Neural representational similarity in episodic and spatial memory.

Ewa Zotow

Submitted for the Degree of Doctor of Philosophy in Cognitive Neuroscience

University College London (UCL)

Supervisor: Prof Neil Burgess

Declaration

I, Ewa Zotow, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Ewa Zotow

Date: 14 September 2018

Abstract

Electrophysiological recordings in rodents and humans show that the contents of spatial and episodic memories are encoded in patterns of activity across neural populations. Here, I applied a mixture of univariate and multivariate techniques to functional magnetic resonance imaging (fMRI) data to investigate memory processes as well as look at how activity across voxels relates to the way memories are represented in the brain.

In the first part of the thesis, I applied the Representational Similarity Analysis (RSA) to investigate the brain representations underlying spatial memories. I examined fMRI data from participants during actual or imagined navigation in virtual environments. Evidence for grid cells, one of the principal types of spatial cells, has been reported in single-cell recordings and in univariate approaches to fMRI data. I found evidence for viewing direction within occipital areas in both modalities, suggesting involvement of the same system in both tasks, but not for the six-fold symmetry characteristic of the grid cell signal. I discuss the potential reasons for this null result.

In the second part I looked at episodic memories, in particular whether similar information undergoes 'pattern separation' during encoding to minimise future interference. I designed tasks using face morphing stimuli and multi-element events with overlapping elements to look for behavioural evidence for pattern separation and interference. The overlapping events showed increased independence in performance across multiple retrievals, suggesting the formation of more independent representations compared to unrelated events. I then looked for neural signatures of this effect using fMRI RSA. I found evidence for a decreased similarity of patterns of overlapping events at retrieval but not at encoding, possibly due to larger influence of perceptual similarity at encoding. Increased pattern separation was not related to decreased interference (improved performance). Lastly, I found evidence of non-target reinstatement, consistently with holistic representation of memory episodes.

Overall, this thesis presents new findings as well as replications of previously observed effects using a range of novel behavioural and analysis techniques to investigate how spatial and episodic memories are represented in the human brain.

Impact statement

Adaptive behaviour in a complex world requires that we are capable of remembering essential details from our lives: what has happened, where and when. Memory for experienced events is not only crucial to our ability to act appropriately and plan the future, but is also fundamental to our sense of identity. Just how important memory is can be seen when it goes awry in old age or disease. Understanding of the processes underlying our ability to encode, store and retrieve information is therefore of great importance at a time when increasing lifespan means that more and more people will develop dementia or other memory impairments at some point in their life.

This thesis investigated some of the mechanisms proposed to support spatial and episodic (autobiographical) memory. I used a novel technique for analysis of brain imaging data, the Representational Similarity Analysis, to look at how memories are encoded in the patterns of brain activity. The findings presented here add to the growing literature on how the distribution of signal (and not only its magnitude, as used in standard approaches to imaging data analysis) can help to understand the characteristics of information processing in different brain areas. I looked at how the brain represents the features of the spatial environment, and how it represents memories for experienced events. I show that by looking at activity patterns we can answer more detailed hypotheses about the internal 'syntax' of the brain than would be available through traditional approaches. I also extensively discuss the constraints of this method and the ways to capitalise on its strengths while avoiding its drawbacks. The findings and issues discussed in this thesis can inform future studies looking to use this method of analysis in cognitive neuroscience as well as other fields.

Finally, as a part of this thesis I adapted a previously used memory task, which allowed for looking at a different memory mechanism in addition to the one measured in the original paradigm (pattern separation of memory representations in addition to pattern completion; both these processes are essential for memory encoding and retrieval). The need for a design of a novel task was brought about by the limitations of the currently used methods which are suboptimal for looking at this memory process. This task presents an alternative (and arguably more suitable) measure which could replace the existing paradigm in the continuously growing field of episodic memory research and potentially within clinical setting.

Acknowledgements

Writing this thesis would not be possible without the help and support of many people. First and foremost, I am immensely grateful for all the guidance, insights and expertise of my primary supervisor, Neil Burgess. I consider myself extremely lucky to have worked and developed as a scientist under his supervision.

Massive thanks go to James Bisby, who put a tremendous amount of time and effort to support me through all stages of my PhD journey, and especially during writing this thesis. I would also like to thank Aidan Horner, who co-supervised my first project and was always really helpful and encouraging.

Thanks also go to my secondary supervisor, Sam Gilbert, for helpful discussions and support, and to Joern Diedrichsen; I learnt a lot in the short time I worked under his supervision.

I would like to thank all members of my research group. Special thanks to Dan Bush for all the advice and help with my analyses, as well as to Lone Hørlyck for emotional support and much needed conversations and to Talfan Evans for being the best desk buddy. Massive thank you to Andrea, Sofie, Alexa, Andrej, Siti, Anika, Thomas, Jesse, Katja and Sebastian; I am very grateful to have worked alongside such amazing scientists and great colleagues.

This thesis would not be possible without the vital help from the staff at the Wellcome Trust Neuroimaging Centre; special thanks to Megan, Clive and Elaine for their help with my scanning project. I would also like to thank Alastair McClelland and Chris McManus for being great statistics mentors during my years as a teaching assistant.

I am so very lucky to have spent the years of my PhD with the amazing colleagues at the ICN. Above all, I want to thank Anne Löffler for being my writing (and climbing) companion during the last months of writing this thesis. I don't think this thesis would exist without her constant support and encouragement (I would also be a much worse climber). My ICN colleagues were not only great people to relax with during breaks, but many of them became some of my best friends. I will always cherish the memories of our many adventures at and outside of the ICN.

Last but not least, I want to thank my parents for their unconditional love and support during writing this thesis and throughout my life.

Table of content

List of figures

List of tables

Abbreviations

BVC	boundary vector cells
CA1 / CA3	*Cornu Ammonis* 1 / 3 (subfields of the hippocampus)
DG	dentate gyrus
EC	entorhinal cortex
fMRI	functional magnetic resonance imaging
FWE	family-wise error
GLM	general linear model
HD	head-direction (cells)
MEC	medial entorhinal cortex
MST	Mnemonic Similarity Task
MTL	medial temporal lobes
OPA	occipital place area
PPA	parahippocampal place area
RDM	Representational Dissimilarity Matrix
ROI	region of interest
RSA	Representational Similarity Analysis
RSC	retrosplenial cortex
SVC	small-volume correction
VR	virtual reality

Chapter 1

Episodic and spatial memory, and the hippocampus

Memories form an essential part of an individual’s life. Knowing facts about the world, acquiring skills and remembering things that have happened to us are some of the most important aspects of our existence. In particular, the memory for events from our own life constitutes a basis for the formation of our self and understanding of our conscious experience, allowing us to adapt and function within the world. Just how important this ability is can be seen when it suddenly fails. For example, patient N.N., after an accident resulting in closed head injury, was reported to exhibit preserved semantic (memory for facts about the world) and working memory while being unable to recall any events from his life (apart from very few remote memories) or envisage future events. When asked about his past or future, he described a "blankness" which felt "like being asleep" (Tulving, 1985). As a result, he lived in "permanent present", unable to relate past experiences to the present or use them to plan the future.

The ability to store and recall the details of our past experiences is therefore extremely important. In everyday life, we often rely on our memory being correct and informative. When remembering our way to work, where we parked our car or what we spoke about at a meeting, we need to be able to retrieve and reconstruct in detail what exactly happened, when and where. This type of memory is referred to as episodic memory as it is related to remembering of specific events or episodes from one's own life. Episodic memory is characterized by our ability to re-experience past events in great detail, allowing us to retrieve the features of an event, such as the people and objects we interacted with, in its rich spatiotemporal context

(Tulving, 1972). This vivid recollective process in which all event elements are retrieved is a hallmark quality of episodic memory. Research over numerous decades has reliably pointed to the hippocampus and surrounding medial temporal lobe structures (MTL) as essential structures in supporting episodic memory. Early investigations of patient H.M. were influential in showing that bilateral resection of the hippocampus, due to intractable epilepsy, resulted in severe amnesia (Scoville & Milner, 1957). Patients with damage to the hippocampus are unable to form new episodic memories despite being able to hold digits in immediate memory, acquire new vocabulary and learn new skills (Cohen & Eichenbaum, 1993; Hassabis, Kumaran, Vann, & Maguire, 2007; Milner, Corkin, & Teuber, 1968; Squire, 1986; Vargha-Khadem et al., 1997). This led to the view that the hippocampus was essential in supporting episodic memory: memory for events from one's life.

Across a broad range of studies and techniques, the hippocampus has consistently been identified in imaging studies that assess encoding of information that is later remembered (Horner, Bisby, Bush, Lin, & Burgess, 2015; Schacter & Wagner, 1999; Wheeler & Buckner, 2004; Woodruff, Johnson, Uncapher, & Rugg, 2005), when making decisions based on past experiences (Barron, Dolan, & Behrens, 2013; Bornstein & Norman, 2017; Gluth, Sommer, Rieskamp, & Büchel, 2015; Wimmer & Shohamy, 2012), remembering the temporal order of experienced events (Deuker, Bellmund, Schröder, & Doeller, 2016; Heusser, Poeppel, Ezzyat, & Davachi, 2016; Kyle, Smuda, Hassan, & Ekstrom, 2015), imagining hypothetical events (Benoit & Schacter, 2015; Schacter et al., 2012), mental simulation of the future (Hassabis & Maguire, 2007; Schacter & Addis, 2007; Thakral, Benoit, & Schacter, 2017), and a multitude of other tasks involved in retrieval of episodic memories (Hassabis & Maguire, 2007; Tulving, 2002). The role of the hippocampus therefore encompasses a wide range of cognitive functions which extend beyond encoding and retrieval of experienced events.

In addition to its function in episodic memory, studies have reliably shown that the hippocampus plays a fundamental role in spatial memory (e.g., Morris, Garrud, Rawlins, & O'Keefe, 1982; O'Keefe, Nadel, Keightley, & Kill, 1975). Stemming from early rodent work showing that single cells in the hippocampus increase their firing rate to a specific location in the environment (O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978), a vast amount of work has shown how cells in the hippocampus and surrounding areas can support spatial memory. Others have gone a step further in proposing that the neural machinery that is involved in providing a spatial representation of an experience can also support the provision of a spatiotemporal context that is a crucial requirement of episodic memory (Byrne, Becker, & Burgess, 2007; Schiller et al., 2015; Tulving, 1985a). The spatial memory system can encode and integrate discrete experiences within a context, which is essential to both representations of space required for navigation as well as episodic memory. For instance, a Byrne, Becker and Burgess (BBB) model, which will be described in greater detail below, proposes that the same mechanisms that guide navigation also support construction of spatially coherent mental images of events, as is necessary for episodic memory recall and imagination (Byrne et al., 2007).

This account of spatial context as a framework for formation of memories is consistent with the idea of 'cognitive map' envisioned by Tolman in 1948, before the modern era of study of the neural mechanisms underlying spatial navigation even began. The map was proposed to be an internal representation of the environment, which can be used to flexibly navigate around it without the constraints of the limited number of learnt stimulus-response sequences. The concept of a cognitive map as an internal representation of space has evolved but the general idea that the hippocampus supports the formation of associations among elements and binding them to a common context has persisted in the models of both spatial and episodic memory (Ekstrom & Ranganath, 2017; Epstein, Patai, Julian, & Spiers, 2017; McNaughton, Battaglia, Jensen, Moser, & Moser, 2006; Schiller et al., 2015).

In the spirit of the role of the hippocampus in supporting episodic and spatial memory, the main aims of this thesis are to investigate some of the underlying mechanisms that are proposed to support memory and particularly how they can flexibly support encoding, recall and imagination. In brief, I will use a mixture of behavioural and neuroimaging techniques to look at some of the properties of the human memory system, in particular whether we can examine the patterns of activity in the human brain to provide evidence for the predictions about how the hippocampus underpins episodic memory. Given the evidence of spatial coding in humans, I will look at whether the same mechanisms can support higher cognitive functions such as imagination (e.g. Byrne, Becker, & Burgess, 2007; Kesner & Rolls, 2015). In particular, I will look at whether we can detect pattern of activity consistent with the presence of grid cells in the human entorhinal cortex (EC) in both navigation in virtual reality (VR) and during imagination of the same routes. I will then investigate the mechanisms involved in storage and retrieval of memories, focusing on how the memory system encodes similar memories as distinct representations in order to prevent their interference, a process called pattern separation (O'Reilly & McClelland, 1994). I will present two behavioural tasks designed to measure the extent to which similar memories are represented as more distinct. Lastly, I will investigate the mechanism of pattern separation in episodic memory using functional magnetic resonance imaging (fMRI) to look for neural correlates of the pattern separation observed in behaviour.

1.1 The role of the hippocampus in episodic memory

Episodic memory is characterized by its associative nature in which different elements from an event, such as the people or objects we interact with or the spatial scene in which it took place, are bound together as a single representation. Whilst regions surrounding the hippocampus such as the perirhinal and parahippocampal areas are thought to store domain specific representations of items and scenes respectively (Aggleton & Brown, 1999; Barense, Henson, Lee, & Graham, 2010; Staresina & Davachi, 2009), the hippocampus is thought to be crucial in binding these elements into a single coherent representation (Cohen & Eichenbaum, 1993; Davachi, 2006; Eichenbaum, Yonelinas, & Ranganath, 2007; O'Keefe & Nadel, 1978). Evidence for this role has been shown in numerous patient studies, demonstrating that hippocampal damage results in impairments in forming associations (Borders, Aly, Parks, & Yonelinas, 2017; Pajkert et al., 2017). More evidence from neurosurgical patients shows that individual neurons represent arbitrary associations between a person and spatial scene after learning (Ison, Quian Quiroga, & Fried, 2015) and that hippocampal activity is observed specifically during contextually-mediated retrieval but not intrusions or deliberations (Long et al., 2017). Additionally, neuroimaging work highlights the hippocampus in storing associations between pairs of items (Kirwan & Stark, 2004) and remembering the context (Ranganath et al., 2004), which suggests the specific involvement in tasks which rely on storing relations between individual elements.

In a similar view of dissociating associative and item-based memory processes, the hippocampus has been shown to support recollection as opposed to familiarity (Uncapher & Rugg, 2005). The recollective experience of remembering is characterised by the retrieval of contextual details (e.g. where and when a given episode was encoded), whereas familiarity is a general sense of knowing that something has been encountered but without any associated detail. Neuropsychological findings (Aggleton et al., 2005; Jeneson, Kirwan, Hopkins, Wixted, & Squire, 2010) and neuroimaging studies support this distinction (Davachi, Mitchell, & Wagner, 2003; Henson, Hornberger, & Rugg, 2005; Yonelinas, Otten, Shaw, & Rugg, 2005).

There is therefore plenty of evidence suggesting the role of the hippocampus in binding elements together to a common context, which can be defined in terms of space but also temporal features (Ezzyat & Davachi, 2014; Howard, 2017; Kragel, Morton, & Polyn, 2015) as well as other behaviourally-relevant dimension such as social hierarchy or stimulus characteristics (Aly & Turk-Browne, 2015; Tavares et al., 2015).

This ability to bind information together and retrieve these important associations has been proposed to also underlie our ability to think about the future (Schacter, Benoit, & Szpunar, 2017) and to construct hypothetical scenarios (Hassabis & Maguire, 2007; Schacter & Addis, 2007, 2009). It has been proposed that the same mechanisms involved in memory recall may be engaged in the simulation of episodic narratives when engaged in imagination, by recalling elements and flexibly manipulating them to form coherent but novel representations which share many commonalities with actual memories for experienced events (Schacter & Addis, 2007). The role of the hippocampus therefore extends beyond simple associative learning and rather it can support more complex, flexible constructive processes (see section 1.3.6 for further detail).

In addition to forming associations that bind elements together into coherent narratives, there is a need for episodic memories to be stored as distinct representations. In particular, similar memories can interfere with each other, causing forgetting, and so need to be distinguished. Meeting these two opposing requirements - to differentiate between memories of different events while preserving the associations that form the representations of each event – is one of the main challenges of the memory system. These two complementary functions, called patter separation and pattern completion, are also thought to rely on the hippocampus and will be outlined in in the next section.

1.2 Mechanisms supporting episodic memory encoding and retrieval

What are the specific mechanisms in which the hippocampus functions to support episodic memory formation and retrieval? Computational models have long proposed that the anatomical characteristics of the hippocampal circuit make it particularly suitable for rapid associative learning whilst reducing interference from similar experiences through the process of pattern separation (Kesner & Rolls, 2015; Marr, 1971; McClelland, McNaughton, & O'Reilly, 1995; Yassa & Stark, 2011). This process is thought to occur in the hippocampus and relies on its ability to decorrelate input signals, creating orthogonalized memory representations. In addition to storing these orthogonalized input patterns, the hippocampus is thought to support memory retrieval via pattern completion by which a subset of cues from a previous experience can re-activate the stored pattern representing that experience (Hopfield, 1982; Marr, 1971; McClelland et al., 1995; Rolls, 2015). These processes therefore give rise to two crucial characteristics of memory in that episodes are stored as independent non-overlapping representations and that all elements of an event memory are retrieved in a holistic manner. These mechanisms and their complementary roles in long-term memory are now described in detail below.

1.2.1 Pattern Separation

As I have described, a major feature of episodic memory is the ability for similar experiences to be stored as separate independent memories thus reducing interference when retrieved. The encoding of new memories, and in particular similar memories, is thought to involve the process of pattern separation. This process is performed by the dentate gyrus (DG) which relays input from the EC to the CA3 subfield of the hippocampus via mossy fibers (Rolls, 2015) (see Figure 1.1). Initially overlapping representations from the EC are decorrelated or orthogonalized, reducing their overlap and interference. The mossy fiber connections are sparse but very powerful and referred to as 'detonator' synapses, capable of producing postsynaptic firings with a single presynaptic spike (Rebola, Carta, & Mulle, 2017). Importantly, they are proposed to be necessary for encoding of new memories but not for their subsequent retrieval. In support of this encoding and retrieval dissociation, Lassalle, Bataille and Halley (2000) found that damage to the mossy fibers in rats selectively impaired storage of spatial memories.

This separation of patterns can be achieved if the number of neurons is large and the pattern of firing is drawn at random (Hopfield, 1982). In the hippocampus, the sparseness of neural activity increases as the signal travels from the EC to the DG and then to CA3, resulting in very low contact probability. In other words, the dilution of connectivity at each stage of processing decreases the probability that a different input pattern will activate the same population of CA3 neurons and lead to incorrect recall (Hunsaker & Kesner, 2013).

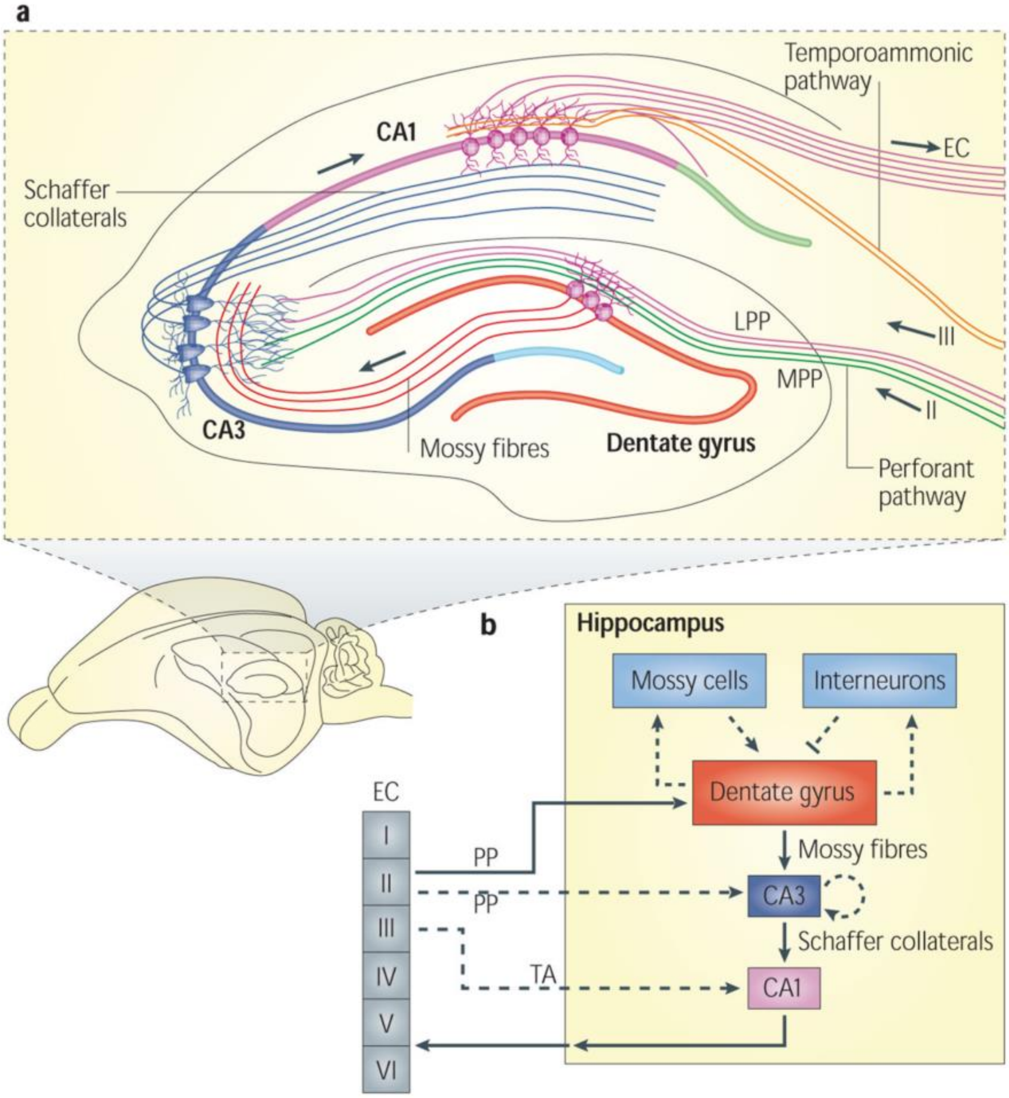


Figure 1.1. Schematic representation of the hippocampal circuitry. (a) Diagram of the hippocampal formation in a rat. (b) Diagram of the hippocampal neural network. The information from the neocortex reaches the hippocampus mainly through the superficial layers of the EC, as well as the parahippocampal gyrus and the perirhinal cortex. From the EC it reaches the CA3 directly through the perforant path and indirectly through the DG via mossy fibres. The DG is proposed to decorrelate similar inputs from the EC (O'Reilly & McClelland, 1994). The pyramidal cells of the CA3 project to other CA3 neurons (recurrent collaterals - this autoassociative network is proposed to underlie pattern completion; Marr, 1971) and to CA1 through Schaffer collaterals. The output back to the neocortex originates in the CA1 and is relayed mainly through the deep layers of the EC and the subiculum. Reprinted from Deng, Aimone, and Gage (2010).

Consistent with its role in pattern separation, DG-lesioned rats show impaired performance on a task where the location of a target object and a lure object is similar (Gilbert, Kesner, & Lee, 2001). Additionally, lesions to dorsal DG impair spatial discrimination of adjacent but not distal locations (McDonald & White, 1995). This is in line with the role of DG in separation of spatial information in order to reduce the interference between proximal spatial locations.

A distinction is often drawn between a neural pattern separation, which can only be inferred when using non-invasive methods, and behavioural pattern separation which is believed to reflect this DG function (e.g. Stark, Yassa, Lacy, & Stark, 2013). In spatial memory, evidence from fMRI studies suggests that different environments are represented as dissimilar patterns in the hippocampus, and correct performance is achieved through the reinstatement of the appropriate pattern (Kyle, Stokes, Lieberman, Hassan, & Ekstrom, 2015; Stokes, Kyle, & Ekstrom, 2013). Consistently, learning of overlapping routes leads to a reduction in similarity in their hippocampal representations (Chanales, Oza, Favila, & Kuhl, 2017).

In episodic memory, pattern separation in humans has mainly been investigated using paradigms in which similar lures need to be distinguished from previously seen items. In the most common version of this task, often referred to as Mnemonic Similarity Task (MST), participants are shown images of everyday objects and are asked to classify them as old (previously seen) or new. Some of the items (the lures) resemble previously seen items in appearance. The correct identification of the lures as novel items is assumed to reflect pattern separation.

Although in humans pattern separation at the level of neurons has not yet been observed, impairment in lure-discrimination has been observed in a rare case of selective DG lesion (Baker et al., 2016). Patient B.L. was impaired at correctly distinguishing similar items in a recognition test while showing normal recognition of non-similar items. This selective impairment suggests inability to form non-overlapping representations of similar items, which is consistent with the role of DG in pattern separation.

Imaging studies using the same task also support the role of the hippocampus, and particularly the CA3/DG subfields (it has not yet been possible to separate them using non-invasive imaging), in pattern separation in that the hippocampus shows increased activity to novel lures (Bakker, Kirwan, Miller, & Stark, 2008; De Shetler & Rissman, 2016; Duncan, Sadanand, & Davachi, 2012; Lacy, Yassa, Stark, Muftuler, & Stark, 2011; Yassa et al., 2010).

However, the novelty signal detected in the imaging studies may reflect a match-mismatch (or recall-to-reject) signal rather than pattern separation (see Chapter 4.1 for a more in-depth discussion of this and other issues). One study provided more direct evidence by showing that the lure-related signal in the DG can be detected for even highly familiar scene stimuli, thus minimizing the need for recall-to-reject (Berron et al., 2016). However, so far most of the tasks used to study pattern separation cannot be considered truly episodic in that they do not rely on the main property of this type of memory to bind information together into coherent episodes. Given the described role of the hippocampus in recollection but not recognition (Uncapher & Rugg, 2005; Yonelinas et al., 2005), continuous recognition tasks might not capture the contextual detail which should accompany recollection.

Being able to measure pattern separation behaviourally in a reliable way would allow for assessment of the functioning of the DG/CA3 which contributes to memory impairment in aging (Stark et al., 2013; Stark, Yassa, & Stark, 2010; Yassa et al., 2011), amnesic Mild Cognitive Impairment (Yassa et al., 2010) and schizophrenia (Das, Ivleva, Wagner, Stark, & Tamminga, 2014). There is therefore a need for a more suitable method which could address these issues. I will present two alternative tasks in Chapters 4 and 5.

1.2.2 Pattern Completion

One of the most crucial features of the hippocampal circuit is the extensive projections between cells within the CA3 subfield (Rebola et al., 2017). Although CA3 receives input from both the EC and the DG, the highest number of connections onto the CA3 neurons are from the CA3 axons themselves (Rolls, 2015). The CA3-CA3 projections, known as recurrent collaterals or associative/commissural (A/C) loop, are proposed to act as an attractor network in which associative memories are stored and recalled (McNaughton & Morris, 1987) (Figure 1.1).

An attractor is a stable state within a system which 'attracts' the incoming signal to move towards that state (Knierim & Zhang, 2012). Because of self-connections that are modified through learning, firing of only a subset of original population can re-activate the entire set of cells associated at encoding (Kesner & Rolls, 2015). This process, called pattern completion, reinstates all associated representations when presented with a partial cue, e.g. with a single element of the original memory (Marr, 1971). This mechanism underlies the ability to store complex associations, which may play a role in connecting events or items to related information in their spatiotemporal context.

The recurrent connections in the CA3 ensure that the recall is symmetric, i.e. any part of the memory can be reactivated by another part (e.g. Day, Langston, & Morris, 2003). This is in contrast to pattern association where the recall is unidirectional; for example, when visual percept is associated with taste, at a later point the visual cue will bring up the taste but not the other way around (Kesner & Rolls, 2015). Pattern completion allows the entire multi-element representation to be recalled from any part of it.

The autoassociative function could have been present immediately in the EC; the fact that it is not may be underlined by the need of pattern-separating circuitry (O’Reilly & McClelland, 1994). However, for the memory reinstatement to take place, the new representation in the

CA3 has to be translated back into the original EC pattern. It has been suggested that this is performed by invertible associations in the CA1 subfield, which allow to both reproduce EC representations to associate with corresponding CA3 representations, and to reinstate them during recall (O'Reilly & McClelland, 1994). Consistently, studies have found that activity patterns in the CA3/DG subfieldss differentiate between similar objects while the CA1 show increased pattern similarity for such objects (Dimsdale-Zucker, Ritchey, Ekstrom, Yonelinas, & Ranganath, 2018; Lacy et al., 2011; Stokes et al., 2013).

One of the main indicators of pattern completion in spatial memory is remapping. Remapping refers to a switching to a different neural representation by activating a different population of cells or changing their firing rates (Colgin, Moser, & Moser, 2008), which allows for simultaneously storing in memory multiple representations which can be recalled when needed (also see section 1.3.2.1). When encountering a previously learnt environment, the currently held representation of space remaps to the previously stored pattern of activity in the CA3 (Wills, Lever, Cacucci, Burgess, & Keefe, 2005). The attractor properties can explain the all-or-nothing response of the hippocampus when exposed to incrementally changing environments; sufficiently dissimilar input leads to remapping, which results in output consistent with one environment or the other, but not a combination of the two (Wills et al., 2005). All cells in the CA3 remap at the same time (Fyhn, Hafting, Treves, Moser, & Moser, 2007; Zugaro, Arleo, Berthoz, & Wiener, 2003), again indicating its role in pattern completion via recurrent connections. Once in the attractor state associated with the recalled environment, the recurrent connections can reinstate the entire internal representation, including the spatial relationships between environmental cues such as boundaries or landmarks as well as any information about the objects encountered in them.

Remapping has been observed in humans in an fMRI study in which participants learnt maps of two distinct environments (Steemers et al., 2016). At retrieval, they were additionally exposed to four intermediate 'morphed' environments. In response to linear alteration of the environment, the activity pattern in the hippocampus showed an abrupt change in representation which was related to the mnemonic decision about the current environment, supporting the presence of attractor dynamics in the human hippocampus and their involvement in rapid remapping.

In a similar way, this network can support storage and retrieval of episodic memories. A continuous attractor, as described above, can store both continuous and discrete representations, making it suitable for representing associations between continuous variables such as space or time with discrete variables such as object identity (Rolls, Stringer, & Trappenberg, 2002). The associative loop of the CA3 is therefore particularly suited for binding multimodal elements into coherent episodes and reinstatement of all elements of an event from any of its parts. Evidence for event reinstatement in the hippocampus has been shown in studies using fMRI (Bosch, Jehee, Fernández, & Doeller, 2014; Horner et al., 2015; Mack & Preston, 2016). There is also evidence that in the human hippocampus, representations of different narratives become progressively more distinct through learning, which is akin to place cell remapping in different environments (Milivojevic et al., 2016).

1.2.3 Complementary roles of pattern separation and completion

The properties of the autoassociative network in the CA3 require that for novel pattern of activity to emerge, the input to the network must be at least as strong as the sum of all the recurrent inputs (Treves, Tashiro, Witter, & Moser, 2008). This is possible through the powerful 'detonator' connections from the DG, as described above (Rebola et al., 2017). During retrieval, on the other hand, the recurrent connections should dominate as this allows for the reinstatement of all associated elements. The two opposing needs – to overcome the recurrent inputs at encoding and to maintain them at retrieval – necessitates that the two processes, encoding and retrieval, are separated in time and potentially also anatomically. It has been proposed that the indirect path through the DG may achieve just that (Treves et al., 2008).

When the differences in the input are very similar, their output representations are also very similar as a result of the strong 'pull' of the attractor state in the CA3. As the inputs become less similar, at some point the two representations fall into two separate and mutually inhibitory attractor basins. The DG is sensitive to even small differences in input, producing difference in output that is bigger than the difference in input, whereas CA3 shows correlated activity in similar environments (consistently with pattern completion) (Neunuebel & Knierim, 2014). Sufficiently distinct input from the EC, on the other hand, results in orthogonalized (independent) representations in the CA3, suggesting that for large enough input there may not be a need for active separations of representations (Leutgeb, Leutgeb, Moser, & Moser, 2007; Vazdarjanova & Guzowski, 2004).

Consistently with this view, the already mentioned patient B.L. with selective damage to the DG showed impaired behavioural discrimination, suggesting the role of the DG in pattern separation of similar inputs (Baker et al., 2016). Additionally, B.L. had a strong bias toward pattern completion on a Memory Image Completion task where an image needs to be recognised as old or new based on partial (degraded) input. When a non-degraded novel image was presented, B.L. performed like controls, while stimuli at even lowest levels of degradation were categorised as 'old' at much higher rates than in control participants. This tendency towards pattern completion is consistent with the preserved EC-CA3 path and its role in the reinstatement of complete memory from partial input.

Moreover, the hippocampus has been found to bind relational information within a context, consistently with pattern completion, whereas it decorrelated representations of contexts that are associated with a different object-reward mapping, suggesting pattern separation of representations that are meaningfully distinct (McKenzie et al., 2015).

The EC-DG-CA3 network therefore supports memory by encoding uncorrelated (pattern-separated) inputs from the indirect path through the DG, and pattern completing direct inputs from the EC at recall. The novelty detection in the DG prevents the system from assuming that the new memories are simply 'bad' copies of the already stored ones and so prevents interference, while the recurrent network of the CA3 allows for even a partial cue to reinstate the entire pattern of associated information.

1.3 A spatial framework for episodic memory

Although spatial and episodic memories seem to substantially differ in kind, both rely on the formation of an abstract representation of the external world and so may be supported by the same mechanisms (Bird & Burgess, 2008; Eichenbaum, Dudchenko, Wood, Shapiro, & Tanila, 1999; Eichenbaum & Lipton, 2008). It has been suggested that a map-like structure may act as a universal code for creating memories, and the term 'event map' has been used to describe the representation of the interrelationships between the elements of an event (Deuker et al., 2016). A cognitive map of space may simply be a special example of a relational map that characterises episodic memories.

The map-like property is less obvious in episodic than in spatial memory. However, placing the experience in the spatial (or spatiotemporal) context supports the holistic way in which memory episodes are stored and their individual elements bound together. Additionally, access to events experienced in the past does require a type of mental time travel and reinstatement of time, place and other information associated with those events (Tulving, 2002). Recall of specific episodes and their relationships to space and time may require a similar kind of mental navigation as, for example, planning a route to a reward.

Both recall of episodic and spatial memories involves creating a mental image of the environment or events encountered in the past, and the ability to replay through them in a structured way. This reconstructive process may also underlie our ability to imagine novel scenes. According to the BBB model of spatial memory, the mechanisms used in formation of spatial maps of the environment may support the episodic recall and imagination (Byrne et al., 2007). During navigation, this system creates a representation of the environment and its features, which later on can be recalled and flexibly manipulated to, for example, plan a novel route or imagine a new viewpoint. The same may apply to episodic memory; events reconstructed from memory or created anew can be dynamically updated by simulating changes in space and time via the same mechanisms that support encoding and retrieval. This would allow for the dynamic, flexible nature of episodic imagination and thinking about the future. The spatial memory system can additionally provide the spatial context that binds the event elements together.

1.3.1 How to create an internal representation of space?

To understand how the spatial memory system can support both memory and imagination, it is first necessary to describe the features that enable this flexibility in storing and constructing the internal representations of space and events. The next section will outline how the specific characteristics of the hippocampal-entorhinal circuit support spatial navigation and formation of cognitive maps or representations of spatial environments (Tolman, 1948). I will also provide a brief overview of the different spatial cells as their specific properties play an important role in the mechanisms underlying both spatial and episodic memory as well as imagination (Bird & Burgess, 2008; Byrne et al., 2007; Eichenbaum & Lipton, 2008).

