the oscillatory phase at 8Hz (Figure 2). The phase modulation index was calculated as described above for each virtual channel in source space, and a mask including left and right hippocampus (from AAL atlas as implemented in FieldTrip, see Figure 2E) was then applied to specifically extract the modulation index from our main region of interest. This was done separately for the phase modulation during retrieval, and the phase modulation during encoding. To directly compare the preferred phase during encoding and retrieval, the bin containing the highest classifier amplitudes was identified in each participant, separately for encoding and retrieval. A Rayleigh test (implemented using *circ_rtest* in the Circular Statistics Toolbox for Matlab) was then used to statistically test the extent to which the distribution of phase angles at encoding and retrieval differed from each other.

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

The third, classifier-locked average analysis was aimed at characterizing the EEG phase of the time points where the classifier showed the highest fidelity. To this end, three criteria were established in order to identify times of maximum fidelity. In order to be considered a maximum, a fidelity value was required to have an amplitude that exceeded the 95th percentile of a baseline constructed from the random-label classifications. For each participant, we drew (with replacement) the fidelity timeseries from random trials 1000 times to obtain the baseline distribution. In addition, a maximum included in the final analysis was also required to remain above the 95th percentile threshold for more than 30ms, and to occur later than 200ms after reminder onset, for the same reasons as mentioned above. The average number of classification maxima extracted per trial was 2.27 (SD = 0.26). The onsets of the classifier maxima in each trial were then marked, and the corresponding time stamps were located in the raw, continuous EEG recordings. New epochs were created that were centred on each classifier maximum and contained 2.5 secs before and after the maximum, which were then cut during preprocessing to 2 secs before and after the maximum. These new epochs were used for all subsequent phase-locking analyses.

A phase-locking analysis was conducted on the new epochs to test whether classifier maxima were related to a consistent phase of a theta oscillation. For every frequency between 1 and 20 Hz, we estimated phase by convolving the

data with a complex Morlet wavelet of 6 cycles. Resulting complex values were then point-wise divided by their magnitude (absolute value or complex modulus), and the mean phase was computed over all trials within each participant. The magnitude of this resulting complex value is a single value (the phase-angle time series) for each time-frequency-channel point averaged over all the trials. The value reflects the consistency of frequency-specific phase across trials and has a minimum of 0 and a maximum of 1, also called phase-locking value (PLV), phase-locking index (PLI) or Intertrial Phase Clustering (ITPC) (Mike X Cohen, 2014).

A baseline was calculated for each trial and each participant by shifting single-trial EEG epochs randomly between 0ms and 150ms (roughly one theta-cycle) forward or backward in time, relative to the centre (i.e., the classifier maxima). By doing so, the temporal structure of the analysed signal was kept intact, but the signal was shifted relative to the classifier maxima. The phase-locking index was calculated as described above for the “real”, non-shuffled data. Shuffling was done 25 times per participant and thereafter averaged together.

First, paired samples t-tests were computed between the real data and the time-shuffled baseline to investigate the difference in phase-consistency when using all maxima. To account for the multiple comparisons problem, the t-statistics for each time point (-500ms to 500ms), frequency band (6 to 14 Hz), and electrode were subjected to nonparametric cluster-based permutation testing, as

implemented in the FieldTrip software. The threshold for the statistical testing was set to an alpha level of 0.025. The minimum number of neighbouring channels that were considered a cluster was set to two. T-values above the threshold of 0.1 were then summed up, and compared against a distribution where condition labels were randomly assigned 5000 times with the Monte-Carlo method, following the standard method implemented in FieldTrip.

Phase consistency is strongly biased by number of trials. For our first analysis comparing all maxima against the time-shuffled baseline, the real data and shuffled baseline contained an equal number of trials. We also ensured that all subsequent comparisons were made between conditions with exactly equal trial numbers, within each participant, including an analysis contrasting classifier maxima of high fidelity and maxima of lower fidelity, and two analyses excluding early maxima (see following two paragraphs). For the analysis contrasting conditions with high and low fidelity values, we additionally controlled the average time of the high and low classifier maxima. This was done by creating 8 time bins of equal size between 200ms and 1500ms post-cue. Fidelity values in each time bin were median split into high and low fidelity values, resulting in two matrices representing high and low fidelity trials, equally distributed across time. To calculate the phase consistency, we then followed the same procedure as described above for all maxima, except that instead of using the shuffled baseline the two groups of trials were directly compared using a non-parametric cluster-based permutation test.

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

4.4.6. Event-related potential analysis

Event-related analyses were mainly conducted as sanity checks, on the one hand to investigate average signal differences between the retrieval of animate and inanimate objects locked to cue onset; and on the other hand to evaluate the average signal differences and their topography/source around the time points at which the classifier showed maximal confidence that the correct category was reinstated. For the classifier-centred analysis, we only used the 20% classifier maxima with the highest fidelity values in each of the to-be-compared classes (i.e., animate and inanimate retrieval trials), in order to enhance signal-to-noise ratio. This latter analysis included on average 48 (SD = 7.10) trials per participant. Cluster-based statistics for ERPs were conducted in

the same way as for phase, except that we here focused on a narrower time window from 200ms pre- until 200ms post-maximum.

4.4.7. Source Analysis

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

brain source reconstructions or the classifier-locked effects are only used to illustrate the most likely sources of the effects observed on scalp level (see above). We do not report additional statistics at source level, since these would be circular relative of the already known effects on scalp level. Labels of MNI coordinates were assigned based on the Talairach atlas (Lancaster et al., 2000).

4.4.8. Distribution of fidelity values across time

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

4.4.9. Time generalisation

To characterise the temporal dynamics of the classifiers, we calculated the full time generalization matrices from encoding and retrieval. These matrices show where in time classification accuracy was maximal, to which degree a classifier

trained at one time point generalises to a different time point, indicating temporal stability of the underlying neural code (King & Dehaene, 2014). All analyses were performed using LDA as implemented in the MVPA-Light toolbox, running on MATLAB (https://github.com/treder/MVPA-Light). Two different analyses were run: training at each time point at encoding and testing at each time point at encoding (Figure S1A); training at each time point at retrieval and testing at each time point at retrieval (Figure S1B). When analysing encoding-to-encoding generalization, data were baseline corrected (-200 to 0ms), and then z-scored per trial before running the classification. We used a k-fold cross-validation approach with 5 folds, which was repeated twice with randomly assigned folds. When training and testing at each time point at retrieval, we did not baseline correct before the classification. However, baseline correction was applied after the classification in both analyses.

4.4.10. Identifying oscillating frequencies

An alternative method for detecting oscillations in time series was used in addition to our FFT approach in order to corroborate our claim that classifier outputs oscillate. This method finds time points of oscillations in the data by investigating the change in phase per unit time. We followed the method detailed in (M. X. Cohen, 2014), with a modification for dynamic filter edges only using minimum and maximum of frequencies exceeding the 1/f distribution, made in line with (Watrous et al., 2018). Briefly, we started with raw time series

data, which in our case was the z-scored fidelity values averaged within participants. Instead of creating a plateau-shaped band-pass filter based on an a priori defined frequency range, the filter was constructed based on the lowest and highest frequencies exceeding the fitted line in log-log space using robustfit in MATLAB (Lega, Jacobs, & Kahana, 2012). The analytic signal was obtained by applying the Hilbert transform to the data, from where we extracted the phase angle time series. To obtain the frequency and phase at each sample, frequency sliding was applied to the data as follows: (sampling frequency*diff(unwrap(signal))/(2*π)). After this step, in order to attenuate "phase slips", we applied median filters with different length in the time domain (50ms to 400ms), wherefrom we took the median, in accordance with (M. X. Cohen, 2014). Frequencies that did not exceed the 1/f-fitted line were then excluded, which gave us a vector for each participant containing the frequencies and time points where an oscillation was present. We then calculated the average probability across time (200 to 1200ms post-cue, as in all other analyses using classifier output) for observing an oscillation in a given frequency between 1 and 15 Hz.

To infer whether the result that we obtained was significantly different from chance, we randomly picked one averaged random label classifier per participant. The same procedure as has been described above was applied. An average of this value was then calculated, and stored. This was done 1000 times, and resulted in an estimated chance distribution. The 95^{th} percentile was

then calculated for each frequency, and compared that to the real data (Figure 2D).

4.5. Quantification and statistical analysis

4.5.1. Behavioural data

N = 24 for all behavioural analyses.

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

Reaction times for the first button press when retrieving animate (Mean = 3.03 secs, SD = .95 secs, min = 1.28 secs, max = 6.01 secs) and inanimate (Mean = 2.96 secs, SD = .77 secs, min = 1.47 secs, max = 4.24 secs) objects did not differ significantly, t(1,23) = .57, p = .58. The time-window used for classification (-200ms to 1500ms around the word cue) thus only minimally overlapped with the time window where participants made a button press, and can be seen on page 4.

4.5.2. EEG data

N = 24 for all EEG analyses.

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

Phase-amplitude coupling between EEG data and classifier output was calculated as described on page 6. The real data and the time-shuffled baseline were subjected to a paired-samples t-test, for hippocampal virtual channels for retrieval, $t(1,23) = 1.8191$, $p < .05$, one-sided t-test (Figure 2E), and encoding, $t(1,23) = 2.7494$, $p < .05$, one-sided t-test (Figure 2E).

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

The p-values for the different analyses were:

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

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

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

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

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

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

One Peak: p = .001, Figure S2C.

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

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

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

Testing for a uniform distribution, the MATLAB function crosstab was used. The function provides a chi-square test, to obtain significant difference between two distributions. The results revealed no significant difference, (χ^2 = (1, N = 6007) = 7376600, p = .375), and can be seen on page 11, Figure S2A.

To identify oscillating frequencies, we implemented the procedure described on page 34. The results were compared to a constructed baseline, and only frequencies exceeding the 95th percentile of the baseline were considered significant (Figure 2D).

To test for difference between the power spectra for all trials and only correct trials, the two matrices were subject to a one-sided paired samples t-test, where we expected higher power for only correct trials in the 7-9Hz frequency range of interest, t(1,23) = 1.9425, p = .03 (Figure S2B).