The precise mechanisms of the spatial system are still a subject of debate; however, over the last few decades a number of different functionally specialised cells have been discovered whose attributes help to understand how the brain constructs the maps of the environment. These are the place cells which fire at particular locations in space, grid cells which fire in a hexagonal lattice spanning the entire floor of the environment, head direction cells which code for the direction the animal is facing, as well as boundary vector cells (BVC) which fire in proximity to navigational boundaries (Hartley, Burgess, Lever, Cacucci, & O'Keefe, 2000; O' Keefe & Burgess, 1996), which will be described in more detail below.

In order to build an internal map of the environment, the egocentric information has to be translated into allocentric coordinates that can then be associated with, for example, a reward or an important object. One evidence for the use of such map is the ability to take shortcuts and plan detours (Tolman, 1948). This is achieved by combining information from several functionally specialized parts of the memory system. The internal coordinate system seems to be largely dependent on the environmental landmarks. Regions particularly involved in the processing of landmarks are the parahippocampal place area (PPA), the retrosplenial cortex (RSC) and the occipital place area (OPA) (Epstein & Kanwisher, 1998; Nasr et al., 2011). The RSC, in particular, seems to be used to orient oneself within the environment (Epstein, 2008; Epstein, Parker, & Feiler, 2007). Boundaries are a special type of landmarks that have a large influence on the orientation of the cognitive map (Bird, Capponi, King, Doeller, & Burgess, 2010; Byrne et al., 2007). Once oriented, the animal can use the information about its heading direction and speed to infer where it is in the environment (Hasselmo, 2009).

1.3.2 Place cells

The precise location of the animal is provided by the place cells in the hippocampus (Neunuebel & Knierim, 2012). Place cells show an increased firing rate at a particular location in a given environment - the cell's place field (O'Keefe & Dostrovsky, 1971), providing the most basic information of the current location in space. Different cells fire at different locations, so that the whole environment is covered by the place fields, and only a small subset of the cells fires strongly at any particular location.

This allows for building of an internal map of the environment, in a form of a Euclidean coordinate system, where different place cells can be associated with objects or other information present at each location.

1.3.1 Remapping

As already briefly introduced, the process of switching to a different representation of space is called ‘remapping’ (Leutgeb, Leutgeb, Treves, Moser, & Moser, 2004; Muller & Kubie, 1987). The firing pattern of place cells tends to stay stable in the same environment across exposures, and it changes completely in a different environment. Different populations of place cells may form representations of different environments (global remapping), or the same cells may be active but at different firing rates (rate remapping) (Colgin et al., 2008). Remapping therefore enables the animal to hold multiple maps of different environments in memory and to recall the correct spatial representation when needed.

Performance on spatial tasks depends on the accurate reinstatement of the place fields present at learning (O’Keefe & Speakman, 1987). In humans, fMRI studies suggest that different neural patterns are associated with different environments, and reinstatement of a pattern associated with an incorrect environment leads to errors (Kyle, Stokes, et al., 2015; Stokes et al., 2013).

1.3.3 Grid cells

Another class of spatial cells that seem to play a crucial role in spatial behaviour are the grid cells. The grid cells are unique in that their receptive fields are located regularly on the vertices of a hexagonal lattice which spans the entire environment (Hafting, Fyhn, Molden, Moser, & Moser, 2005)(Figure 1.2). Originally observed in the medial entorhinal cortex (MEC), they have also been found in pre- and parasubiculum (Boccara et al., 2010).

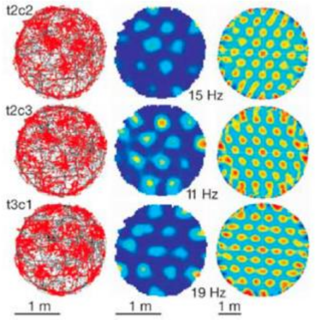


Figure 1.2. Firing fields of three grid cells recorder in the EC of a rat moving in a circular enclosure. Left: trajectory of the rat with superimposed spike location. Middle: rate maps of the cells, where blue is minimum and red is maximum. Right: spatial autocorrelation for each rate map. Reprinted from Hafting et al. (2005).

Although grid cells, like place or head-direction cells, are anchored to distal landmarks, the regularity of their firing in terms of spacing and field sizes is context-independent and consistently emerges in every new environment (Hafting et al., 2005).

Grid cells have been reported in primates (Killian, Jutras, & Buffalo, 2012b), bats (Yartsev, Witter, & Ulanovsky, 2011) and rodents (Fyhn, Hafting, Witter, Moser, & Moser, 2008). Recordings from neurosurgical patients during navigation in virtual reality environment confirmed the presence of place cells in the hippocampus and the grid cell activity in the human EC (Ekstrom et al., 2003; Jacobs et al., 2013; Miller, Fried, Suthana, & Jacobs, 2015), suggesting that the spatial navigation may be largely preserved across species. Imaging evidence also suggests presence of grid-like signal in the human EC during navigation (Doeller, Barry, & Burgess, 2010) as well as during visual search, with the grid spanning the visual space (Julian, Keinath, Frazzetta, & Epstein, 2018; Nau, Navarro Schröder, Bellmund, & Doeller, 2018). This suggests that the same mechanism may support updating of the allocentric representations of both the body and the gaze position based on self-motion and eye-motion signal.

The grid cells are arranged in subsets which share the same (or very similar) scale and orientation, while the cells within each subset vary in their phase or offset (Barry, Hayman, Burgess, & Jeffery, 2007; Stensola et al., 2012; Towse, Barry, Bush, & Burgess, 2014). Using the combination of information from multiple scales, it is possible to estimate the precise location of the animal in an abstract space of any spatial resolution. It is also this property that allows the detection of grid-like signal in humans with fMRI (Doeller, Barry, & Burgess, 2012; also see Chapter 3).

1.3.4 Head-direction and boundary vector cells

Additionally, the head-direction (HD) cells provide the allocentric representation of heading orientation (Taube, Muller, & Ranck, 1990). HD cells are found in the presubiculum (Taube et al., 1990), subiculum (Olson, Tongprasearth, & Nitz, 2016) and the EC (Sargolini et al., 2006) as well as outside of the hippocampal formation (e.g. thalamic nucleus and retrosplenial cortex; (Jacob et al., 2016)). Different cells are tuned to different allocentric directions. Finally, the boundary vector cells (BVCs) fire at a specific distance in relation to a barrier (e.g. a wall or an edge) in a consistent allocentric direction, regardless of the shape of the environment (Lever et al., 2002; O'Keefe & Burgess, 1996). BVCs were found in the subiculum (Barry et al., 2006), MEC (Solstad et al., 2008) and presubiculum and parasubiculum (Boccara et al., 2010) – but not in the hippocampus.

1.3.5 Internal representations and imagination

The already briefly mentioned BBB model proposes that the same mechanism involved in spatial navigation, as described above, are used to retrieve maps of the environment from memory and to flexibly manipulate them (Byrne et al., 2007). During recall or imagination, the hippocampus reactivates the relevant representations from across the cortex, re-creating the learnt cognitive map using the information about boundaries from the parahippocampal cortex/dorsal stream (Epstein & Kanwisher, 1998) and information about objects and their features from the perirhinal cortex/ventral stream (Buckley & Gaffan, 1997). The egocentric representation is created in the precuneous; it can be based on perception or reconstruction from memory (Byrne et al., 2007; Lambrey, Doeller, Berthoz, & Burgess, 2012). In order to imagine how a remembered scene looks from a certain viewpoint, the allocentric representation from the parahippocampal cortex is combined with information from the head direction cells and translated into an egocentric view in the parietal lobes and retrosplenial cortex (Lambrey et al., 2012). To 'move' within the imagined scene in order to, for example, imagine it from a different perspective, actual movement is simulated in the prefrontal cortex and used to update the scene. Consistently, the patterns of activity in the human entorhinal/subicular region are the same when thinking about a goal direction and facing the same direction, which suggests a shared mechanism for simulation of and actual navigation (Chadwick, Jolly, Amos, Hassabis, & Spiers, 2015). Additionally, the same mechanisms for navigation and formation of spatial representation of the environment are found in humans when navigating in VR (Doeller et al., 2010; Wolbers, Wiener, Mallot, & Buchel, 2007) as well as in rats in VR (Harvey, Collman, Dombeck, & Tank, 2009), suggesting that no actual movement in space is needed in order to engage the same processes.

The main message here is that the spatial representation can be created and manipulated by both real or imagined perception of scene and updated by a real or 'mock' idiothetic signal. The same system is therefore capable of supporting real-time spatial navigation, the recall of this information, and flexible manipulation of this information to plan a route or imagine a novel viewpoint. Additionally, this may support the ability to recall episodic memories and construct novel episodic scenes.

1.3.6 Recall and imagination in episodic memory

Neuropsychological evidence suggests that the ability to use our internal representations to reconstruct and manipulate a recalled scene may be involved in the simulation of nonexperienced events. Patient N.N. described at the beginning of this chapter experienced the same difficulty when asked to either think about his past or about hypothetical future (Tulving, 1985), which may suggest that his ability to construct the type of representation needed for both recall and simulation of events was impaired. Without being able to create a mental image, it is impossible to recall or simulate episodes.

The relationship between encoding, recall and construction of novel representations in episodic memory may be akin to the spatial memory account as described in the BBB model of spatial memory (Byrne et al., 2007) in that the same mechanisms may be utilized during encoding and during reconstruction from memory or from imagination. The constructive episodic simulation hypothesis (Schacter & Addis, 2007, 2009) proposes that both episodic recall and future thinking rely on the same constructive process, where different elements are combined to form a coherent narrative. This theory treats memory as a simulation rather than direct replay and states that the ability to flexibly manipulate and recombine information to simulate possible futures is the crucial function of the memory system. Similarly, the concept of self-projection proposed by Buckner and Carroll (2007) refers to projections of oneself to another time, place or perspective, and this can refer to both past, future or hypothetical scenarios.

Similar activity during remembering and imagining has been observed in the MTL and other areas including frontal lobes, cingulate cortex, retrosplenial cortex and parietal areas (Addis, Wong, & Schacter, 2007; Szpunar, Watson, & McDermott, 2007), which overlap with the default network (Raichle et al., 2001). It has been speculated that the activity in the default network (defined as increase in activity during passive rest as compared to task) may in fact reflect engagement in some form of mental simulation while the participants 'rest' (Schacter et al., 2012). Overall, there is high consensus that similar processes are engaged in many forms of mental simulations of episodic events, whether they have been experienced or not (Addis et al., 2007; Bar, 2007; Hassabis & Maguire, 2007; Schacter, Addis, & Buckner, 2007).

1.4 Organisation of the thesis

A major theme within this thesis is to examine the patterns of activity supporting spatial and episodic memory using fMRI. I will investigate whether some of the mechanisms outlined above can be observed in the human brain by looking at how different aspects of memory are represented in the hippocampus and other memory structures using a Representational Similarity Analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008).

In Chapter 2, I provide a technical introduction to the RSA and explain how looking at the patterns of activity can inform our understanding of the structure of memory.

The four experimental chapters represent two sub-themes. In the first part (Chapter 3), I will focus on the spatial memory. Using a VR setup, I will look at whether we can use patterns of activity in the fMRI signal to detect representations of space during both navigation and imagination. In particular, I will investigate whether the same mechanisms are used in both. In the second part (Chapters 4-6), I will investigate mechanisms underlying episodic memory, specifically the process of pattern separation of similar memories (see Section 1.2.1). I will first present two tasks I designed to provide a behavioural measure of pattern separation. The first task (Chapter 4) failed to show any evidence for the investigated mechanism; I will discuss potential explanations of the lack of predicted results. The second behavioural study (Chapter 5) provides evidence that similar memories are represented as more distinct, which is consistent with the pattern separation account. Finally, in Chapter 6 I will use this paradigm to look for evidence of pattern separation in the brain using fMRI.

Chapter 2

Methods: Representational Similarity Analysis

Recent years have seen a move away from the univariate approaches to imaging data, which look at the differences in the magnitude of signal in response to experimental manipulation, to methods that look at changes in patterns across voxels in order to make inferences about the underlying computations. These multivariate methods not only tell us which brain areas are involved in the task, but also tell us something about how task information is reflected in spatial variations of activity.

The patterns of activity across populations of neurons can be used to infer things about the type of cognitive representations they support. This functional interpretation of a neural activity pattern as a 'representation' implies that a given feature is not only encoded in the pattern but also, importantly, the way it is represented serves a functional purpose (Diedrichsen & Kriegeskorte, 2016). For example, we might expect that behaviourally important differences in two cognitive representations will be reflected in greater differences between their corresponding patterns of activity across neural populations – so that these differences stand out strongly from differences due to stochastic variation or 'noise'. In other words, differences in population patterns are interpretable in terms of their effects on cognitive processes and behaviour. Methods looking at multivoxel patterns of activity may be seen as intermediate step between single neurons recordings and standard univariate methods.

These methods are therefore particularly suitable for looking at more fine-grained hypotheses about the way in which memory is represented in the brain, providing a common ground for the observed behaviour, the measurements of brain activity, and the computational models of cognition. In the context of this thesis, the use of multivariate analyses will provide a novel way to extend and enrich the findings obtained using different methods as well as help to answer more nuanced questions about the characteristics of memory representations encoded in different brain areas.

This thesis will extensively use one of the multivariate methods: the Representational Similarity Analysis (RSA; Kriegeskorte et al. 2008). In the next sub-sections, I will outline the main features of the RSA, followed by the details of the processing stages involved in this analysis.

2.1 Representational geometry

The RSA looks at the relationships (similarities) across different patterns, allowing for making conclusions about the structure of neural representation of a given feature. In other words, RSA can tell us not only which feature is represented in an area, but also which aspects of this feature the neuronal population is sensitive to.

This RSA is an extension of another popular multivariate approach – pattern classifier analysis. Classifier analysis can investigate the discriminability of pre-defined features. For example, fear conditioning (associating certain items with an electric shock) has been found to lead to a reorganization of similarity pattern in the frontal areas; the representation of original categories (faces/houses) shifted to a representation of affective (threat/ no threat) categories, while the visual areas remained sensitive to the original representational structure (Visser, Scholte, Beemsterboer, & Kindt, 2013; Visser, Scholte, & Kindt, 2011). Pattern classifiers have also been successfully used in memory research to show that category-specific activity patterns are separable and can be used to predict the free-recalled category by a few seconds (Polyn, Natu, Cohen, & Norman, 2005).

However, pattern classification only tells us whether there are reliable differences between representations but doesn’t provide further insight into the relative sizes of those differences, i.e. the geometrical structure of those differences. The RSA can be used to look at these precise representational geometries. For instance, in the visual domain it was found that although the activity patterns in V1 most accurately decode different colour categories, the perceptual colour space is best represented geometrically in V4 (Brouwer & Heeger, 2009). The RSA can therefore answer different and often more parametric research questions beyond showing that the patterns of activity produced by two experimental conditions can be distinguished from each other in a given region.

2.2 Patterns of activity

The vector of activity measurements across all channels (in the case of fMRI data, across voxels) is called the activity pattern. Each condition (e.g. each stimulus) has its associated activity pattern. These spatial patterns are the fMRI representations of the mental states associated with a given condition/stimulus in a given area.

In univariate analyses, the average activity in a given region across participants is compared to zero or another average activity measure. The RSA, on the other hand, treats the activity patterns as a random variable with a distribution across voxels. Instead of a mean value, it uses a measure of the distribution of the patterns of activity, thus combining information across space and time. Therefore, unlike the traditional univariate approach, it can account for the inter-subject variability in the pattern of activity; the precise signal (in terms of its magnitude and spatial distribution across voxels) does not need to be preserved across different people, as long as the relationships (i.e. relative differences) between activity patterns associated with different features are maintained.

2.3 Representational Dissimilarity Matrices

The distances between neural activity patterns can be expressed in terms of a correlation distance (one minus the correlation value), a Euclidean distance, or a Mahalanobis distance (spatially-prewhitened Euclidean distance). The Euclidean distance corresponds to distances in the physical world and uses Pythagoras' theorem (i.e. distance is the square root of the sums of squared differences between the coordinates) extended to a P-dimensional space (where P is equal to the number of voxels).

These results can be arranged in a matrix called the Representational Dissimilarity Matrix (RDM), which summarises the distances between the patterns for all conditions. The off-diagonals of these matrices are the dissimilarities (or distances) between each condition (each activity pattern) and all other conditions.

RSA is primarily used to compare these dissimilarity matrices (i.e. the 'neural' RDMs) to model RDMs. The model RDMs can be either theoretically derived matrices with hypothesized similarity structure (e.g. distances based on semantic similarity of stimuli), can be based on brain RDMs from another area/subject/species, on behavioural measures (e.g. accuracy, frequency), or on other computational models. This allows for great flexibility as the precise measurements used to construct the matrices do not need to be the same and instead any descriptions of similarities between patterns that can be translated into a dissimilarity matrix can be used.

2.4 RSA analysis

In this thesis, the data was analyzed using custom Matlab scripts, which were a part of or adapted from the freely available RSA toolbox (Nili et al., 2014). The first- and second-level estimates were obtained with SPM8 or SPM12 (www.fil.ion.ucl.ac.uk/spm). The following sub-sections will outline the stages involved in the analysis. This section is aimed at giving a general overview of the processing steps, while the precise details of the analyses used in the studies presented in this thesis will be found in the relevant chapters.

2.4.1 First-level GLM

The first-level regressors are obtained using the ‘standard’ SPM processing steps and include the conditions of interest, head motion parameters, cardiac and respiratory regressor (if available), and any other parameter which may explain some of the variance in the data. The regressors are modelled with a boxcar function and convolved with a hemodynamic response function in SPM. Importantly, the first-level RSA uses unsmoothed data as the main analysis is performed separately for each subject. Otherwise, the pre-processing stages remain the same as in univariate analyses.

The conditions of interest are sets of trials whose activity patterns are compared to each other to estimate the brain RDMs, for example corresponding to different stimuli, features or different intensities of stimuli.

2.4.2 Estimating activity patterns

The regression coefficients, or beta-weights, from the first-level GLM are used as activity estimates for each condition. A vector with beta values over voxels is the activity pattern for a given condition (Figure 2.1).

The activity pattern vectors can be based on predefined regions of interest (ROIs; with one RDM describing the relationship between activity patterns per ROI) or on searchlights of a specified size (with one RDM per searchlight). In searchlight analyses, the activity patterns are calculated for all search spheres for all conditions.

The choice between ROI and searchlight-based analysis is not straightforward. On one side, the choice of ROIs is usually informed by previous research, making it more likely that a given representation can be detected. Additionally, unlike a searchlight analysis, it does not require a correction for multiple comparisons. Although only a small portion of the ROI may represent a given feature, if there is a lot of spatial variability between subjects, the ROI analysis may still be able to detect it on the group level. Similarly, a larger-than-searchlight ROI may be more effective when weak information is distributed across the whole ROI. On the other hand, however, any uninformative voxels lead to poorer discrimination (Diedrichsen, Ridgway, Friston, & Wiestler, 2011). If a given feature is represented in only a subset of voxels within an ROI, a smaller searchlight may be more able to detect it (unless spatial variability is high, as explained above). It is also not always possible to have a priori knowledge of the location of expected effects. Searchlight can also continuously map the presence of a given feature across the brain volume, providing additional and more fine-grained information to a single-value result from an ROI analysis (Kriegeskorte, Goebel, & Bandettini, 2006). In this thesis, I will use a mixture of searchlight analyses and ROI analyses wherever there is a strong hypothesis for a location of a specific representation. The searchlight sizes were in line with other studies looking at memory representations (Chadwick, Hassabis, Weiskopf, & Maguire, 2010; Dimsdale-Zucker & Ranganath, 2019; Libby, Hannula, & Ranganath, 2014; Stokes et al., 2013). However, as optimal size of a searchlight depends on the size and shape of the region that represents the feature, a number of different sphere sizes were tested; the reported results are for the searchlight size which provided most interpretable results.

2.4.3 Calculating distances

These activity patterns are then compared to the activity pattern of each of the other k conditions using a chosen distance measure (correlation, Euclidean or Mahalanobis), creating a k-by-k matrix of distances between all pairs of conditions (k = number of conditions). Two identical patterns would have a distance of zero (Figure 2.1).

The steps below describe how to estimate distances among all pairs of conditions for each search sphere or ROI.

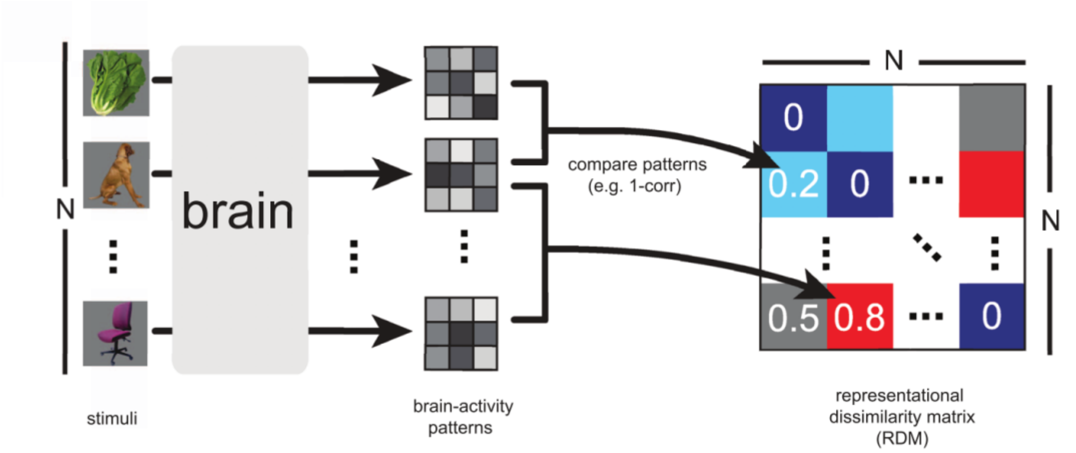


Figure 2.1. Computing the representational dissimilarity matrix. Brain activity related to each condition of the experiment (e.g. different stimuli) is recorded. For each region of interest or searchlight, we estimate a pattern of activity across voxels. Then, we calculate a dissimilarity measure (in this example, a correlation distance) between each pair of conditions. The dissimilarities are then arranged in a k-by-k representational dissimilarity matrix (RDM; k = number of conditions). These matrices can then be compared to models, between different brain regions or different individuals or species. Reprinted from Nili et al. (2014).

2.4.4 Multivariate Noise Normalization

The noise in fMRI data is never randomly distributed across voxels but rather tends to show some degree of spatial structure (Diedrichsen & Shadmehr, 2005). There are multiple reasons for this: spatially-smooth noise processes, anatomical proximity to the vascular system, grey/white matter composition, and as a result of motion correction. This is problematic as the distance measures weigh each voxel equally, independently of its variance. Ideally, the data should be spatially prewhitened by weighting each voxel based on the spatial structure of the noise, as estimated by the residuals from the first-level GLM. Unlike univariate noise normalization, which scales each voxel by its variance (the diagonals of the error covariance matrix), the multivariate method used in the RSA also controls for noise covariance (the off-diagonals), thus accounting for the noise structure. In other words, noisy voxels or highly correlated clusters of voxels are down-weighted. This is typically done on local regions of interest rather than the whole brain.

In order to perform this step, the raw time-series data is used; i.e. the spatial structure of noise is calculated on the first-level estimates and then used to 'adjust' the weighting of each voxel in the raw data, before calculating new, prewhitened first-level estimates.

2.4.5 Crossvalidation

Crossvalidation is an optional step aimed at bringing the estimated distances closer to the true distances between representations. In crossvalidation, the data set is divided into a number of ‘partitions’ (usually each partition represents an imaging run, as estimations of noise across runs is assumed to be independent), and the activity patterns for all conditions are calculated separately for each partition. Although ‘crossvalidation’ in a pattern-recognition context involves leaving out a partition to check the unbiased performance of a classifier trained on the other partitions, in this context it is used to avoid the positive bias caused by noise in squared distance measures like Mahalanobis distance. Without controlling for this, the presence of noise would make the Mahalanobis distance between two identical patterns (i.e. true distance equal to zero) higher than zero (if the data points have some spread then mean squared distances been them will be positive; Walther et al., 2015). ‘Crossvalidation’ here implies multiplying the distance in one partition by the corresponding distance in another partition (rather than with itself). Thus, when the distances are calculated between all conditions for all possible pairs of partitions, and then averaged, they are closer to the true distances between pairs of patterns assuming noise across partitions is independent. Crossvalidated Mahalanobis distance (or crossnobis) has been shown to have higher reliability than other dissimilarity measures (Walther et al., 2015). As a result of crossvalidation, some distances, especially the small ones, can become negative. This is because the true distance can be both over- and underestimated.

However, crossvalidation requires that at least two partitions (independent estimates) exist for each condition. This is not always possible as some conditions may only appear in the study once. This is especially problematic in memory studies where sometimes a repeated presentation of the same stimulus may not be possible (when, for example, looking at single-trial learning). It may also be inappropriate if there is any reason that the representation has changed from one imaging run to the next, for example as a result of learning. In the analyses in this thesis, the crossvalidation was therefore used only when possible and appropriate. The precise steps of the RSA pipeline used in any given analysis (pattern estimation, spatial noise normalization, crossvalidation, etc.) will be provided in their relevant methods sections.

2.5 Comparison of brain RDM to model RDMs

The estimates of distances between all pairs of activity patterns (i.e. the brain RDMs) are calculated for each search sphere or ROI, for each participant. In order to see whether two conditions have the same pattern of activity, it would be enough to compare their distance against zero. Alternatively, it is also possible to look at the overall representational structure across all conditions by comparing the entire brain RDM to pre-specified representational models, i.e. the model RDMs. Model RDMs may be based on computational models, behavioural data or brain RDMs from different studies, conditions, subjects or brain regions. Conditions are expected to be similar if they belong to the same category or share a feature.

Comparison of RDMs to the models can be achieved by calculating correlations between them. The use of correlation coefficients accounts for subject-specific scaling of the dissimilarities that results from individual variability due to physiological responsiveness, head movements, etc. (Diedrichsen & Kriegeskorte, 2016).

2.5.1 Group-level analysis

The above steps result in a searchlight map of measures of relatedness between the neural and model RDMs for every subject, or a single correlation value for a given ROI for every subject. These correlation values are then spatially smoothed (in. searchlight analysis) and used in a group-level analysis where a one-sample t-test is performed to compare the average correlations across subjects against zero. This comparison tests whether the model RDM provides a good fit to the representational similarity observed in the data (brain RDM).

Chapter 3

fMRI study of grid-like representation of the spatial environment

3.1 Introduction

An internal map of the environment allows for flexible navigation without the constraints of using a limited number of learnt sequences. The hippocampal-entorhinal circuit of spatial cells is thought to support this type of memory representation. By combining information about the environmental boundaries from boundary vector/border cells (Barry et al., 2006; Lever, Burton, Jeewajee, O'Keefe, & Burgess, 2009; Solstad et al., 2008), the current position from place cells (O'Keefe & Dostrovsky, 1971), the direction the animal is facing from the head direction cells (Taube et al., 1990), the changes in position from the speed cells (Moser et al., 2014) as well as other complementary information from spatial and non-spatial cells, the EC plays a central role in estimating and continuously updating the representation of the current location (Hasselmo, 2009). In particular, grid cells are thought to play an important role in path integration: using self-motion information to update one's location (Burgess, Barry, & O'Keefe, 2007; Hasselmo, 2009; McNaughton et al., 2006). The same mechanisms may also be used for planning and imagination of routes (Byrne et al., 2007).

As described in Chapter 1, grid cells have a unique firing pattern, exhibiting multiple receptive fields organised in a hexagonal pattern which consistently emerges in all new environments (Hafting et al., 2005) (Figure 1.2). Cells with grid-like properties have been observed in the EC in recordings from neurosurgical patients during VR navigation (Jacobs et al., 2013; Miller et al., 2015). The grid cell-like signal in the human EC was also detected using functional magnetic resonance imaging (fMRI) (Doeller et al., 2010; Horner, Bisby, Zotow, Bush, & Burgess, 2016; Kunz et al., 2015). Despite the relatively poor spatial resolution of fMRI images, it is possible to provide evidence for the hexagonal firing structure based on two assumptions (Doeller et al., 2010): (i) the orientation of the grid is constant across all grid cells (Barry et al., 2007; Stensola et al., 2012; Towse et al., 2014), (ii) 'conjunctive' grid cells, whose firing is modulated by running direction and running speed (Sargolini et al., 2006), have 'preferred' firing directions aligned with one of the grid axes. Assumption (i) means that there will be different neural dynamics when running along grid axes versus between them (some cells firing a lot and others firing a little versus all cells firing at an intermediate rate). The non-linear coupling of neural activity to BOLD signal means that there may be differences in the signal generated by these different dynamics, even if the total firing rates of the grid cell population do not differ, and this difference will also depend on running speed. Assumption (ii), in addition to assumption (i), implies that there might be differences in the total firing rates of the conjunctive grid cell population according to running direction and running speed, and thus differences in the BOLD signal (independent of its precise coupling to neural firing rates). These fMRI studies (Doeller et al., 2010; Horner et al., 2016) used the differences in the BOLD signal between direction on- and off-axes of the hexagonal grid to look for evidence of 60° rotational symmetry (grid signal; Figure 3.1). Importantly, no evidence for four-, five-, seven- and eight-fold rotational symmetry was observed, suggesting a presence of specifically hexagonal spatial structure. Behaviourally, the coherence of the orientation of the grid cell signal across voxels positively correlated with memory performance (Doeller et al., 2010).

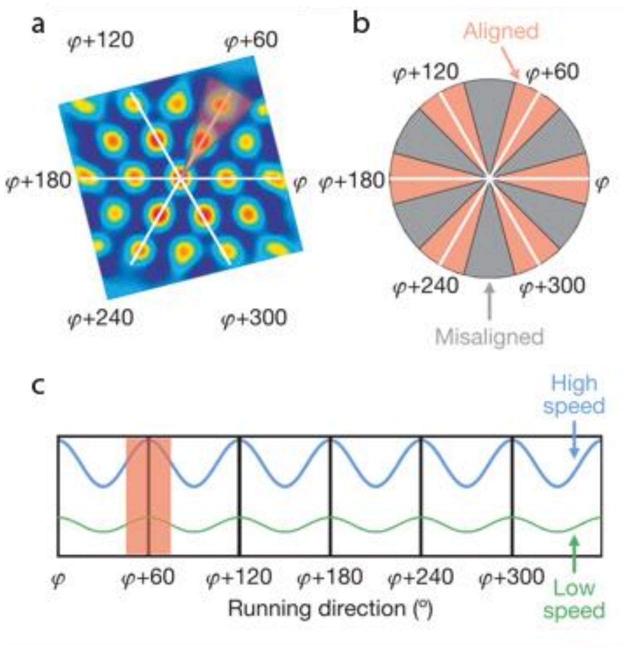


Figure 3.1. Experimental logic from Doeller et al. (2010). (A) Spatial autocorrelation of a typical grid cell with a 30° sector aligned with the grid. (B) Schematic of running directions aligned (red) and misaligned (grey) with the grid. (C) Predicted modulation of BOLD signal by running direction showing 60°periodicity, with a stronger effect for fast than slow speed. Adapted from Doeller et al. (2010).

The grid-like signal has been detected in humans in a number of different tasks. Using the same analysis method as in Doeller et al. (2010) described above, Kim and Maguire (2018) found a hexagonal signal in a 3D virtual environment spanning the entire volumetric space, showing that the grid-like representation is not limited to 2D navigation tasks that are most commonly used in this line of research. Recent work has also suggested that grid cells may be used to code for more abstract 'spaces'. A grid-like signal has been found when navigating an abstract space defined by two non-spatial dimensions (length of a bird's legs and length of its neck); the signal was stronger when the transitions in this space was aligned with the grid as compared to misaligned (Constantinescu, O'Reilly, & Behrens, 2016). Grid cells in the EC were also found to represent sound frequencies in rats (Aronov, Nevers, & Tank, 2017) and gaze direction in humans (Julian et al., 2018; Nau et al., 2018). The same mechanism may therefore have been adapted to deal with more complex situations that rely on capturing map-like relationships between stimulus dimensions.

Taken together, the above studies show strong evidence that the involvement of grid cells in spatial navigations and other non-spatial tasks can be detected using neuroimaging techniques in humans. Their role in spatial imagination, however, is still largely unknown. An efficient memory system requires that future events can be planned through simulation of potential outcomes, and the network of temporoparietal areas containing spatial cells was shown to be involved in spatial imagery based on remembered information (Burgess, Maguire, Spiers, & O'Keefe, 2001). The BBB model outlined in Chapter 1.3 proposes that the mechanisms employed during navigation and encoding of spatial information are also engaged during retrieval of this information as well as during imagination of novel routes and viewpoints by updating the internal representation with 'mock' idiothetic signals (Byrne et al., 2007). Neuropsychological and neuroimaging evidence supports the proposed overlap in the networks (including the MTL) used for remembering the past and imaging the future (Schacter et al., 2007). This suggests that the spatial memory system not only represents the current location and spatial information about the environment but is also involved in simulation of potential routes and scenarios.

In the present study, fMRI was used to look for evidence of grid cell activity resulting from both actual and imagined navigation within a VR environment. The same task design has previously been used to show evidence for grid-like signal during navigation (Doeller et al., 2010). Using representational similarity analysis (RSA; Kriegeskorte et al. 2008), the differences in neural activity patterns across angular directions and modalities (real vs. imagined) were assessed for 60° periodicity and similarity of grid orientation. A recent study found evidence for grid-like signals in the EC using RSA (Bellmund, Deuker, Schröder, &

Doeller, 2016). In a large-scale VR environment, participants imagined different directions from multiple locations. Directions that were 60° apart had more similar pattern of activity than other directions. However, this task measured the activity related to imagined facing direction from a stationary viewpoint. The current study, on the other hand, aims to look at the role of the EC in mental simulation of planned trajectory. Because of the overlap of mechanism involved during navigation and imagination, we should expect to see the same or similar representation of space in both modalities. The analyses were not limited to the MTL due to the previous findings suggesting a wider network of brain areas involved in the representation of space and running direction, including the visual cortex and medial prefrontal cortex (Doeller et al., 2010) as well as during imagination, such as frontal lobes, cingulate cortex, retrosplenial cortex and parietal areas (Addis et al., 2007; Szpunar et al., 2007),

The results reported below are a part of a larger project in which, using a univariate analysis of the same data, the presence of grid-like activity within the EC has been detected in both navigation and imagination (Horner et al., 2016). As in Doeller et al. (2010), the orientation of the grid was estimated from half of the data in each participant, and the second half was used to detect six-fold symmetry (as well as other control models). The grid-like signal was observed in the EC during movement but not stationary periods. Crucially, the 60° periodicity was also observed during periods when participants were asked to imagine navigating towards the previously memorized location. These results show that grid-like activity of the same orientation as during navigation is present during imagination and can be detected in the EC using fMRI. The results below extend these findings by using a multivariate analysis approach to look for patterns of activity consistent with the 60° rotational symmetry. As in Bellmund et al. (2016), the grid-like representation was predicted to result in higher pattern similarity for directions similarly aligned to the grid (i.e. separated by multiples of 60°) than directions where one of them lays 'on-axis' and the other one 'off-axis' (i.e. separated by 30°, 90°, 150°).

(Author contribution: analysis and interpretation of data presented here)

3.1 Methods

3.1.1 Participants

26 neurologically healthy participants (17 males, 9 females) of mean age 23.7 (*SD* = 4.2) were recruited through the UCL ICN Subject Database. All participants had normal or corrected-to-normal vision and were right handed. The study was approved by the University College London Research Ethics Committee (1825/003) and informed consent was obtained before the session.

3.1.2 Design and procedure

The VR environment was created with Unity (https://unity3d.com/) and comprised of a flat circular plane surrounded by a cliff and more distal background of mountains rendered at infinity, as well as trees and clouds, which provided orientation cues. The proximal cues were not part of the original setup (Doeller et al., 2010) and were added. here to make it easier for participants to orient in the environment as the task required that participants were highly accurate in the trajectories they took. Participants were required to learn the locations of two sets of objects, presented in one of three pseudo-randomised arrangements. Within each set, the locations of six objects were arranged with equal spacing around the centre of the environment, i.e. with angular distance of 60°. The main axes on which the objects were located were shifted by 15° in the second set, allowing for sampling of 24 angular directions (every 15°) across both sets (Figure 3.2). The environment was unchanged across blocks.

Before the scanning session, participants were familiarised with the task and the environment until they were able to understand and perform the task accurately. During scanning, participants lay in a supine position with the visual display reflected from a back projector by a mirror attached to the head coil. Participants moved in the VR using a button box with three keys with their right hand (left turn, right turn, move forward) and signalled their readiness using a single button with their left hand.

During the encoding phase, each object was shown at its respective location and participants had to ‘collect’ the item by running over it using the three-key pad. Each trial started at the end location (the object location) of the preceding trial. Each object was collected five times, each time approached using one of the five possible routes, giving 30 trials in total in each set. Each encoding phase took approximately five minutes.

In the subsequent retrieval phase, participants were cued with a centrally-presented item and asked to orient themselves in the direction where they believed the object was located, and signal with a left-hand button press once they believed they were facing the correct way (cue period, see Figure 3.2B). They were then asked to close their eyes and imagine navigating in a straight line towards this object, again pressing the left key when this action was complete (imagination period). After a variable jitter period during which they could not rotate or move (2-6 seconds; wait period), they were instructed to navigate to the location where they believed the object was placed and to signal with a left-hand press when in the correct location (navigation period). At this point the object appeared in the correct location and, if needed, participants had to navigate towards it to collect it (feedback period). The overall structure and the structure of trials in the retrieval phase are presented in Figure 3.2B. For every object, each of the five possible routes was used twice, giving 60 trials in total in each set. Each retrieval phase took approximately 20 minutes.

Participants then completed the second encoding-retrieval set with another set of items in new locations (shifted by 15°; Figure 3.2A), following the same procedure.

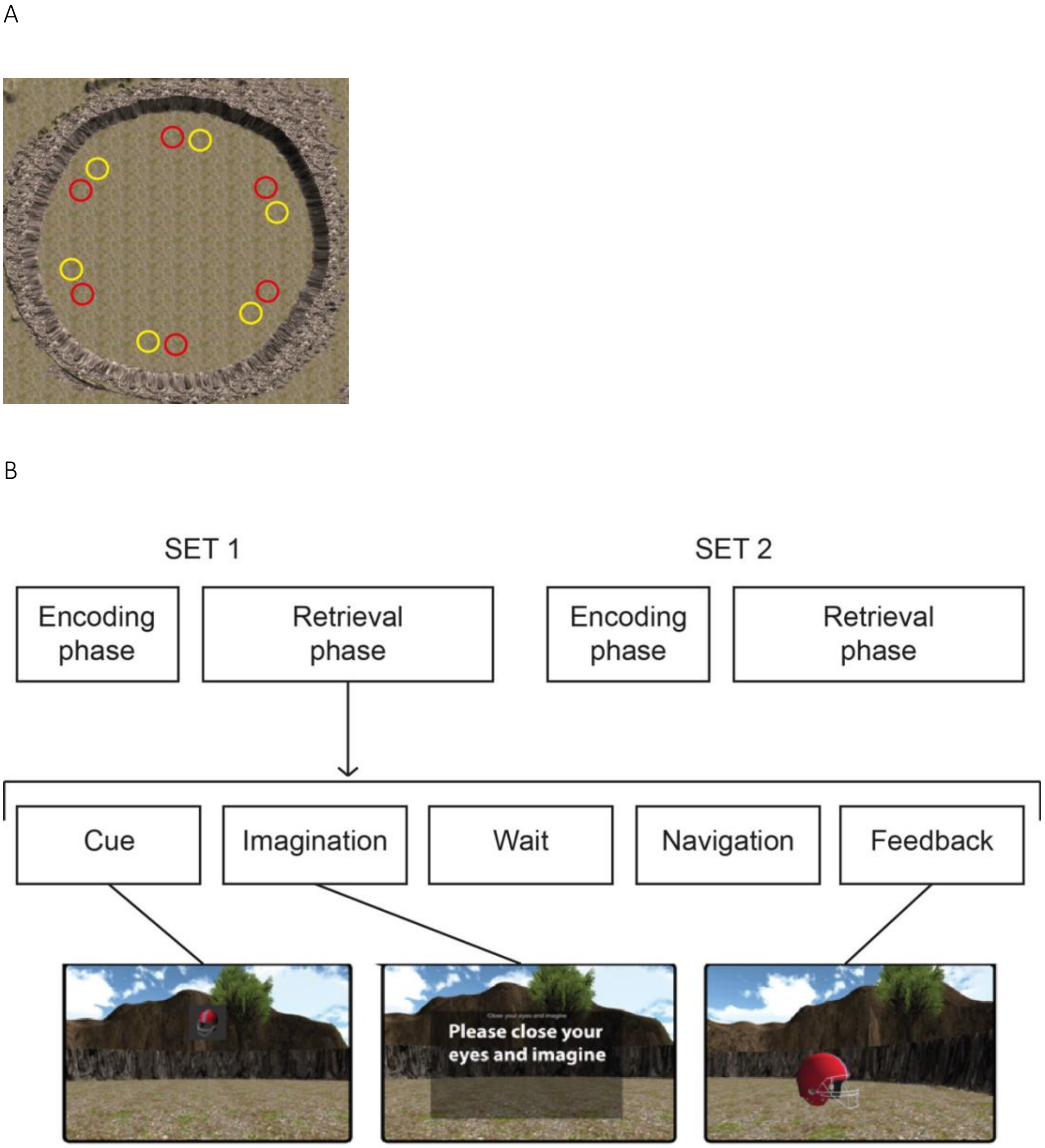


Figure 3.2. Experimental design. (A) The virtual reality environment. The arena presented from a birds-eye view, with target-object locations for each encoding-retrieval set represented as red and yellow circles. The target axes are shifted by 15 °in relation to each other. (B) Structure of the experimental session. Each trial within the retrieval phase was further divided into five periods of variable duration. Examples of screen shots for the cue, imagination and feedback periods are provided.

3.1.3 fMRI acquisition

Imaging data were acquired using echo-planar imaging (EPI) on a 3T Trio (Siemens) scanner with a 32-channel head-coil. The parameters were: repetition time (TR) = 3,360ms, echo time (TE) = 30ms, voxel size / resolution = 3x3mm, 48 slices per volume. A double-echo FLASH field-map for distortion correction of the EPI volumes was acquired, as well as a three-dimensional MDEFT structural image (1mm^3 resolution) for normalisation to the MNI template.

3.1.4 Data preprocessing

The images were bias-corrected, unwarped, realigned to the first image, slice-time corrected and normalised using SPM8 in Matlab (Mathworks; www.fil.ion.ucl.uk/spm). The first five volumes were discarded to allow for T1 equilibration.

3.1.5 Data analysis

The data was analysed using the RSA with custom Matlab scripts, some of which were a part of the RSA toolbox (Nili et al., 2014).

3.1.6 1st level GLM

The predictors for each set of analyses consisted of analysis-specific regressors (i.e. the conditions used as a basis for creating dissimilarity matrices in the RSA; see below for details), and regressors of no interest: six head motion parameters, regressors for button presses (one for right hand and one for left hand) as well as regressors coding for each period which did not include the analysis-specific regressors. These regressors were modelled with a boxcar function for the duration of movement during the navigation period, for the duration of the entire period for all periods which did not include navigation (i.e. entire cue period, imagination period; wait period, and feedback period; see Figure 3.2B for task structure), or for duration of a button press, and analysed using a General Linear Model (GLM). The regressors were temporarily convolved with the hemodynamic response function.

3.1.7 RSA

3.1.1.1 Dissimilarity structure in data

The beta estimates from the first-level analysis were used to obtain condition-specific activity patterns (for specific conditions in each analysis see section 3.2 below) in a searchlight-based RSA. A whole-brain searchlight analysis was used, using search spheres with three-voxels radius from the central voxel. As explained in detail in Chapter 2, the beta estimates of each condition were used to calculate correlation distances between all pairs of conditions, resulting in a brain RDM (dissimilarity matrix based on brain activity) for every search sphere for every participant.

3.1.1.2 Comparison of brain RDM to model RDMs

A brain RDM from each search sphere was compared to pre-specified model RDMs (i.e. matrices the size of the brain RDM with predicted distances between all pairs of conditions). The specific model RDMs used in the current analyses will be described below (see Figure 3.3 for examples). The brain RDMs were compared to their model RDM using partial correlation (partialling out other models from a given analysis; see below).

3.1.1.3 Group-level analysis

The above steps resulted in a searchlight map of measures of similarity between the brain and model RDMs for every participant. These correlation values were then smoothed with a 5mm FWHM Gaussian filter and used in the group-level analysis which compared the average correlations across subjects against zero with a one-sample t-test.

3.2 Analyses

All analyses were performed either on the two encoding blocks or the two retrieval blocks. In the analyses presented below, each model was tested against data while removing the effects of all other models in the respective analyses. Additionally, a model that predicted higher similarity within a block than across blocks to account for increased similarity within blocks was partialled out in all analyses. This was achieved through calculating a partial correlation coefficient between the model of interest and the data after removing the effects of all other models/predictors.

The purpose of the main analyses was to look for patterns consistent with the grid cell-like activity and with the representation of heading direction. In additional analyses, I looked at other representations of space, namely the representations of the current location and the goal location. For the ease of reading, however, the analyses below are split based on the specific periods within the task (navigation during encoding, navigation and imagination during retrieval, and cue period during retrieval).

3.2.1 Encoding blocks

Within each encoding block, the objects to be collected could appear at six possible goal locations (see Figure 3.2). This resulted in 30 different routes between starting and goal locations. Each condition (i.e. each regressor in the first level analysis) represented one of the 30 routes, resulting in a total of 60 conditions across both sets. The next sub-sections will describe how, on the basis of these conditions, the model RDMs were constructed to look for evidence of different representations.

3.1.1.4 Grid-like signal

I looked for a representation of a spatial structure with 60° periodicity. The locations of the objects in the two encoding blocks were offset by 15°, resulting in routes sampled every 15° across both sets. Based on the angular direction of the routes, pairs of conditions separated by the multiples of 60° (i.e. similarly aligned to the grid, e.g. 0° and 60°) were maximally similar (distance = 0), pairs differently aligned to the grid (e.g. 0° and 30°) were maximally dissimilar (distance = 1), while pairs where one of the bins was aligned with the grid and the other was in between the ‘on’ and ‘off’ grid axes (e.g. 0° and 15°) had distance equal to 0.5.

3.1.1.5 Heading direction

The same conditions based on angular heading directions in each trial as above were used as a basis for the current model. The model RDM for heading direction assumed that similarity linearly decreases as the angular difference in heading direction increases (distances = 0 – 1), reaching maximum distance at 60°. The decision to include a maximum angular distance at which dissimilarity in patterns should plateau instead of using full 180° range was motivated by the findings that the firing of head direction cells shows a tuning curve, with decreasing firing as a function of the distance from the preferred direction (e.g. Taube, 1998, 2007; Zhang, Heuer, & Britten, 2004).

3.1.1.6 Goal location

Each object (and therefore each of the six goal locations) was reached five times during each encoding block. The model RDM for goal location assumed that the activity patterns for the same goal would be maximally similar for the same location (distance = 0) and then gradually increase (distances = 0.2 – 1) as the angular distance between goal locations increases, reaching maximum distance at 60° .

3.2.2 Retrieval blocks: navigation and imagination periods

At retrieval, participants first imagined to navigate and then navigated to a location where they believed the target object was located. The continuous metric of heading direction had to be discretised in order to represent it as a dissimilarity matrix; the possible movement directions (0-360°) were divided into twelve 30° conditions per set, with each condition bin centred on one of the direction axes used during the corresponding encoding set (in other words, the condition bins were based on the 'ideal' straight-line trajectories between starting and goal locations needed to obtain the objects). Each navigation and imagination trial was classified as the condition that was the closest to the mean navigation direction during that trial. As the two sets were offset by 15°, the resulting direction-bins across both blocks were overlapping. In the analyses described below I used either only navigation or only imagination periods (resulting in 24 conditions in each).

3.1.1.7 Grid-like signal

This model was based on the same principles as the grid signal models in the encoding period. Pairs of conditions separated my multiples of 60° (similarly aligned to the grid, e.g. 0° and 60°) had distance equal to 0, pairs differently aligned to the grid (e.g. 0° and 30°) had distance equal to 1, while pairs where one of the bins was aligned with the grid and the other was in between the 'on' and 'off' grid axes (e.g. 0° and 15°) had distance equal to 0.5 (Figure 3.3A).

3.1.1.8 Heading direction

In the same way as in the encoding blocks, this model was defined based on angular differences in heading direction during either navigation or imagination periods, with increasing angular differences represented by increasingly dissimilar patterns (with distance measure ranging from 0 to 1, reaching maximum at 60° angular distance; see Figure 3.3B).

3.1.1.9 Navigation vs. imagination

I looked for areas that showed more similar patterns of activity within modality (navigation or imagination, or in other words during movement and stationary period) than across. The model RDM predicted distance of 0 for same modality and 1 for different modality.

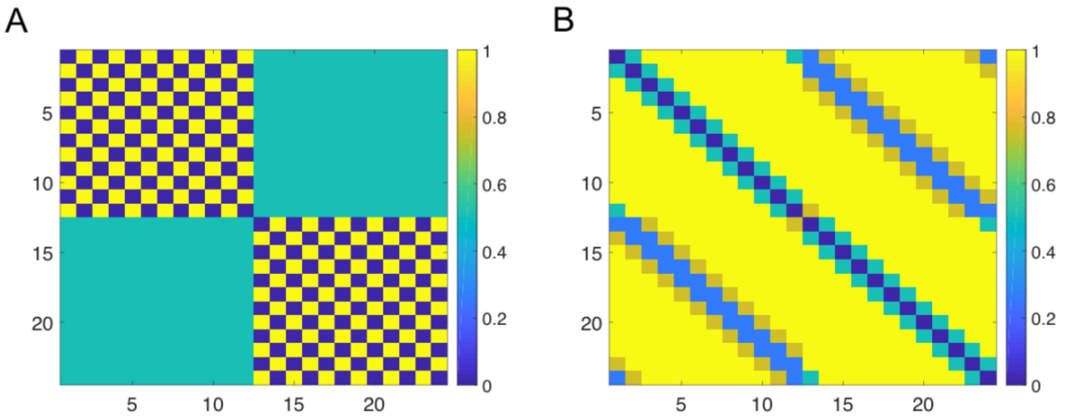


Figure 3.3. Examples of model RDMs (Representational Dissimilarity Matrices with predicted distances between all pairs of conditions). (A) Model of grid-like signal from the analysis of retrieval blocks: conditions based on heading direction. The conditions represent 24 direction bins of 15° each (12 in each set). The similarity depends on the direction of movement either off- or on-axis. (B) Model of heading directions from the analysis of retrieval blocks: conditions based on heading direction. The conditions represent 24 direction bins of 15° each (12 in each set). The similarity varies with angular difference between heading directions. The colour scale represents the level of predicted dissimilarity for any pair of conditions. For details of each model, see text.

3.2.3 Retrieval blocks: cue period

In another analysis of the retrieval blocks, I defined conditions as all possible combinations of starting locations and the goal locations (30 combinations in total). These were estimated during the cue period (i.e. when the object to be retrieved is presented on the screen and the participants are asked to turn around in the VR until they face the direction where they believe the object is).

3.1.1.10 Starting location

As in the analysis of encoding blocks, the model RDM was based on similarity of the starting location across trials, with distance measure ranging from 0 to 1, reaching maximum at 60°

angular distance (measured from the location in the centre of the arena; the angular direction similarity can therefore be thought of as simply distance-based similarity).

3.1.1.11 Goal location

Here the model RDM was based on the similarity of the goal location (i.e. the location of the object to be picked up on this trial; noting that this was not necessarily the same as the location the participants navigated towards), with same locations represented as more similar (distance = 0 – 1, maximum at 60° angular distance).

3.3 Results

3.3.1 Angular error

Participants were successful at learning the locations of the objects within the environment, with median angular error across both retrieval blocks = 8° (see Figure 3.4).

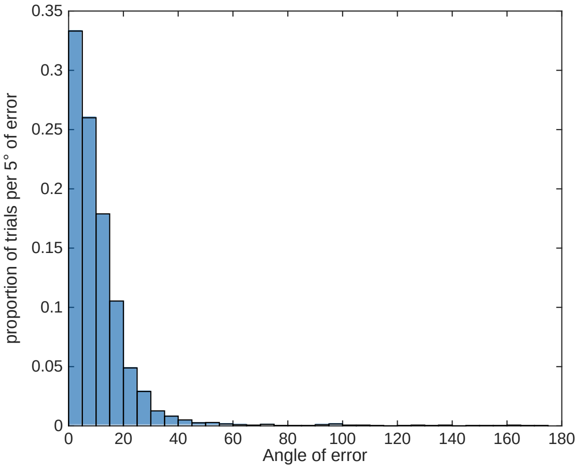


Figure 3.4. Histogram showing the proportion of trials per 5° of angular error across both retrieval blocks.

3.3.2 Grid stability

Based on the grid orientation estimated from the univariate results (reported in Horner et al., 2016), the differences in grid orientation across the two retrieval blocks (Figure 3.5A,B) and across navigation and imagination periods from the same block (Figure 3.5C) were calculated for every participant. The average difference in orientation between the two navigation blocks was 20.27° (*SD* = 16.96), between the two imagination blocks: 20.00° (*SD* = 16.19), and between corresponding navigation and imagination periods: 18.31° (*SD* = 15.35).

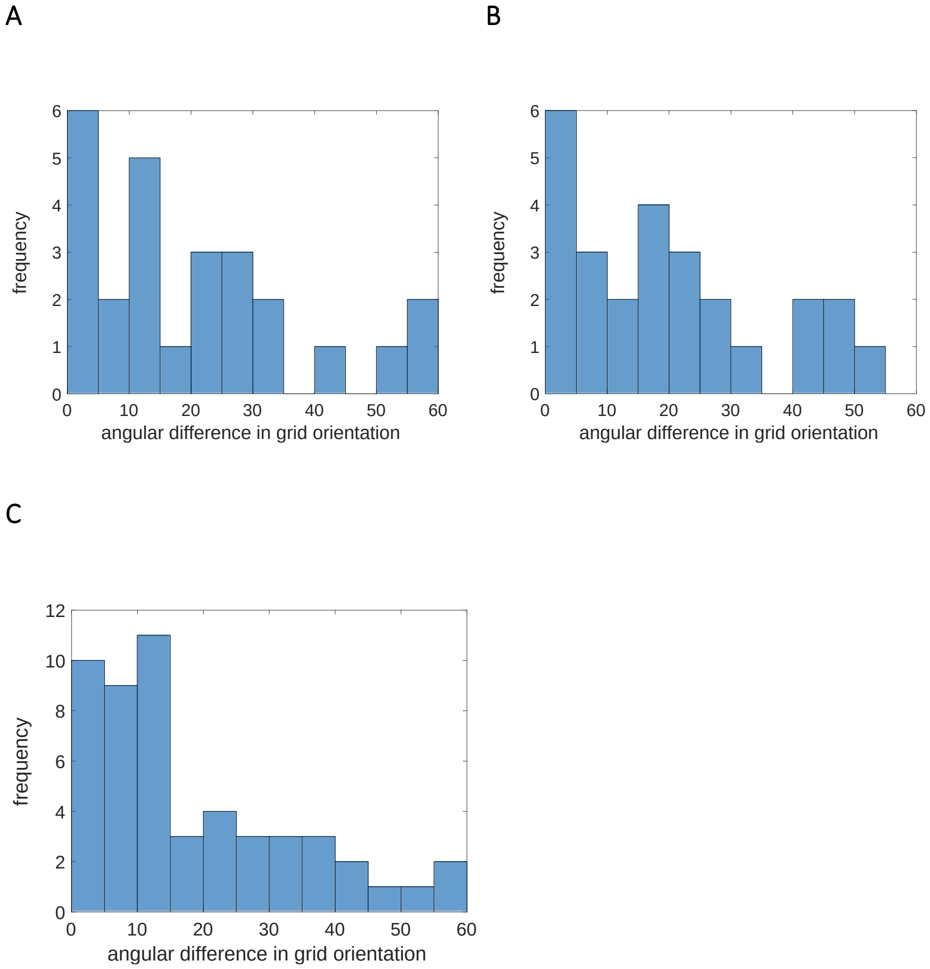


Figure 3.5. Stability of the grid orientation across participants estimated using univariate analyses on the same data, reported in Horner et al. (2016). (A) Angular differences in grid orientation across navigation periods. (B) Angular differences in grid orientation across imagination periods. (C) Angular differences in grid orientation between navigation and imagination periods from the same block, for both retrieval blocks.

3.3.3 Encoding blocks

3.3.1 Grid-like signal

I found no areas consistent with the representation of spatial structure with 60° periodicity consistent with grid-like signal (*p* > .05 FWE-corrected for the whole brain and for the bilateral EC).

3.3.2 Heading direction

In order to look for heading direction representation, I compared the data to a model which assumed that similarity decreases as the angular difference in heading direction decreases. I found areas in the bilateral occipital cortex and left inferior parietal lobe that followed this prediction (*p* < .05 FWE for the whole brain; Figure 3.6).

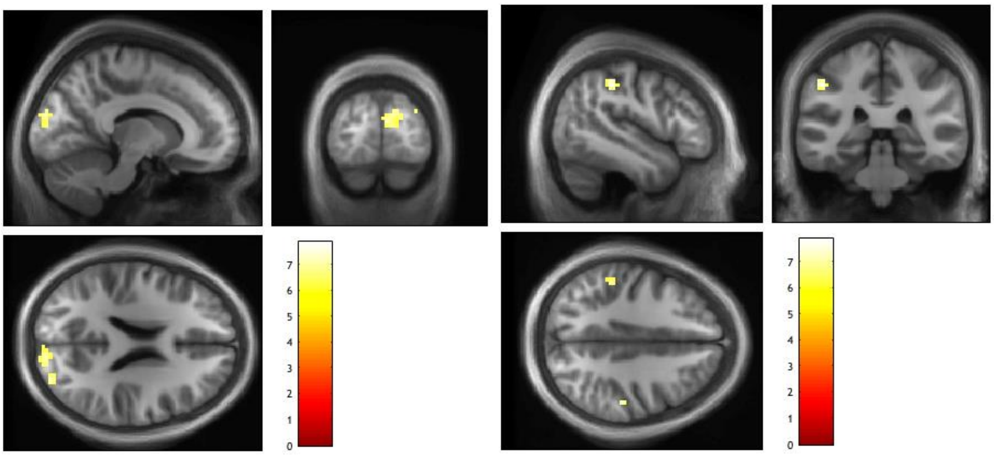


Figure 3.6. Representation of heading direction during encoding. Searchlight results showing clusters in the occipital lobes (3, -78, -3) and left inferior parietal lobe (-48, -33, 42; p < .05 FWE-corrected for the whole brain).

3.3.3 Goal location

I found no areas corresponding to the prediction that more proximal goal locations are represented with more similar activity patterns ($p > .05$ FWE for the whole brain).

3.3.4 Retrieval blocks: navigation and imagination periods

3.3.1 Grid-like signal

The main prediction regarding the grid-like representation of space was tested at retrieval. I found no evidence for this representational structure during navigation or during imagination phases in full-brain analyses and when using small-volume correction for bilateral EC.

3.3.2 Heading direction – navigation

Activity consistent with the representation of heading direction was found in the occipital lobes ($p < .05$ FWE-corrected for the whole brain; Figure 3.7).

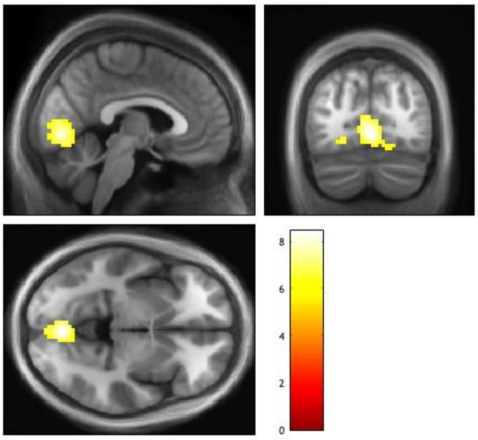


Figure 3.7. Representation of heading direction during navigation period at retrieval. Searchlight results showing clusters in the occipital lobes (6, -87, 18; $p < .05$ FWE-corrected for the whole brain).

3.3.3 Heading direction – imagination

The same model as above was used to look for heading direction representation during imagination. With a FWE-correction for the whole brain, I found a single significant voxel in the occipital cortex. At more liberal threshold, approximately the same area as the one during the navigation period (see above) was observed (Figure 3.8).

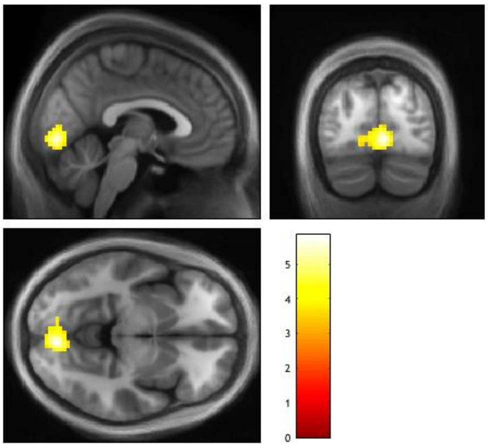


Figure 3.8. Representation of heading direction during imagination period at retrieval. Searchlight results showing clusters in the occipital lobes (6, -81, -6; for display purposes the image is threshold at p < .001 uncorrected).

3.3.4 Navigation vs. imagination

In this exploratory analysis, I looked at which areas show the highest difference between within-modality and across-modalities similarity. At whole-brain-corrected threshold, the entire brain showed higher similarity within a modality (navigation or imagination) than across modalities. At increased threshold ($p < 10^{-9}$ FWE-corrected for the whole brain), the areas that showed the greatest fit to this model were the occipital lobes extending to posterior parietal cortex, and bilateral PPA (Figure 3.9).

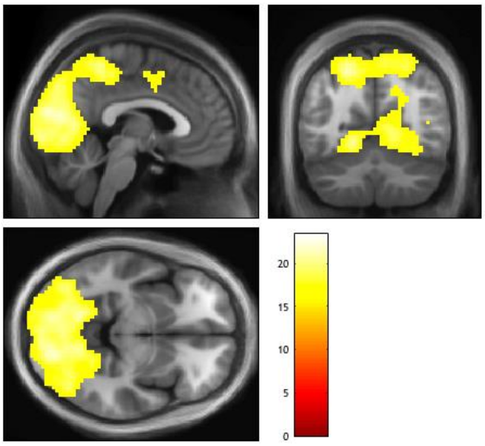


Figure 3.9. Areas differently engaged during navigation and imagination. Searchlight results showing clusters in the occipital lobes (-15, -81, 24; $p < 10^{-9}$ FWE-corrected for the whole brain).

3.3.5 Retrieval blocks: cue periods

3.3.1 Starting location

These additional analyses of the cue periods were used to further explore the representations of space during stationary periods where it was also possible to define the current (starting) locations (during encoding, the time spent at the starting location was very limited). I found activity consistent with the representation of starting locations (i.e. higher similarity when starting from proximal locations) in the right hippocampus and right parahippocampal cortex, which may correspond to the parahippocampal place area (PPA) ($p < 10^{-6}$ FWE-corrected for the whole brain; Figure 3.10).

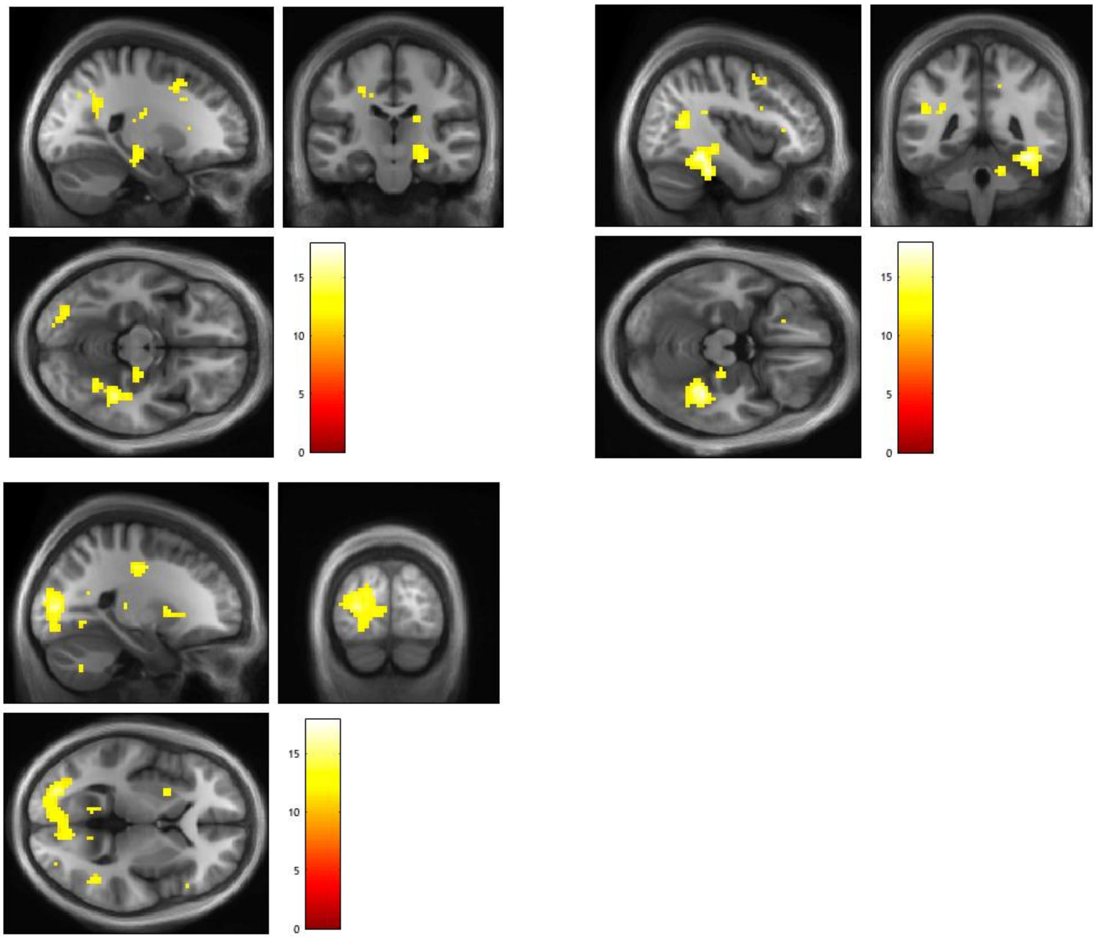


Figure 3.10. Representation of starting location during cue period at retrieval. Searchlight results showing clusters in the right hippocampus (24, -21, 15), right parahippocampal place area (42, -33, -30), and occipital lobes (-27, -81, 6; $p < 10^{-6}$ *FWE-corrected for the whole brain).*

3.3.2 Goal location

I found that the pattern of activity in the left PPA was consistent with the representation of the goal location (i.e. trajectories to the same goal more similar to different goals) ($p < 10^{-7}$ FWE-corrected for the whole brain; Figure 3.11).

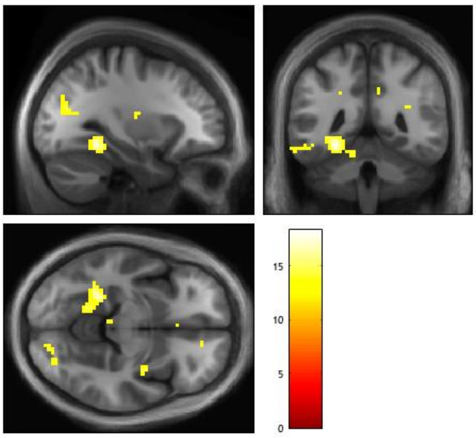


Figure 3.11. Representation of goal location during cue period at retrieval. Searchlight results showing clusters in the left parahippocampal place area (-27, -45, -12; $p < 10^{-7}$ FWE-corrected for the whole brain).

3.4 Discussion

This study investigated the grid-like signals during navigation and imagination in a VR environment. Previous studies using univariate analyses have shown that the differences in the BOLD signal between heading directions consistent and inconsistent with the main axes of the grid can be used to decode the grid orientation from the fMRI signal (Doeller et al., 2010; Horner et al., 2016). Here I used RSA to estimate the similarity of the activity patterns associated with navigating and imagining to navigate in different directions within a virtual environment.

Grid analyses

The main aim was to decode the grid-like signal which I expected to be characterized by high similarity of heading directions separated by the multiples of 60° (similarly aligned with the grid) and low similarity of the directions where one of them lays ‘on-axis’ and the other one ‘off-axis’ (i.e. separated by 30°, 90°, 150°). In particular, I wanted to see whether the same representational structure is observed in both navigation and imagination. I found no evidence for this prediction. However, in retrospect this method may have been unsuitable for testing of this model. I will argue that the grid-like signal detectable with univariate fMRI analyses could not be detected with the RSA due to the differences in how the grid cells firing affects the estimates used by these methods.

As mentioned in the introduction, the univariate signal consistent with 60° periodicity in the EC has been found previously in navigation (Doeller et al., 2010) and both navigation and imagination periods on the same data set as presented here (Horner et al., 2016), showing that the grid cell activity can be detected in humans using fMRI. The analysis method used to detect this signal relies on the two assumptions outlined in the introduction to this chapter.

First, the 'regular' grid cells across the EC have consistent orientation (Barry et al., 2007; Stensola et al., 2012). Although their firing is not modulated by the running direction, the neural dynamics of the cells differ, with some cells firing a lot and others a little when running along axis or all firing at an intermediate rate when running in-between axes. This, combined with non-linear mapping of neural activity to the BOLD signal, could produce differences in the univariate signal. However, although the overall amount of firing might differ with running direction, the pattern of firing across voxels would not be affected.

Second, the conjunctive grid cells are modulated by heading direction in a unimodal way, with preferred firing directions aligned to one of the grid axes (Sargolini et al., 2006). This again could produce a univariate signal that differs between aligned and misaligned to the grid directions. However, as before, there would be no difference in the pattern of activity across voxels for aligned versus misaligned runs; even if the cells with different preferred directions were un-evenly distributed across voxels, this would produce variation in activity patterns between differently aligned directions, but not between aligned versus misaligned directions.

There is therefore no reason to suggest that different grid cells become activated in the same environment as a result of differences in behaviour such as the heading direction. Rather, the same grid cells should be active regardless of the direction of navigation, and heading direction should only affect the dynamics of firing but not the pattern. Thus, different heading directions should result in differences in the univariate signal but not in the pattern of neural activity, which is what is assumed to be detected with the multivoxel analysis methods such as the RSA. The RSA may therefore be unsuitable for looking at activity consistent with the grid cells firing; in principle, it should not be possible to detect any differences in patterns across conditions above the univariate differences in the BOLD signal.