5. Acknowledgements

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

6. Author Contribution

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

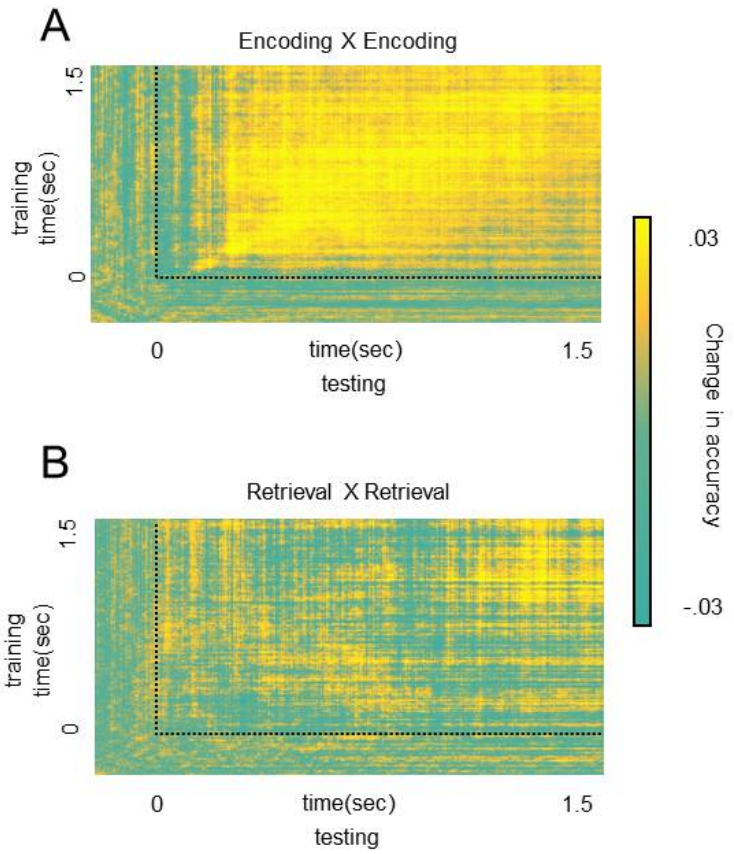
7. Declaration of Interests

The authors declare no competing financial interests.

8. Data and software availability

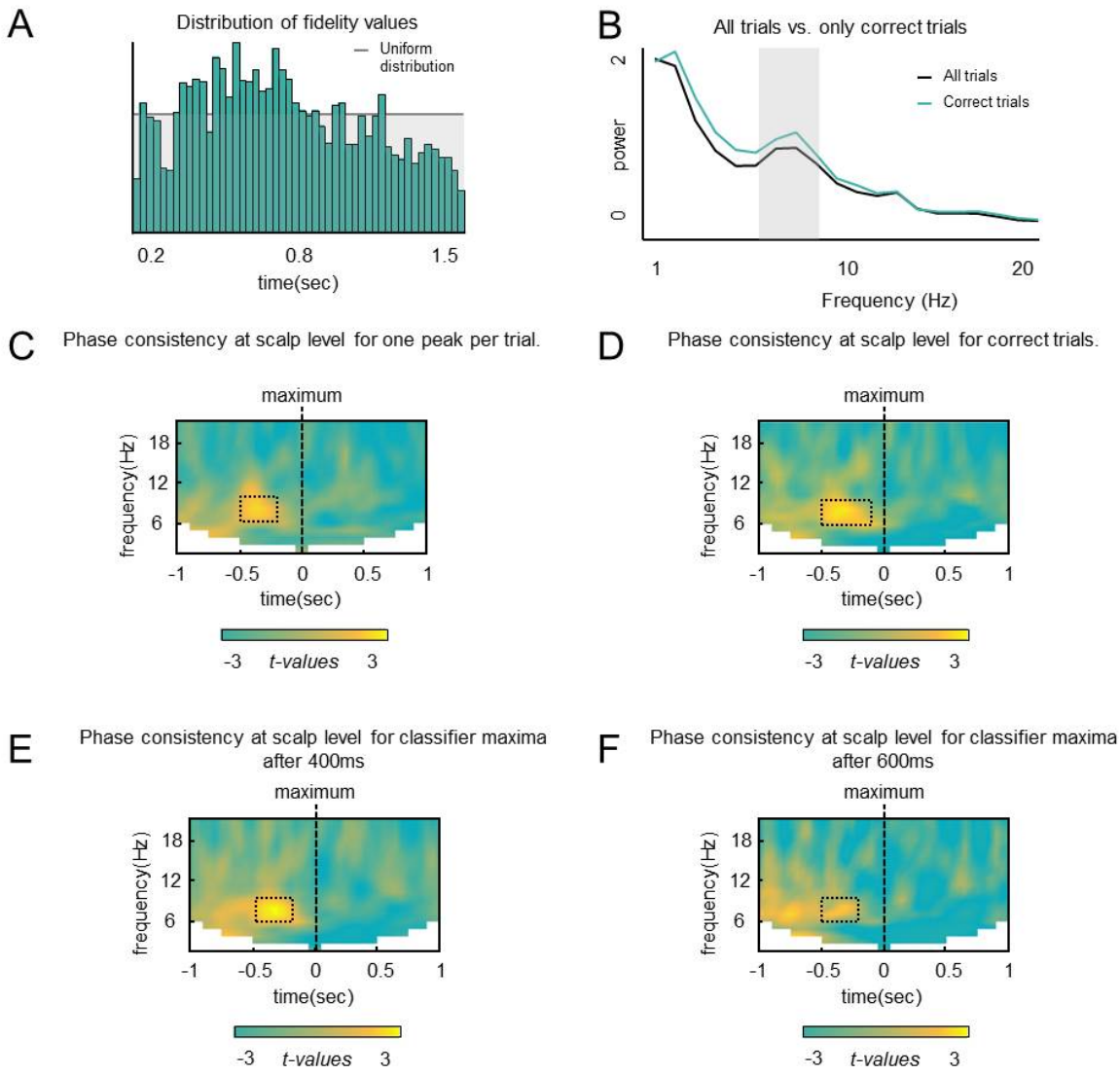
Custom MATLAB code as well as data additional to the already published on http://dx.doi.org/10.17632/h4vcpxt4sr.1 will be made available upon request (fulfilled by Lead Contact, C.Kerren@pgr.bham.ac.uk). Since consent for sharing data at the level of the individual participant was not received originally,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

9. Supplementary figures



S. Chapter 6. Supplementary Figure 1.

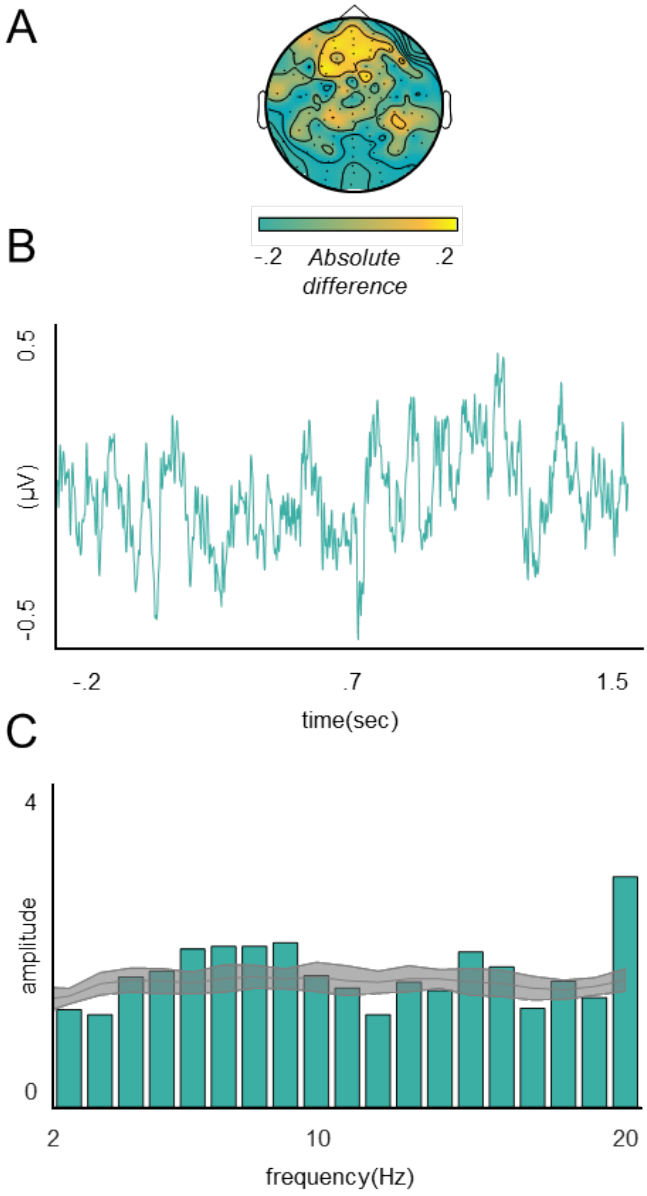
Figure S1. Time generalisation matrices for encoding and retrieval, with time zero indicating the onset of the object during encoding, and the onset of the reminder word during retrieval, related to STAR Methods. (A) Training and testing at encoding showed sustained high classifier accuracy from approximately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