On the other hand, one study did find evidence for grid-like representational structure of the environment using the RSA. In that study, participants mentally simulated different facing directions in a familiar VR environment, and the pattern similarity corresponded to the representational structure analogous to the one used in the current study (Bellmund et al., 2016). The reasons for this discrepancy in findings are unclear. The design in both studies is quite different; in the current study, I looked at spatial representations during actual and imagined movement, while Bellmund et al. looked at directional imagination in the absence of (real or imagined) movement. Potentially, the analysis in the study by Bellmund et al. picked up on another process which showed some similarity to the 'on vs. off grid' representation but which was different to the grid cells signal. For instance, the detected differences could reflect firing of directional cells with multi-modal modulation by direction; tri-modal modulation could result in the similarity structure consistent with the 60° periodicity. This type of modulation is occasionally seen (N. Burgess, personal communication) but these types of responses have not been documented in significant numbers. It is also unclear why such a signal would have been present in Bellmund et al., but not the present study. Perhaps the signal in Bellmund et al. was somehow more reliable, possibly reflecting the larger environment used there, or other elements of design which mitigated against our study (see below). Indeed, the grid-like signal in Bellmund et al. was relatively weak – only reaching significance in a small region of interest corresponding to the homologue of medial EC, and only in one hemisphere.

Another alternative is that the grid orientation does vary across voxels to an extent that could result in pattern similarity consistent with the hexadirectional model used here and in Bellmund et al. (2016). Although previous research showed that the orientation is generally shared across grid cells, there is some variability (although lower than would be expected if orientations were random; Barry et al., 2007a; Stensola et al., 2012). Additionally, variability in orientation between grid cell 'modules' (clusters of grid cells with similar properties such as grid scale) is higher between than within modules (Stensola et al., 2012). This may suggest that the differences in patterns observed in Bellmund et al. might have reflected differences in grid orientation between the modules of different scales. If, as an example, the orientations across modules differed by 30°, running directions that are 30°, 90° and 150° apart would be aligned to the grid for some modules and misaligned for the other, resulting in expected pattern similarity structure; although differences in grid orientation of this magnitude have not been observed, the same argument could potentially work for lower variation. Potentially, the smaller environment used here was insufficient to sample enough different grid scales (and thus also different grid orientations) to show any differences in activity patterns. The presence of grid-like signal in the univariate signal (Doeller et al., 2010; Horner et al., 2016) could still be detected despite the variation in grid orientation as long as it was relatively low.

In addition to the smaller environment, the discrepancy between the current results and Bellmund et al. might have resulted from the low power in the current analyses; for in-depth discussion of the aspects of the data and task design that might have contributed to this issue (e.g. low number of examples in each condition, continuous as opposed to discretised environment), see limitations sub-section below as well as section 7.7 of the general discussion.

Heading direction analyses

Although it was not possible to detect any grid-like signal using the RSA, the pattern consistent with the representation of the heading direction was observed in the occipital lobes during both encoding and retrieval phases. Head-direction signal has been previously decoded using multivariate pattern analysis of fMRI data from the EC, pre- and parasubiculum (Chadwick et al., 2015) and in the posterior parahippocampal gyrus (Bellmund et al., 2016). The current results likely do not reflect the activity of the head-direction cells, which outside of the hippocampal formation have been found only in the thalamus and the retrosplenial cortex (Jacob et al., 2016; Olson et al., 2016; Sargolini et al., 2006; Taube, 2017; Taube et al., 1990). This area may instead represent the similarity of the visual scene which, rendered at infinity, was more alike for the more proximal heading directions. The same areas were identified in both actual navigation and imagination, although at relatively lower level during the imagination period, suggesting that the imagined route is reinstated in the cortex as would be predicted by the BBB model (Byrne et al., 2007). The patterns of activity were similar, but not identical, possibly due to qualitative differences between the two modalities, with imagination potentially containing fewer details and less rich or intense content than perception during navigation. The representation of heading direction in the occipital cortex was observed in the original study using the same paradigm to look for grid signal during navigation (Doeller et al., 2010), as well as during both perception and imagery of scenes, with stronger responses observed during perception, consistently with the current finding (Ganis, Thompson, & Kosslyn, 2004).

Despite the overlap in the neural mechanisms involved in both perception during navigation and imagination of movement (Byrne et al., 2007), areas that differed between the two modalities were also identified. Consistently with previous research (e.g. Ganis et al., 2004), I found that regions in the occipital cortex show the highest between-modality distances, which may reflect the differences in sensory processing involved in perception and imagery. This area extended to regions previously implicated in mental navigation: parieto-occipital junction (Ino et al., 2002) and precuneus (Ghaem et al., 1997). Finally, the PPA is known to play a role in both perception (Epstein & Kanwisher, 1998) and imagery of spatial scenes (O'Craven & Kanwisher, 2000), however the current results suggest that the specific representational structure in this area is more similar within- than between-modalities. Although imagery recapitulates perception to some extent, it cannot match the richness and resolution present during actual perceptual experience.

Starting location analyses

During the stationary cue period, when the object to be found was displayed on the screen, the representation of the starting location (i.e. the currently occupied location), with more proximal locations represented by more similar patterns of activity, was found in the right hippocampus, right PPA and the occipital lobes. This is broadly in line with the role of these regions in representation of space and visual scenes. Hippocampus is known to represent the location of the animal in the environment (O'Keefe & Dostrovsky, 1971). The allocentric representation of space relies on the ability to represent distances (Wilson & McNaughton, 1994), which may be achieved through the attractor properties of the hippocampal network (Marr, 1971; McNaughton & Morris, 1987; Samsonovich & McNaughton, 1997), as described in Chapter 1. Representation of distances in terms of strengths of synaptic connections results in a continuous representation of space (Muller, Kubie, & Saypoff, 1991), which could explain the pattern of similarity observed in the current study, with more proximal locations represented as more similar. This is consistent with a fMRI study which showed that closer landmarks are represented with more similar patterns of activity in the hippocampus (Morgan, Macevoy, Aguirre, & Epstein, 2011). The parahippocampal place areas represent the local visual scenes, particularly the spatial layout (Aguirre, Detre, Alsop, & D'Esposito, 1996; Bastin et al., 2013; Epstein & Kanwisher, 1998) in an observed-cantered reference frame (Epstein, 2008). In the current study, both the involvement of the PPA and the occipital cortex during the cue period may reflect the higher similarity of the visual scene from the same viewpoint than from others. The occipital cluster extended to the traverse occipital sulcus (TOS), also referred to as the occipital place area (OPA), also involved in scene perception (Dilks, Julian, Paunov, & Kanwisher, 2013; Hasson, Harel, Levy, & Malach, 2003; McAvoy & Sahay, 2017).

Alternatively, the similarity of starting locations could reflect similarity of the visual input. To test this, it would be beneficial to exclude trials with high visual similarity. However, as the starting location in a given trial was the end location of the previous trial, the facing direction would most likely always show some similarity at similar locations (because of facing "outward" from the arena, and likely never facing towards the centre of the arena).

Goal location analyses

The goal location during cue periods was found to be represented in the left PPA which may, again, correspond to the similarity of the visual scene when facing proximal locations, or processing of the spatial layout of the environment from an egocentric viewpoint (Epstein, 2008; Epstein & Kanwisher, 1998). In this analysis, unlike in the starting location analysis, I used the actual locations of the objects in the VR which may have not corresponded to the retrieved direction, and thus these results should be treated with caution - although the model would be appropriate even if the remembered item-locations were incorrect in absolute terms, as long as the relative position of each object in relation to other objects was correctly remembered. I did not observe patterns related to the goal in the EC/subiculum as would be expected from previous studies showing involvement of these areas in representing goal and facing directions (Bellmund et al., 2016; Chadwick et al., 2015). This could be because during at least some part of the cue period participants rotated around, which may have obscured any signal related to facing direction.

Limitations

There were several limitations to this study. Some of them can be attributed to the design of the experiment, which originally was not intended to be analysed with the RSA. First, the dissimilarity matrices assessed in the RSA require that the stimuli or experimental factors can be split into categorical features or conditions. The continuous environment used in this study, while allowing for sampling of a wide range of directions, was not optimal for this purpose. In particular during the retrieval blocks, participants were likely to take routes which did not follow the intended trajectories. Despite this, the RSA analysis required that the environment is split into heading direction bins which allowed for classification of behaviour into one of the resulting conditions. The conditions themselves were fairly arbitrary in that they were based on the 'ideal' straight-line trajectories between starting and end locations in each set, which were not necessarily the directions taken by the participants; however, they were assumed to be the most likely directions and alternative ways of discretising the environment were found to be even less appropriate (as will be explained below).

Another issue stemmed from the fact that the two sets of encoding-retrieval blocks were offset by 15°, and so the condition-bins (each spanning 30°) were inconsistent across the sets. For example, if the first set included a condition with directions between 30° and 60°, the second set would have included condition between 15° and 45°, and between 45° and 75°. The same heading direction in two different sets would therefore be a part of a different condition when comparing similarity across blocks. This, combined with the instability of the grid orientation, could have affected the fit of the models to the data. Arguably, a task design with a more structured environment would make it easier to construct appropriate conditions. One example of a paradigm that achieved this comes from a study looking at representations of current location and heading direction in a 3D environment (Kim, Jeffery, & Maguire, 2017); the environment consisted of a structure with four floors and four-by-four arrangement of 'nodes' on each floor, with every node connected to its neighbours in both vertical and horizontal directions by a narrow path. The lack of ceilings or walls allowed for perceiving of the entire continuous environment but the paths constrained movement to pre-specified directions. Nonetheless, there was still not enough sampling of direction to assess the presence of grid-like patterns in that study, which emphasises the difficulty of constructing an appropriate environment for this type of analyses. In the current study, limiting the available paths was partially achieved in the encoding blocks where, because the items to be collected were visible in the environment, participants were more likely to reach them in a straight line, providing some control over heading directions. Reassuringly, in these blocks corresponding areas as in the retrieval blocks were observed to show the similarity pattern corresponding to heading direction. However, unlike during retrieval blocks, I failed to detect any signal consistent with the similarity of goal locations, which may be a result of different cognitive processes involved at encoding and retrieval of item locations. Possibly, encoding may rely more on processing of the object location in an allocentric manner, in relation to the landmarks and other remembered objects, while retrieval may be dominated by placing it in the egocentric reference frame in order to decide on the required direction; although both frames of reference are likely used in both tasks.

Another attempted solution to the overlapping direction bins in our study was to use consistent bins across both sessions (despite differences in starting-goal locations) or smaller (15°) bins that were consistent across blocks, however in these analyses (not reported here) some direction bins were not used at all in a large proportion of participants, and so the dissimilarity matrices could not be calculated.

This highlights another issue with the current design which made it suboptimal for the RSA; a small number of examples of behaviour in each condition resulted in relatively poorly estimated beta coefficients in the first level GLM analysis, which in turn would have affected the estimates of distances. This might have contributed to the differences in results between encoding and retrieval blocks; for example, the corresponding analyses of heading direction showed higher number of significant voxels in the analysis of the retrieval blocks than of the encoding blocks, potentially due to the higher number of examples used to estimate each of the conditions at retrieval (each route was taken once at encoding and twice at retrieval). The study was optimised to sample a wide range of possible directions, while for the RSA ideally there should be fewer and more constrained movement directions with multiple examples of each.

Small number of trials per conditions also limited the complexity of models; for instance, it was not feasible to compare representational structures between trials with different levels of angular error. Additionally, I could not combine the analyses of heading direction and start-end locations during retrieval because there were not enough trials to provide examples of all combinations of these conditions.

The grid orientation, as estimated using univariate analyses, was not stable across blocks. Research in rodents has found the grid orientation to be stable across time (Hafting et al., 2005; Sargolini et al., 2006), even when the context changes and the hippocampal place cells undergo rate remapping (Fyhn et al., 2007), as well as in healthy people using intracranial recordings (Kunz et al., 2015). It is therefore unclear what could underlie the instability of the grid observed in the current data. Possibly, this may be a result of switching between real-world and virtual reference frames which were inconsistent due to the supine position in the scanner (Moser et al., 2014). Although it is not clear how the change to VR would cause changes in orientation within the same VR environment across runs, it is clear that the environmental orientation references provided in VR will be less strong than provided by real environmental and vestibular inputs in the real world (e.g. actual head-rotation is required to see head-direction cell firing in rodent VR; Chen, King, Lu, Cacucci, & Burgess, 2018). Although, as explained above, the RSA may not be appropriate for the analysis of grid-like signal, other analyses (e.g. of heading direction) could in principle be affected by the unstable orientation signal as the orientations of both head-direction and grid cells are similarly anchored to the environment (Hafting et al., 2005; Mizumori & Williams, 1993); this may explain why the 'heading-direction' signal detected here was only observed in the visual cortex, as similarity of the visual input would be more stable across runs than the sense of direction.

3.4.1 Conclusions

I did not find any evidence for grid-like neural representations during virtual or imagined navigation using an RSA analysis of fMRI data. Overall, in retrospect the main aim of the current analyses may have been misaligned with the analysis technique I attempted to use. Because of the nature of the grid cell signal, where heading direction should only affect the magnitude of the BOLD signal but not the pattern of activity across voxels, it may not be possible to detect evidence for grid-like representation of space in either navigation or imagination (although it may be detectable with larger environments; see Bellmund et al., 2016). I did, however, find evidence for the representation of the heading direction within occipital lobes in both navigation and imagination, supporting the idea that the same mechanisms support both real navigation, retrieval of spatial memories and imagination (Byrne et al., 2007). Patterns of activity reflecting the similarity of starting locations were found in the hippocampus and the right PPA, supporting a continuous representation of space in this brain region, as well as areas involved in scene processing. Lastly, patterns of activity reflecting the similarity of goal locations was found in the left PPA, but only during retrieval and not at encoding, possibly reflecting the differences in cognitive processes during encoding and retrieval of goal locations.

Chapter 4

Behavioural study of the effects of category membership on pattern separation

4.1 Introduction

As I described in Chapter 1.2, a proposed function of the hippocampus is the separation of memory representations during encoding in order to reduce their interference and maximise the capacity of the memory storage network (McClelland et al., 1995; McNaughton & Morris, 1987; O'Reilly & McClelland, 1994). This process, pattern separation, is thought to be performed by the DG, which is proposed to orthogonalise the inputs from the EC to the hippocampus (Marr, 1971; McClelland et al., 1995; Rolls, 2013). Pattern separation is proposed to operate specifically on similar inputs, decorrelating them and producing more distinct outputs to be stored in the autoassociative network of CA3 of the hippocampus (Leutgeb et al., 2007).

The task most commonly used to measure pattern separation in humans is the Mnemonic Similarity Task (MST) in which participants are presented with a continuous sequence of items and asked to decide whether the item is new, old (previously seen) or similar (a 'lure' perceptually similar to but not identical to an old item). Successful recognition of lure items within this task is used as an index of pattern separation. Studies using these tasks have shown that the hippocampal signal is consistent with its ability to distinguish between original items and similar lures; the hippocampal responses to similar lures resemble those to novel but not to repeated items (Bakker et al., 2008; Kirwan et al., 2012; Kirwan & Stark, 2007; Lacy et al., 2011; Stark et al., 2013; Yassa et al., 2010).

Other studies have measured the differences in patterns of activity as a proxy of pattern separation. Higher dissimilarity of neural activity patterns in the hippocampus has been found to predict later memory performance for individual items, suggesting that items with more distinct activity patterns are better remembered (LaRocque et al., 2013). This is consistent with the proposed role of pattern separation in the reduction of memory interference. Moreover, patients with hippocampal damage are impaired at correctly identifying novel but similar items, again suggesting the role of the hippocampus in distinguishing similar memories (Kirwan et al., 2012). Likewise, the reduced ability to differentiate lures from original items observed in people with Mild Cognitive Impairment was associated with DG/CA3 hyperactivity in fMRI, as well as reduced DG/CA3 volume (Yassa et al., 2010).

These studies provide preliminary evidence consistent with hippocampal (and specifically the DG) involvement in pattern separation of similar memories. However, the MST suffers from a number of limitations which makes it suboptimal for testing of this function. First, most of the studies using MST do not include any relational aspect; i.e. the stimuli are standalone items and not more complex, interconnected associations. Arguably, associative processing is the crucial function of the hippocampus (e.g. Eichenbaum & Cohen, 2014) and so any task exploring one of its functions should require some form of association (Hunsaker & Kesner, 2013; Liu, Gould, Coulson, Ward, & Howard, 2016). A study combining both imaging and lesion-behaviour approaches concluded that the hippocampus is involved in contextual memory, but not in memory for individual items (Horner et al., 2012). 'Pure' item memory, on the other hand, may be processed in other medial temporal lobe structures (e.g. perirhinal cortex) (Aggleton & Brown, 1999; Eichenbaum et al., 2007; Horner et al., 2012; McClelland et al., 1995; Rugg et al., 2012).

Another potential issue with continuous recognition tasks such as the MST is the inability to distinguish between pattern separation and pattern completion, two complementary processes that utilize overlapping circuits within the hippocampus, allowing for both encoding of novel information and retrieval of associated representations. As the former operates at encoding and the latter at retrieval, it is suggested that an appropriate task should be one where the process of interest (separation or completion) is the most appropriate strategy (Hunsaker & Kesner 2013; Liu et al. 2016). The 'similar-lures’ paradigms, however, require continuous ‘recall-to-reject’ in addition to intentional encoding, thus blurring the distinction between the encoding and retrieval, and so also between pattern separation and completion. Put simply, the ability to detect the similar lures might only reflect accurate encoding and retrieval of the memorised items, rather than specifically requiring a process of pattern separation during encoding.

One way of dealing with this issue is the use of an incidental-learning paradigm where, for example, participants may judge the stimuli based on dimensions other than similarity (e.g. whether it is an outdoor or indoor item) while their brain activity is being measured. This has been successfully implemented by studies which did not require an explicit old/new/lure response. These studies still found evidence for pattern separation in the hippocampal responses to the ‘standard’ MST stimuli (Bakker et al., 2008; Lacy et al., 2011; LaRocque et al., 2013); however, these tasks did not produce a behavioural measure that can be interpreted without a reference to the brain activity at the time of encoding.

In the experiments presented in this chapter I attempted to address some of these limitations and thus try to design a task more appropriate for testing pattern separation. To do so, I broadly followed the recommendations provided by Liu et al. (2016), which outline the features of 'ideal' measures of pattern separation and completion. Liu et al. advise that a task designed to measure pattern separation should include an associative aspect and separate encoding and retrieval phases, as I already mentioned, as well as using stimuli which allow for parametric manipulation of similarity in order to look at graded changes in responses. An fMRI study used different degrees of similarity of lures to 'old' items and showed a change in activity to even small changes in the stimulus in CA3/DG, while CA1 showed more graded changes in response (Lacy et al., 2011). Manipulation of the degree of similarity may therefore provide more insight into the conditions under which pattern separation operates.

Based on these recommendations, I designed a task where I parametrically manipulated the similarity of facial stimuli. I created novel, unique facial stimuli that were equidistant to each other in terms of similarity (see the methods section of this chapter). For each pair of faces, I created morphs (linear combinations of the original faces) with different levels of similarity to the 'prototype' faces. Participants learnt to recognize each prototype face as either a friend or a foe. At retrieval, they were presented with either a previously memorized face or a morph and asked to identify it as either 'old' (seen during encoding; a prototype) or 'new' (a morph). Importantly, the participants did not know in advance that the test will involve recognition of old and new stimuli; rather, they were likely to assume that they will be tested on whether each face was a friend or a foe. This was to encourage a formation of representations that are more distinct for stimuli from different categories than for faces within the same category. This therefore resembled the 'similar-lures' paradigms as described earlier, but it included an associative aspect (the category to which each face belonged) and separated encoding and retrieval phases.

It was hypothesized that if the pair of original faces is of the same category (e.g. both faces are friends or both are foes), their corresponding mnemonic representations will be more similar (i.e. less separated) compared to when each member of the pair is from a different category. This is because pattern separation should operate mainly along the dimensions that need to be distinguished, in this case the valence of faces. In the case of opposite-category pairs of stimuli, insufficient pattern separation could lead to incorrect responses and so the representations of such stimuli should be more distinct.

Although it is not clear whether pattern separation operates selectively on behaviourally relevant dimensions, there is some evidence suggesting that the similarity of representations in the hippocampus changes with the behavioural requirements of the task. For example, Aly and Turk-Browne (2015) showed that the same stimuli in a virtual environment resulted in different similarity of patterns in the hippocampus depending on whether the participants had to attend to geometrical features of the environment or to the painting in each virtual room. Tavares et al. (2015) asked participants to interact with virtual characters of varying levels of power and affiliation and found that the hippocampal activity correlated with the position in a 'social' space' defined by these two attributes. This provides some evidence that the hippocampal representations may be based on behaviourally-relevant dimensions and not exclusively on perceptual similarity (as used in the MST).

It is predicted that, overall, correct classification of morphs as novel will decrease with increased similarity to the prototypes. As a result of the need to distinguish between faces of different categories, departures from the original 'prototype' representations in the current study should be more easily perceived for the opposite-category (pattern-separated) faces, and so their morphs should be classified as 'new' at relatively higher levels of similarity to the originals. Conversely, lures created from same-category pairs will need to be more dissimilar to the original faces to be recognised as novel. For a schematic illustration of the predicted pattern of results see Figure 4.1. I will present two experiments designed to explore this prediction, each using a slightly altered version of the task that differently manipulate the way the faces are encoded (individually or in relation to other faces).

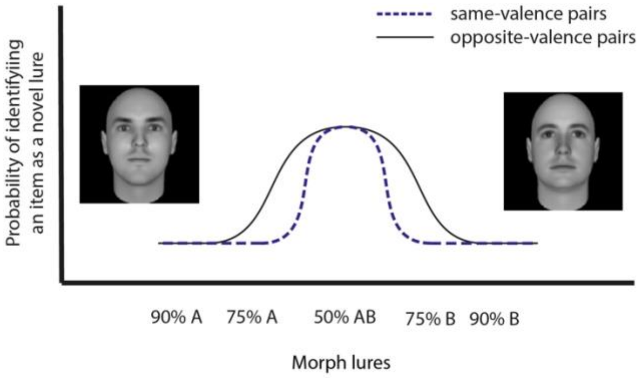


Figure 4.1. Diagram showing the expected behavioural pattern. Opposite-valence pairs create a greater need for pattern separation, resulting in higher ability to identify similar lures as novel. The gradient on the x-axis is not the one used in the studies and is only shown for illustration purposes.

4.2 Experiment 1

4.2.1 Methods

4.2.1 Participants

23 participants were recruited through the UCL ICN Subject Database, with one excluded from the analyses due to failure to complete the learning phase of the experiment within an hour. The remaining 22 participants (13 females, 9 males) had a mean age of 26.18 (*SD* = 6.58) and normal or corrected-to-normal vision. None of the participants suffered from prosopagnosia (face blindness).

4.2.2 Materials

Basel Face Model (Paysan, Knothe, Amberg, Romdhani, & Vetter, 2009) was used to create original faces and their morphs. BFM is a publicly available 3D Morphable Model (3DMM; http://faces.cs.unibas.ch/bfm/main.php?nav=1-0&id=basel_face_model). It was created by training the model on 100 male and 100 female mostly European faces of people scanned using a coded light system. Faces were co-registered in the 3D domain by progressively fitting a triangulated template mesh to the surface of the face. Each face was then represented by an *m*-dimensional vector (m being the number of vertices: 160,470). Principal Component Analysis was used to fit the data, creating a parametric face model. New faces can be created by the user through linear combinations of the principal components (PCs), with *N-1* PCs needed to create *N* faces.

The faces in the BFM were additionally represented by colour/texture; however, the faces used in the experiments presented here were kept at constant colour parameters (i.e. the mean of all faces' colour at a given vertex). The resulting face colour was that of a Caucasian person, reflecting the sample on which the model was trained.

I created ten equidistant facial stimuli by manipulating the PCs used to define face model. The first PC was not used as it corresponded to the overall size or 'masculinity' of the face, as opposed to local features, and had a disproportionate influence on the appearance of the stimuli.

Faces were created by generating the coordinates of *N+1* equidistant points in *N*-dimensional space defined by the PCs using the Pythagorean theorem. Although this could not ensure that the faces were equally similar to each other perceptually (as some PCs may be more perceptually 'meaningful' than others), it resulted in a reasonable measure of similarity or distance between two faces, which in turn allowed to create morph faces (combinations of two faces) with the same 'spacing' along this distance. All faces were presented looking straight ahead and on a black background (see Figure 4.2).

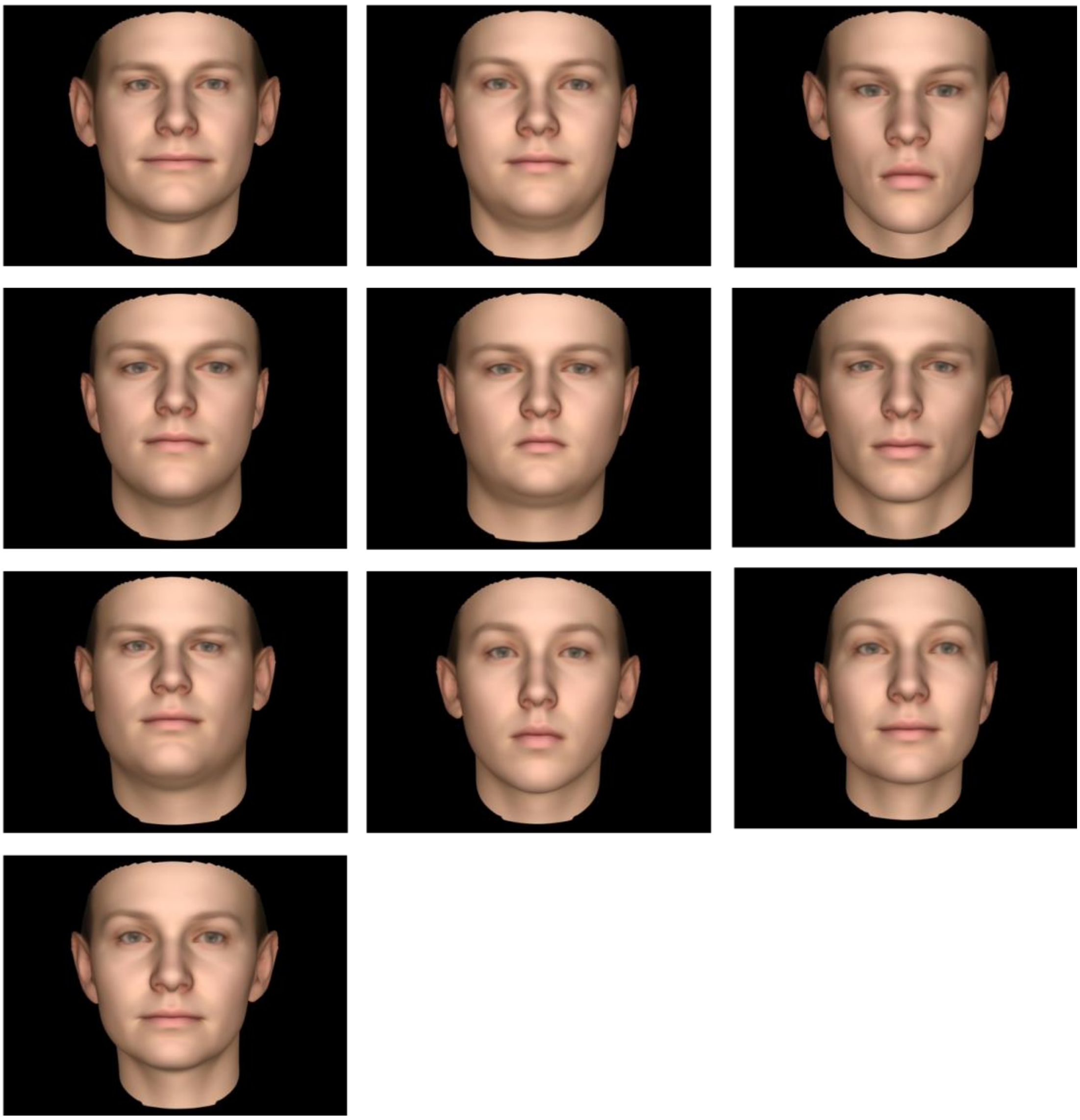


Figure 4.2. Experimental stimuli. Ten prototype faces created with Basel Face Model. The morph faces were created by manipulating similarity between the prototype faces (see text).

For each A-B pair of prototype faces (45 pairs in total from ten prototype faces), nine 'morphs' were generated, with the similarity to face A increasing incrementally from 10% to 90%, in 10% intervals (and correspondingly decreasing for face B). The morphs were generated by calculating the PC coordinates of the points lying on a Euclidean line connecting the coordinates of the original face pairs (A and B) at equal intervals and creating facial stimuli with these coordinates. This resulted in 405 morphs for a total of 415 facial stimuli. The stimuli were presented via the Cogent 2000 toolbox in Matlab (Mathworks).

4.2.3 Procedure

Encoding phase

Participants were shown one face at a time with a label ‘friend’ or ‘foe’ displayed above the centrally presented stimulus. Half of the faces were friends, and half were foes. The division of faces into the friend and foe categories was random for each participant. There were four encoding blocks; each face was shown once in each block in a randomized order but always with the same label. Each stimulus was presented for 5 seconds, preceded by a fixation cross (1 second) and followed by a blank screen (0.5 seconds). In total, there were 40 encoding trials (each of the 10 prototype faces was presented four times).

Learning phase

The learning phase was introduced to ensure that the participants correctly memorized the associations between the faces and their labels. Participants were shown one face at a time and asked to indicate with a key press whether the face was a friend or a foe. The face was presented for 3 seconds during which the response was to be made, preceded by 0.5 seconds fixation cross and followed by 0.5 seconds blank screen. Each response was followed by feedback: correct responses were indicated by a green background, and incorrect by a red background. During the feedback period (1 second), the correct label for a given face was also displayed. A lack of response was treated as an incorrect response. Each face was presented once within a block, and the order was randomized within each block. The learning phase finished when a participant correctly labelled all faces within the block (accuracy 100%) and so the number of blocks varied between participants. Every four blocks, participants were given an opportunity to take a break and continue when ready.

Retrieval

During the retrieval phase, participants were shown either faces they had previously encountered during the two preceding phases, or morph faces. The task was to indicate whether the face was ‘old’ (i.e. seen previously) or ‘new’ (a morph). Participants were informed that the new faces were created by ‘combining pairs of the old faces’ and so expected that all stimuli would to some extent resemble the previously learnt faces. The maximum time to respond was 5 seconds. Each face was preceded by a 0.5 seconds fixation cross and followed by a 0.5 seconds blank screen. There were 415 trials overall (10 prototype faces and 405 morphs).

4.2.2 Results

On average, participants took 14 learning blocks to achieve the threshold in accuracy (*SD* = 12, range: 2-43).

4.2.1 Comparison of original and morph faces

First, I compared recognition performance between original faces and all morph faces, regardless of their composition; however, it needs to be stressed that this comparison is shown here as a way to summarise the data, while the main effect of interest is the distribution of responses and not the mean accuracy, which disregards the dimension of similarity of morphs to the prototype faces (see section 4.2.2.2). Accuracy (proportion correct) in correctly categorizing the stimuli as ‘old’ or ‘novel’ significantly differed between original and morph faces, $t(21) = 2.98$, $p = .007$, $d = 0.64$. The original faces were correctly classified more often than the morph faces (Figure 4.3). However, this effect is largely driven by the classification of morphs with high similarity to the prototypes as ‘old’ (see Figure 4.4).

D-prime (d') scores showed that the ability to correctly classify stimuli was significantly greater than chance; $d' = 0.56$ (SD = 0.66), $t(21) = 3.99$, $p < .001$, $d = 0.85$, suggesting that participants were able to discriminate the two types of faces, although there was also a significant bias towards 'old' responses as suggested by the decision criterion value, *C* = 0.36 (*SD* = 0.59), *t*(21) = 2.84, *p* = .010, *d* = 0.55.

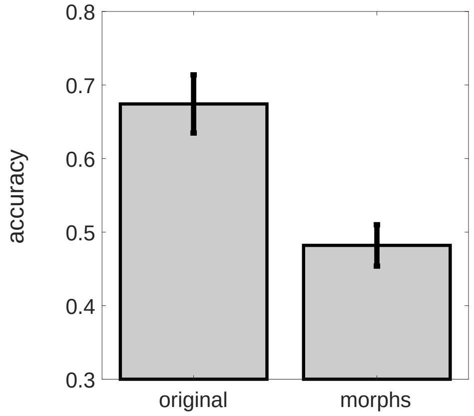


Figure 4.3. Average recognition accuracy between original and morph faces during the retrieval phase. Error bars represent +/-1SD.

4.2.2 Comparing response profiles

I tested whether the categorisation of faces as novel as a function of similarity to the original faces depends on whether the pair of original faces were of same or different category. I used Kolmogorov–Smirnov test to compare the two distributions of 'new' responses to all stimuli between same- and different-category faces for each participant. Each profile consisted of six data points corresponding to the mean probability of classifying each 'level' as novel, including the original faces (100% similarity) and all intermediate stages (10-50% similarity), across all face pairs. The test revealed no significant difference in the two profiles, *D* = 0.33, *p* = .810 (Figure 4.4).

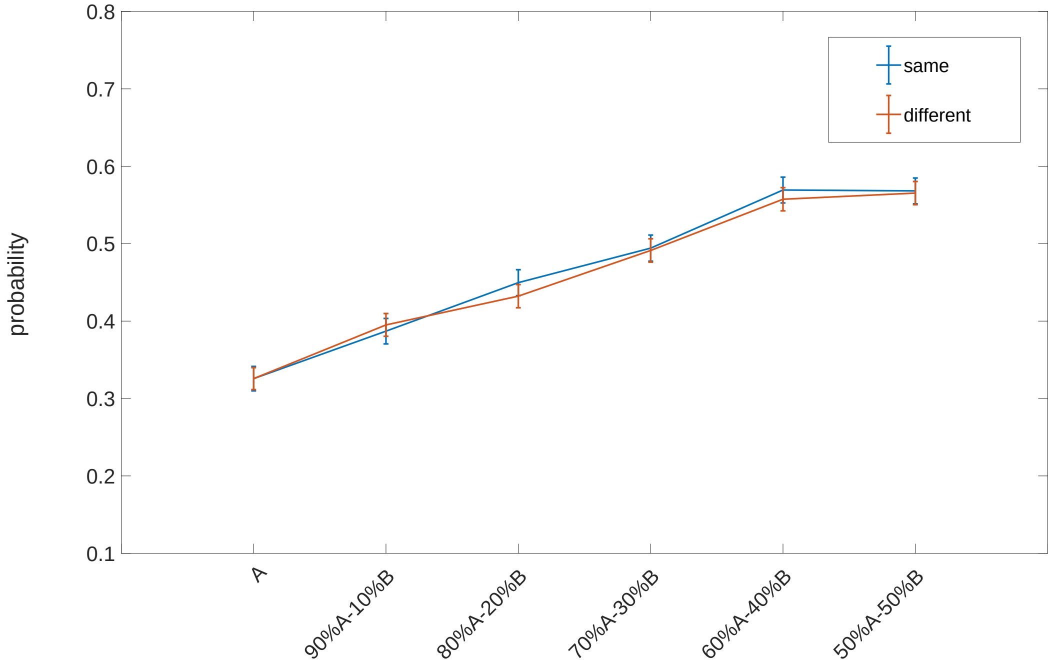


Figure 4.4. Probability of classifying face stimuli as 'new' for original faces (A and B) and morphs created by combining the original faces (from 90%A-10%B to 50%A-50%B), separately for morphs of same-category and different-category faces. Error bars represent +/- 1 SEM.

4.3 Experiment 2

The profile of responses in Experiment 1 was the same for morphs created from same- and different-category prototype faces. This does not support the prediction that the similarity of representations of pairs of faces would be lower for different-category faces, which in turn would lead to a correct identification of morphs as 'new' at lower 'distances' from the original faces.

The accuracy results for the morph stimuli in the Experiment 1 were close to the chance level (around 50%) and the number of blocks needed to achieve a desirable level of performance was high, suggesting that the task was too difficult for participants – although the sensitivity measure (*d'*) showed that participants were able, to some extent, to distinguish between old and new faces. The chance-level accuracy on morph faces were largely driven by the morphs at the highest levels of similarity to the prototypes. However, the performance was still very low and the range in the number of blocks it took to achieve the required level of accuracy was high, which meant that some participants were exposed to the same stimuli much more than others. The task therefore seemed too difficult, and so it was adapted and made easier in order to improve the overall performance.

First, we used a lower number of prototype faces (eight instead of ten) to make the encoding easier and decrease the number of trials. Second, the procedure was changed in a way to encourage the formation of representations of faces in relation to other representations. One possible explanation of the results from the Experiment 1 could be that during the encoding and learning phases, the face-category associations for each stimulus were learnt in isolation and this potentially led to emergence of equally distinct representations of prototype faces, regardless of whether they belonged to the same or different category. In the next experiment, I modified the task so that each stimulus was presented and had to be considered in comparison to other stimuli (see methods). This was intended to lead to pattern separation along the dimension relevant for the task (i.e. pattern separation of faces from the opposite categories).

At encoding, all faces of the same category were presented on the screen at the same time. This was intended to result in formation of individual representations with relatively high degree of similarity within each category due to an absence of any behavioural need to differentiate between them. This was repeated several times. During the learning phase, two faces were presented at the same time and participants were asked to decide whether the faces were from the same or different category (regardless of which category it was). It was intended that through this alteration to the procedure, the need for behavioural separation (or absence of it) would affect the process of formation of representations of each face, making the representations of face pairs with different category membership more distinct.

These changes to number of stimuli and to the procedure were expected to make the task easier and to encourage formation of pattern-separated representations of different-category faces. Other details of stimulus creation and procedure were the same as in Experiment 1.

4.3.1 Methods

4.3.1 Participants

22 participants were recruited through the UCL ICN Subject Database; four were excluded from the analysis as they failed to complete the learning phase of the experiment within an hour. The remaining 18 participants (10 females, 8 males) had a mean age of 26.56 (*SD* = 5.93) and had normal or corrected-to-normal vision. None of the participants suffered from prosopagnosia (face blindness).

4.3.2 Materials

Eight equidistant faces were created using the methods described in Experiment 1. This resulted in 28 face-pairs, which were used to create 252 morphs for a total of 260 facial stimuli.

4.3.3 Procedure

Encoding

Four out of the eight prototype faces were presented on a screen at a time. One group of faces was labelled ‘friends’ and the other one ‘foes’. The label for a given group was displayed above the face stimuli. Each group was presented for 15 seconds, eight times in alternating order, with the same faces always comprising each group but with their location on the screen (four corners) randomized. The display of each group was preceded by 1 second fixation cross and followed by 0.5 seconds blank screen.

Learning

Two faces were shown on the screen, on either side of the screen. Participants were asked to indicate whether the faces belong to the same category (both friends or both foes) or to a different category. Each pair was presented for 4 seconds during which a response was required. Each response was followed by feedback: green background for correct response and red background for incorrect. Additionally, the correct label for each face was displayed above it for 2 seconds. Each pair of faces was presented once per block, giving 28 trials per block. Due to the higher number of responses required within each block in this experiment (28 versus 10 in Experiment 1), the threshold was lowered; the learning phase finished when the accuracy within a block exceeded 85% (i.e. 24/28 correct responses).

Retrieval

The retrieval phase was the same as in the Experiment 1 but with fewer trials (260). The task was again to decide whether the face was an 'old' face or a 'new' face (a morph). The instructions given to the participants were as in Experiment 1.

4.3.2 Results

On average, participants took 14 learning blocks to achieve the threshold in accuracy (*SD* = 13, range: 2-48).

4.3.1 Comparison of original and morph faces

There was a significant difference in the recognition accuracy between the original and morph faces, $t(17) = 4.92$, $p < .001$, $d = 1.16$ (Figure 4.5). Here again, the effect was driven by an incorrect classification of morphs that were highly similar to the prototype faces (see Figure 4.6).

The d-prime showed that the correct classification was above chance level, *d'* = 0.48 (*SD* = 0.40), *t*(17) = 5.06, *p* < .001, *d* = 1.19. Again, there was a bias towards 'old' responses, *C* = 0.41 (*SD* = 0.35), *t*(17) = 5.03, *p* < .001, *d* = 1.19.

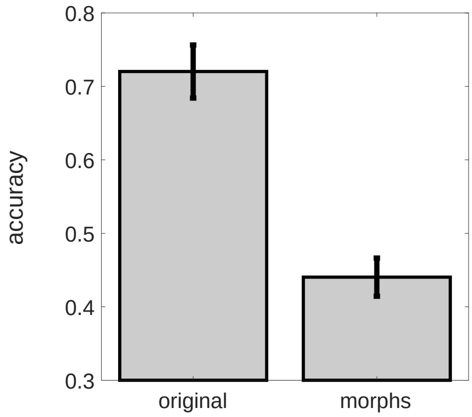


Figure 4.5. Average recognition accuracy between original and morph faces during the retrieval phase. Error bars represent +/-1SD.

4.3.2 Comparing response profiles

Finally, I compared the profiles of responses. As in Experiment 1, I calculated the average probability of classifying the stimuli as novel for all pairs of original faces and their morphs, proceeding from 100% face A to 50A-50%B in increment of 10%. Again, the Kolmogorov-Smirnov test revealed that the profile of responses did not differ between stimuli created from same-category and different-category faces, *D* = 0.17, *p* = 1 (Figure 4.6).

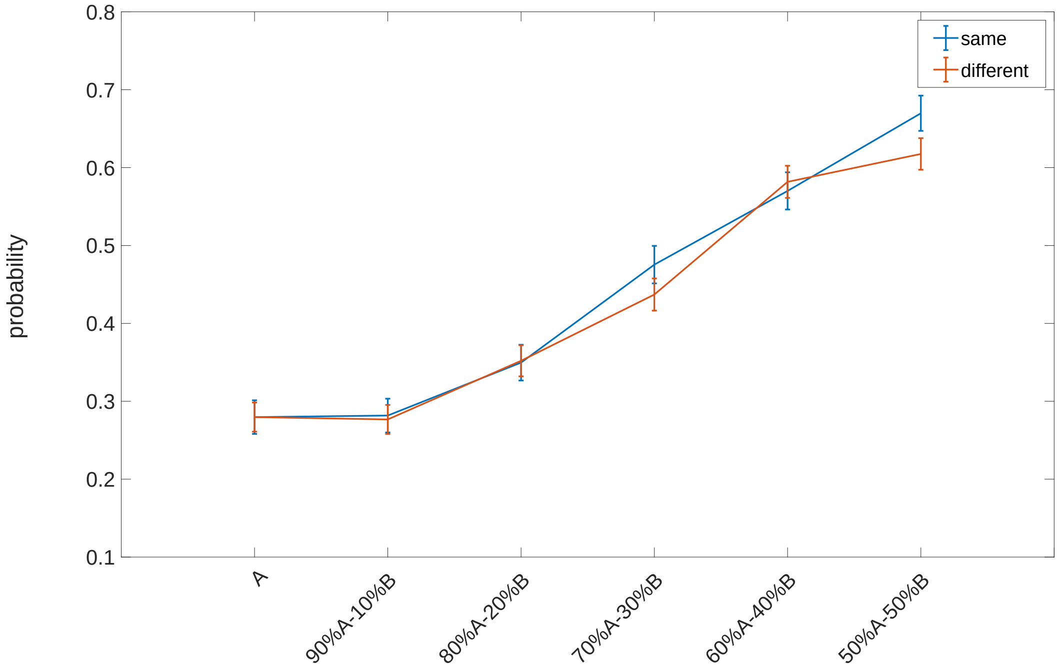


Figure 4.6. Probability of classifying face stimuli as 'new' for original faces (A and B) and morphs created by combining the original faces (from 90%A10%B to 10%A90%B), separately for morphs of same-category and different-category faces. Error bars represent +/- 1 SEM.

4.4 Discussion

In these two experiments, I attempted to manipulate the need for pattern separation by inducing the need to distinguish between some pairs of faces but not others. The faces were categorised in terms of valence: participants learnt that half of the stimuli were friends, and half were foes. I hypothesised that pattern separation would be increased along the behaviourally-relevant dimension (i.e. the dimension which had to be distinguished: the category membership) and, as a result, different-category stimuli will be represented as more distinct to each other than stimuli within each category. However, I found no evidence for this. The ability to correctly recognize lures showed the same distribution as a function of similarity to the original faces regardless of whether the two prototype faces were of the same or different category. However, it needs to be noted that the task was probably too difficult (despite efforts to make it easier in Experiment 2), and so any differences could have been masked by the floor effect – although the d-prime sensitivity measure suggested that 'old' and 'new' stimuli were distinguishable at an above-chance level.

The task used at retrieval was a similar-lures task where lures were morph faces created by linear combination of pairs of original (prototype) faces. As previous studies have shown (Lacy et al., 2011; Norman & O'Reilly, 2003), the ability to recognize lures as novel decreases as a function of similarity to the original stimulus. I hypothesised that this decrease will be less extreme for 'pattern-separated' pairs of stimuli, i.e. pairs of faces from different categories, due to higher distinctiveness of different-category representations. Contrary to this prediction, lures created from different-category faces were not recognized as novel any more readily than morphs of same-category faces.

There are multiple possible reasons for this pattern of results. The predictions were built on a number of assumptions that need to be addressed. First, it was assumed that different category membership would lead to higher pattern separation. The process of pattern separation is proposed to originate from the need to distinguish between representations in order to avoid their interference. Therefore, if two stimuli lead to different outcomes, their representations should diverge over learning (Gluck & Myers, 1993). It was therefore assumed that the stimuli are likely to be differentiated based on their category membership.

There is extensive evidence suggesting that category training leads to better perceptual discriminability along the category-relevant dimension, a phenomenon called "acquired distinctiveness" (Soto & Ashby, 2015; see Goldstone & Hendrickson, 2010 for review). However, in these studies the categories are related to the perceptual features that define them. Here, on the other hand, categories were arbitrary and the prototype faces assumed to be equally similar to one another regardless of category membership.

Within memory research, in most studies on pattern separation it is the similarity of representations that is assumed to induce interference. Here, on the other hand, similarity (i.e. belonging to the same category) is not the factor that could impair performance. Quite the opposite, being unable to distinguish between items from the same category would not affect one's success on the task in the learning phase, especially in the Experiment 2 in which participants only had to recognize whether the two faces are from the same or different category, without specifying which category it is. It was therefore the stimuli that differed in their behavioural 'meaning' that were intended to require increased pattern separation. It is not clear how this could affect pattern separation, and if separation of categories at all depends on this hippocampal process and not, for example, on the MTL cortex which extract similarity-related information for category building (Huffman & Stark, 2014).

It has also been argued that pattern separation and other hippocampal mechanisms do not operate on representations of individual items at all (Horner et al., 2012). Rats with a lesioned DG were only found impaired at item-context but not item memory (Dees & Kesner, 2013). However, the already mentioned paradigms using perceptually similar lures do suggest that pattern separation may also operate on representations without any associative aspect (Bakker et al., 2008; Kirwan et al., 2012; Kirwan & Stark, 2007; Lacy et al., 2011; Stark et al., 2013; Yassa et al., 2010).

A study with a relatively similar task to the one used in the current studies also found no evidence for pattern separation based on a behavioural need for discrimination (Favila, Chanales, & Kuhl, 2016). Participants were presented with pairs of highly similar stimuli (e.g. pictures of two similar bridges), which could be associated with either the same face, two different faces, or nothing. Like in the current experiments, it was assumed that only the representations of pairs linked to the two different faces would need to be pattern-separated. They found increased separation of patterns in the hippocampus as a result of pairing with facial stimuli (as compared to the no-face condition) but there was no difference between pairs associated with same or different faces. The authors argue that the fact that differentiation occurred also for same-outcome stimuli could be explained if presentation of one such stimulus would lead to a reactivation of the other, causing the hippocampus to resolve this competition by eliminating overlapping features, making the representations more distinct as a result. It is not clear whether this could explain the current results, especially as here groups consisted of more than two stimuli.

The second assumption for our prediction was that higher pattern separation of representations of the original faces would affect the ability to correctly recognize morphs as different to those faces. However, it is only speculative whether more distinct patterns would affect the intermediate stages between the learnt stimuli.

The proposal that memories are stored as discrete attractors (Marr, 1971; Hopfield, 1982; Wills, Lever, Cacucci, Burgess, & Keefe, 2005) implies that when moving in between the representations, at some point the network will abruptly change from one attractor to another. On the other hand, using MVPA Bonnici et al. (2012) found that when making a choice of whether the stimulus corresponds to one scene or another, the neural representations of morphs of these scenes were similar to other morphs but not to the original scenes, suggesting that another 'morph' representation emerged in the hippocampus. The authors propose that this argues against the idea of discrete attractors and instead suggests that the ensembles in the hippocampus can occupy intermediate configurations, although they do not provide any evidence that these representations are in fact 'intermediate' (and not fully distinct from the originals). In any case, the current task did not require a binary choice between the two original faces but rather relied on the ability to 'resist' the pull of the attractor of the learnt representation in order to recognize a lure as novel, and so the requirements of the two tasks make them difficult to compare.

If morphs are represented as intermediate network configuration, as Bonnici et al. suggest, the similarity of the original representations would likely play a role in how the morphs are perceived. However, it would still be unclear how this could affect the recognition of morphs. On one side, 'stronger' (more distinct) representations of the original faces may improve the ability to recognize morphs as novel because these representations would be more defined and so easier to distinguish from similar but non-identical ones – as was the prediction of the current study. On the other hand, however, defined representations may decrease accuracy on morph trials because of the strong 'pull' towards the original attractor state, which would result in pattern completion to that representations even at lower levels of similarity. However, the experiments presented here did not find evidence for either of these two possibilities as both same- and different-category pairs showed the same profile of responses (although again, the floor effect might have masked any differences between the two conditions).

The very high range of learning blocks that was needed by the participants to achieve the appropriate level of performance was also very high, suggesting very pronounced individual differences in memory or in motivation to perform well. Although inter-individual variation in memory is to be expected, the fact that the participants who took the longest to learn were also the ones most exposed to the experimental stimuli likely affected the pattern of results in unpredictable way. Additionally, the stimuli were equidistant in terms of the underlying principal components but not necessarily in terms of perceptual characteristics. It is possible that some were more readily recognised as members of friend or foe categories; as the order of presentation was random, it could have introduced bias for some participants. A pseudo-random presentation could have alleviated this limitation.

Although this task was designed to address some of the limitations of previously used tasks, some of its aspects may still have been suboptimal. For example, although associative information was used at encoding in order to engage hippocampal processing, at retrieval this information was not explicitly used. This, retrospectively, goes against the original intention to use a 'hippocampal' task as opposed to one that only relies on recognition memory and which may be supported by another part of the memory system. The retrieval task involved a recognition judgement which may be supported by the MTL cortex rather than the hippocampus (Norman & O'Reilly, 2003)

Additionally, I intended to separate the encoding from retrieval in order to maximise the probability that the correct mechanism (pattern separation or completion) will be engaged (Liu et al., 2016). However, because none of the 'new' faces were fully unique, it is possible that the continuous exposure to morphs over retrieval trials somehow have changed the original representations of the memorized faces over the course of retrieval blocks. This might have affected the strength or precise features of the learnt representations. This was made even more likely by the fact that the morphs were presented much more frequently than the original faces. This may explain why the performance on the original faces was much lower than 100% despite the fact that learning continued until a consistently high level of performance was achieved and participants were assumed to have memorized the faces.

4.4.1 Conclusions

Overall, it is unclear whether the results reflected incorrect assumptions about the nature of pattern separation of different-category stimuli and how it affects categorization of intermediate stimuli, or whether the task was inappropriate and incapable of eliciting behaviours consistent with these mechanisms, either because of the nature of the stimuli or the floor effect in performance. In the next study in this thesis, I will present another task with a new design that addressed some of the issues described in this discussion.

Chapter 5

Behavioural evidence for pattern separation in human episodic memory

5.1 Introduction

In this chapter I took a different approach to look at how pattern separation is manifested in behaviour. I adapted a task used previously to show evidence for pattern completion (Bisby, Horner, Bush, & Burgess, 2018; Horner et al., 2015; Horner & Burgess, 2013, 2014; Trinkler, King, Spiers, & Burgess, 2006). This task is particularly suitable for investigating episodic memory as it makes use of the associative nature of hippocampal representations to look at the underlying mechanisms of memory formation. It uses the coherence of within-event associations as an index of pattern completion, a measure which I adapted to look for evidence of pattern separation of representations of similar events, as will be explained in more detail below.

The flexibility of using the same measure to assess both mechanisms was made possible by the complementary and opposing roles of pattern completion and separation; while the elements of each event should be bound together into coherent narratives, the patterns associated with different events should be sufficiently different to prevent interference.

As already explained in detail in Chapter 1, the binding of event elements into a single representation (i.e. an event engram; Tulving 1985) is thought to be performed by the hippocampus, which acts as a convergence zone, binding items and their context together (Cohen & Eichenbaum, 1993; Davachi, 2006; Eichenbaum et al., 2007; O'Keefe & Nadel, 1978). This mechanism underlies the holistic nature of episodic memory; when retrieving one aspect of a memory, other elements tend to be recalled as well. It is proposed that through the process of pattern completion to a previously stored event representation, all elements of an event can be retrieved from a partial cue, such as one element of that event (Marr, 1971).

The dependency measure mentioned above explored the holistic nature of episodic memories. The properties of memory engrams imply that performance in retrievals of their constituent elements should be statistically more dependent than for retrievals of elements from different episodes. In a series of studies, Horner and colleagues tested memory for imagined person-object-location events. At recall, participants were cued with one element from the event (e.g. a person) and asked to recall either the location or the object associated with it. The probability of retrieving any given association was found to depend on the probability of retrieving other associations from the same event, suggesting that the memories for the events were stored as coherent wholes and not a series of independent pairwise associations (Horner and Burgess, 2013, 2014; Horner et al., 2015; Bisby et al., 2018). At retrieval, a reinstatement of cortical activity associated with all elements was also observed, including that of the item which did not have to be recalled on that trial, and the degree of reinstatement was correlated with the hippocampal activity (Horner et al., 2015), which is consistent with the pattern completion account.

Here I adapted this measure to assess how similar events are encoded. As in the original studies, in the current study participants were presented with three-element episodes (person, location and object). The episodes were either unique (unrelated to any other episodes) or had one element in common with one other event (e.g. the same person appeared in two events). This created pairs of overlapping events of the type assumed to require successful pattern separation in order not to interfere with each other. For instance, the same person could be associated with a different set of locations and objects – this is akin to everyday memories where some elements can be common to more than one event.

As the role of pattern separation is to reduce the interference of overlapping memories by making their representations less similar, I hypothesized that the statistical dependency in retrieval of elements from these overlapping memories will be even lower than for non-overlapping (unrelated) memories. Because representations of unrelated memories already produce sufficiently different neural patterns, their associated patterns do not need to be actively orthogonalized in this way (Leutgeb et al., 2007). Therefore, I hypothesized that (i) the recall of associations within each event should be highly dependent, as shown in the previous studies (Horner and Burgess, 2013, 2014; Horner et al., 2015) due to the holistic representation of episodic memories; (ii) the recall of associations from unrelated events should be independent; and (iii) the recall of associations from paired overlapping events should show lower dependency than the unrelated events, due to the presence of active pattern separation of similar events.

5.2 Methods

5.2.1 Participants

39 participants were recruited from the University College London student population. Seven participants were excluded due to poor overall task accuracy (below 25%) and two due to too high accuracy (above 90%). The exclusion criteria was necessary as with very good or very poor performance, the low variability in accuracy made the calculations of the dependency unreliable. The remaining 30 participants (25 females, 5 males) had a mean age of 25.61 (*SD* = 6.22), had normal or corrected-to-normal vision, were fluent English speakers and were familiar with Western culture including major celebrities and politicians (self-reported). The study was approved by the UCL Research Ethics Committee and all participants provided written informed consent prior to taking part in the study. Following completion of the test, participants were debriefed and paid for their time.

5.2.2 Materials

Word stimuli included 40 famous people (e.g. Tom Cruise, Barack Obama), 40 locations (e.g. supermarket, kitchen) and 40 objects (e.g. necklace, bottle). Stimuli were combined to create 45 three-element events with each event consisting of a person, location and object. Fifteen of these events were unique from each other in that they shared no common elements (i.e. non-overlapping events). The remaining 30 events were combined to create 15 pairs of 'overlapping' events which shared one common element (Figure 5.1A). For overlapping events, the common element (person, location or object) was counterbalanced across event pairs. Novel randomized events were created for each participant. The stimuli for the task were presented via the Cogent 2000 toolbox in Matlab.

5.2.3 Procedure

The experiment consisted of a study phase and a test phase. During the study phase, participants encoded a total of 45 event triplets with each comprising a person, location and object. Each trial began with a 0.5 seconds fixation period, after which participants were presented with one of the event triplets (Figure 5.1B). All three elements remained on the screen for 8 seconds and participants were instructed to imagine all event elements interacting together as vividly as possible. The screen location of each element type (person, location, object) was randomized across encoding trials. The trial ended with a blank screen presented for 1.5 seconds. Events were only shown once during the study phase. The pairs of overlapping events had at least five trials gap between them.

At test, associative accuracy was assessed using a six alternative forced choice test. On each trial, following a 0.5 seconds fixation period, participants were presented with one of the previously encoded items at the top of the screen (cue) and six possible 'target' items were presented below. They were then instructed to decide which of the targets had previously appeared in the same event as the cue. All targets on a single test trial were of the same category. For example, if cueing with the person to retrieve the associated location, all six options were of locations. Participants were given 8 seconds to select their option via button press. All possible cue-target pairs were tested in both directions (e.g., cue with the person to retrieve the location, cue with the location to retrieve the person), giving a total of 270 trials. The order of trials was randomized. Each trial terminated with a 1.5 seconds blank screen.

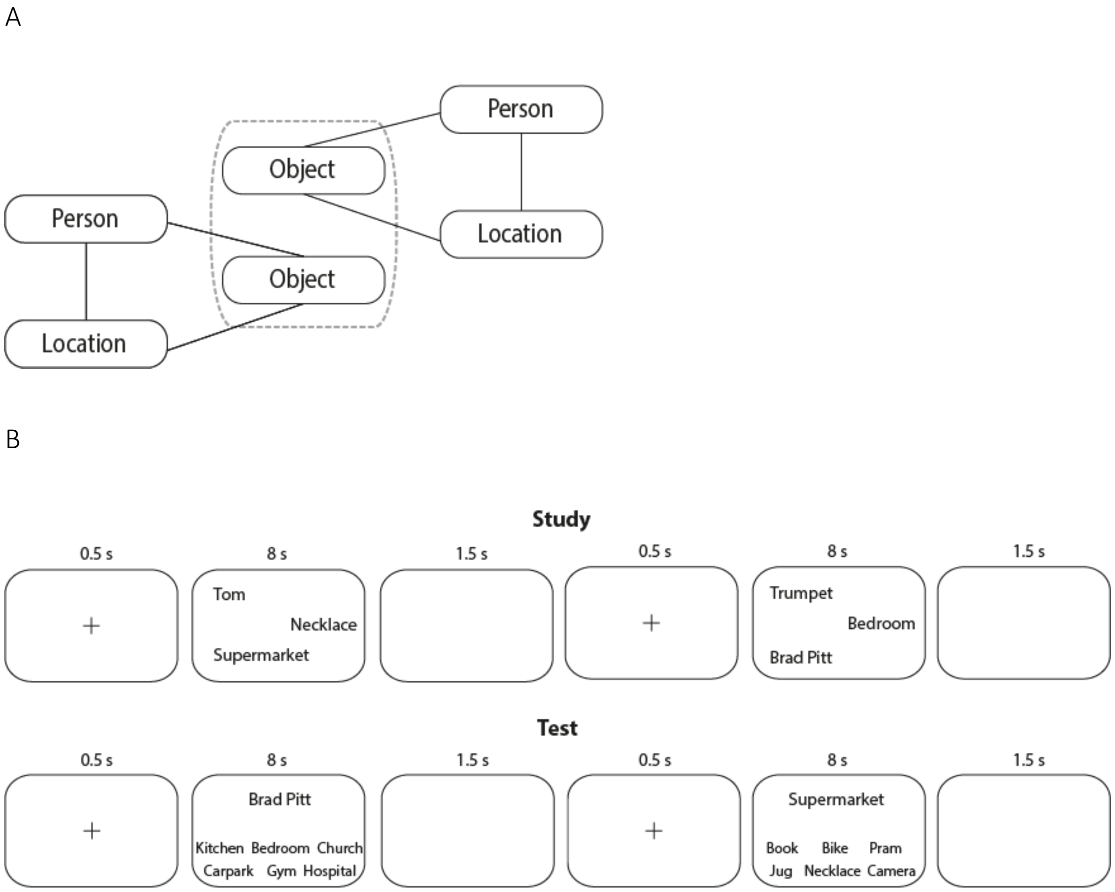


Figure 5.1. Experimental stimuli and an example of a trial sequence. Experimental stimuli: an example of the associative structure of two matched, overlapping events where the object is the common element (i.e. the same object appears in both events). Non-overlapping events consisted of the same elements but shared no elements with other events. (B) Example of the trial sequence and timing for the study phase (top) and the test phase (bottom).

5.3 Analyses

5.3.1 Analysis of within-event dependency

I assessed the statistical dependency in retrieval of different associations from the same event, for instance the probability of retrieving both a person and a location when cued by an object (see Horner & Burgess, 2013; 2014).

The dependency measure reflected how often these associations are retrieved either both correctly or both incorrectly. I created 2 x 2 contingency tables for each participant. For within-event dependency, the contingency tables were based on (i) the probability of retrieving two items from the same event when cued by the remaining item from that event (AbAc; e.g. retrieving either a person 'b' or a location 'c' when cued by an object 'A'), and (ii) the probability of retrieving an item when cued by the two remaining items from the event (BaCa; e.g. retrieving a person 'a' when cued either by a location 'B' or an object 'C').

This measure was then compared to an independent model and a dependent model calculated individually for each participant. The models estimated the level of dependency based on the average performance and level of guessing. The independent model assumed that the retrieval of any two elements from the same event is completely independent and is contingent only on the overall accuracy level. It was calculated by multiplying the probabilities of separately retrieving two elements from the categories in question.

The dependent model additionally adjusted the predicted level of dependency by an event-specific 'episodic factor' – a measure of the average performance on a given event (across all retrieval trials for that event) relative to the overall performance. The probability of retrieving any association from an event was weighted by the episodic factor for that event. The dependent model also accounted for the level of guessing; the episodic factor affected the probability of intentional retrieval but not the probability of correct guessing, which is assumed to be independent.

For each participant, I calculated the measure of dependency as well as the independent and dependent models separately for events of non-overlapping and overlapping types. This was to check whether both types are retrieved as complete events (i.e. both show high within-event dependency).

Table 5.1. Contingency tables for independent and dependent models giving the frequency (over events) of the four combinations of correct or incorrect retrievals of elements B and C when cued with element A.

	Retrieval of element (B)	
Retrieval of element (C)	Correct (P_{AB})	Incorrect (1-P_{AB})
Independent Model		
Correct (P_{AC})	$\sum_{i=1}^{N} P_{AB}P_{AC}$	$\sum_{i=1}^{N} P_{AC}(1 - P_{AB})$
Incorrect (1-P_{AC})	$\sum_{i=1}^{N} P_{AB}(1 - P_{AC})$	$\sum_{i=1}^{N} (1 - P_{AB})(1 - P_{AC})$
Dependent Model		
Correct (P_{AC})	$\sum_{i=1}^{N} \acute{P}^{i}_{AB}\acute{P}^{i}_{AC}$	$\sum_{i=1}^{N} \acute{P}^{i}_{AC}(1 - \acute{P}^{i}_{AB})$
Incorrect (1-P_{AC})	$\sum_{i=1}^{N} \acute{P}^{i}_{AB}(1 - \acute{P}^{i}_{AC})$	$\sum_{i=1}^{N} (1 - \acute{P}^{i}_{AB})(1 - \acute{P}^{i}_{AC})$

Dependent model replaces the probability of correctly recalling B when cued with A (across all events; P_{AB}) with $\acute{P}^{i}_{AB} = E^{i}_{AB}(P_{AB} - P_G/c) + P_G/c$ where the episodic factor E^{i}_{AB} reflects performance on event *i* relative to other events (based on retrievals other than B and C cued by A), P_G is the probability of guessing, and $c = 6$ is the number of choices in the test trial. P_{AC} is replaced similarly. The dependency model equates to the independent model if the episodic factors are set to 1.

5.3.2 Analysis of across-events dependency

The contingency tables for dependencies across pairs of overlapping events were based on the probability of correctly or incorrectly retrieving associations from both events. Unlike in the 'standard' within-event dependency described above, where both compared associations (e.g. Ab and Ac) came from the same event, each association came from a different event from a pair (event_1 and event_2). Here the contingency tables were based on (i) how the probability of retrieving a specific association depends on the probability of retrieving the same type of association from the corresponding paired event (Ab_1Ab_2), (ii) how the retrieval of an item from one event depends on the retrieval of a different-type item from the corresponding event when cued by the same-type items (Ab_1Ac_2), and (iii) the probability of retrieval of the same-type items when cued by different-type items in each event (Ba_1Ca_2).

Importantly, for the calculation of the dependency measure and models I excluded the trials in which the overlapping element was a cue, as these types of trials had two possible correct answers, and potentially a retrieval of the incorrect pair could interfere with the retrieval of the appropriate item.

All analyses were performed using custom written scripts in Matlab (Mathworks).

5.4 Results

5.4.1 Associative accuracy

Overall accuracy across all trials was good (66.42%, *SD* = 17.66) and well above chance (chance would be 16.7% given the six test options). Analysis of performance across cue-type (collapsed across retrieval-type) using a 2 x 3 repeated-measures ANOVA with factors of event type (overlapping, non-overlapping) and cue-type (person, location, object) showed a significant main effect of event type ($F(1,29) = 13.41$, $p = .001$, $\eta^2 = 0.32$) with slightly higher accuracy in overlapping events. There was no significant main effect of cue type ($F(2, 58) = 2.56$, $p = .09$, $\eta^2 = 0.08$) nor a cue-type x event type interaction ($F(2,58) = 0.07$, $p = .93$, $\eta^2 < 0.01$). A similar 2 x 3 ANOVA on retrieval-type (collapsed across cue-type) showed no main effect of retrieved type ($F(2,58) = 2.31$, $p = .11$, $\eta^2 = 0.07$). The main effect of event was the same as in the cue-type analysis. There was no interaction between retrieved type and event type ($F(2,58) = 1.96$, $p = .15$, $\eta^2 = 0.06$) (Table 5.2).

Table 5.2. Proportion correct (SD) for associative memory performance across non-overlapping and overlapping events for each cue and retrieval type.

	Retrieval type					
	Person		Location		Object	
Cue type	Non-Overlap	Overlap	Non-Overlap	Overlap	Non-Overlap	Overlap
Person	n/a		.63 (.17)	.68 (.21)	.65 (.15)	.70 (.20)
Location	.64 (.19)	.68 (.22)	n/a		.59 (.19)	.66 (.23)
Object	.62 (.18)	.72 (.22)	.62 (.20)	.64 (.24)	n/a	

5.4.2 Interference

First, I looked at whether performance on the overlapping pairs of events is related in order to see whether memory for one of the events interferes with the memory for the other. Negative correlation would be a sign of interference. I calculated a correlation between performance accuracy on retrieval trials of overlapping events and averaged them across all pairs. I then compared the average correlation scores across all participants against zero. The average correlation was negative, consistent with interference (r = -0.10), however the trend did not reach significance, t(29) = 1.83, p = .078, d = 0.33. The average correlations for overlapping events were, however, significantly lower than for unrelated events, t(29) = 2.09, p = .045, d = 0.38. The average correlation between unrelated events were close to zero (r = .01), t(29) = 1.46, p = 0.156, d = 0.27.

To investigate whether the order of presentation of the overlapping events affects accuracy, I compared the average accuracy between first and second event for the overlapping pairs of events. I found no difference in performance, t(29) = 1.37, p = .182, d = 0.25. The dependency results described below therefore cannot be attributed to effects of prospective or retrospective interference of overlapping memories on retrieval performance.

5.4.3 Within-events dependency

As described above, dependency was assessed by constructing contingency tables for retrieving two elements when cued with the third element, and retrieving one element when cued by the other two elements across retrieval trials. I then calculated within-event dependency (D) in the data by taking the proportion of events where elements were both correctly or incorrectly retrieved. This dependency was then compared to the amount of dependency predicted if retrievals from the same event were completely independent (Di) or dependent (Dd; see Methods for more information on how the models were constructed). I compared dependency in the data with both independent and dependent models separately for non-overlapping and overlapping events. A 2 x 3 repeated-measures ANOVA with factors of event type (non-overlapping, overlapping) and dependency measure (D, Di, Dd) showed a significant interaction between event type and dependency ($F(1.52,9.82) = 9.82$, $p = .001$, $\eta^2 = 0.25$; Greenhouse-Geisser corrected).

To further assess the interaction, I first analysed overlapping and non-overlapping events separately, comparing the data to independent and dependent models. For non-overlapping events, I found greater within-event dependency in the data compared to the independent model (D>Di, $t(29) = 7.41$, $p < .001$, $d = 1.33$) and lower than the dependent model (D<Dd, $t(29) = 4.06$, $p < .001$, $d = 0.75$). Similarly for overlapping events, dependency in the data was greater than the independent model (D>Di, $t(29) = 9.00$, $p < .001$, $d = 1.66$) and was lower than the dependent model (D<Dd, $t(29) = 10.46$, $p < .001$, $d = 1.67$) (Figure 5.2).

I next performed a direct comparison of dependency between non-overlapping and overlapping events. There was no difference between non-overlapping and overlapping events in the amount of dependency relative to the independent model (D-Di, $t(29) = 1.87$, $p = .077$, $d = 0.34$). A comparison of dependency relative to the dependent model showed a significant difference between events (D-Dd, $t(29) = 3.04$, $p = .005$, $d = 0.55$) with less dependency in the data compared to the dependent model for overlapping events. Importantly, both non-overlapping and overlapping events showed greater dependency in the data than predicted by the independent model.

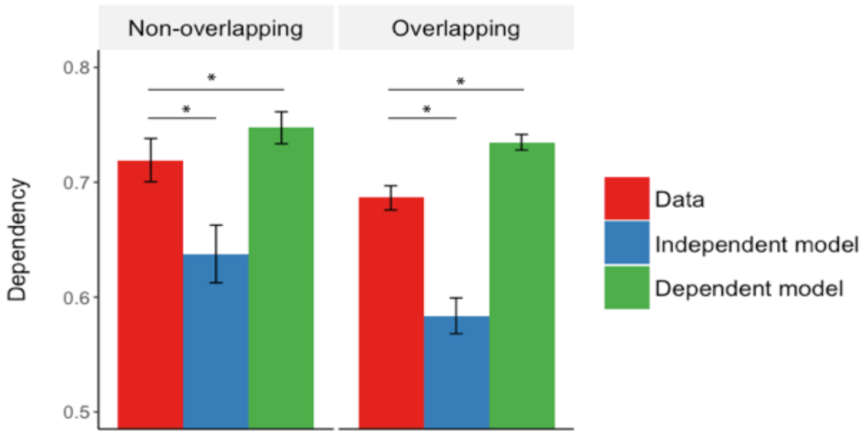


*Figure 5.2. Mean dependency for the observed data, the independent model and the dependent model for the within-event analysis. The error bars represent +/- 1 SEM. * p < .001*

5.4.4 Across-events dependency

As I have proposed, successful pattern separation between events that share a common element would likely reduce dependency due to the decorrelation of neural representations of those events. To examine this prediction, I examined the amount of dependency across matched overlapping events that shared a common element (i.e. whether retrieval success of an element from one event is dependent on the retrieval success of an element from an overlapping event) and compared this to the amount of dependency across non-overlapping events.