T. Chapter 6. Supplementary Figure 2

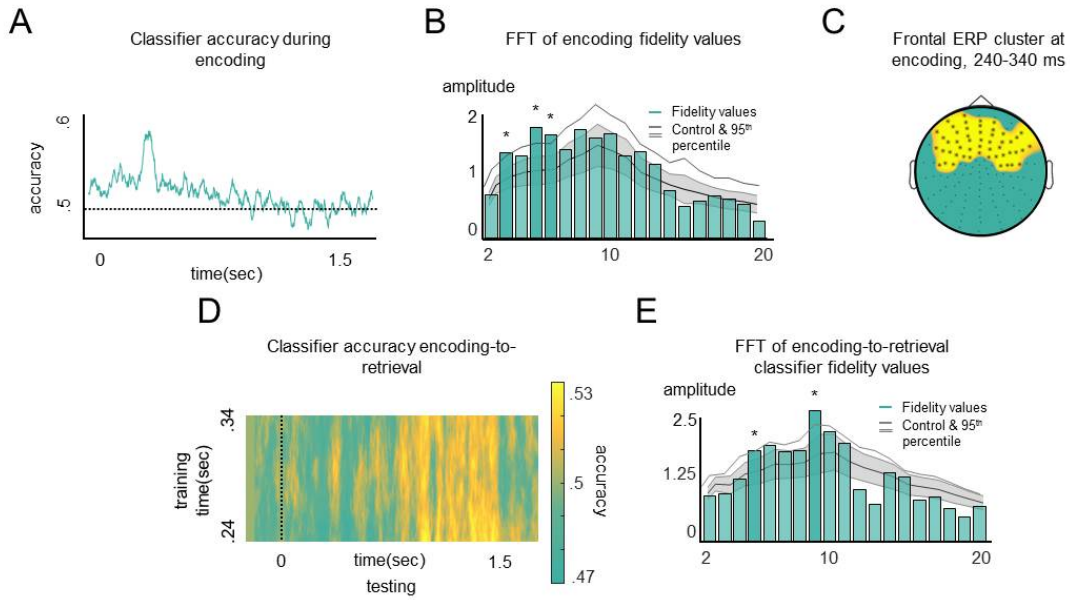
Figure S2. Distribution of classifier maxima across participants and time, behavioural relationship to power spectra, and phase consistency for various control analyses, related to Figure 3 and STAR Methods. (A) The distribution of classifier maxima, accumulated across participants, showed no significant deviation from a uniform distribution, indicating that the maxima were evenly distributed across the entire retrieval period, with a noticeably increased density around 400-800ms. This is in line with previous studies showing strongest memory reinstatement in the recollection period. **(B)** To evaluate the relationship between the power spectra and memory performance, we compared the power spectra for all trials and correct trials only for 7-9 Hz, which revealed a significantly stronger effect for correct trials compared to all trials.

Note that a direct comparison between correct and incorrect trials was not possible due to a low number of incorrect trials in the cued recall task. **(C)** Classifier-locked averages showing phase consistency when using only the highest maximum per trial, and thus excluding all overlapping epochs. As expected, the phase consistency is less temporally smeared, with a cluster from -500ms to -150ms pre-maximum. **(D)** Same analysis as shown in main Figure 3, but limited to correct trials, showing a cluster of significant phase consistency 500-150ms before the classifier maxima. **(E)** Removing the first 400ms of classifier maxima did not change the phase consistency effect, neither did removing the first 600ms of classifier maxima **(F)**. When using only very late maxima, an even earlier cluster of 7-8Hz phase consistency becomes evident, with the later cluster at -500ms to -250ms remaining significant. This results likely reflects several cycles of a 7-8Hz oscillation.



U. Chapter 6. Supplementary Figure 3

Figure S3. Average difference between animate and inanimate object retrievals shown in the time and frequency domain, related to Figure 4 and STAR Methods. (A) Topographies of the absolute EEG difference between the recall of animate and inanimate objects, showing a frontal maximum during retrieval (600-1200ms). **(B)** Average difference signal between animate and inanimate objects during retrieval, interestingly showing a visible rhythmicity. **(C)** Applying the Fourier-transform, we can see above baseline power increases in spectral frequencies between 6-9Hz, the same frequencies that also show power increases in the classifier time series.



V. Chapter 6. Supplementary Figure 4

Figure S4. Encoding for encoding and encoding for retrieval analyses, related to Figure 2 and STAR Methods. (A) Training and testing at encoding revealed a peak of classifier performance at ~300ms, a time window commonly seen when investigating encoding activity for semantic memory. **(B)** The averaged fidelity values were subjected to a Fourier Transformation, and showed a peak in the lower frequencies. **(C)** At encoding, a frontal cluster survived a non-parametric cluster-based permutation test, indicating an overlap with retrieval activity seen in Figure S3A. **(D)** Using the time points from 240-340ms during encoding, where animate vs inanimate differences showed an overlapping topography compared with retrieval (see Fig. S4A), a fluctuating pattern is also visible in the time generalisation matrix of a classifier trained on encoding and tested during retrieval. **(E)** The power spectra of this encoding-retrieval classifier revealed a peak at 9 Hz.

Chapter 7: General discussion

1. A brief summary of the main objectives and how they were addressed

In this doctoral thesis we explored two essential questions regarding the temporal dynamics of episodic memory retrieval. On the one hand, we explored how different components of memory representations (i.e., lower-level perceptual and higher-level conceptual information) unfold over time, and how this temporal sequence evolves compared to visual encoding. On the other hand, our main second goal was to examine whether our memories are reactivated in a specific oscillatory rhythm. In this section, I will recap our alternative hypotheses and the approaches followed to investigate these questions.

First, we tested what we named the reverse reconstruction hypothesis. Does the hierarchical processing cascade of perceptual and semantic information found in object recognition (T. Carlson et al., 2013; Cichy et al., 2014) reverse when an object representation is retrieved from memory? According to our hypothesis, we expected that during the visual perception of an everyday item, its low-level details (i.e., as colours or lines) would be processed before its high-level semantic aspects (i.e., whether the item represents an animal or not). However, we predicted that during retrieval the opposite temporal order would be found, such that semantic information should be reactivated before low-level aspects. These predictions were explored through a series of four behavioural and two electrophysiological studies. In all of them we used an encoding-retrieval paradigm that we slightly modified depending on the goals of each

experiment. During encoding, participants learned a set of arbitrary associations between verbal cues and everyday object images. Later, in a retrieval phase, participants were asked to vividly visualise the images associated when the cue was presented on the screen. During retrieval, we also asked participants to answer questions about perceptual and semantic details of these remembered representations.

One version of this paradigm was used to behaviourally test the reverse reconstruction hypothesis, measuring RT and accuracy when participants answered questions about perceptual and semantic details of currently perceived or mentally reinstated visual representations. In this series of experiments (Experiment 1, 2, 5 and 6), we predicted that, during retrieval, participants would be more accurate and faster when responding to questions about semantic information compared to perceptual inquiries, showing the opposite pattern compared to visual processing. Our central prediction was also tested at a neural level in two electrophysiological experiments (EEG in Experiment 3 and iEEG in Experiment 4). Using time-resolved decoding analyses, we measured on a trial-by-trial level at which moment the brain signal was maximally associated with perceptual and semantic processing of these images. This multivariate approach allowed us to compare the temporal pattern of both types of representational features during encoding and retrieval. In these experiments we expected to find evidence for the reactivation of semantic features before perceptual reactivation during retrieval. The reverse sequence

between low and high-level processing was predicted when the object was visually presented.

Secondly, based on animal models (Hasselmo et al., 2002; Kunec et al., 2005), we tested whether in humans memory retrieval is an oscillatory process that is modulated by a specific phase of a theta oscillation. We explored this premise using the same encoding-retrieval paradigm described above in an EEG experiment. Via decoding analyses, we again obtained an index that indicated the classifier's fidelity to identify a memory's reactivation on a single trial level. If theta oscillations open time windows for retrieval in the hippocampus, we expected that this memory reactivation index would fluctuate in a theta rhythm and that the peak of this index would be modulated by a specific phase of theta.

In the next section the principal findings obtained in this series of experiments will be recapitulated and integrated with the previous literature.

2. A summary of the most relevant findings: the reverse reconstruction effect and the oscillatory nature of episodic memory retrieval

Two behavioural studies (Experiment 1 and Experiment 2) indicated that, when objects are visually presented, participants are faster detecting an object's low-level perceptual details than conceptual (semantic) features of these items. However, when these objects were not visually presented but retrieved from memory, subjects significantly faster remembered the objects' semantic details than perceptual information. This reverse pattern between visual processing

and retrieval was also confirmed in both experiments by a significant interaction between the type of task (i.e., visual or recall task) and kind of feature (perceptual or semantic). In line with these response latency patterns, an identical pattern was found in terms of participants' accuracy in both experiments. When the object was visually perceived, participants showed a significantly higher accuracy when responding about perceptual details compared to semantic information. Conversely, the opposite trend was found when they retrieved objects' features from memory: semantic information was significantly better remembered than perceptual attributes. In addition, the same significant interaction obtained for RTs was found when analysing accuracy profiles. Altogether, these two first behavioural experiments confirmed our predictions and suggested the existence of semantic prioritization over low-level features when past memories are retrieved.

Experiment 3 fully corroborated these findings in an EEG study that used the same material and a similar task as Experiments 1 and 2. Here we used a time-revolved decoding approach to observe when perceptual and semantic information were maximally decodable using scalp EEG signals during object perception and object retrieval. These multivariate analyses gave us a measure (*d* values) of classifier fidelity in selecting the correct perceptual category (i.e., line drawing or photograph) or semantic category (i.e., animate or inanimate) for each single item. Importantly, we compared the temporal distance between both perceptual and semantic *d* value peaks on a single trial level. During encoding (or object visual perception) perceptual *d* value peaks were found

approximately 100ms before the semantic peaks. On the other hand, the opposite pattern of results was found when objects were reactivated from memory: classification peaks for semantic information were obtained approximately 300ms before the perceptual peaks. Additionally, we also tested the reverse reconstruction hypothesis through more traditionally approaches as ERP analyses and same pattern of results were obtained: during object presentation the maximum perceptual ERP cluster (i.e., comparing line drawing vs. photograph trials) appeared around 100ms before the semantic ERP cluster (i.e., comparing animate vs. inanimate trials). However, during retrieval, the perceptual cluster occurred around 400ms after the semantic one. In summary, results from Experiment 3 supported previous behavioural results and our hypothesis about the temporal order of memory reconstruction. Importantly, the similarity between these results and findings obtained in Experiments 1 and 2 emphasizes that our behavioural approach can be used to tap into neural processing speed using RT analyses.

Preliminary iEEG results testing the reverse reconstruction hypothesis were presented in Chapter 4 (Experiment 4). In this case study, we had the opportunity of analysing iEEG signals from electrodes located intracranially along the ventral visual stream and the hippocampus while a participant performed the same task as in Experiment 3. Following a similar time-resolved decoding approach, results showed that when object images were retrieved from memory, their semantic features were reactivated significantly earlier than their perceptual elements. In this single case study, no significant differences

were found during encoding. This temporal benefit in retrieving semantic aspects over perceptual elements was found in electrode contacts located along the ventral visual stream, but also in those contacts closest to the hippocampus. Although these results represent preliminary work and further analyses and a bigger sample size are needed, the prioritization of conceptual details during memory reconstruction in electrodes placed on areas of interest represents a promising finding in line with the reverse reconstruction hypothesis.

Overall, the first series of behavioural and electrophysiological experiments revealed a highly consistent pattern of results supporting our idea that the visual processing hierarchy is reversed when visual memories are retrieved. However, in all these four experiments we maintained the same low-level and semantic categories using the same set of stimuli. To manipulate perceptual features, objects were presented either as a photograph or a line drawing. On the other hand, all items were either an animate or an inanimate object, allowing us to control the conceptual properties of these stimuli. In two follow-up behavioural experiments we explored whether the reverse reconstruction effect can be replicated with different perceptual and semantic manipulations. Instead of using animacy, semantic categories were defined by naturalness (i.e., whether the objects represent something natural or manmade), and as perceptual categories we used the retinal size (Experiment 5) and object shape (Experiment 6). In both cases, the task procedure was the same as used in Experiment 1. In general, accuracy and RT analyses again revealed a

significant interaction between type of task (visual or memory task) and type of question (perceptual or semantic), suggesting that the processing of low and high-level information reverse depending on whether the representation is visually perceived or retrieved from memory. Apart from this effect replicated with different feature manipulations, some effects changed depending on the type of perceptual manipulation. RT and accuracy results showed that participants discriminated low-level features faster and more accurately than semantic elements when an object is visually perceived (at least when the object' real size was not controlled in Experiment 5). When a representation of the object was retrieved from memory, participants performed better and faster remembering semantic than perceptual information in Experiment 5 (when object's real size was controlled). However, no significant differences at retrieval were found when perceptual manipulation was based on shape (Experiment 6). One potential explanation for this lack of differences during retrieval in Experiment 6 could be the closer distance along the ventral stream between shape and conceptual processing (a more elaborated explanation can be found in Chapter 5).

In summary, the evidence presented above corroborated our two main alternative hypotheses about the temporal features of memory retrieval. Our predictions were based on a series of widely accepted assumptions about the memory system in human and rodents. Starting from these assumptions, we tested novel ideas regarding the time course of the retrieval processes. In this section, I will discuss how these results can be integrated with the previous

body of evidence in the field. I will also sketch the new questions about retrieval processing that are opened by our findings, and how these findings in the future could allow for a general model of retrieval to be elaborated.

The reverse reconstruction effect suggests that, during retrieval, memories are reassembled in a specific hierarchical order, from gist-like information to perceptual details. Although future research should explore the consistency and boundaries of this effect, the prioritization of semantic information over perceptual details is consistent with some hierarchical memory system models (Henson & Gagnepain, 2010), with recent findings that indicate that visual imagery does not rely on early perceptual representations (Dijkstra et al., 2018), and is also coherent with a widely accepted phenomenon in the visual field that suggest that visual learning is a top-down process that advances from high to low-level domains (Ahissar & Hochstein, 2004).

Our results indicate that memory retrieval is biased towards conceptual information. This preference to emphasize specific aspects of an event representation reflects two fundamental properties of our episodic memory system: (i) recollection is not impartial and accurate but a reconstructive, highly biased process (Schacter, 2012); and, second, (ii) that memory engrams are in continuous transformation (Dudai, 2012). The preference to reactivate high-level information is also consistent with some models and predictions about episodic memory that underlie the semantic nature of retrieved past representations. For instance, in order to form stable long-term memories (system-level consolidation), the neocortical system is thought to be able to

extract the meaningful, statistical regularities from our memories in order to form stable, but more gist-like representations (Dudai et al., 2015; Moscovitch, 2008). Moreover, it has been suggested that the neocortical reactivation of the past episodes could support memory consolidation by facilitating the formation of more semantic or gist-like memories (Antony et al., 2017). Both notions are highly compatible with our findings demonstrating that retrieval prioritizes those meaningful, high-level aspects of an event that are eventually consolidated in our long-term store.

In addition, the idea of a reverse processing cascade between perception and retrieval is not only intuitive and neurologically plausible, but also indirectly related with existing findings regarding the communication between the hippocampus and adjacent neocortical areas. The change in information directionality is coherent with results that demonstrated that during retrieval, entorhinal neurons project back to those neocortical areas that delivered input to the hippocampus during memory encoding (Lavenex & Amaral, 2000; Witter et al., 2000). The reverse reconstruction proposal is also consistent with findings indicating that the information flow (in term of neural connectivity) between hippocampus and neocortical areas changes its direction between encoding and retrieval (Fell et al., 2016). Additionally, a fMRI study by Staresina et al. showed that, along the temporal lobe, brain areas involved in object and scene processing are active during retrieval but with a delay that is consistent with a reversed information flow (Staresina et al., 2013). However, follow-up experiments should investigate the specific information pathway associated to this reverse reconstruction effect.

All previous experiments explored how memory retrieval is not an all-or-none process but a reactivation cascade where each different aspect of a past episode is remembered following a hierarchical stream. In this series of 6 experiments we focused on studying the temporal dynamics between low and high level features of episodic representation during encoding and retrieval. In Experiment 7 we continued investigating the temporal characteristics of episodic memory retrieval in an EEG study. In particular, based on computational models of the hippocampus, we were interested in whether memory retrieval in human is an oscillating process that fluctuates in a theta rhythm and is associated to a specific phase of this frequency band. Using a time-resolved decoding approach similar to the one applied in Experiments 3 and 4, we obtained a memory reactivation index (*d* value) for each single trial. This value allowed us to identify when the EEG signal strong reactivation of episodic information, above a certain noise threshold. Confirming our predictions, first analyses showed that this reactivation index fluctuated at 7-8Hz. Also, further evaluations suggested that classifier fidelity on each trial was significantly modulated by the phase of an 8Hz oscillation that was source localised to virtual channel in the hippocampus. Importantly, the phase of peaks occurred with a difference of 188 degrees when encoding and retrieval trials were compared. Finally, we also found that when EEG recordings were analysed time-locked to *the classifier fidelity* peaks, a significant phase alignment at 7-8Hz ocurred 200-300ms before these peaks. This finding indicates that memory reactivation, as reflected in classifier fidelity, is tightly

time locked to theta phase, and likely happens in brain regions that are active 200-300ms after retrieval has been initiated.

Evidence from computational models and rodent experiments suggests that retrieval fluctuates depending on a hippocampal theta oscillation. In these models, theta oscillations shift the hippocampal system into different processing modes, one that is optimal for encoding which occurs at one particular phase of the oscillation, and a state that is optimal for memory reconstruction at the opposing phase of the oscillation (Hasselmo et al., 2002; Kunec et al., 2005). Results from Experiment 7 represent the first direct evidence that suggests a connection between theta phase and retrieval in human long-term memory. They are consistent, however, with previous human studies indicating a similar phenomenon in working memory. Firstly, the role of theta phase has been investigated in earlier working memory experiments that, in line with some of our results, pointed to how theta phase, at least at very early moments of a trial, shifts between encoding and retrieval (Rizzuto et al., 2006). Secondly, our results are highly consistent with a working memory study that used gamma oscillations as a proxy for memory reinstatement (Fuentemilla et al., 2010). Using this approach they showed that working memory reactivation is repeated several times during a time-window of 5 seconds, and this reactivation during working memory maintenance was also phase-locked to a theta oscillation.

Overall, our results are thus in line with previous work on learning and retrieval in working memory and long-term memory. More importantly, however, they go beyond the existing literature by shedding light onto the sub-second temporal

dynamics of memory retrieval. We developed a number of new approaches to tap into these temporal processing sequences at the representational level. First, we designed reaction time experiments that can provide an index of how rapidly different aspects of a previously encoded memory – perceptual or conceptual – come online during retrieval. We validated these reaction time experiments with time-resolved pattern classification analyses, yielding a very consistent pattern of results. To the best of our knowledge, the series of behavioural and EEG data provides the first direct evidence for a reversal of the information processing cascade between encoding and retrieval. Using a novel approach of linking classifier-based indices of memory reinstatement to brain oscillations, we also report the first human evidence for a tight functional relationship between theta phase and memory retrieval. Although these outcomes are promising, some important questions related to the main findings presented here currently remain open, and should be addressed in follow-up investigations. I will shortly discuss the most important questions in the following section.

3. Possible functional relationships between reverse reconstruction and theta phase

One question very directly following from the present results is how the two main findings of this doctoral thesis (i.e., the reverse reconstruction effect and the oscillatory nature of memory retrieval) could be integrated in a general model of episodic memory. Although it is not possible to fully address this point with the findings presented, I will introduce some initial predictions.

First of all, it is important to highlight that the reactivation index used in Experiment 7 was based on a semantic classifier (i.e., categorizing animate vs. inanimate memory representations). In this sense, results from this experiment could be rephrased concluding that the reactivation of semantic features of a mnemonic representation oscillates in a theta rhythm, and is modulated by a specific theta phase. What would happen with respect to the reactivation of perceptual information remains unclear. Based on the results presented above (i.e., from Experiment 1 to 6), we can expect that perceptual details are also independently reactivated during memory recollection, and that this type of information is retrieved with a delay compared with semantic aspects. The first key point regarding perceptual information is whether the reactivation of low-level information also oscillates in a theta rhythm and also whether it is modulated by a specific theta phase. Based on the literature previously commented, we could expect that low-level feature reactivation would also follow an oscillatory pattern, whose maxima moment of reactivation will appear after the maxima for high-level information. Although this hypothesis is interesting per se, it is even more important to go a step beyond and wonder how the hippocampus could orchestrate the reactivation of high and low-level features via oscillatory mechanisms.