I calculated dependency in the data and corresponding independent and dependent models across pairs of matched, overlapping events (see Methods for how this was achieved).

First, I looked at whether dependency in retrievals from overlapping events differed from non-overlapping events. I ran a 2 x 3 repeated-measures ANOVA with factors event type (non-overlapping, overlapping) and dependency measure (D, Di, Dd). It is important to note that the measures used here are the within-event dependency measures for the non-overlapping events and the across-events dependency measures for the overlapping events. As expected I found a significant interaction (*F*(1.46,42.24) = 37.91, *p* < .001, η^2 = 0.57, Greenhouse-Geisser corrected).

To test the hypothesis that the overlap leads to a lowered dependency across matched events, I compared the amount of across-event dependency to its independent model for overlapping, matched events. I tested whether the retrieval of associations from one event depends on the retrieval of associations from its matched event. As predicted, this across-events dependency was lower than the independent model (D<Di, *t*(29) = 2.70, *p* = .011, *d* = 0.48), and lower than the dependent model (D<Dd, *t*(29) = 6.48, *p* < .001, *d* = 1.17). This pattern is different to the one observed in the within-events dependency comparisons where the dependency in data was higher than predicted by the independent model. The lower than baseline (independent model) dependency supports the hypothesis that the representations of similar events are actively separated.

Across participants, the level of relative dependency also was highly correlated with the interference measure, i.e. the correlation between accuracy on overlapping pairs of events (*r* = .80, *p* < .001).

Next, I directly compared the dependency between non-overlapping and overlapping events. The level of dependency relative to the independent model was lower for the overlapping events (D-Di, *t*(29) = 6.90, *p* < .001, *d* = 1.26). A comparison of dependency relative to the dependent model also showed a significant difference between events (D-Dd, *t*(29) = 4.66, *p* < .001, *d* = 0.85), with lower dependency as compared to the dependent model in the overlapping events.

Since I didn’t find any evidence that accuracy on the pairs of overlapping events depends on the order in which they are shown (see comparison between first and second event above), the lowered dependency reported here cannot be explained by effects of pro- or retrospective interference on retrieval performance for the overlapping representations. In other words, the dependency results cannot be simply attributed to the loss of the representation of one event as a result of subsequent encoding of its overlapping pair.

5.4.5 Comparison of across-events relative dependencies

To directly compare dependency across conditions, I first calculated a difference score between each participant's dependency and the independent model (D - Di). This reflected a *relative dependency* level; positive scores indicated higher dependency level than what was predicted by the average performance taken as independent, and vice versa.

Crucially, I wanted to see whether the relative dependency score for the overlapping pairs of events differed from unrelated pairs. If matched events are stored more independently than what would be expected from unrelated memories, this would provide evidence for pattern separation of overlapping episodic memories. To test this, I randomly paired unrelated events (separately for events of non-overlapping and overlapping type). Unrelated events of overlapping type were the events that did have a matched pair, but here were paired up with another (unrelated) event to calculate the across-events dependency. For example, if event_1 overlapped with event_2, in this analysis it could be randomly paired with any other event from the set of overlapping paired events (e.g. event_3) but not with the event_2. This created pairs of events that were of unrelated, overlapping type with no shared elements (in contrast to pairs of matched, overlapping events which shared one common element).

I then calculated the across-events dependency measures for these pairs in the same way as for the matched pairs of events (see above). This was to ensure that the lowered dependency is not a result of the event type, but rather that it is specific to pairs of matched events with a common element. In other words, I wanted to show that it is specifically the need for active separation of two related episodes that leads to their 'negative' relative dependency.

As each analysis used a randomly created set of unrelated event pairs, the specific pairs selected in any given analysis could have affected the results. I therefore ran each analysis a 1,000 times and used a Brown's method (i.e. an extension to the Fisher's combined probability test; Brown 1975; Poole et al. 2016) which provides a combined p-value for non-independent tests.

Each analysis compared the relative dependency in the matched, overlapping pairs (which was the same in all 1,000 analyses, as these pairs were not randomly selected) to the relative dependency in the same number of unrelated pairs of events, with different unrelated pairs selected each time.

As would be predicted by a pattern separation account, the matched overlapping events showed lower relative dependency than unrelated events of non-overlapping type, $p < .008$, and of overlapping type, $p < .006$ (Figure 5.3). The two types of unrelated events were not significantly different from each other, $p = .532$, and neither was significantly different from their baseline, i.e. the independent model ($p = .497$ for unrelated non-overlapping events and $p = .475$ for unrelated overlapping events; Figure 5.3).

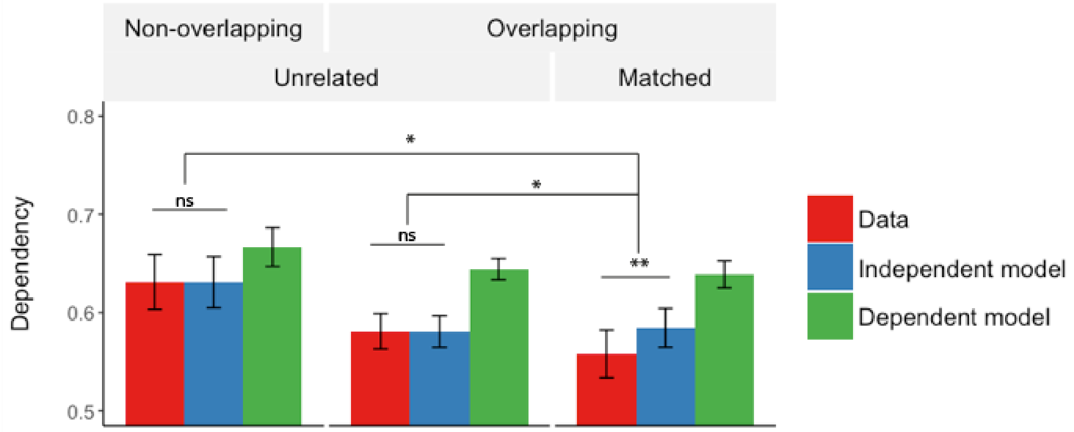


Figure 5.3. Mean across-event dependency for pairs of unrelated events (first set of three bars: unrelated, non-overlapping events; second set of three bars: unrelated, overlapping events) and pairs of matched, overlapping events (third set of three bars). The data for unrelated pairs was obtained by randomly selecting pairs of unrelated events (separately for overlapping and non-overlapping events) and calculating the dependency and the models, and repeating this procedure 1,000 times, each time with a different set of random pairs. The p-values were obtained using Brown's method for combining tests of significance (Brown, 1975). The bars represent the mean relative dependency, and the error bars represent the mean standard error across the 1,000 tests (unrelated events) or +/- 1 standard error (matched events). $^{*} p < .01$; $^{**} p < .001$.

5.5 Discussion

Computational models of hippocampal function propose that similar memories are stored as distinct non-overlapping representations through a process of pattern separation (e.g. Chavlis & Poirazi, 2017; McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & McClelland, 1994; Rolls, 2013). To test this hypothesis, I examined whether retrievals across events that share a common element would be represented as more distinct. The lower correlation between average accuracy for pairs of overlapping events than for unrelated events suggested some level of interference. A closer look at the data revealed that this interference effect may be underlay by the need to separate the representations of similar events on one hand, and binding together of their elements on the other. I show that, whilst retrievals from within the same event show dependency, retrievals from overlapping events were 'negatively' dependent, i.e. they showed even lower dependency than would be expected from unrelated events. This supports the proposal that pattern separation serves to decorrelate similar inputs by transforming their representations into orthogonal patterns (e.g. McClelland et al., 1995; O'Reilly & McClelland, 1994; Treves & Rolls, 1992).

Our results are consistent with the view that events are stored as single coherent representations and retrieved in a holistic manner. In accordance with previous studies, I found greater dependency in the retrieval of elements from the same event (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2013, 2014), supporting the holistic way in which episodic memories are stored and retrieved (Tulving, 1983). I assume that within-event dependency reflects the associative structure in which event elements are bound together as a single representation. When cued by a single event element all associated within-event elements would be reinstated. This is consistent with computation models of hippocampal pattern completion and the way in which event elements are stored in an autoassociative network in which presentation of a partial cue will cause reinstatement of all associated event elements (Hopfield, 1982; Marr, 1971; McClelland et al., 1995). Neuroimaging evidence, using a similar task and dependency measure as I used here, is also complementary to our results in suggesting that within-event elements are reinstated in neocortical areas and this reinstatement is supported by the hippocampus (Horner et al., 2015).

Within-event dependency was also seen in events even when they overlapped with other events (i.e. shared a common element). This finding is important as it suggests that pattern separation affects specifically the representations that may interfere with each other due to their similarity (i.e. representations of overlapping events), while the constituent elements of these events, which needs to be bound into coherent narratives, are not separated but instead show pattern completion. This is consistent with findings from rats which showed that the hippocampus binds information encountered in the same context together while separating events from different contexts (McKenzie et al., 2014).

This supports the proposed role of the different hippocampal mechanisms in storage of episodic memories (Marr, 1971; McClelland et al., 1995). The autoassociative network of CA3 is proposed to store learnt associations in its recurrent connections, which allows for linking all elements of a memory together and for their reinstatement through pattern completion at retrieval. When the events to be remembered are similar, the dentate gyrus decorrelates their representations by 'selecting' a different population of CA3 neurons for their storage (Rolls, 2013; Treves et al., 2008). In this way, the overlapping representations are pattern separated while their constituent elements are bound together. This shows the complementary functions of pattern separation and completion in supporting episodic memory.

Because similar memories are more likely to interfere with each other, it may be that similar memories will have less similar representations than unrelated memories do on average. As the neural patterns associated with different stored memories in an autoassociative network become more similar, the retrieval accuracy falls (Hopfield, 1982; Treves & Rolls, 1992), and so there is a need for a mechanism that actively decorrelates similar inputs to avoid subsequent interference, driving their respective representations further apart.

This is what may underlie the differences in across-event dependency, the extent to which remembering an association from one event depends on the retrieval of associations from another event. If events are represented as separate engrams, we would expect no dependency in the retrieval of their respective associations. This prediction was reflected in the independent models which assumed that the amount of dependency corresponds simply to the average probability of correctly recalling any association, based on performance across trials. For unrelated pairs of events this was indeed found to be the case, suggesting their storage as separate memory engrams.

Interestingly and crucially for our hypothesis, retrievals across matched overlapping events (i.e. events with one element in common) showed a level of dependency that was significantly lower than predicted by the independent model. This may be because the hippocampal representations of these events are actually less similar to each other than memory representations are on average due to the need to actively reduce their similarity and, consequently, reduce their interference. Representations reflecting commonalities across stimuli may be stored elsewhere (e.g. the MTL; McClelland, McNaughton and O'Reilly, 1995), but the current task was designed to tap specifically into the hippocampal type of processing (see below). The role of pattern separation may be therefore to accentuate the differences between similar input patterns, as distinct memories already produce sufficiently different output and do not suffer from interference (Lacy et al., 2011; Leutgeb et al., 2007; Vazdarjanova & Guzowski, 2004). This leads to even lower dependency than what is observed across independent events and what is predicted by the independent model.

For unrelated pairs of events, the results were the same regardless of whether these events overlapped with another event or not – i.e., the across-event dependency was the same for unrelated, non-overlapping events and for unrelated, overlapping events. Both showed higher dependency than the matched events, and their dependency was no different to the independent model. The finding that the dependency was 'negative' only for the matched, overlapping pairs but not for unrelated, overlapping events suggests that the results are not simply due to the fact that events that are not fully unique are processed in a different way; rather, the increased separation occurs only for the specific pairs of events which share an element in common and not for all events of this kind.

It is noteworthy that the lowered dependency across overlapping events in the current study resembles the pattern of results consistently observed (although never reaching significance) in within-event associations of open-loop events in previous studies using the original paradigm adapted here (Horner et al., 2015; Horner & Burgess, 2014). In these studies, closed-loop events were identical to the events used in the current study; i.e., all within-event elements were linked to one another. In the open-loop events, on the other hand, the chain of associations was never 'closed': e.g. person was associated with a location, location with an object, and object with an animal, but participants never learnt the person-animal association which would 'close the loop'. The current results suggest that possibly the lowered dependency observed in the open-loop events may be explained by pattern separation of overlapping but, crucially, unrelated associations (i.e. associations that were not bound together into a holistic episodic representation).

Our paradigm extends upon research examining pattern separation processes in humans by using a task that would engage the hippocampus, and focuses on the nature of the stored representations of similar or unrelated event memories. Previous studies providing evidence for behavioural pattern separation have used tasks that require participants to make discrimination judgments on whether items were previously seen or similar but not previously seen (Bakker et al., 2008; Lacy et al., 2011). In these studies, successful pattern separation is inferred from being able to mnemonically distinguish a previously seen item from a similar but non-identical lure. The observed brain activity during those tasks corresponds to the prediction of the pattern separation account. The activity in the hippocampus distinguishes between new items, repeated items, correctly recognized lures and false alarms (lures mistakenly categorized as 'old') (Bakker et al., 2008; Kirwan & Stark, 2007; Lacy et al., 2011). Additionally, patients with hippocampal damage were shown to have a spared overall recognition ability while being impaired at recognition of similar lures (Kirwan et al., 2012), again supporting the role of the hippocampus in pattern separation of similar representations.

However, arguably a task which involves an associative aspect, as the one used in our study, is more appropriate for testing of hippocampal functions. The hippocampus is proposed to be involved in contextual memory but not in memory for individual items (Horner et al., 2012), which instead may be processed in other medial temporal lobe structures (e.g. perirhinal cortex) (Eichenbaum et al., 2007; McClelland et al., 1995). The reduced similarity of patterns associated with overlapping spatial memories has been previously found using fMRI in the hippocampus (Chanales et al., 2017), however so far evidence for pattern separation of complex (non-spatial) episodic memories is lacking.

Our study therefore offers a novel way of investigating behavioural pattern separation. Another advantage of the current task is the segregation of the encoding and retrieval phases, i.e. tasks which involve the encoding of overlapping and unrelated memories are more appropriate for assessing pattern separation than ones which only consider distinguishing memories from similar lures at retrieval. Pattern separation and pattern completion operate at, respectively, encoding and retrieval, and so it has been suggested that an appropriate task should be one where the process of interest (separation or completion) is the most appropriate strategy (Hunsaker & Kesner 2013; Liu et al. 2016). The mnemonic similarity paradigms require continuous ‘recall-to-reject’ in addition to intentional encoding, making it difficult to separate the two processes, an issue recognised by the authors (Kirwan & Stark, 2007). This is problematic as the responses to new or old items in such paradigms have been shown to lead to a bias towards pattern separation or completion in following trials (Duncan et al., 2012). Our task, on the other hand, does not require any recall at encoding or encoding at retrieval, potentially making it more 'process-pure' as a result. The current task may therefore provide another way of looking at behavioural pattern separation which overcomes some of the limitations of pattern separation/completion tasks as outlined by Liu *et al.* (2016) and others (Hunsaker & Kesner, 2013; Kirwan & Stark, 2007).

Another alternative would be to use an incidental learning paradigm in order to avoid the need for a continuous recall (Liu et al., 2016). This was successfully implemented in two imaging studies (Lacy et al., 2011; LaRocque et al., 2013); however, neither study provided a behavioural measure of pattern separation and instead they were used to look for separation-related signal in the hippocampus and MTL.

Although in the current task it cannot be determined that the first episode from each pair was not recalled during the presentation of the second episode from that set, this possibility was minimized as the task did not require any recall during the study phase. Rather, participants were motivated to focus on the episode at hand to improve their memory for this specific set of associations. There was also no difference in accuracy on associations between first and second event, making it unlikely that the first event was recalled and somehow affected the encoding of the second one.

One potential alternative explanation could be that the reduction of dependency is due to the effect of recall when cued by the overlapping item; in other words, when presented with a cue originally associated with two separate events, the 'dominant' event may be recalled and block the retrieval of the correct associations. However, the dependency measure was calculated with these associations excluded and so this cannot explain the pattern of results.

It is also important to note that the evidence of behavioural pattern separation does not necessarily imply that pattern separation in the strict sense is taking place; for this it would need to be demonstrated that the neural representations associated with overlapping events diverge during encoding as information reaches the hippocampus (specifically its CA3 subfield through the DG) from more upstream areas; i.e. more similar inputs result in more dissimilar outputs.

5.5.1 Conclusions

This study presents a novel and relatively process-pure way of behaviourally investigating pattern separation. I propose that the lowered ('negative') dependency in retrieval of associations from overlapping events results from the need to differentiate their neural representations through pattern separation. This finding is specific to pairs of events with a common element and not to pairs of unrelated events (whose retrieval was found to be independent from each other), which confirms the predictions regarding how similar episodic memories are represented.

Chapter 6

Pattern separation in human episodic memory: fMRI study

6.1 Introduction

The previous chapter provided evidence that overlapping multi-element memories are represented in a way that is more distinct than unrelated memories. The lowered statistical dependency of performance across multiple retrievals from overlapping memories compared to unrelated ones is proposed to reflect pattern separation of overlapping memories at encoding. This in turn is thought to result from the need to differentiate the stored representations of similar inputs in order to avoid subsequent interference between them during retrieval (see Chapter 1.2). In this chapter, I use the same task in a fMRI study to explore the hippocampal contribution to memory for similar episodes. I look at whether the patterns of neural activity associated with different memory episodes matches the observed behaviour; i.e., whether an overlapping element leads to more distinct hippocampal representations of the two multi-element memories. Following the hypotheses from the previous chapter, it is predicted that the activity patterns for overlapping memories will be less similar than patterns for unrelated memories.

As already described in Chapter 1, rodent studies support the role of the hippocampus, and specifically the DG, in pattern separation of representations of similar locations (Gilbert et al., 2001; Leutgeb et al., 2007; McDonald & White, 1995). In humans, a DG lesion was found to lead to a selective impairment in recognition of similar items (Baker et al., 2016), providing evidence for the DG involvement in pattern separation. More support comes from neuroimaging studies, although here due to the limited spatial resolution it has not yet been possible to functionally separate the CA3 and the DG.

Studies using adaptation, where the BOLD response to a repetition of a stimulus is reduced compared to its first presentation, found evidence that a similar lure was treated as a novel item and not a repetition in the combined CA3/DG region (Bakker et al., 2008; Kirwan & Stark, 2007), even at relatively high degrees of similarity to the original item (Lacy et al., 2011; Motley & Kirwan, 2012). However, the hippocampal activity detected in these studies may reflect 'match-mismatch' processing (Kumaran & Maguire, 2007) rather than pattern separation.

Recently several studies used differences in patterns across voxels, a proxy for neural pattern similarity, to look at pattern separation. This method is particularly suitable as it allows to test the main prediction of the pattern separation account: that similar events are represented as more distinct patterns of activity.

LaRocque *et al.* (2013) showed that subsequent performance on the old/new recognition paradigm was predicted by a greater pattern dissimilarity (i.e. higher distinctiveness) between the item and other items from the same category in the hippocampus, while the opposite was true for the perirhinal and parahippocampal cortices. This suggests that successful pattern separation at encoding predicts later memory. However this, as well as previously mentioned studies, used standalone objects as stimuli, and not more complex scenes or associations whose memory is proposed to specifically require support by the hippocampus (Aggleton & Brown, 1999; Cohen & Eichenbaum, 1993; Davachi, 2006; Eichenbaum et al., 2007).

More complex stimuli have been used to show evidence of pattern separation in the spatial domain. Similar environments were found to be represented with distinct hippocampal patterns (Bonnici et al., 2012; Stokes et al., 2013), and more distinct patterns predicted later recall of the layout of the environment (Kyle, Stokes, et al., 2015). When presented with linear combinations of two previously learnt environments, the pattern of activity in the hippocampus showed an abrupt change in representation which was related to the mnemonic decision about the current environment, suggesting that the environments were separated from each other and then the activity pattern was completed to the more similar of the two stored scenes (Steemers et al., 2016). In a particularly good demonstration of pattern separation, participants learnt real-world routes which were either completely independent or included overlapping segments. The representations of the overlapping parts of the routes became less similar to each other over learning (Chanales et al., 2017). This 'reversal effect' was particularly strong for segments that were most difficult to differentiate behaviourally, as tested with a separate sample. The similarity of unrelated routes, on the other hand, remained the same. This is consistent with the role of pattern separation in reducing interference from similar memories by creating decorrelated representations.

A divergence in activity patterns for multi-element stimuli over time has also been found in the non-spatial domain. When watching a film with two interleaved stories, the patterns of activity in the hippocampus associated with people, locations and narratives belonging to the two stories became less similar (Milivojevic et al., 2016). In another study, using green-screen technology, two narratives and two background scenes were combined to create four highly overlapping movies (Chadwick, Hassabis, & Maguire, 2011). Multivariate pattern classifiers could distinguish between the movies in the hippocampus despite the overlapping content, but not in other MTL regions. These studies again suggest that hippocampal representations of different events become more distinct over time, particularly when these representations are similar and could lead to interference.

Pattern separation can therefore be detected in the hippocampus for complex memory episodes. However, so far no study has looked at whether similar episodic memories are represented with more distinct patterns of activity than unrelated memories in the same way as representations of overlapping spatial routes are more dissimilar to non-overlapping routes as shown by Chanales *et al.* (2017). Additionally, it is not clear how the separation of activity patterns associated with different multi-element episodic memories affects subsequent memory performance.

In the current study, I will use the task presented in the previous chapter to look for evidence of neural pattern separation (as proxied by dissimilarity of patterns of BOLD activity across voxels). I will look at whether the patterns of activity across voxels are more dissimilar for overlapping events than non-overlapping events in the hippocampus and other parts of the memory system, at both encoding and retrieval. I will also test whether successful pattern separation reduces interference and leads to better memory performance. Finally, I will look at whether non-target items are reinstated at retrieval (i.e. an item from the same event as the target but that is not the target of retrieval), as would be predicted by the pattern completion account.

6.2 Methods

6.2.1 Participants

33 neurologically healthy participants were recruited from the student population at the University College London. Three participants were excluded due to low performance (<20% accuracy). Compared to the study in Chapter 5, the exclusion criterion here was lowered to reflect the overall lower performance in this study (see Results below), potentially due to different experimental setting. The remaining 30 participants (10 males) had a mean age 24.43 (SD = 3.87), normal or corrected-to-normal vision and were right handed. Participants were paid £20 for the scanning session. The study was approved by the University College London Research Ethics Committee (1825/003) and informed consent was obtained before the session. Following completion of the test, participants were debriefed and paid for their time.

6.2.2 Materials

The same stimuli were used as in Chapter 5. Word stimuli included 30 famous people (e.g. Tom Cruise, Barack Obama), 30 locations (e.g. supermarket, kitchen) and 30 objects (e.g. necklace, bottle). In this study, participants selected famous people they were familiar with from a list of 60 candidates before attending the scanning session; out of all familiar to the participant people, 30 were chosen at random (all participants were familiar with at least 30 people from the full list). Stimuli were combined to create 36 three-element events with each event consisting of a person, location and object.

In the behavioural task presented in Chapter 5, the main results (comparison between overlapping and unrelated events) were the same for fully non-overlapping (fully unique) events and for pairs of unrelated events of overlapping type (i.e. ones that shared an element with a matched event but which were paired with another, unrelated one for the purpose of the analyses). In the current study, for time efficiency I therefore used only overlapping events (i.e. no event was fully unique). Each event was paired with another event with which it shared one element (person, location or object). The type of the common element was counterbalanced across event sets. Novel randomized events were created for each participant. The stimuli for the task were presented via the Cogent 2000 toolbox in Matlab.

6.2.3 Procedure

Before the scanning session, participants were familiarized with the task until they were able to understand the task. During scanning, participants lay in a supine position with the visual display reflected from a back projector by a mirror attached to the head coil. During the test phase, participants responded using three keys with their right hand and three keys with their left hand, and signalled their readiness to start each block using any button.

The main part of the procedure was the same as in the behavioural study (Chapter 5) and consisted of a study phase and a test phase. For the purpose of the analyses, before the first encoding phase, participants were presented with all stimuli words one at a time, and were asked to imagine each item (person, location or object) in turn. Each item was presented in the centre of the screen for 3 seconds, preceded by 0.5 seconds fixation cross and followed by 1.5 seconds blank screen. This 'item' phase was repeated twice, with all 90 items in each. The order of stimuli was randomised in each run.

This was followed by the study phase, where participants encoded a total of 36 events, each comprising a person, location and object. Each trial began with a 0.5 seconds fixation, after which participants were presented with one of the event triplets. All three elements remained on the screen for 0.5 seconds and participants were instructed to imagine all event elements interacting with each other as vividly as possible ("Imagine this person interacting with this object at this location"). The screen location of each element type (person, location, object) was randomized across encoding trials. The trial ended with a blank screen presented for 1.5 seconds. Each event was shown once during the study phase.

At test, associative accuracy was assessed using a six alternative forced choice test. On each trial, following a 0.5 seconds fixation cross, participants were presented with one of the previously encoded items at the top of the screen (cue) and six possible ‘target’ items were presented below in two rows (three targets per row). They were then instructed to decide which of the targets had previously appeared in the same event as the cue. All targets on a single test trial were of the same category. For example, if cueing with the person to retrieve the associated location, all six options were of locations. Participants were given 6 seconds to select their option via a button press. All possible cue-target pairs were tested in both directions (e.g., cue with the person to retrieve the location, cue with the location to retrieve the person, etc.), giving a total of 216 trials. Each trial terminated with a 0.5 seconds blank screen. The test phase was divided into two scanning blocks with 108 trials in each. Pairwise associations for each event were split so that half were tested in the first run and half in the second run. The order within each run was randomised.

For the purposes of the RSA analyses, the same stimuli as in the encoding block (i.e. all events) were presented again after the test blocks. Participants were asked to imagine the same interaction between the event elements as they did in the first encoding block. This time the three elements remained on the screen for 4 seconds, and were preceded by 0.5 seconds fixation cross and 1.5 seconds blank screen. The order of presentation and location of each item-type on the screen was again randomised. For an example of the trial sequence with timings, see Figure 6.1.

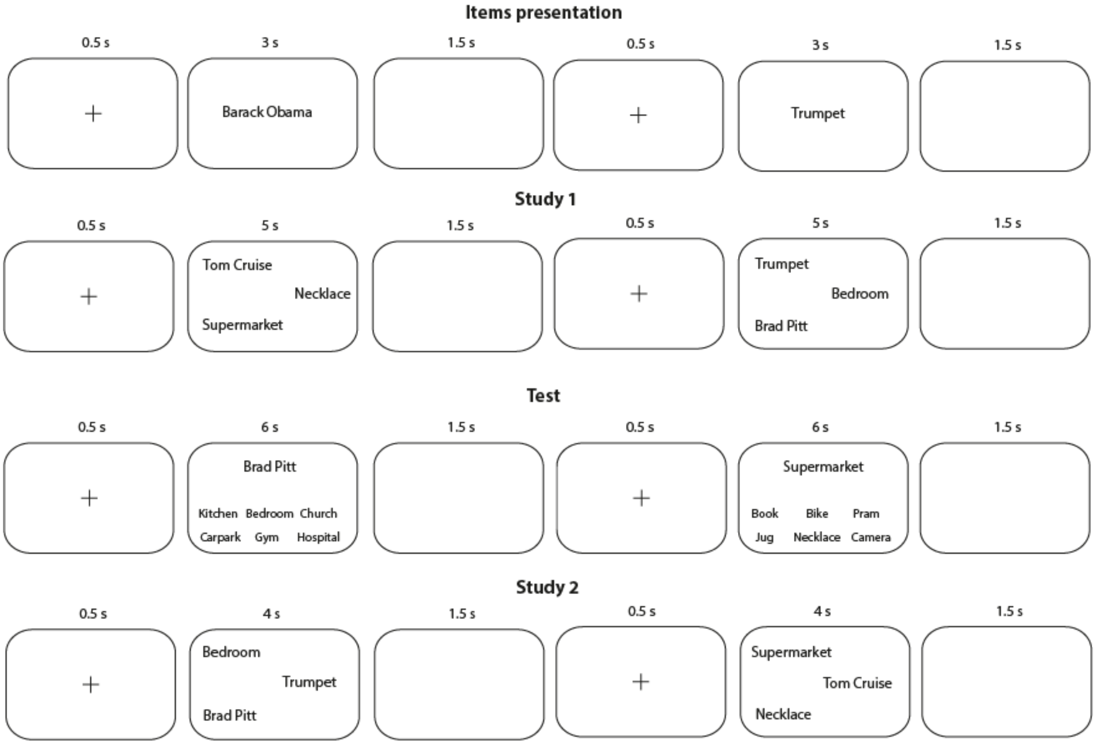


Figure 6.1.Example of the trial sequence and timing for the item presentation phase (top), the study phase (second row), the test phase (third row), and the second study phase (bottom row).

6.3 fMRI acquisition

Imaging data were acquired using echo-planar imaging (EPI) on a 3T Trio (Siemens) scanner with a 64-channel head-coil. The parameters were: repetition time (TR) = 3,360ms, echo time (TE) = 30ms, slice thickness = 2mm, resolution = 2x2mm, 48 slices per volume. A double-echo FLASH field-map for distortion correction of the EPI volumes was acquired, as well as a three-dimensional MDEFT structural image (1mm^3 resolution) for normalisation to the MNI template.

6.4 Analyses

6.4.1 Analysis of behavioural data

All analyses of behavioural data were the same as described in Methods section of Chapter 5, with the exception that here I did not use any events of non-overlapping type.

6.4.2 Analysis of fMRI data

The images were bias-corrected, unwarped, realigned to the first image, slice-time corrected and normalized using SPM12 in Matlab (www.fil.ion.ucl.uk/spm). The first five volumes were discarded to allow for T1 equilibration. In univariate analysis, normalised functional images were spatially smoothed with an isotropic 6mm FWHM Gaussian kernel.

The time series were modelled with a boxcar from the stimulus onset to end and analysed using a General Linear Model (GLM). The regressors were temporarily convolved with the hemodynamic response function. In addition to the regressors of interest, each model included six regressors of movement parameters obtained from the realignment procedure and one regressor of cardiac and respiratory signals (collected using a pulse oximeter and respiratory belt during the scanning session).

The parameter estimates from each participant from the first level analysis were included in the second level univariate analyses, or first analysed with RSA at the level of single subjects and then smoothed and used in the second level analysis in SPM.

The anatomical masks for region of interest (ROI) analyses were created using the WFU PickATlas toolbox, with each region defined by the Automated Anatomical Labelling atlas. ROIs used were: left and right hippocampus, and left and right parahippocampal cortex. Masks for areas specific to different types of items (person, location and object) were created based on functional data from univariate analyses of the item presentation phase (see below for detail).

6.4.1 Univariate analyses

Univariate analyses were performed in addition to the main RSA analyses in order to look at the relationship between overall level of activity and memory performance, as well as to define areas specific to the processing of different item types.

Item type. At item presentation, the trials were split based on the type of item (person, location or object). Each of the three item-type regressors modelled the 3-seconds item presentation period with a boxcar function, and two regressors modelled the 2.5-seconds ITI period and 0.5-seconds fixation cross. The contrasts for the three conditions were used in a second level analysis where I looked for areas where activity was associated with each item type by contrasting the ‘beta’ GLM coefficient for the regressor for one item type with those for the remaining two item types with a paired-samples t-test (e.g. person > (location + object)/2).

Subsequent memory effect. To assess subsequent memory effects, I used a regressor for all encoding trials during the first encoding phase (modelled with a boxcar function for the entire 6-seconds encoding period), and two regressors modelled the 2.5-seconds ITI period and 0.5-seconds fixation cross. The encoding regressor was parametrically modulated by the number of correct responses across all retrieval trials for this event (a value between 0 and 6). The parametric modulator was compared to zero with a one-sample t-test.

Retrieval. To look at memory effects at retrieval, I created one regressor each for correctly and incorrectly retrieved trials. A boxcar function was used to model each retrieval period (of length from trial onset to button press or 6-seconds if no response was given) and two other regressors modelled the 2.5-seconds ITI period and 0.5-seconds fixation cross. The beta coefficients (from the first level GLM analysis) for correct and incorrect retrieval regressors were compared to each other with a paired t-test.

At retrieval I also looked at whether the signal differs between trials which do and do not include the overlapping item as either cue or target. I created two regressors for each type (and ITI and fixation regressors) and contrasted their betas (GLM coefficients) with a paired-samples t-test at the second level.

Finally, I looked for the changes in BOLD signal related to the item type. I used 12 regressors, each corresponding to a particular combination of cue-target type (e.g. person-location, person-object, location-object, etc.) and two regressors modelled the 2.5-seconds ITI period and 0.5-seconds fixation cross. At the second level, I contrasted trials where a given element was a cue or target with trials where it was neither, and repeated this for all element types. This method was used in the study using the original version of this task (Horner et al., 2015) to look for representations of different item types.

6.4.2 RSA analyses

In order to look at more fine-grained hypotheses regarding memory representations, I used the RSA. In these analyses, the first-level GLM analysis was performed on unsmoothed images. Each trial was modelled as a separate regressor with a boxcar function for the length of this trial. The neural RDMs for each analysis were obtained by calculating distances between the patterns of beta coefficients for all pairs of conditions (specific conditions will be defined below) from first-level estimates, which were then spatially prewhitened and, where indicated, crossvalidated within each searchlight (see Chapter 2 for the explanation of RSA procedure). Search spheres had a two-voxel radius from the central voxel. The neural RDMs were then correlated with the appropriate model RDMs; relevant models will be explained below. Depending on the analysis, this resulted in a searchlight map of correlation values or with a single correlation value for each ROI for each participant. Searchlight maps were smoothed with a 6mm FWHM Gaussian kernel. The resulting images were compared against zero with a one-sample t-test in a second-level GLM analysis in SPM12 unless indicated otherwise. Specific conditions and comparisons are explained in the next sub-sections. All analyses were performed using custom written scripts in Matlab (Mathworks), some of which used functions provided in the RSA toolbox (Nili et al., 2014).

Pattern separation at encoding. The main analyses investigated the similarity of patterns for overlapping and unrelated events. At encoding, each encoding trial (i.e. each event) was modelled as a separate regressor. The neural RDMs were then obtained by calculating distances between the beta coefficients for each pair of events. The model RDM for this analysis predicted a higher distance for pairs of overlapping events than for pairs of non-overlapping events.

Pattern separation and interference. In order to see whether increased pattern separation, as indicated by lower similarity of patterns, led to lower interference, dissimilarity scores were calculated for each pair of overlapping events at encoding. These scores were then correlated with the average behavioural accuracy at retrieval across all twelve retrieval trials for that pair of events.

Additionally, I looked at whether overall dissimilarity to other events (regardless of overlap) predicts performance. The calculated distances between each event and all other events at encoding (i.e. one 'row' from a neural RDM) were averaged and compared to the mean performance on that event.