For several memory models, theta rhythms seem to be key to trigger the reconstruction of previous representations (Hasselmo, 2005; Klimesch et al., 2001; Kunec et al., 2005; Parish et al., 2018). One possibility is that the

reactivation of high and low-level features is time-locked to different phases of the same theta oscillation in the hippocampus, reflecting the relevance of these different mnemonic features (for relevancy-dependent phase coding in working memory, see Bahramisharif, Jensen, Jacobs, & Lisman, 2018). However, predicting a separate theta phase for each memory component is not the most parsimonious option: in fact, for complex real life memories, such a coding scheme would almost certainly result in a scenario where the available number of non-overlapping theta phases is not sufficient to code for all the separate elements that constitute a memory representation. Therefore, a second possibility is that all components of a memory representation that were previously indexed by the hippocampus are triggered simultaneously in the same theta phase. In this scenario, this theta oscillation would merely trigger an early (e.g. ecphoric, see Tulving, Voi, Routh, & Loftus, 1983) stage of the retrieval process, while the reactivation of the various elements of our memories (including high and low-level information) then unfolds along the neocortex. The reactivation of different pieces of information in neocortex (likely linked to higher frequencies; Parish et al., 2018) would be produced with certain time delays depending, for instance, on the effectiveness of the connection from the hippocampus, or even depending on previous encoding processes (e.g. what information was attended most). Unfortunately, answering all these points is beyond the objectives of this doctoral thesis, and future research projects should explore these possibilities.

4. Future directions for investigating the temporal dynamics of episodic memory

First, the body of findings presented in this thesis highlights the fact that memory representations are gradually evolving on a time scale of several hundred millisecond, where the processing hierarchy of their elements reverses between visual perception and memory. As mentioned above, it is widely accepted that memory representations are modified with the passage of time and further offline processing (including sleep), and that their cortical consolidation depends on extracting regularities between memories, accentuating their semantic details (Moscovitch, 2008). This way, it could be expected that, with each successful retrieval of a particular past episode, its semantic aspects will be strengthened relatively more than its low-level perceptual details (Antony et al., 2017). In this respect, an important question is how the reverse reconstruction effect changes when episodic memories are recalled repeatedly. Based on the findings presented in Chapters 1 to 6, it can be predicted that through repeated memory reactivation, the prioritization of semantic information over low-level details would be further enhanced. For instance, using a behavioural paradigm similar to Experiment 1, but testing each episode several times, we could expect that the benefit of remembering semantic features over perceptual details in RT and accuracy will become more pronounced with each successful test. Similarly, when using decoding approaches with neural signals (like in Experiments 3 and 4) it could be expected that the distance for perceptual and semantic classifier peaks would increase after each successful memory reactivation. Ultimately, repeated

reactivations might selectively strengthen conceptual aspects and “wash out” perceptual detail, in line with previous suggestions that retrieval aids the creation of consolidated, gist-like memories (Antony et al, 2017).

Second, the findings presented in this doctoral thesis mainly focus on temporal dynamics. To completely understand the memory reactivation stream, it is necessary to also identify the spatial dynamics of this sequential process: how the hippocampus orchestrates different cortical areas and that represent high and low-level information, and how the recruitment of these representational areas maps onto the temporal sequence that was found in our electrophysiological experiments. For this reason, further experiments combining brain imaging techniques that allow to record brain activity with a high spatial and temporal resolution (e.g. iEEG, MEG or simultaneous EEG-fMRI) are fundamental to continue shedding light onto the retrieval process.

Third, episodic memories in real life are complex representations where conceptual and perceptual information are just a small part of a bigger picture. Moving away from simple visual stimuli, it might turn out to be difficult to fragment memory representations into all of their constituent elements, but there are certainly other memory components that are special interest. Due to their essential role in the memory system, contextual (i.e., spatial) and emotional aspects of episodic memories have been in the spotlight of memory research for decades (for instanse, see Dunsmoor, Murty, Davachi, & Phelps, 2015; Leal, Tighe, Jones, & Yassa, 2014; Moser, Kropff, & Moser, 2008; Yang

& Wang, 2017). In this sense, future research should investigate how such elements are reactivated over time, which elements are key for triggering a memory, and importantly, explain how all these different memory components are ultimately integrated into a coherent mental representation and a subjective sense of recollection.

5. Conclusions

Understanding how fragments of our life are encoded and retrieved from our vast memory system is fundamental for neuroscience. Throughout the last century and past decades, our knowledge about human memory has undergone an unprecedented growth thanks to important technical and theoretical advances. However, for such a young discipline as cognitive neuroscience there are still innumerable questions regarding episodic memory. The findings presented here are the result of our work over the last years, exploring the mechanisms behind memory retrieval. We believe these are a small but crucial step in comprehending how the human brain manages to bring back our personal past to our present.

In terms of focus, this doctoral thesis departs from traditional investigations of retrieval processes (e.g., familiarity or recollection) to instead study the memory representations per se, and to explore the temporal dynamics of their various perceptual and semantic components. In this series of experiments, our findings emphasize that memory engrams are multi-layered representations and the

reactivation of their low- and high-level components occur in a biased and sequential manner. Based on the body of evidence presented, access to past representations depends on a cascade of multiple reactivations from semantic to perceptual details in order to retrieve different pieces of information. But, additionally, this work suggests that the reactivation of past events occurred repeatedly following an oscillatory rhythm (7-8Hz). Thus, although our memories can appear in our “internal eye” as clear images, they are not simple snapshots from the past that include all information at once. Fortunately, episodic memory reconstruction seems quite more complex and exciting, raising new questions and thrilling challenges for future research.

References

References

Aggelopoulos, N. C. (2015). Perceptual inference. *Neuroscience and Biobehavioral Reviews*, *55*, 375–392. http://doi.org/10.1016/j.neubiorev.2015.05.001

Aggelopoulos, N. C., & Rolls, E. T. (2005). Scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. *The European Journal of Neuroscience*, *22*(11), 2903–16. http://doi.org/10.1111/j.1460-9568.2005.04487.x

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406. http://doi.org/10.1038/387401a0

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. http://doi.org/10.1016/j.tics.2004.08.011

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(15), 7041–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8041742

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

Ashby, F. G. (2000). A Stochastic Version of General Recognition Theory. *Journal of Mathematical Psychology*, *44*(2), 310–329. http://doi.org/10.1006/jmps.1998.1249

Bahramisharif, A., Jensen, O., Jacobs, J., & Lisman, J. (2018). Serial

representation of items during working memory maintenance at letter-selective cortical sites. *PLoS Biology*, *16*(8), e2003805. http://doi.org/10.1371/journal.pbio.2003805

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape Similarity, Better than Semantic Membership, Accounts for the Structure of Visual Object Representations in a Population of Monkey Inferotemporal Neurons. *PLoS Computational Biology*, *9*(8). http://doi.org/10.1371/journal.pcbi.1003167

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). http://doi.org/10.1016/j.jml.2012.11.001

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

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology*, *23*(3), 393–419. http://doi.org/10.1016/0010-0285(91)90014-F

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components--a tutorial. *NeuroImage*, *56*(2), 814–25. http://doi.org/10.1016/j.neuroimage.2010.06.048

Born, R. T., & Bradley, D. C. (2005). STRUCTURE AND FUNCTION OF VISUAL AREA MT. *Annual Review of Neuroscience*, *28*(1), 157–189. http://doi.org/10.1146/annurev.neuro.26.041002.131052

Bosch, S. E., Jehee, J. F. M., Fernandez, G., & Doeller, C. F. (2014).

Reinstatement of Associative Memories in Early Visual Cortex Is Signaled by the Hippocampus. *Journal of Neuroscience*, *34*(22), 7493–7500. http://doi.org/10.1523/JNEUROSCI.0805-14.2014

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897X00357

Bridson, N. C., Fraser, C. S., Herron, J. E., & Wilding, E. L. (2006). Electrophysiological correlates of familiarity in recognition memory and exclusion tasks. *Brain Research*, *1114*(1), 149–160. http://doi.org/10.1016/j.brainres.2006.07.095

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

Brooks, K. R., Morris, T., & Thompson, P. (2011). Contrast and stimulus complexity moderate the relationship between spatial frequency and perceived speed: Implications for MT models of speed perception. *Journal of Vision*, *11*(14), 19–19. http://doi.org/10.1167/11.14.19

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, *33*(3), 325–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11832222

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction Time for Object Categorization Is Predicted by Representational Distance. *Journal of Cognitive Neuroscience*, *26*(1), 132–142. http://doi.org/10.1162/jocn_a_00476

Carlson, T., Tovar, D., Alink, A., & Kriegeskorte, N. (2013). Representational

dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–19. http://doi.org/10.1167/13.10.1.doi

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

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570. http://doi.org/10.1073/pnas.1402594111

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2016). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. http://doi.org/10.1038/nn.4450

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and CA1 mismatch detection. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *18*(8), 523–8. http://doi.org/10.1101/lm.2135211

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

Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition. *Cerebral Cortex*, *26*(8), 3563–3579.

http://doi.org/10.1093/cercor/bhw135

Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning From Vision. *Trends in Cognitive Sciences*, *19*(11), 677–687. http://doi.org/10.1016/j.tics.2015.08.008

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

Cohen, M. X. (2014). Fluctuations in Oscillation Frequency Control Spike Timing and Coordinate Neural Networks. *Journal of Neuroscience*, *34*(27), 8988–8998. http://doi.org/10.1523/JNEUROSCI.0261-14.2014

Colgin, L. L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., … Moser, E. I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, *462*(7271), 353–7. http://doi.org/10.1038/nature08573

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, *250*, 126–136. http://doi.org/10.1016/j.jneumeth.2015.01.010

Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, *17*(2), 140–147. http://doi.org/10.1016/j.conb.2007.03.002

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized Color Modules in Macaque Extrastriate Cortex. *Neuron*, *56*(3), 560–573. http://doi.org/10.1016/j.neuron.2007.10.008

Cowey, A., & Weiskrantz, L. (1967). A Comparison of the Effects of

Inferotemporal and Striate Cortex Lesions on the Visual Behaviour of Rhesus Monkeys. *Quarterly Journal of Experimental Psychology*, *19*(3), 246–253. http://doi.org/10.1080/14640746708400099

Dandolo, L. C., & Schwabe, L. (2018a). Correction: Time-dependent memory transformation along the hippocampal anterior-posterior axis (Nature Communications (2018) DOI: 10.1038/s41467-018-03661-7). *Nature Communications*, *9*(1), 1–11. http://doi.org/10.1038/s41467-018-04516-x

Dandolo, L. C., & Schwabe, L. (2018b). Time-dependent memory transformation along the hippocampal anterior–posterior axis. *Nature Communications*, *9*(1), 1205. http://doi.org/10.1038/s41467-018-03661-7

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

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*(6), 693–700. http://doi.org/10.1016/j.conb.2006.10.012

Deuker, L., Doeller, C. F., Fell, J., & Axmacher, N. (2014). Human neuroimaging studies on the hippocampal CA3 region - integrating evidence for pattern separation and completion. *Frontiers in Cellular Neuroscience*, *8*(March), 64. http://doi.org/10.3389/fncel.2014.00064

Dickerson, B. C., & Eichenbaum, H. (2010). The episodic memory system: Neurocircuitry and disorders. *Neuropsychopharmacology*, *35*(1), 86–104. http://doi.org/10.1038/npp.2009.126

Dijkstra, N., Mostert, P., Lange, F. P. de, Bosch, S., & van Gerven, M. A.

(2018). Differential temporal dynamics during visual imagery and perception. *ELife*, *7*, 1–16. http://doi.org/10.7554/eLife.33904

Douchamps, V., Jeewajee, A., Blundell, P., Burgess, N., & Lever, C. (2013). Evidence for encoding versus retrieval scheduling in the hippocampus by theta phase and acetylcholine. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(20), 8689–704. http://doi.org/10.1523/JNEUROSCI.4483-12.2013

Dudai, Y. (2012). The restless engram: consolidations never end. *Annual Review of Neuroscience*, *35*, 227–47. http://doi.org/10.1146/annurev-neuro-062111-150500

Dudai, Y., Karni, A., & Born, J. (2015). The Consolidation and Transformation of Memory. *Neuron*, *88*(1), 20–32. http://doi.org/10.1016/j.neuron.2015.09.004

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, *22*(3), 389–398. http://doi.org/10.1002/hipo.20933

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*. http://doi.org/10.1038/nature14106

Duvernoy, H. M., Cattin, F., & Risold, P.-Y. (2013). *The Human Hippocampus*. Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-33603-4

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural

representations that underlie declarative memory. *Neuron*, *44*(1), 109–120. http://doi.org/10.1016/j.neuron.2004.08.028

Eichenbaum, H. (2016). Still searching for the engram. *Learning & Behavior*, *44*(3), 209–222. http://doi.org/10.3758/s13420-016-0218-1

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. *Annual Review of Neuroscience*, *30*(1), 123–152. http://doi.org/10.1146/annurev.neuro.30.051606.094328

Fabiani, M., Gratton, G., & Federmeier, K. (2007). Event-Related Brain Potentials: Methods, Theory and Applications. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology*. Cambridge: Cambridge University Press.

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews. Neuroscience*, *12*(2), 105–18. http://doi.org/10.1038/nrn2979

Fell, J., Wagner, T., Staresina, B. P., Ranganath, C., Elger, C. E., & Axmacher, N. (2016). Rhinal-Hippocampal Information Flow Reverses Between Memory Encoding and Retrieval (pp. 105–114). http://doi.org/10.1007/978-3-319-46687-3_11

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierachical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. http://doi.org/10.1093/cercor/1.1.1

Ferrera, V., Nealey, T., & Maunsell, J. (1994). Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *The Journal of Neuroscience*, *14*(4), 2080–2088.

http://doi.org/10.1523/JNEUROSCI.14-04-02080.1994

Fox, P. T., Mintun, M. a, Raichle, M. E., Miezin, F. M., Allman, J. M., & Van Essen, D. C. (1986). Mapping human visual cortex with positron emission tomography. *Nature*, *323*(6091), 806–9. http://doi.org/10.1038/323806a0

Friese, U., Supp, G. G., Hipp, J. F., Engel, A. K., & Gruber, T. (2012). Oscillatory MEG gamma band activity dissociates perceptual and conceptual aspects of visual object processing: a combined repetition/conceptual priming study. *NeuroImage*, *59*(1), 861–71. http://doi.org/10.1016/j.neuroimage.2011.07.073

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Düzel, E. (2010). Theta-coupled periodic replay in working memory. *Current Biology : CB*, *20*(7), 606–12. http://doi.org/10.1016/j.cub.2010.01.057

Gallivan, J. P., & Goodale, M. A. (2018). The dorsal “action” pathway. *Handbook of Clinical Neurology*, *151*, 449–466. http://doi.org/10.1016/B978-0-444-63622-5.00023-1

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

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 694–9. http://doi.org/10.1073/pnas.98.2.694

Gruber, W. R., Klimesch, W., Sauseng, P., & Doppelmayr, M. (2005). Alpha

phase synchronization predicts P1 and N1 latency and amplitude size. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(4), 371–7. http://doi.org/10.1093/cercor/bhh139

Hanert, A., Weber, F. D., Pedersen, A., Born, J., & Bartsch, T. (2017). Sleep in Humans Stabilizes Pattern Separation Performance. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *37*(50), 12238–12246. http://doi.org/10.1523/JNEUROSCI.1189-17.2017

Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing Visual Scenes in Cortex: Gradients of Object and Spatial Layout Information. *Cerebral Cortex*, *23*(4), 947–957. http://doi.org/10.1093/cercor/bhs091

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

Hasselmo, M. E., Bodelón, C., & Wyble, B. P. (2002). A Proposed Function for Hippocampal Theta Rhythm: Separate Phases of Encoding and Retrieval Enhance Reversal of Prior Learning. *Neural Computation*, *14*(4), 793–817. http://doi.org/10.1162/089976602317318965

Hasselmo, M. E., & Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Networks : The Official Journal of the International Neural Network Society*, *18*(9), 1172–90. http://doi.org/10.1016/j.neunet.2005.08.007

Hasselmo, M. E., & Schnell, E. (1994). Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1:

computational modeling and brain slice physiology. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *14*(6), 3898–914. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8207494

Hebart, M. N., & Baker, C. I. (2017). Deconstructing multivariate decoding for the study of brain function. *NeuroImage*, 0–31. http://doi.org/10.1016/j.neuroimage.2017.08.005

Hebart, M. N., Bankson, Harel, Baker, C. I., & Cichy, R. M. (2017). Representational dynamics of task context and its influence on visual object processing, 0–22. http://doi.org/10.1101/153684

Helmholtz, H. (1924). *Treatise on physiological optics. Optical Society of America (1924–5), English translation*. (J. P. C. Southall, Ed.). Rochester: Optical Society of America. http://doi.org/10.1037/13536-000

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–26. http://doi.org/10.1002/hipo.20857

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), 1374–80. http://doi.org/10.1038/nn.4374

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*(1), 7462. http://doi.org/10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2014). Pattern Completion in Multielement Event Engrams. *Current Biology*, *24*(9), 988–992. http://doi.org/10.1016/j.cub.2014.03.012

Horner, A. J., & Doeller, C. F. (2017). Plasticity of hippocampal memories in humans. *Current Opinion in Neurobiology*, *43*, 102–109. http://doi.org/10.1016/j.conb.2017.02.004

Howard, M. W., & Eichenbaum, H. (2013). The hippocampus, time, and memory across scales. *Journal of Experimental Psychology: General*, *142*(4), 1211–1230. http://doi.org/10.1037/a0033621

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4966457

Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *The Journal of Comparative Neurology*, *177*(3), 361–379. http://doi.org/10.1002/cne.901770302

Huerta, P. T., & Lisman, J. E. (1993). Heightened synaptic plasticity of hippocampal CA1 neurons during a cholinergically induced rhythmic state. *Nature*, *364*(6439), 723–5. http://doi.org/10.1038/364723a0

Hyman, J. M., Wyble, B. P., Goyal, V., Rossi, C. A., & Hasselmo, M. E. (2003). Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to the peak of theta and long-term depression when delivered to the trough. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(37), 11725–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14684874

Jafarpour, A., Fuentemilla, L., Horner, A. J., Penny, W., & Duzel, E. (2014). Replay of Very Early Encoding Representations during Recollection. *Journal of Neuroscience*, *34*(1), 242–248.

http://doi.org/10.1523/JNEUROSCI.1865-13.2014

Jamalabadi, H., Alizadeh, S., Schönauer, M., Leibold, C., & Gais, S. (2016). Classification based hypothesis testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. *Human Brain Mapping*, *37*(5), 1842–1855. http://doi.org/10.1002/hbm.23140

James, W. (1890). *Principles of Psychology*. New York: Henry Holt and Company.

Jensen, O. (2006). Maintenance of multiple working memory items by temporal segmentation. *Neuroscience*, *139*(1), 237–49. http://doi.org/10.1016/j.neuroscience.2005.06.004

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, *16*(4), 200–6. http://doi.org/10.1016/j.tics.2012.03.002

Jensen, O., Kaiser, J., & Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317–324. http://doi.org/10.1016/j.tins.2007.05.001

Jiang, Y., Lee, M. T., & Rosner, B. (2017). Wilcoxon Rank-Based Tests for Clustered Data with R Package clusrank.