Changes to event representations over learning. Although in the current study learning was only required during the first encoding phase, it was still possible to look at how the representations of different events change over the course of the study by comparing the first and second encoding blocks. For each pair of events, the distances in the first encoding block were subtracted from those in the second block; positive scores meant that the distance for this pair increased. These values were then averaged separately for overlapping and non-overlapping pairs of events, resulting in two searchlight maps of average difference scores per participant. In the second-level analysis, they were compared to each other with a paired-samples t-test or to zero with a one-sample t-test.

Pattern separation at retrieval. The main hypothesis was also explored at retrieval. I looked for areas where associations from overlapping events are represented as less similar than associations from non-overlapping events. The distances were calculated between patterns of all correctly retrieved associations (maximum 216; six associations per event). Because of the high number of conditions and the size of resulting dissimilarity matrix (216x216), I limited this searchlight analysis to the hippocampal and parahippocampal cortex. The resulting neural RDMs were compared to a model which predicted higher dissimilarity for associations from overlapping events than for those from non-overlapping events. As the retrieval phase was divided into two blocks, a model which predicted higher similarity for associations within- than between blocks was used in this analysis to partial out the effects of within-block similarity. This analysis was also repeated for left and right hippocampal ROIs.

Encoding-retrieval similarity (ERS). I tested whether the similarity of patterns at encoding and retrieval can predict the success on a given trial. I used a whole-brain searchlight analysis. For each retrieval trial (i.e. for each association tested at retrieval), I correlated the pattern of activity at retrieval with the pattern of its corresponding event at encoding. I then averaged the correlation values separately for correct and incorrect trials for each event, and calculated a difference score between correct and incorrect trials. This was repeated for each searchlight, resulting in a searchlight map of difference score for each participant.

Non-target reinstatement. The holistic retrieval of episodic memories implies that all elements of an event are retrieved during recall of only a subset of its elements (e.g. one of its constituent associations). I therefore looked at whether the item from an event that is not a cue or a target (the non-target) in any given trial would also be recalled. For example, in a trial where person cues location, the non-target would be the object from the same event. In this analysis I used item-specific ROIs of person-, location, and object-related areas from the univariate analyses of item types during item presentation phase (see above). Non-target reinstatement was measured as a correlation between patterns at encoding and retrieval in its type-specific ROI. The correlations between encoding and retrieval in the appropriate ROI were calculated for each retrieval trial. A mixed-effect model (with random intercept for each subject and each event, with event effect nested within subject effect) was used to see whether the correlation values (the measure of non-target reinstatement) differed from zero, which would indicate that the non-target pattern at encoding is observed at retrieval. Additionally, I looked at whether the level of non-target reinstatement across all within-event associations can predict the mean accuracy on a given event; I compared this model to a null model which assumed no such relationship.

6.5 Behavioural results

6.5.1 Associative accuracy

Overall accuracy was slightly lower than in the behavioural study (Chapter 5) at 54.85% (*SD* = 19.43) but still well above the chance level of 16.7%. Analysis of performance across cue-types (collapsed across retrieval-types) using a one-way repeated-measures ANOVA showed no effect of cue type (*F*(2,58) = 0.03, *p* = .967, η^2= .001), and analysis of performance across retrieved-type (collapsed across cue-type) also showed no effect of retrieved-type (*F*(2,58) = 1.60, *p* = .212, η^2= .09). The means and standard deviations for accuracy across different cue and retrieval types for both conditions are in Table 6.1.

Table 6.1. Proportion correct (SD) for associative memory performance across events for each cue and retrieval type.

	Retrieval type		
Cue type	Person	Location	Object
Person	n/a	.55 (.19)	.55 (.21)
Location	.55 (.20)	n/a	.56 (.19)
Object	.55 (.20)	.54 (.21)	n/a

6.5.2 Interference

Interference was calculated by correlating performance on the overlapping pairs of events and averaging the correlation values across all pairs. I then compared the average correlation scores across all participants against zero. The average correlation was negative, consistent with interference (*r* = -0.11) and it was significantly lower than zero, *t*(29) = 2.28, *p* = .030, *d* = 0.42. The average correlations between non-overlapping events were positive but very low and showed trend towards significance (*r* = .02), *t*(29) = 2.03, *p* = 0.052, *d* = 0.37. The average correlations for overlapping events were significantly lower than for unrelated events, *t*(29) = 2.65, *p* = .013, *d* = 0.48.

To investigate whether the order of presentation of the overlapping events affects the accuracy, the average accuracy was compared between first and second event for the overlapping pairs of events. The first overlapping event in a pair was better remembered than the second event, *t*(29) = 2.37, *p* = .025, *d* = 0.43.

6.5.3 Within-events dependency

As in Chapter 5, the within-events dependency was assessed by constructing contingency tables for retrieving two elements when cued with the third element, and retrieving one element when cued by the other two elements across retrieval trials (see Methods of Chapter 5). The within-event dependency (D) in the data was calculated as the proportion of events where elements were both correctly or incorrectly retrieved. This dependency measure was then compared to the amount of dependency predicted if retrievals from the same event were completely independent (Di) or dependent (Dd; see Methods of Chapter 5 for more information on how the models were constructed).

The within-event dependency in the data was greater than the independent model (D>Di, *t*(29) = 7.38, *p* < .001, *d* = 1.35) and lower than the dependent model (D<Dd, *t*(29) = 11.09, *p* < .001, *d* = 2.02) (Figure 6.2).

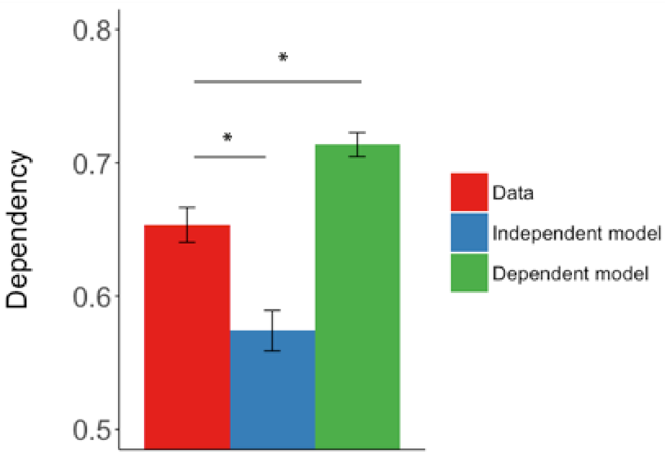


*Figure 6.2. Mean dependency for the observed data, the independent model and the dependent model for the within-event analysis. The error bars represent +/- 1 SEM. * p < .001*

6.5.4 Across-events dependency

As in Chapter 5, I examined the amount of dependency across matched overlapping events that shared a common element (i.e. whether retrieval success of an element from one event is dependent on the retrieval success of an element from an overlapping event). I calculated dependency in the data and corresponding independent and dependent models across pairs of events (see Methods of Chapter 5).

Consistently with the behavioural study, the across-events dependency was lower than the independent model (D<Di, *t*(29) = 3.04, *p* = .001, *d* = 0.56), and lower than the dependent model (D<Dd, *t*(29) = 8.54, *p* < .001, *d* = 1.56) (Figure 6.3).

6.5.5 Comparison of across-events relative dependencies

As in Chapter 5, I calculated a relative dependency measure (difference score between each participant's dependency and its independent model: D - Di). I looked at whether the relative dependency score is lower for overlapping events than for pairs of unrelated events.

As before, in order to create unrelated pairs, I randomly selected sets of two events with no shared elements. I then calculated the across-events dependency measures for these pairs in the same way as for the matched pairs of events (see Methods for Chapter 5). Because each analysis selected a random set of unrelated pairs, I repeated the procedure 1,000 times, with different unrelated pairs selected each time, and used the Brown's method to combine the probability values (Brown, 1975; Poole et al., 2016).

Supporting the results from Chapter 5, I found that overlapping events had lower relative dependency than the unrelated events (p = .001). The dependency in unrelated events was not different from baseline, i.e. the independent model (p = .289) (Figure 6.3). The main behavioural results of increased dependency for within-event associations and decreased dependency for specifically the pairs of overlapping events were therefore fully replicated in this study.

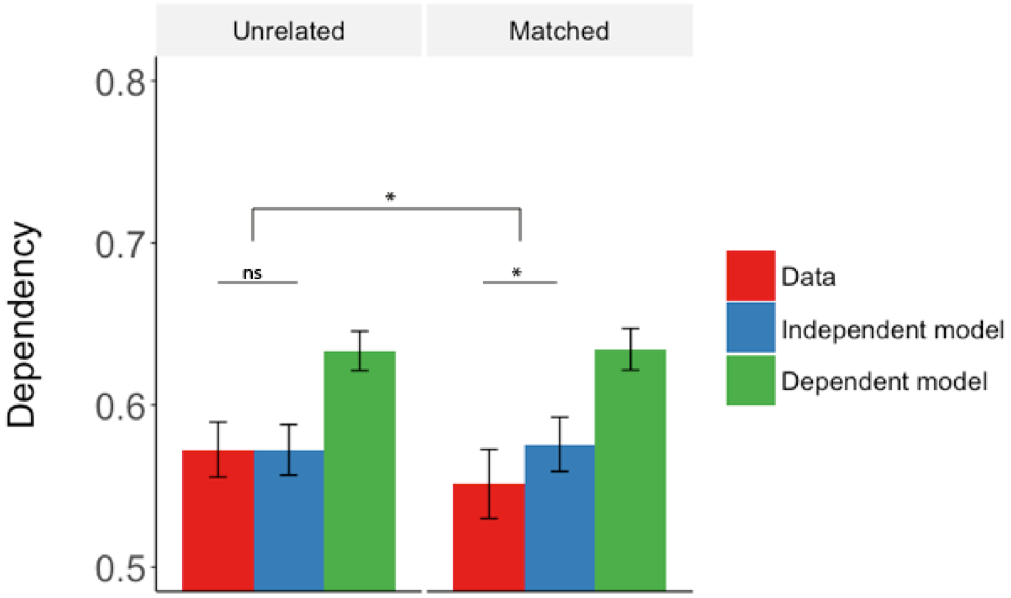


*Figure 6.3. Mean across-event dependency for pairs of unrelated events and pairs of matched, overlapping events. The data for unrelated pairs was obtained by randomly selecting pairs of unrelated events and calculating the dependency and the models, and repeating this procedure 1,000 times, each time with a different set of random pairs. The p-values were obtained using Brown's method for combining tests of significance (Brown, 1975). The bars represent the mean relative dependency, and the error bars represent the mean standard error across the 1,000 tests (unrelated events) or +/- 1 standard error (matched events). * p = .001*

6.6 fMRI analysis results

6.6.1 Univariate results

Encoding. These analyses looked for evidence of univariate memory effects to accompany the main RSA findings. First, I looked for areas where activity was predictive of subsequent memory. The analysis of the parametric modulator of encoding trials (based on number of correctly retrieved associations from an event) revealed a cluster in the left parahippocampal gyrus after small-volume correction (SVC) within bilateral hippocampus and parahippocampal gyrus (-30, -42, -6; *p* < .05 FWE small-volume corrected) (Figure 6.4).

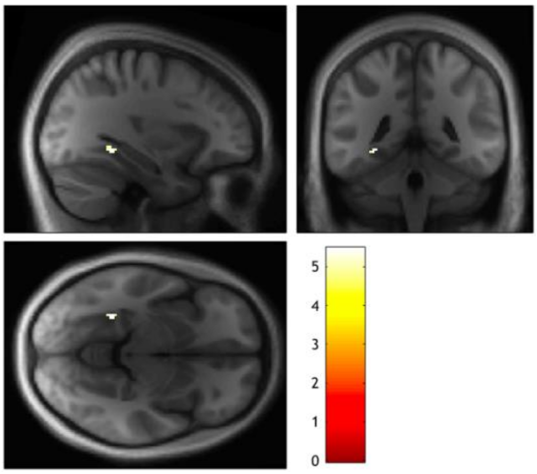


Figure 6.4. Regions showing subsequent memory effect (p < .05 FWE using small-volume correction for the hippocampus and parahippocampal gyrus).

Retrieval. In order to investigate memory effects, I looked for locations where activity differs between correct and incorrect retrieval trials. I saw a large cluster in the left parietal cortex (*p* < .05 FWE; -42, -62, 46), nucleus accumbens bilaterally (*p* < .05 FWE; -12, 8, -6 and 14, 8, -6). I also saw a small cluster in the right hippocampus (*p* < .05 FWE; -22, -34, -4). With small volume correction using a bilateral hippocampal mask, I saw increased activity for correct trials bilaterally. For the contrast between retrieval trials with and without the overlapping item, I saw no significant difference in activity in any voxels (*p* > .05 FWE-corrected for both whole brain and SVC for the hippocampus/parahippocampal cortex) (Figure 6.5).

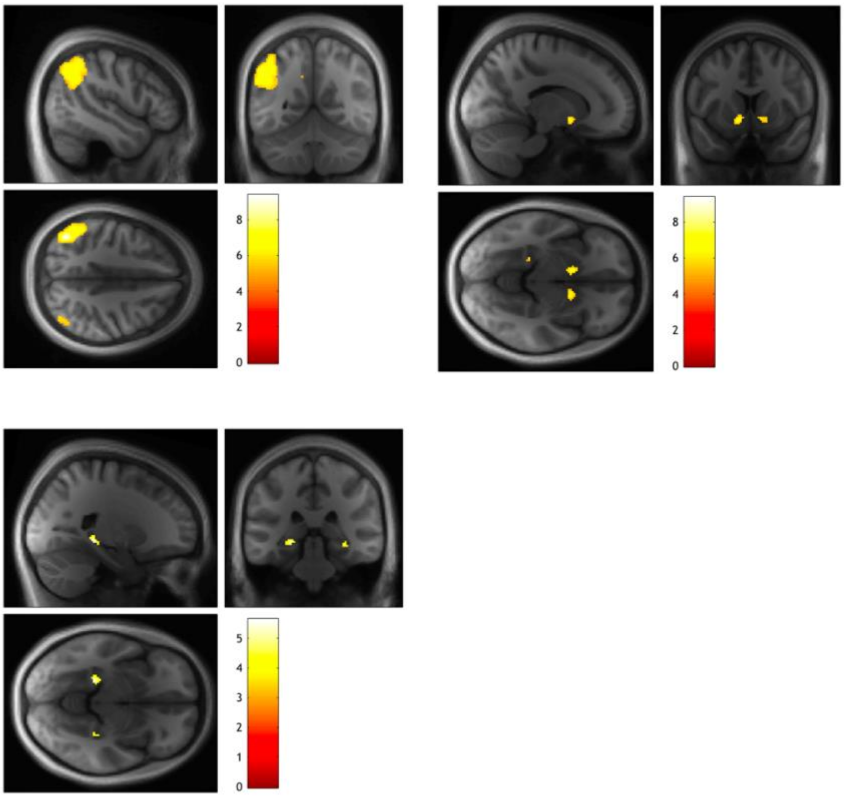
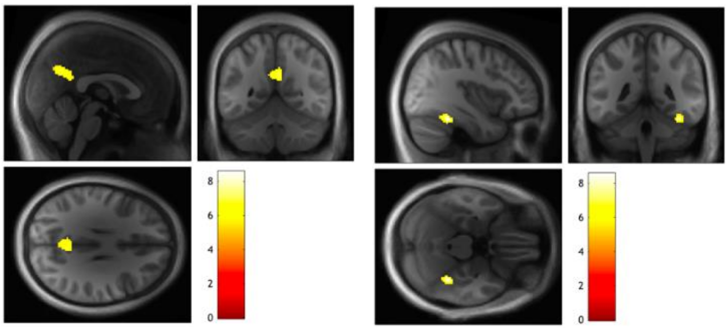


Figure 6.5. Regions showing greater activity for correct than incorrect trials included left parietal cortex, bilateral nucleus accumbens (p < .05 FWE for the whole brain, top row), and posterior hippocampus (p < .05 FWE SVC for bilateral hippocampus, bottom row).

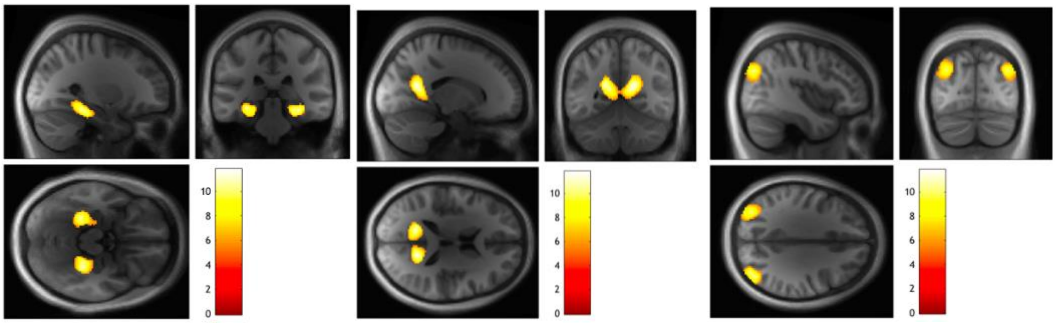
Item presentation. To investigate where the activity differs between different item types, I looked at areas involved in imagery of different element types (person, location or object) during the item presentation period. I compared activity related to one type to the activity of the two remaining types (e.g. person > location + object). In trials including people, I saw increased activity in the bilateral occipital lobe (L: -18, -92, -6; R: 18, -90, 0), right fusiform gyrus (fusiform face area; FFA) (R: 38, -42, -20), as well as bilateral precuneous/posterior cingulate cortex (2, -60, 34). In trials including locations, I saw regions in bilateral parahippocampal place area (PPA) (L: -28, -38, -10; R: 30, -34, -14), bilateral retrosplenial cortex (L: -16, -56, 18; R: 16, -54, 18), and bilateral posterior parietal cortex (L: -36, -76, 34; R: 42, -72, 34). In trials including objects, I found clusters in the left anterior inferior parietal cortex (-58, -24, 38) and lateral temporo-occipital area overlapping with the lateral occipital complex (-52, -64, 2) (all $p < .05$ FWE for the whole brain; Figure 6.6).

These areas were used to define type-specific ROIs for item types. ROIs for person- and object-specific areas were thresholded at alpha = .05 (FWE), while for location-specific areas I used a more stringent alpha level of .001 FWE to restrict the size of this ROI as too large ROIs may lead to a loss any spatial information across the region.

A) Person



B) Location



C) Object

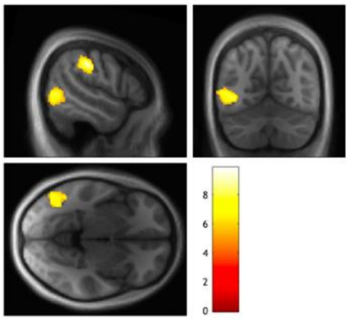
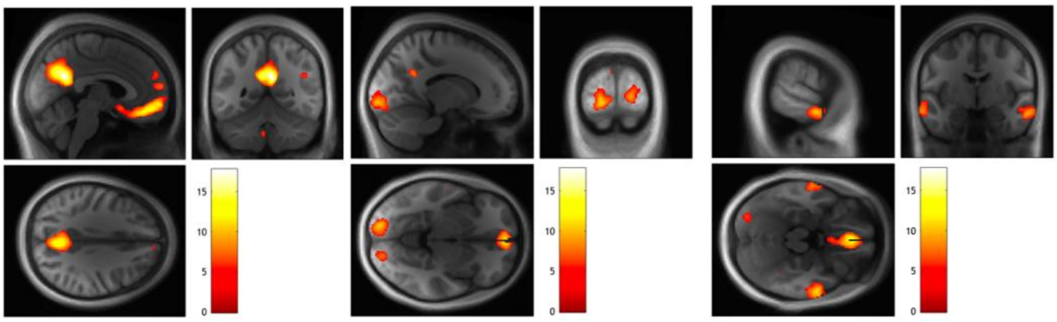


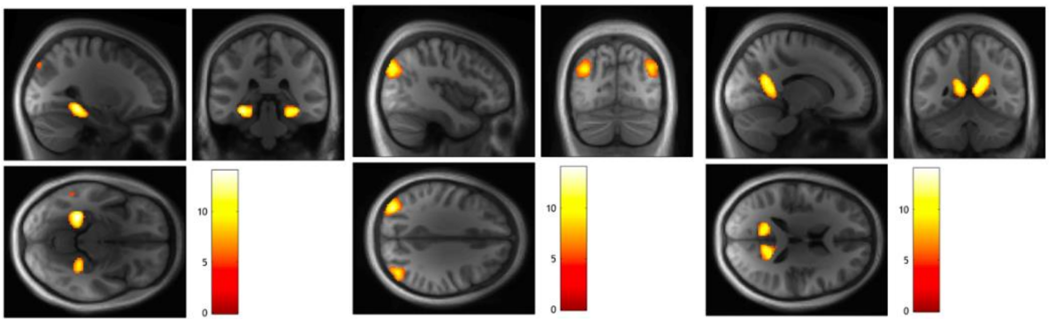
Figure 6.6. Regions showing greater activity at item presentation for (A) people, (B) locations and (C) objects (all p < .05 FWE-corrected for the whole brain).

Retrieval: element type. Next, I investigate whether the same areas show item type-related activity at retrieval as were observed at encoding. To look at item type-specific areas, I used regressors corresponding to each possible combination of cue-target type, and then contrasted trials where a given element was a cue or target with trials where it was neither, and repeated this for all element types. For person, the greatest difference was observed in bilateral precunous / cingulate region (4, -52, 28), medial prefrontal cortex (2, 50, -12), bilateral occipital (L: -18, -92, -4; R: 18, -92, 2) and bilateral lateral temporal areas (L: -64, -6, -16; R: 62, -2, -18). For location, I saw areas in the bilateral parahippocampal place area (L: -30, -38, -10; R: 30, -32, -14), bilateral retrosplenial cortex (L: -14, -56, 20; R: 12, -54, 18), and the posterior parietal cortex (L: -38, -78, 34; R: 42, -72, 32). For object, I identified areas in the left lateral temporo-occipital area / lateral occipital complex (-58, -54, -4), anterior inferior parietal cortex (-50, -44, 50) and the left inferior frontal gyrus (-48, 40, 14) (all $p < .05$ FWE for the whole brain; Figure 6.7).

A) Person



B) Location



C) Object

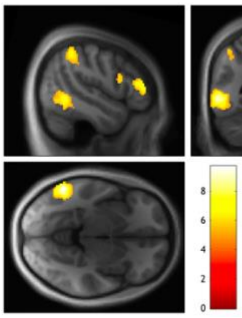


Figure 6.7. Regions showing greater activity at retrieval for (A) people, (B) locations and (C) objects (all p < .05 FWE for the whole brain).

6.6.2 RSA analyses

Pattern separation at encoding. The first RSA analysis explored the main hypothesis that overlapping events should lead to increased separation of patterns. I looked at whether the patterns of activity associated with two overlapping events are less similar than those of two non-overlapping events during the first encoding phase. I performed a whole-brain searchlight analysis and found no evidence that overlapping events are represented as less similar (p >

FWE for the whole brain and p > .05 FWE SVC for the hippocampus and parahippocampal cortex). On the other hand, I found areas in the bilateral hippocampus and some parahippocampal areas where overlapping events were represented as more similar than non-overlapping events (30, -10, -12 and -34, -36, -8; p < .05 FWE SVC for the hippocampus and parahippocampal cortex). When using hippocampal ROIs, I found higher similarity for overlapping events in the right hippocampus ($t(29) = 2.55$, $p = .017$, $d = 0.46$) and marginally for the left hippocampus ($t(29) = 2.04$, $p = 0.051$, $d = 0.37$).

For the second encoding phase (after the retrieval blocks) as well as for a combination of the first and second encoding phases using the crossvalidation technique (see Chapter 2), there were no areas showing differences in similarity between overlapping and non-overlapping events in either positive or negative direction ($p > .05$ FWE for the whole brain and $p > .05$ FWE SVC for the hippocampus and parahippocampal cortex; Figure 6.8).

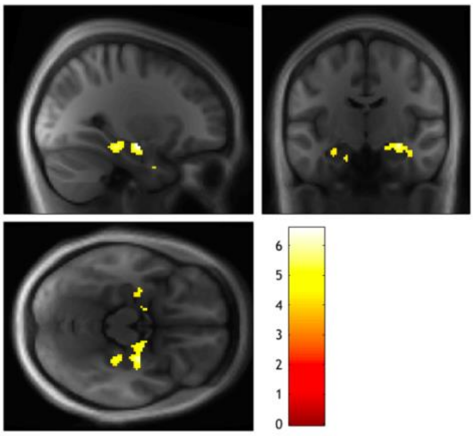


Figure 6.8. Areas where the patterns of activity corresponding to pairs of overlapping events were more similar than those corresponding to non-overlapping events in the whole-brain searchlight analysis of the first encoding block (p < .05 FWE SVC for bilateral hippocampus and parahippocampal cortex; masked for the ROI).

Pattern separation and interference. In addition to the primary analysis of event patterns at encoding, I also explored the association between pattern similarity and behavioural performance. To look at whether increased pattern separation reduces interference of similar memories, I correlated the dissimilarity measure for patterns of activity for overlapping pairs of events at encoding (first encoding block) with the average accuracy score on both events. I found no voxels that showed significant correlation (positive or negative) between similarity of patterns and behavioural accuracy ($p > .05$ FWE-corrected for both whole brain and SVC for the hippocampus/parahippocampal cortex). ROI analyses in the left and right hippocampi were also non-significant ($p > .05$).

I then looked at whether events with a more distinct pattern overall (when compared to all other events, and not only their matched event) led to better memory. The searchlight analysis within the hippocampus and parahippocampal cortex again failed to find regions where more dissimilar patterns were predictive of better performance ($p > .05$ FWE SVC for the hippocampus and parahippocampal cortex). ROI analysis in the left and right hippocampus (using one brain RDM per ROI), on the other hand, revealed that overall higher pattern distinctiveness of an event at encoding is associated with higher average performance on that event in the left hippocampus ($t(29) = 2.27$, $p = .031$, $d = 0.41$) but not in the right hippocampus, although there was a trend towards significance ($t(29) = 1.93$, $p = 0.064$, $d = 0.35$).

Changes to event representations over learning. In order to see whether the event patterns become more distinct over time, I compared similarity of patterns between first and second encoding phase. In a searchlight analysis limited to the hippocampus, I calculated a difference score (difference in similarity between first and second encoding phase) for overlapping and non-overlapping events and compared them with a paired-samples t-test. I found no areas where the difference in similarity was higher for the overlapping events than for non-overlapping events (p > .05 FWE SVC for the hippocampus and parahippocampal cortex), suggesting that the representations of overlapping events did not diverge over the course of the study more than the representations of non-overlapping events.

When looking at the two types of events separately, I found that the distances between events increased significantly for both overlapping pairs of events (p < .05 SVC for the hippocampus and parahippocampal cortex), suggesting that the representations of individual events became more distinct over time, but this did not differ between overlapping and non-overlapping event pairs (Figure 6.9).

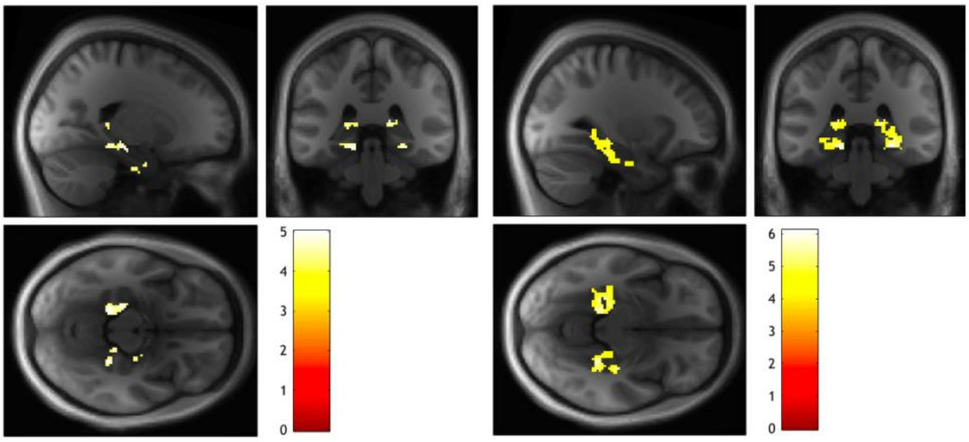


Figure 6.9. Areas where the distances between pairs of events increased between first and second encoding phase for overlapping pairs of events (left) and non-overlapping pairs of events (right) (p < .05 FWE SVC for hippocampus and parahippocampal cortex; masked for the ROI).

Pattern separation at retrieval. The main prediction that patterns of overlapping memories should show higher dissimilarity (separation) was also tested at retrieval. In a searchlight analysis limited to the hippocampus and the parahippocampal cortex, I looked for areas where associations from overlapping events are represented as less similar than associations from non-overlapping events. One cluster (16 voxels) was found in the right hippocampus (36, -14, -24; $p < .05$ FWE SVC for the bilateral hippocampus) (Figure 6.10). To look at this effect in more detail, I separately analysed associations which did and which did not include the overlapping item, to see if, potentially, one type of association drives the results. I found that none of the two sub-groups showed any significant effects of overlap (both showed mainly positive correlations within the hippocampus and the parahippocampal cortex, however these did not reach significance), suggesting that the result based on all associations was not disproportionally driven by associations that did (or did not) include the overlapping item.

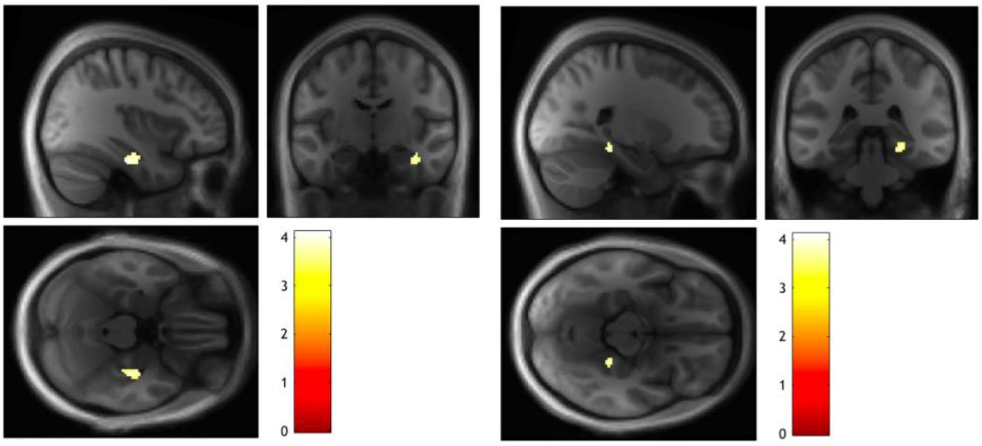


Figure 6.10. Areas showing more dissimilar patterns of activity for associations from overlapping events than for associations from non-overlapping events at retrieval (p < .001 uncorrected for display purposes).

Encoding-retrieval similarity (ERS). Additional analyses explored other memory effects by comparing similarity of events (and later items) at retriavl and encoding. In a whole-brain searchlight analysis, for each retrieval trial I correlated the pattern of activity at retrieval with pattern of its corresponding event at encoding (the ERS score), and averaged the correlation values separately for correctly and incorrectly recalled trials for each event. For each searchlight, I calculated a difference score between the ERS scores for correct and incorrect trials. In the group analysis, I compared these searchlight maps to zero. I found no voxels showing a difference in ERS dependant on the retrieval accuracy (*p* > .05 SVC for relevant ROI). I additionally compared the ERS score for correct events only against zero but I also found no significant voxels (*p* > .05 SVC for relevant ROI).

Non-target reinstatement. In each retrieval trial, I looked for reactivation of the pattern associated with the non-target, i.e. the item from the same event that was neither cue nor target in this trial. I compared the patterns for the non-target item in its type-specific ROI (from the univariate analysis) between encoding and retrieval phases.

First, I tested the correlation value from all retrieval trials against zero. I used a mixed-effects model with random intercept for each subject and each event, with event effect nested within subject effect. I compared the intercepts from this model against zero with a one-sample t-test. I found that the correlation between patterns at encoding and retrieval in item type-specific areas was significant, *t*(29) = 2.40, *p* = .023, *d* = 0.44.

I then used a mixed effects regression to see whether the level of non-target reinstatement across all within-event associations predicts mean accuracy on a given event (with random intercept for each subject and each event, with event effect nested within subject effect). I compared it to a null model which predicted no relationship between average non-target reinstatement and overall performance on the triads. A likelihood ratio test failed to reject the null model, χ^2 (1) = 0.81, *p* = .776.

6.7 Discussion

I investigated the hypothesis that similar (overlapping) memories are represented as more distinct neural representation as a result of pattern separation, a process which decorrelates similar inputs at encoding in order to prevent their interference. As in the experiment presented in Chapter 5, participants memorised multi-element events that were overlapping (had an element in common) or did not share any elements. I replicated the behavioural findings showing lowered dependency in retrieval of associations from overlapping events, and extended these results by showing that the overlapping memories corresponded to more distinct patterns of neural activity (Figure 6.10). This may reflect the neural mechanisms behind the reduced dependency at the behavioural level. However, the decreased similarity of patterns associated with overlapping events was observed only at retrieval, but not at encoding. I will discuss possible reasons for this below.

Evidence of pattern separation at retrieval

Although I found no evidence of pattern separation at encoding, the processing stage at which this mechanism is proposed to operate (Marr, 1971; O'Reilly & McClelland, 1994; Treves & Rolls, 1992), the results from the retrieval phase showed that the associations from overlapping events were represented with more dissimilar patterns of activity than the associations from non-overlapping events in the right anterior hippocampus. This reduced similarity was observed despite the fact that these associations shared an element and so could be expected to be represented as more similar – as they were at encoding (see below). These results were not driven by associations which included (or did not include) the overlapping element, as suggested by similar results in both types when analysed separately; although neither associations with or without the overlapping items showed a significant effect of overlap (as observed in the full dataset), both were in the predicted direction and of similar magnitude. Potentially, the decrease in the effect size could be due to the lower number of conditions used in these analyses. Overall, these results support the pattern separation hypothesis; they suggest that the events which are most likely to interfere because of a shared element are represented as more distinct than unrelated events, and this applies to all associations of these events and not, for example, only to the associations with the overlapping element.

This finding is in line with previous studies which showed evidence for pattern separation in the hippocampus (Bakker et al., 2008; Bonnici et al., 2012; Chadwick et al., 2011; Chanales et al., 2017; Kyle, Stokes, et al., 2015; Lacy et al., 2011; LaRocque et al., 2013). Studies using searchlight-based pattern analyses found similar clusters in the right hippocampus (Milivojevic et al., 2016; Stokes et al., 2013). In the current study, however, we were not able to separate the hippocampal subfields and look for evidence that pattern separation is reflected specifically in the CA3/DG activity and not the CA1 and other MTL areas (Lacy et al., 2011; Stokes et al., 2013).

An alternative explanation of the reduced pattern similarity for overlapping events could be retrieval inhibition, a mechanism which suppresses competing memories, leading to adaptive forgetting (Kuhl, Dudukovic, Kahn, & Wagner, 2007; Wimber et al., 2008). One study showed that the pattern associated with a memory was progressively suppressed when its competitor was repeatedly recalled, leading to lower pattern similarity between competing memories than between unrelated memories (Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015), which resembles the pattern of results observed here. However, in Wimber et al. the below-baseline suppression effect only emerged after several retrieval trials whereas it was absent (or even reversed; i.e. the competitor was reactivated) during the first retrieval. This repeated presentation potentially indicated that this association (rather than its competitor) is to be remembered. In the current experiment, on the other hand, the forgetting of the non-recalled memory would not be adaptive as both overlapping events were repeatedly recalled in an interleaved fashion.