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, Familiarity, and Cortical Reinstatement: A Multivoxel Pattern Analysis. *Neuron*, *63*(5), 697–708. http://doi.org/10.1016/j.neuron.2009.08.011

Johnson, J. D., Price, M. H., & Leiker, E. K. (2015). Episodic retrieval involves

early and sustained effects of reactivating information from encoding. *NeuroImage*, *106*, 300–10. http://doi.org/10.1016/j.neuroimage.2014.11.013

Kent, B. A., Hvoslef-Eide, M., Saksida, L. M., & Bussey, T. J. (2016). The representational-hierarchical view of pattern separation: Not just hippocampus, not just space, not just memory? *Neurobiology of Learning and Memory*, *129*, 99–106. http://doi.org/10.1016/j.nlm.2016.01.006

Kesner, R. P., & Rolls, E. T. (2015). A computational theory of hippocampal function, and tests of the theory: New developments. *Neuroscience & Biobehavioral Reviews*, *48*, 92–147. http://doi.org/10.1016/j.neubiorev.2014.11.009

Ketz, N., Morkonda, S. G., & O'Reilly, R. C. (2013). Theta coordinated error-driven learning in the hippocampus. *PLoS Computational Biology*, *9*(6), e1003067. http://doi.org/10.1371/journal.pcbi.1003067

King, J., & Dehaene, S. (2014). Characterizing the dynamics of mental representations : the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. http://doi.org/10.1016/j.tics.2014.01.002

Klimesch, W., Doppelmayr, M., Yonelinas, A., Kroll, N. E. A., Lazzara, M., Röhm, D., & Gruber, W. (2001). Theta synchronization during episodic retrieval: Neural correlates of conscious awareness. *Cognitive Brain Research*, *12*(1), 33–38. http://doi.org/10.1016/S0926-6410(01)00024-6

Konkle, T., & Oliva, A. (2012). A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. *Neuron*, *74*(6), 1114–1124. http://doi.org/10.1016/j.neuron.2012.04.036

Kowalczyk, A., & Chapelle, O. (2005). An analysis of the anti-learning phenomenon for the class symmetric polyhedron. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3734 LNAI*, 78–91. http://doi.org/10.1007/11564089_8

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 4. http://doi.org/10.3389/neuro.06.004.2008

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences*, *108*(14), 5903–5908. http://doi.org/10.1073/pnas.1016939108

Kuhl, B. a, Bainbridge, W. a, & Chun, M. M. (2012). Neural reactivation reveals mechanisms for updating memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(10), 3453–61. http://doi.org/10.1523/JNEUROSCI.5846-11.2012

Kumaran, D., Hassabis, D., & Mcclelland, J. L. (2016). What Learning Systems do Intelligent Agents Need ? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, *20*(7), 512–534. http://doi.org/10.1016/j.tics.2016.05.004

Kunec, S., Hasselmo, M. E., & Kopell, N. (2005). Encoding and Retrieval in the CA3 Region of the Hippocampus: A Model of Theta-Phase Separation. *Journal of Neurophysiology*, *94*(1), 70–82.

http://doi.org/10.1152/jn.00731.2004

Kurth-Nelson, Z., Barnes, G., Sejdinovic, D., Dolan, R., & Dayan, P. (2015). Temporal structure in associative retrieval. *ELife*, *2015*(4), 1–18. http://doi.org/10.7554/eLife.04919

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., … Mazziotta, J. C. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping*, *5*(4), 238–42. http://doi.org/10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., … Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120–31.

Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus*, *10*(4), 420–30. http://doi.org/10.1002/1098-1063(2000)10:4<420::AID-HIPO8>3.0.CO;2-5

Leal, S. L., Tighe, S. K., Jones, C. K., & Yassa, M. a. (2014). Pattern separation of emotional information in hippocampal dentate and CA3. *Hippocampus*, *1155*(April), 1146–1155. http://doi.org/10.1002/hipo.22298

Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*, *22*(4), 748–61. http://doi.org/10.1002/hipo.20937

Lehky, S. R., & Tanaka, K. (2016). Neural representation for object recognition in inferotemporal cortex. *Current Opinion in Neurobiology*, *37*, 23–35.

http://doi.org/10.1016/j.conb.2015.12.001

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, *56*(2), 387–399. http://doi.org/10.1016/j.neuroimage.2010.11.004

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*(August), 1171. http://doi.org/10.1016/j.brainres.2009.05.091

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *262*(841), 23–81. http://doi.org/10.1098/rstb.1971.0078

Marr, D., & Nishihara, H. K. (1978). Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society B: Biological Sciences*, *200*(1140), 269–294. http://doi.org/10.1098/rspb.1978.0020

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *ELife*, *7*. http://doi.org/10.7554/eLife.31873

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, *49*(5), 1127–1147. http://doi.org/10.1152/jn.1983.49.5.1127

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning

and memory. *Psychological Review*, *102*(3), 419–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7624455

Mecklinger, A. (2006). Electrophysiological Measures of Familiarity Memory. *Clinical EEG and Neuroscience*, *37*(4), 292–299. http://doi.org/10.1177/155005940603700406

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), 1–27. http://doi.org/10.1371/journal.pbio.1002528

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

Milner, A. D. (2017). How do the two visual streams interact with each other? *Experimental Brain Research*, *235*(5), 1297–1308. http://doi.org/10.1007/s00221-017-4917-4

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785. http://doi.org/10.1016/j.neuropsychologia.2007.10.005

Mishkin, M., Vargha-Khadem, F., & Gadian, D. G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus*, *8*(3), 212–216. http://doi.org/10.1002/(SICI)1098-1063(1998)8:3<212::AID-HIPO4>3.0.CO;2-L

Moscovitch, M. (2008). The Hippocampus As a “Stupid,” Domain-Specific Module: Implications for Theories of Recent and Remote Memory, and of

Imagination. *Canadian Journal of Experimental Psychology*, *62*(1), 62–79. http://doi.org/10.1037/1196-1961.62.1.62

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, *67*(1), 105–134. http://doi.org/10.1146/annurev-psych-113011-143733

Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, *31*, 69–89. http://doi.org/10.1146/annurev.neuro.31.061307.090723

Murty, V. P., DuBrow, S., & Davachi, L. (2018). Decision-making Increases Episodic Memory via Postencoding Consolidation. *Journal of Cognitive Neuroscience*, 1–10. http://doi.org/10.1162/jocn_a_01321

Nadel, L., & Peterson, M. A. (2013). The hippocampus: Part of an interactive posterior representational system spanning perceptual and memorial systems. *Journal of Experimental Psychology: General*, *142*(4), 1242–1254. http://doi.org/10.1037/a0033690

Nakashiba, T., Buhl, D. L., McHugh, T. J., & Tonegawa, S. (2009). Hippocampal CA3 Output Is Crucial for Ripple-Associated Reactivation and Consolidation of Memory. *Neuron*, *62*(6), 781–787. http://doi.org/10.1016/j.neuron.2009.05.013

Norman, K. A. (2006). Episodic Memory, Computational Models of. In *Encyclopedia of Cognitive Science*. Chichester: John Wiley & Sons, Ltd. http://doi.org/10.1002/0470018860.s00444

Norman, K. A. (2012). Revisiting the Complementary Learning Systems model,

20(11), 1217–1227. http://doi.org/10.1002/hipo.20855.How

Norman, K. A., Newman, E., Detre, G., & Polyn, S. (2006). How inhibitory oscillations can train neural networks and punish competitors. *Neural Computation*, *18*(7), 1577–610. http://doi.org/10.1162/neco.2006.18.7.1577

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), 611–46. http://doi.org/10.1037/0033-295X.110.4.611

Norman, K. a, Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–30. http://doi.org/10.1016/j.tics.2006.07.005

Nyhus, E., & Curran, T. (2009). Semantic and perceptual effects on recognition memory: Evidence from ERP. *Brain Research*, *1283*, 102–114. http://doi.org/10.1016/j.brainres.2009.05.091

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1023–35. http://doi.org/10.1016/j.neubiorev.2009.12.014

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. http://doi.org/10.1038/nn.3248

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*(3), 317–30. http://doi.org/10.1002/hipo.450030307

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

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

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus*, *4*(6), 661–82. http://doi.org/10.1002/hipo.450040605

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*(12), 505–510. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12475710

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. http://doi.org/10.1155/2011/156869

Orban, G. A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *Journal of Neurophysiology*, *56*(2), 462–80. http://doi.org/10.1152/jn.1986.56.2.462

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, *3*(JAN), 1–8. http://doi.org/10.3389/fpsyg.2012.00620

Parish, G., Hanslmayr, S., & Bowman, H. (2018). The Sync/deSync Model: How a Synchronized Hippocampus and a Desynchronized Neocortex Code Memories. *The Journal of Neuroscience*, *38*(14), 3428–3440. http://doi.org/10.1523/JNEUROSCI.2561-17.2018

Pasupathy, A., & Connor, C. E. (2001). Shape Representation in Area V4: Position-Specific Tuning for Boundary Conformation. *Journal of Neurophysiology*, *86*(5), 2505–2519. http://doi.org/10.1152/jn.2001.86.5.2505

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. http://doi.org/10.1038/nrn2277

Pavlides, C., Greenstein, Y. J., Grudman, M., & Winson, J. (1988). Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. *Brain Research*, *439*(1–2), 383–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3359196

Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *9*(5), 1749–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2723748

Poggio, T., & Ullman, S. (2013). Vision: are models of object recognition catching up with the brain? *Annals of the New York Academy of Sciences*, *1305*(1), 72–82. http://doi.org/10.1111/nyas.12148

Poldrack, R., & Packard, M. (2003). Competition among multiple memory

systems: converging evidence from animal and human brain studies. *Neuropsychologia*, *41*, 245–251. http://doi.org/10.1016/S0028-3932(02)00157-4

Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience*, *13*, 587. Retrieved from http://dx.doi.org/10.1038/nrn3251

Raffi, M., Persiani, M., Piras, A., & Squatrito, S. (2014). Optic flow neurons in area PEc integrate eye and head position signals. *Neuroscience Letters*, *568*, 23–28. http://doi.org/10.1016/j.neulet.2014.03.042

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42. Retrieved from http://dx.doi.org/10.1038/nrn.2016.150

Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*, *98*, 262–268. http://doi.org/10.1016/j.cortex.2017.10.020

Richter, F. R., Chanales, A. J. H., & Kuhl, B. A. (2016). Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *NeuroImage*, *124*, 323–335. http://doi.org/10.1016/j.neuroimage.2015.08.051

Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: insights from functional brain imaging. *Annual Review of Psychology*, *63*, 101–28. http://doi.org/10.1146/annurev-psych-120710-100344

Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging Object

Representations in the Visual System Predict Reaction Times for Categorization. *PLoS Computational Biology*, *11*(6), 1–19. http://doi.org/10.1371/journal.pcbi.1004316

Rizzuto, D. S., Madsen, J. R., Bromfield, E. B., Schulze-Bonhage, A., & Kahana, M. J. (2006). Human neocortical oscillations exhibit theta phase differences between encoding and retrieval. *NeuroImage*, *31*(3), 1352–8. http://doi.org/10.1016/j.neuroimage.2006.01.009

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–81. http://doi.org/10.1038/26475

Rolls, E. T. (1991). Functions of the primate hippocampus in spatial and nonspatial memory. *Hippocampus*, *1*(3), 258–261. http://doi.org/10.1002/hipo.450010310

Rolls, E. T. (1996). A theory of hippocampal function in memory. *Hippocampus*, *6*(6), 601–20. http://doi.org/10.1002/(SICI)1098-1063(1996)6:6<601::AID-HIPO5>3.0.CO;2-J

Rolls, E. T. (2010). A computational theory of episodic memory formation in the hippocampus. *Behavioural Brain Research*, *215*(2), 180–196. http://doi.org/10.1016/j.bbr.2010.03.027

Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, *7*(October), 74. http://doi.org/10.3389/fnsys.2013.00074

Rolls, E. T. (2015). Pattern separation, completion, and categorisation in the hippocampus and neocortex. *Neurobiology of Learning and Memory*,

(2015). http://doi.org/10.1016/j.nlm.2015.07.008

Rolls, E. T. (2017). The storage and recall of memories in the hippocampo-cortical system. *Cell and Tissue Research*, 1–28. http://doi.org/10.1007/s00441-017-2744-3

Rossi, S., Pasqualetti, P., Zito, G., Vecchio, F., Cappa, S. F., Miniussi, C., … Rossini, P. M. (2006). Prefrontal and parietal cortex in human episodic memory: an interference study by repetitive transcranial magnetic stimulation. *The European Journal of Neuroscience*, *23*(3), 793–800. http://doi.org/10.1111/j.1460-9568.2006.04600.x

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. http://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., & Vilberg, K. L. (2014). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*(2), 255–260. http://doi.org/10.1016/j.conb.2012.11.005.Brain

Schacter, D. L. (2012, March). Constructive memory: past and future. *Dialogues in Clinical Neuroscience*. France.

Schacter, D. L., Guerin, S. a, & St Jacques, P. L. (2011). Memory distortion: an adaptive perspective. *Trends in Cognitive Sciences*, *15*(10), 467–74. http://doi.org/10.1016/j.tics.2011.08.004

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. *Current Biology*, *22*(17), 1622–1627. http://doi.org/10.1016/j.cub.2012.06.056

Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *The Journal of Neuroscience*, *10*(10), 3369–3389.

Schönauer, M., Alizadeh, S., Jamalabadi, H., Abraham, A., Pawlizki, A., & Gais, S. (2017). Decoding material-specific memory reprocessing during sleep in humans. *Nature Communications*, *8*, 15404. http://doi.org/10.1038/ncomms15404

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4447–52. http://doi.org/10.1073/pnas.0800431105

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429. http://doi.org/10.1073/pnas.0700622104

Siegel, M., Warden, M. R., & Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences*, *106*(50), 21341–21346. http://doi.org/10.1073/pnas.0908193106

Solomon, S. G. (2005). Chromatic Gain Controls in Visual Cortical Neurons. *Journal of Neuroscience*, *25*(19), 4779–4792. http://doi.org/10.1523/JNEUROSCI.5316-04.2005

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. http://doi.org/10.1016/j.cub.2017.09.057

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). the Medial Temporal Lobe. *Annual Review of Neuroscience*, *27*(1), 279–306. http://doi.org/10.1146/annurev.neuro.27.070203.144130

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

Staresina, B. P., Cooper, E., & Henson, R. N. (2013). Reversible Information Flow across the Medial Temporal Lobe: The Hippocampus Links Cortical Modules during Memory Retrieval. *Journal of Neuroscience*, *33*(35), 14184–14192. http://doi.org/10.1523/JNEUROSCI.1987-13.2013

Staresina, B. P., & Davachi, L. (2008). Selective and Shared Contributions of the Hippocampus and Perirhinal Cortex to Episodic Item and Associative Encoding. *Journal of Cognitive Neuroscience*, *20*(8), 1478–1489. http://doi.org/10.1162/jocn.2008.20104

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), 18150–6. http://doi.org/10.1523/JNEUROSCI.4156-12.2012

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. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*(13), 5373–84.

http://doi.org/10.1523/JNEUROSCI.4198-14.2015

Staudigl, T., Zaehle, T., Voges, J., Hanslmayr, S., Esslinger, C., Hinrichs, H., … Richardson-Klavehn, A. (2012). Memory signals from the thalamus: Early thalamocortical phase synchronization entrains gamma oscillations during long-term memory retrieval. *Neuropsychologia*, *50*(14), 3519–3527. http://doi.org/10.1016/j.neuropsychologia.2012.08.023

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

Suzuki, W. A., & Amaral, D. G. (1994). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *14*(3 Pt 2), 1856–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8126576

Suzuki, W. L., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *The Journal of Comparative Neurology*, *350*(4), 497–533. http://doi.org/10.1002/cne.903500402

Tallon-Baudry, & Bertrand. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, *66*(1), 170–189. http://doi.org/10.1152/jn.1991.66.1.170

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory.

Behavioral Neuroscience, *100*(2), 147–54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3008780

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

Tort, A. B. L., Komorowski, R., Eichenbaum, H., & Kopell, N. (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of Neurophysiology*, *104*(2), 1195–210. http://doi.org/10.1152/jn.00106.2010

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1), 190–198. http://doi.org/10.1111/nyas.12680

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391. http://doi.org/10.1002/hipo.450040319

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*. http://doi.org/10.1017/S0140525X00047257

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Oxford University Press.

Tulving, E. (1984). Relations among components and processes of memory. *Behavioral and Brain Sciences*, *7*(02), 257. http://doi.org/10.1017/S0140525X00044617

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, *53*(1), 1–25.

http://doi.org/10.1146/annurev.psych.53.100901.135114

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

Tulving, E., Voi, M. E. L., Routh, D. A., & Loftus, E. (1983). Ecphoric Processes in Episodic Memory [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *302*(1110), 361–371. http://doi.org/10.1098/rstb.1983.0060

Tyler, L. K., Chiu, S., Zhuang, J., Randall, B., Devereux, B. J., Wright, P., … Taylor, K. I. (2013). Objects and categories: feature statistics and object processing in the ventral stream. *Journal of Cognitive Neuroscience*, *25*(10), 1723–35. http://doi.org/10.1162/jocn_a_00419

Van de Nieuwenhuijzen, M. E., Backus, A. R., Bahramisharif, A., Doeller, C. F., Jensen, O., & van Gerven, M. A. J. (2013). MEG-based decoding of the spatiotemporal dynamics of visual category perception. *NeuroImage*, *83*, 1063–1073. http://doi.org/10.1016/j.neuroimage.2013.07.075

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 Bio-Medical Engineering*, *44*(9), 867–80. http://doi.org/10.1109/10.623056

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*(4654), 1260–1262. http://doi.org/10.1126/science.6539501

Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., & Gazzaley, A. (2015). Age-Related Changes in 1/f Neural Electrophysiological Noise. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*(38), 13257–65. http://doi.org/10.1523/JNEUROSCI.2332-14.2015

Waldhauser, G. T., Braun, V., & Hanslmayr, S. (2016). Episodic Memory Retrieval Functionally Relies on Very Rapid Reactivation of Sensory Information. *The Journal of Neuroscience*, *36*(1), 251–260. http://doi.org/10.1523/JNEUROSCI.2101-15.2016

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

Watrous, A. J., & Ekstrom, A. D. (2014). The Spectro-Contextual Encoding and Retrieval Theory of Episodic Memory. *Frontiers in Human Neuroscience*, *8*. http://doi.org/10.3389/fnhum.2014.00075

Watrous, A. J., Miller, J., Qasim, S. E., Fried, I., & Jacobs, J. (2018). Phase-tuned neuronal firing encodes human contextual representations for navigational goals. *ELife*, *7*. http://doi.org/10.7554/eLife.32554

Westmacott, R., Black, S. E., Freedman, M., & Moscovitch, M. (2004). The contribution of autobiographical significance to semantic memory: evidence from Alzheimer's disease, semantic dementia, and amnesia. *Neuropsychologia*, *42*(1), 25–48. http://doi.org/10.1016/S0028-3932(03)00147-7

Westmacott, R., Freedman, M., Black, S. E., Stokes, K. A., & Moscovitch, M. (2004). Temporally graded semantic memory loss in Alzheimer's disease: cross-sectional and longitudinal studies. *Cognitive Neuropsychology*, *21*(2), 353–78. http://doi.org/10.1080/02643290342000375

Wikenheiser, A. M., & Redish, A. D. (2015). Decoding the cognitive map: ensemble hippocampal sequences and decision making. *Current Opinion in Neurobiology*, *32*, 8–15. http://doi.org/10.1016/j.conb.2014.10.002

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. http://doi.org/10.3758/BRM.42.3.671

Williams, M., Hong, S. W., Kang, M.-S., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. *Psychonomic Bulletin & Review*, *20*(2), 348–55. http://doi.org/10.3758/s13423-012-0354-3

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. http://doi.org/10.1038/nn.3973

Wimber, M., Maaß, A., Staudigl, T., Richardson-Klavehn, A., & Hanslmayr, S. (2012). Rapid memory reactivation revealed by oscillatory entrainment. *Current Biology : CB*, *22*(16), 1482–6. http://doi.org/10.1016/j.cub.2012.05.054

Witter, M. P., Naber, P. A., van Haeften, T., Machielsen, W. C., Rombouts, S. A., Barkhof, F., … Lopes da Silva, F. H. (2000). Cortico-hippocampal

communication by way of parallel parahippocampal-subicular pathways. *Hippocampus*, *10*(4), 398–410. http://doi.org/10.1002/1098-1063(2000)10:4<398::AID-HIPO6>3.0.CO;2-K

Xiao, R., & Ding, L. (2015). EEG resolutions in detecting and decoding finger movements from spectral analysis. *Frontiers in Neuroscience*, *9*(11), 308. http://doi.org/10.3389/fnins.2015.00308

Yang, Y., & Wang, J.-Z. (2017). From Structure to Behavior in Basolateral Amygdala-Hippocampus Circuits. *Frontiers in Neural Circuits*, *11*(October), 1–8. http://doi.org/10.3389/fncir.2017.00086

Yassa, M. a., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. http://doi.org/10.1016/j.tins.2011.06.006

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

Yonelinas, A. P., Aly, M., Wang, W.-C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178–1194. http://doi.org/10.1002/hipo.20864

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(17), 6594–611. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10964965