Pattern separation at encoding

At encoding, I found no regions which showed less similar pattern of activity for pairs of overlapping events than for pairs of unrelated events. On the contrary, the representations of overlapping events in the hippocampus and parahippocampal gyrus were found to be more similar. This is inconsistent with previous studies which showed that activity patterns associated with different events in the hippocampus can be distinguished from each other despite overlapping elements (Chadwick et al., 2011; Milivojevic et al., 2016). However, unlike the current study, these studies did not look at whether similar episodic memories are represented with more distinct patterns of activity than unrelated memories.

One study which did show a decreased similarity of patterns for specifically the overlapping parts of navigational routes during learning (but not for the non-overlapping parts or for different routes), only observed this effect when comparing the first and second half of the learning phase (Chanales et al., 2017). Similarly, another study compared two overlapping narratives and also only showed a difference in representations of the two episodes in the final stage of the task (Milivojevic et al., 2016). A gradual divergence of environment-specific activity patterns has also been observed in hippocampal place cells in rats (Lever et al., 2002). One possible explanation may therefore be that the divergence of patterns associated with overlapping memories develops over time.

However, this would not be consistent with the proposed mechanism of pattern separation which is proposed to immediately orthogonalise overlapping inputs at encoding, resulting in decorrelated patterns to be stored in the hippocampus (McClelland et al., 1995; McNaughton & Morris, 1987). This is also not consistent with the findings from studies using continuous recognition paradigms where evidence for pattern separation is observed immediately at encoding (Bakker, Albert, Krauss, Speck, & Gallagher, 2015; Kirwan & Stark, 2007; Lacy et al., 2011). However, the findings from Chanales et al. and Milivojevic et al. may suggest that more complex episodic representations, like the ones used here, take longer to diverge to the point that is detectable in the pattern of activity.

I tested this hypothesis by comparing the similarity structure between the first and second encoding phase. I found that the representations of both overlapping and non-overlapping events became less similar, but the representations of overlapping events did not diverge (i.e. the distances did not increase) to a higher extent than of non-overlapping pairs of events. It needs to be stressed, however, that the design of the current study was different to the two studies mentioned above in that these studies used multiple consecutive learning blocks while here the two 'encoding' blocks were separated by the retrieval phase, which could have affected the representations of the events and/or the processes engaged during the task.

The results may be a consequence of the encoding period being dominated by perceptual similarity. Any signal related to pattern separation of overlapping events may be weaker than the effects of similarity, producing output consistent with increased representational similarity of overlapping memories. This may explain why the higher similarity observed during the first encoding phase was no longer present in the second encoding phase; potentially, the first phase was largely influenced by the perceptual characteristics of the stimuli while in the second phase the representations might have been more defined. However, I still did not see any evidence for pattern separation in this phase, although there are a number of possible reasons for this: representations changed as a result of retrieval, participants were fatigued or not motivated at this stage (as they knew there would be no memory test following the second encoding task), or potentially the balance between representing perceptual similarity versus event distinctiveness shifted but not to the point at which pattern separation could be detected.

Alternatively, the higher similarity of patterns associated with overlapping events at encoding could have reflected a recall of the previously encountered event when encoding its matching pair. This would result in activity patterns consistent with pattern completion as opposed to pattern separation; the former operates at retrieval and the latter at encoding, but their complementary roles mean that they are often difficult to distinguish. It has been recommended that to measure pattern separation, the desirable strategy should be the most appropriate for the task at hand, and therefore the most likely to use (Hunsaker & Kesner 2013; Liu et al. 2016). Consistently with this recommendation, previous findings showed that different instructions can bias processing towards either encoding or retrieval, as indicated by their different neural signatures (Richter, Chanales, & Kuhl, 2016). The current task was therefore designed to decrease the likelihood that any retrieval would take place during the encoding phase, and vice versa. However, it is still possible that participants re-imagined the matching event, either consciously as a way of rehearsal or automatically as a result of pattern completion to the already stored overlapping memory. The signal detected at encoding could therefore be a mixture of pattern separation of new episodes and pattern completion to the matching representations.

There is also a possibility that the overlapping items were not in fact the same items; for example, the kitchen imagined as an element of one event might not have been the same kitchen as when imagined in another event. This would require no pattern separation at encoding and, potentially, the mechanism of retrieval inhibition (Wimber et al., 2015) would suppress the shared features at retrieval. However, the fact that participants were asked to imagine individual elements before the memory task hopefully minimised this possibility and created fixed representations that were then used throughout the task. The reasons why the retrieval inhibition effect was unlikely to be observed at retriavl have been explained in the previous section.

Pattern separation and performance

The main role of pattern separation is to reduce interference between memories (McClelland et al., 1995; McNaughton & Morris, 1987; O'Reilly & McClelland, 1994). I therefore investigated the effect of pattern similarity on performance. LaRocque et al. (2013) previously showed that higher distinctiveness of an item-specific pattern (when compared to all other items) predicts later memory for that item, suggesting that higher representational dissimilarity leads to better memory performance. I found no evidence for this prediction in pairs of overlapping events; there was no relationship between pattern similarity and performance across all constituent retrieval trials in that pair.

However, consistently with LaRocque et al., I found that in the right hippocampus the representations of events that were overall more distinct (as compared to all other events and not only their matched pair) were associated with higher accuracy at retrieval - although it did not reach significance in the whole brain searchlight analysis. This again seems inconsistent with the role of pattern separation in reducing representational overlap between specifically these memories that are likely to interfere with each other, as would be expected from the behavioural results. As speculated already, the inability to detect pattern separation-like signal at encoding may have resulted from retrieval of representation of the overlapping events, potentially masking any separation-related activity.

Encoding-retrieval similarity

Many models propose that the hippocampus binds the associated elements during learning (McClelland et al., 1995; Norman & O'Reilly, 2003). This allows for retrieval of an entire memory through pattern completion, a reinstatement of the signal observed at encoding when presented with a partial cue, e.g. one of the elements of the event (Hopfield, 1982; Marr, 1971; McClelland et al., 1995). This implies that the same neural states should be observed at both encoding and retrieval. Studies have shown overlap of areas at encoding and retrieval (reviewed in Rugg *et al.*, 2008) and ability to predict the type of retrieved information from a classifier trained at encoding (reviewed in Rissman and Wagner, 2012). I therefore looked at whether the pattern of activity present at encoding is reinstated at retrieval. However, I found no areas which showed higher encoding-retrieval similarity of patterns for correct than for incorrect trials. This is at odds with other neuroimaging studies which suggest a common neural signature associated with encoding and retrieval. However, the encoding and retrieval task in the current study were markedly different: participants imagined interaction between three elements at encoding, while at retrieval they judged single paired associates. This is unlike other studies showing overlap in the patterns at encoding and retrieval where the elements to be retrieved directly corresponded to the encoded stimuli (e.g. Bird, Keidel, Ing, Horner, & Burgess, 2015; Ritchey, Wing, Labar, & Cabeza, 2013). The differences in the type of representations used during the two phases of the current task might have therefore underlain the lack of correspondence in patterns at encoding and retrieval.

In the univariate analyses, on the other hand, I observed increased average activity in highly similar areas during item presentation and retrieval (because all item-types were presented simultaneously at encoding, it was not possible to look for item-specific areas during this phase). The areas involved in location processing were consistent with previous literature implicating the PPA and interior calcarine area in scene perception (Epstein, Graham, & Downing, 2003; O'Craven & Kanwisher, 2000), posterior parietal cortex in spatial cognition (Burgess, Jeffery, & O'Keefe, 1999; Sack, 2009), and retrosplenial cortex in scene integration (Park & Chun, 2009). The people-specific areas were also in line with studies showing the role of the fusiform gyrus in representing faces (O'Craven & Kanwisher, 2000), while the posterior cingulate is involved in face perception (Pourtois, Schwartz, Spiridon, Martuzzi, & Vuilleumier, 2009) as well as autobiographical memory of familiar people (Maddock, Garrett, & Buonocore, 2001). Finally, the left anterior inferior parietal cortex is known to represent pragmatic properties of tools and objects (Kellenbach, Brett, & Patterson, 1984; Rumiati et al., 2004) and the lateral temporo-occipital area is involved in perception of tools (Valyear, Cavina-Pratesi, Stiglick, & Culham, 2007) and may be a part of the lateral occipital complex which supports object recognition (Grill-Spector, Kourtzi, & Kanwisher, 2001; Pourtois et al., 2009). Their role in the current task may reflect the imagery of object-affordances, as all object stimuli were intended to be easily handled or manipulated.

Similar areas were detected during both item presentation and retrieval phases, suggesting that the same neocortical areas are engaged during imagination of items from these categories and during their retrieval, which is consistent with the proposal that retrieval is accompanied by reactivation of signal present at encoding or during passive viewing (Bosch et al., 2014; Norman & O'Reilly, 2003; Nyberg, Habib, McIntosh, & Tulving, 2000; Wheeler, Petersen, & Buckner, 2000). The current results suggest therefore that although the same areas are used for imagination of individual elements and for their retrieval, the specific cortical representation is not reinstated, contrary to previous findings (Gordon, Rissman, Kiani, & Wagner, 2014; Kuhl, Rissman, Chun, & Wagner, 2011; Ritchey et al., 2013; Wing, Ritchey, & Cabeza, 2014).

Non-target reinstatement

Another consequence of the holistic representation of episodic memories and pattern completion at retrieval is the cortical reinstatement of neural signatures associated with all elements of an episode, including elements that are not the target of retrieval. The non-target reinstatement has been previously shown by Horner *et al.* (2015) in the original paradigm used in the current study; they showed that when the events were encoded in a holistic way, the cortical areas specific to the non-target showed increased activity during retrieval of the remaining two elements from the event. Here I extended this finding by showing that the specific neural representation of the non-target, i.e. its pattern at encoding, is observed in item type-specific areas at retrieval. This, however, was not related to the overall accuracy on an event, as would be expected if pattern completion of all event elements (as indicated by reinstatement) was the main factor predictive of the strength of the representation of this event.

Increased activity and memory performance

Finally, the univariate results showed evidence for a subsequent memory effect only in the left posterior parahippocampal gyrus but not in the hippocampus. This is different to previous findings showing increased activity to later-remembered items in the hippocampus (Bisby, Horner, Hørlyck, & Burgess, 2016; Horner et al., 2016; Staresina & Davachi, 2009). However, parahippocampal cortex is consistently implicated in encoding of novel stimuli (Stern et al., 1996; Tulving, Markowitsch, Kapur, Habib, & Houle, 1994), and specifically in memory for contextual (especially spatial) detail (Eichenbaum et al., 2007). A similar posterior cluster has previously been found during a source recollection task (Davachi et al., 2003) and was related to an encoding-retrieval similarity match in a word-scene association task (Staresina, Henson, Kriegeskorte, & Alink, 2012). The area identified here also lies close to the parahippocampal place area (PPA) which is involved in scene perception (Epstein & Kanwisher, 1998) and its relationship to performance may be explained by the role of spatial context in binding the memory elements together (Robin, Wynn, & Moscovitch, 2016).

During successful retrieval, I observed increased activity in lateral parietal areas which have been implicated in successful retrieval (Gilmore, Nelson, & McDermott, 2015; Okada, Vilberg, & Rugg, 2012). Similar left-lateralized retrieval-related signal in superior parietal lobe has also been observed during successful recognition judgement (Henson, Rugg, Shallice, Josephs, & Dolan, 1999). Additionally, I found increased activity in the bilateral nucleus accumbens, which is normally associated with rewards processing and motivation (Ikemoto & Panksepp, 1999; Schultz, 1998) but has also been suggested to play a role in some aspects of declarative memory (Setlow, 1997). Damage to the basal forebrain is known to cause amnesia (Zola-Morgan & Squire, 1993), and a case has been reported where a selective lesion to the nucleus accumbens led to amnesia (Goldenberg, Schuri, Grömminger, & Arnold, 1999). The univariate results are therefore broadly in line with previous research on the neural substrates of successful memory encoding and retrieval.

6.7.1 Conclusions

The evidence supporting the main prediction of this study, that overlapping memory episodes are represented by less similar neural activity patterns as a result of pattern separation, was found during retrieval of associations from multi-element events. At retrieval, activity patterns in the hippocampus were less similar for overlapping than for non-overlapping events. I did not, however, found any evidence for pattern separation at encoding, the processing stage at which pattern separation is proposed to operate (O'Reilly & McClelland, 1994; Treves & Rolls, 1992), inconsistently with previous evidence that pattern separation of episodic memories can be detected during encoding (e.g. Chadwick et al., 2011; LaRocque et al., 2013; Milivojevic et al., 2016). Potentially, this may be because the effect of perceptual similarity of overlapping events dominated any (smaller) effect of pattern separation. There was also no relationship between the degree of pattern separation of overlapping events and overall performance on the two events, suggesting that neural dissimilarity of overlapping events does not affect how well they are remembered. On the other hand, higher dissimilarity of an event to all other events did predict higher performance on that event in the left (and marginally in the right) hippocampus, suggesting that overall distinctiveness improves memory. I also found no evidence of reinstatement of encoding activity patterns at retrieval, despite showing that the same areas are involved in both imagination and retrieval of different item-types. I did, however, found some evidence that a non-target item (item from an event that is not currently the target of retrieval) is activated in type-specific areas, consistently with pattern completion to a complete event representation when presented with a partial input (Marr, 1971; McClelland et al., 1995). Overall, I found partial support for the predictions regarding pattern separation of overlapping memories and pattern completion of within-event elements.

Chapter 7

General discussion

7.1 Summary of findings

The overall aim of this thesis was to apply univariate and multivariate analysis methods to functional magnetic resonance imaging (fMRI) data to investigate the brain representations underlying spatial and episodic memories.

Using Representational Similarity Analysis (RSA), Chapter 3 looked for evidence of a grid-like representation of space during navigation and imagined movement. However, the results did not show any evidence of grid-like signal, despite the success of related univariate analysis of the same dataset (Horner et al., 2016). Potential reasons for this will be discussed below (also see Chapter 3). I did, however, find evidence of representational structures consistent with heading direction in the occipital lobes, which most likely reflected the similarity of the visual scene; this was preserved across actual navigation and imagined movement, showing that the same representation is used in both modalities. I also found evidence for representation of current spatial location in the hippocampus, right parahippocampal place area (PPA) and representation of the goal location in the left PPA. These findings are broadly consistent with previous literature on spatial navigation in humans and extend univariate results from other studies using the same paradigm (Doeller et al., 2010; Horner et al., 2016).

The second part of the thesis investigated the process of pattern separation of episodic memories, a mechanism through which representations of similar memories are decorrelated in order to prevent their interference (Marr, 1971; McClelland et al., 1995; Treves & Rolls, 1992). The general aim was to design a task which would address some of the limitations of the currently used paradigms in the pattern separation literature which, for reasons described in Chapters 1 and 4, are suboptimal for looking at this mechanism. I first attempted to induce the need for pattern separation by training participants on a task which required them to distinguish between facial stimuli belonging to two different categories (Chapter 4). At retrieval, I presented participants with these stimuli or with 'morph' faces created by linear combinations of the original faces, with varying level of similarity to the originals. I found that the ability to distinguish between 'old' (prototype) and 'new' (morph) faces decreased with similarity, but, contrary to the predictions, this did not depend on whether the prototype faces belonged to the same or different category. However, the overall poor performance levels in this study limits any strong conclusions as the absence of predicted results may be a consequence of floor effects.

Extending this method of examining pattern separation, Chapter 5 used more life-like memory events comprising multiple associated elements. Using this paradigm enables investigation of the coherence of the event representations by examining the statistical dependency of performance in retrieving different associations from the same event. I showed that, while each event is coherently represented, pairs of similar events (overlapping events with one element in common) show the opposite pattern: showing less dependency between retrievals from paired events than from unrelated events (or models of independent storage). This was interpreted as evidence of pattern separation; the overlapping memories may be stored as decorrelated representations, reducing the dependency in their retrieval. Despite this lowered similarity of representations of overlapping events, the holistic retrieval of their within-event elements (an index of pattern completion) was unaffected by the overlap, suggesting that the binding of individual items into a coherent memory is independent of the pattern separation of their respective memory engrams.

Finally, Chapter 6 investigated the pattern separation of multi-element events at the brain level. It was hypothesised that, just as retrievals from overlapping events are less statistically dependent than retrievals from unrelated events, the neural representations of overlapping events should be less similar to each other. I used the task designed in Chapter 5 to look for evidence that overlapping memories are indeed associated with more dissimilar patterns of activity in the hippocampus than memories that are unrelated. In the RSA analyses. Univariate results showed memory effect (greater activity for correctly remembered associations) in the parahippocampal cortex at encoding and in a network of areas (hippocampus, nucleus accumbens, left posterior parietal cortex) at retrieval. I also found areas which showed increased activity to each of the three item types (people, locations and objects) that were highly similar in task phases when items were presented in isolation, and when they were presented in association with another element at retrieval, suggesting that the same mechanisms may be involved during both encoding/imagination and retrieval. The areas included the FFA, PPA and LOC for people, locations and objects, respectively. I found a representational structure consistent with pattern separation at retrieval (as indexed by lower similarity of patterns of associations from overlapping events), however I found no evidence of higher neural pattern separation for overlapping events as compared to unrelated events at encoding. On the contrary, the hippocampal representations of overlapping events at encoding were more similar. This may be because the encoding phase is dominated by the effects of perceptual similarity (which is greater for overlapping events) so that smaller effects of pattern separation cannot be detected.

7.2 Using RSA to examine grid cells

A main aim of the first part of the thesis was to look for evidence of grid-like representation of space in both navigation and imagination using RSA analyses. Despite the presence of such signals in univariate analysis of the same data (Horner et al., 2016), I found no such RSA effects in the EC or elsewhere in the brain. Multi-voxel analyses are often considered to be more sensitive than univariate methods, and are often employed for this reason when univariate methods fail. However, in retrospect, I argue that multi-voxel pattern analyses may be less appropriate for looking at this particular research question than univariate methods. Due to the firing characteristics of grid cells, changes in heading direction should affect the univariate BOLD signal (as shown in Doeller, Barry, & Burgess, 2010 and Horner, Bisby, Zotow, Bush, & Burgess, 2016), but not necessarily the pattern of activity across voxels.

The presence of 60° periodicity in the univariate signal relies on (i) different dynamics of grid cell firing when running on- and off-axes of the grid, or (ii) the modulation of firing of the conjunctive cells by heading direction, as their preferred firing directions are aligned with grid axes (Doeller et al., 2012). These two conditions result in detectable difference in the univariate BOLD signal between movement along and in-between grid axes (Doeller et al., 2010; Horner et al., 2016). However, these on/off axis differences are general to all populations of grid cells or conjunctive cells, and would affect all grid cell containing voxels proportionally to the number of cells they contain. Thus, heading direction may modulate fMRI signal as a common factor across voxels, but changes to a multiplicative factor would not affect correlation-based assessments of the similarity of the patterns across voxels for on- or off-axis heading (correlations are independent of variations of the overall scale of one or the other vector).

Might there be an alternative mechanism that could support an RSA on/off axis effect? Supposing that conjunctive grid cells varied in the distribution of preferred firing directions across voxels, this would produce dissimilarity of patterns that unimodally reflects heading direction, but not a hexadirectional effect. Nonetheless, a grid-like representation of facing direction was observed in another study using the RSA (Bellmund et al., 2016). One explanation of that effect would be populations of cells that fire differently according to on- or off-axis movement. It is possible that directional cells with multiple preferred directions could produce such patterns across voxels, or other cell types so far unreported in rodents, although it is not clear why this signal would be absent in the study presented here. Potentially, the larger and more complex environment in Bellmund et al. allowed for detection of this relatively weak signal, which even in that study was observed only in one small region in the EC.

Alternatively, if the grid orientation varies across voxels (although evidence suggests that the variability is quite low, and high variability would be inconsistent with the assumptions of the univariate analyses as described above; Barry, Hayman, Burgess, & Jeffery, 2007; Doeller et al., 2010; Stensola et al., 2012), it would be possible to detect the expected differences in patterns provided that enough grid cell modules were sampled. The grid cell modules (cluster of cells of similar characteristics) show higher consistency in orientation within than between modules. The differences between modules of different scales could produce the results in Bellmund et al., whereas the smaller environment used here would not sample a sufficient range of grid scales to show corresponding effects. It is therefore possible that the variability in grid orientation, although low, could be used to successfully detect a pattern similarity structure consistent with 60° periodicity but only in larger environments. The null findings in the current study could have also been due to some of the limitations described in the Discussion of Chapter 3 and in section 7.7 below.

7.3 Behavioural evidence of pattern separation

The main aim of the second part of this thesis was to investigate pattern separation of similar episodic memories, and the conditions under which it operates. First, I looked at whether a behavioural need to distinguish between two categories affects pattern separation of representations of their respective members. It was hypothesised that the representations of different-category faces would need to be differentiated to a higher extent than of same-category faces, in turn affecting the ability to recognise their linear combination (their 'morphs') as novel as a function of whether they were from the same category or not. The results did not support this claim, which might suggest that either the category membership does not affect pattern separation of the prototype faces, or that the distinctiveness of representations of the prototype faces (as a result of pattern separation) does not affect the ability to discriminate their 'morphs'. However, as already mentioned, the task was difficult and, although participants' ability to distinguish between old and new stimuli was above chance, the low performance could have masked any potential effects of pattern separation on discrimination of similar stimuli.

With a more ecologically valid memory task using multi-element events that were either overlapping (with a common element) or fully unique, I found evidence of behavioural pattern separation of overlapping memories in Chapters 5 and 6. I used a measure of dependency (adapted from Horner & Burgess, 2013), which reflects the extent to which the retrieval of associations from one event depends on the retrieval of other associations from the same event (i.e. within-event coherency) compared to other associations from different events. I found that, although each event was coherently represented, retrieval of associations from one event had lower dependency on retrievals from an overlapping event than retrievals from unrelated events. This reduction in dependency is interpreted as evidence of pattern separation - a mechanism proposed to be performed by the dentate gyrus of the hippocampus (Marr, 1971) to decorrelate the neural representations of similar memories in order to prevent their interference.

This extends the results from other studies which have interpreted the ability to distinguish visually similar items as an index of pattern separation (Bakker et al., 2008; Kirwan et al., 2012; Kirwan & Stark, 2007; Lacy et al., 2011; Stark et al., 2013; Yassa et al., 2010). The present studies add to this literature by showing that more complex multi-element memories that share a common item also induce the need for pattern separation, possibly resulting in more dissimilar representations. In addition, it was possible to directly address the original hypothesis: of pattern separation of the encoding of two memories as a function of their similarity, rather than relying on an indirect measure of how well a single memory can be distinguished from a similar novel item at test.

7.4 Assessing pattern separation with RSA

The neural mechanisms behind the behavioural pattern separation of multi-element events were investigated using RSA. The RSA was deemed the most suitable analysis method for looking at this mechanism as it allowed for direct testing the hypothesis that patterns of activity associated with overlapping events are less similar than of unrelated events. As each event is both overlapping with another and unrelated to all other events, it would not be possible to use univariate methods or pattern classifiers to test this prediction. Instead, I looked at the overall similarity structure across all events to see whether its geometry fits with the proposed characteristics of the cognitive representations underlying similar memories.

At encoding, the overlapping events were represented with more similar activity patterns, contrary to the pattern separation account which predicts decreased representational overlap of similar memories. This is likely due to the higher perceptual similarity which may be more salient at encoding and which could have masked any smaller effects of pattern separation.

Additionally, dissimilarity of patterns of overlapping events in the hippocampus and parahippocampal cortex at encoding was not predictive of average performance on these events while the overall dissimilarity of event patterns was, suggesting that improved performance was related to the overall distinctiveness of a representation and not the degree of pattern separation from a similar event. I also explored whether the representational structure changed over time; I compared the similarities across events between first and second encoding phases (the second phase took place after the retrieval; see Methods of Chapter 6). I found that event representations diverged over time, but the magnitude of this change did not differ between overlapping and non-overlapping events, contrary to the proposal that the effects of pattern separation on the similarity of specifically the overlapping representations may emerge over time.

At retrieval, on the other hand, I found evidence of pattern separation of similar memory episodes: associations from overlapping events had less similar patterns of activity than associations from unrelated events. This did not seem to be driven by either associations which did (or which did not) include the overlapping item, suggesting that the lowered pattern similarity similarly affects all associations from overlapping events and not only those that overlap. Finally, inconsistently with the proposal that patterns at encoding are reinstated at retrieval, I found no encoding-retrieval pattern similarity for correctly retrieved associations anywhere in the brain. However, this may be a consequence of the differences in stimulus presentation at encoding and retrieval: participants memorised three-element triads while recalled pairwise associations. On the other hand, I found evidence for reactivation of the pattern associated with the non-target (i.e. the item from the same event that was neither cue nor target in this trial) in its type-specific area, which is in line with the pattern completion account that all event elements are retrieved when presented with a partial cue.

7.4.1 Neural mechanisms of pattern separation

The similarity structure observed in the activity patterns is assumed to reflect pattern separation, a mechanism which operates at encoding and facilitates the formation of distinct memory representations, as well as other factors such as perceptual similarity. Pattern separation is proposed to rely on the DG which receives similar inputs from the EC and transforms them into more sparse representations to be stored in the CA3 subfield of the hippocampus, thus minimising the overlap between patterns of activity (Marr, 1971; McClelland et al., 1995; Treves & Rolls, 1992). In humans, pattern separation at the cellular level has not yet been observed, although a lesion restricted to the DG produced behavioural performance consistent with impaired pattern separation (Baker et al., 2016). The current results provide further evidence for the involvement of the (right) hippocampus in patterns separation. It was not, however, possible to separate the hippocampal subfields; future work with e.g. a 7T scanner allowing for higher image resolution may be able to narrow down the precise anatomical substrate of this effect at retrieval, as well as to distinguish between the different mechanisms which might have contributed to the overall output at encoding.

7.4.1 Neurogenesis

The DG is also one of only two known regions within the brain that continues to create new neurons throughout life (the other one being the olfactory bulb; Sahay, Wilson, & Hen, 2011), a process called adult neurogenesis, suggesting that this mechanism may be involved in reducing overlap of similar memories and preventing interference through pattern separation (e.g. Aimone, Wiles, & Gage, 2006; Anacker & Hen, 2017; Johnston, Shtrahman, Parylak, Gonçalves, & Gage, 2016). Neurogenesis has been found to have an effect on discrimination of similar stimuli and encoding of novel experiences (Clelland et al., 2009; Kirschen et al., 2017; Sahay, Scobie, et al., 2011), which are proposed to depend on pattern separation. The role of neurogenesis in human memory, however, is still very much a point of debate. A recent paper by Sorrells et al. (2018) showed that neurogenesis decreases to undetectable levels during childhood, and so the human hippocampus may work differently to that of other species. At approximately the same time as these findings, another paper published conflicting results showing a continued neurogenesis in humans into adulthood (Boldrini et al., 2018) – although the opposite findings may be a consequence of differences in methodology (Kempermann et al., 2018). Its precise role, if any, is yet to be established. Development of new methods, such as the task developed here, which may provide behavioural measures of pattern separation is therefore beneficial in further advancement of this line of research.

7.5 Mental imagery in spatial and episodic memory

One of the aims of this thesis was to explore the role of a well characterised human memory system in mental simulation, both at encoding and retrieval. In Chapter 3, I looked at how the brain represented space during imagination of navigation. Chapters 5 and 6, although not specifically designed to compare 'experienced' autobiographical memory with imagination, also used imagined events as examples of real life-like memories. My results provide further evidence (also see Addis, Wong, & Schacter, 2007; Hassabis, Kumaran, & Maguire, 2007; Horner et al., 2016) that the representations used during encoding, retrieval and imagery may rely on the same or highly overlapping mechanisms.

Chapter 3 demonstrated that representations reflecting heading direction in occipital areas during both actual and imagined navigation towards a goal location were highly similar. These representations likely reflected the egocentric layout of the spatial scene from the different viewpoints, hence their dependence on the heading direction. This correspondence across modalities consistent with the idea that the spatial memory system is involved in navigation, memory retrieval and imagination. As proposed by the BBB model (Byrne et al., 2007), during imagination the relevant cortical representations are reactivated by the hippocampus, recreating the remembered scene. Movement within this scene can be simulated with 'mock' idiothetic signals, which then updates this representation, to re-experience a remembered route or imagine a novel one. The evidence for highly similar representations reflecting heading direction in navigation and imagination in Chapter 4 provides support for this model.

The task used in Chapters 5 and 6 used simulated event episodes. Participants were asked to imagine an interaction between an object, a person and a location as a way of creating life-like multi-element episodes. They then had to retrieve pairwise associations from each event. The results suggested that the storage and retrieval of these imagined scenarios rely on the same memory mechanisms (pattern separation and completion) and share the same neural signatures as encoding and retrieval of actual autobiographical memories (e.g. Addis, Wong, & Schacter, 2007; Hassabis, Kumaran, & Maguire, 2007). This may be because creating a mental image of this kind and then retrieving it requires a constructive process which is common to both memory for experienced events and imagination (Burgess, Becker, King, Keefe, & O'Keefe, 2001; Hassabis & Maguire, 2007).

These results are consistent with the role of the spatial/episodic memory system in constructing a rich contextual 'experience' (Byrne et al., 2007; Schacter & Addis, 2007, 2009; Tulving, 2002) as opposed to imagery of simple objects (which can be retained in autobiographical amnesia; Rosenbaum, McKinnon, Levine, & Moscovitch, 2004). This thesis therefore extends current knowledge on the flexibility of the memory system and shows that the use of imagination can be a useful tool in investigation of spatial and episodic memory.

7.6 RSA versus univariate approaches to fMRI data analysis

This thesis extensively used a relatively novel method of imaging data analysis, the RSA. This as well as other multivariate analysis techniques are slowly gaining popularity due to their ability to look at novel research questions. In addition to looking at whether a given brain area is involved in a given task, as in univariate analyses, RSA can additionally give us metric information on task information is represented.

Univariate approaches allow us to look at cognitive processes underlying behaviour. Processes are different from representations in that they are mechanisms that act upon the representations, but they do not store any information (Davis & Poldrack, 2013). The increase (or decrease) in the BOLD signal can be interpreted as an engagement of a voxel or a brain area in a given process. The only representations that can be identified using this approach are those that vary along a single dimension (e.g. stimulus intensity) or representations that affect a cognitive process (e.g. different models of memory representations may predict differential effects on the activity during retrieval; Davis & Poldrack, 2013). It cannot, however, look at multidimensional representations across multiple voxels and so its use is limited – although the univariate single-dimension effects can be used to infer information about representations, as shown for example in the univariate grid signal analyses (Doeller et al., 2010; Horner et al., 2016) or in studies looking at fMRI adaptation to make inferences about the nature of represented information (Grill-Spector, Henson, & Martin, 2006).

The notion of representation is important, because it carries a functional interpretation: the function is not only encoded in the area but also is read out by upstream areas (Decharms & Zador, 2000). Although RSA cannot directly provide evidence of such cause-effect relationship, by defining the representational geometry it specifies the representational content to much greater extent and with higher flexibility than is typical of univariate approaches.

Typical univariate approaches rely on the localisation of a function and so can only look at regions that as a whole represent the content, but they cannot distinguish between representations encoded as a pattern of activity across voxels. The RSA can look at this type of representation, while also removing the need to define the specific spatial correspondence between the neural activity and the BOLD signal across voxels; as long as the information can be decoded from the pattern, the precise spatial distribution does not matter. This allows for inter-subject variability in the spatial distribution of neural responses and for looking at details which would be lost when averaging activity across participants in univariate analyses.

The RSA therefore offers a novel way of looking at very specific hypotheses, especially suitable for condition-rich designs. Looking at the patterns of activity across voxels provides a middle ground between univariate approaches and single-cell recordings, thus bringing the human research closer to the animal studies. Additionally, as fMRI capabilities increase and the size of voxels decreases, it will become more and more meaningful to look at patterns of activity across voxels. The downside of this advantage is that RSA studies require more elaborate task designs, with particular attention needed to ensure that the data is optimal for the analyses, as will be discussed below.

Additionally, as seen in this thesis, not all representations are more effectively captured by the RSA than univariate approaches. The representations of overlapping and unrelated episodic memories (Chapter 6) lent themselves to this approach, as the main question of interest regarded specifically the similarity of the underlying neural pattern. The grid-like representation of space (Chapter 3), on the other hand, could not be captured by this technique while univariate analyses were able to detect differences in the BOLD signal consistent with a grid-like representation of space (Doeller et al., 2010; Horner et al., 2016). The choice of the appropriate approach should therefore be determined by the particular research question and type of data to fully capitalise on the strengths of the respective methods.

7.7 Practical considerations for design of RSA-appropriate tasks

The studies highlighted a number of practical issues related to the use of RSA on fMRI data to look at cognitive representations. The RSA has an advantage over other multivariate analysis methods of fMRI data such as pattern classifier approaches in that it goes beyond simple discrimination of features; rather, it can look at the representational structure of cognitive processes, providing deeper insight into the nature of the underlying computations. When designing an RSA-appropriate task, however, a number of requisites need to be considered.

In the first experiment presented in this thesis, one challenge resulted from the continuous metric used to measure the behaviour, which had to be translated into RSA-appropriate conditions. The movement directions had to be artificially split into directional bins (thus combining a range of different directions into one condition), which additionally were not consistent across blocks due to the way the task environment was set up (see Discussion section of Chapter 3). This highlights the need for task designs with more structured ways of measuring or constraining behaviour, ideally by using a categorical as opposed to a continuous metric.

Another challenge of applying the RSA to the data presented in this thesis was the low number of examples of behaviour within each condition. In univariate analyses, the same condition or behaviour is often measured repeatedly over multiple scanning runs in block- or event-related designs. The main advantage of RSA, however, is the ability to look at the representational structure of the cognitive function of interest, which usually entails calculating distances between patterns associated with several or even dozens of different conditions. Higher number of conditions allows for more elaborate models, but also means that fewer instances of behaviour can be used to estimate any one of the conditions.

On the other hand, some paradigms are not suited for the use of multiple examples of each condition. In some memory experiments, in particular, it may not be desirable to, for example, present the same stimulus more than once. The pilot studies used during the design of the multi-element event memory task presented in Chapters 5 and 6 (not reported here) showed that presenting the same events more than once resulted in too-high performance to detect any variability in dependency across the retrieval trials related to the task manipulation (overlapping or non-overlapping events). Multiple retrieval trials would also not be suitable in this paradigm. The potential benefits of multiple examples of each condition must also be weighed against fMRI-related constraints such as the time that can be spent in the scanner by the participants and the length of the BOLD responses (and therefore of the inter-trial intervals) which impose a limit on the study length and the number of trials. The decision regarding number and kind of conditions is therefore not straightforward and requires consideration of multiple, often opposing demands.

7.8 Theoretical design considerations for RSA

The analysis of the data with multi-voxel methods rely on some assumptions that have to be kept in mind when interpreting the result. First, any searchlight analyses assume that information can be decoded from all brain regions equally well; in other words, that the mapping between the neuronal activity and the brain signal is the same for all regions. This may not be the case (e.g. Bhandari, Gagne, & Badre, 2018). Additionally, the correlation between two patterns decreases with addition of any uninformative voxels (Diedrichsen et al., 2011), making the decision about the size of searchlights a crucial one, and yet there are almost no guidelines advising on the appropriate sphere size – and it is likely that these would be highly dependent on the very specific cognitive process of interest and the brain region that represents it.

There is also a question of whether the use of patterns of activity across voxels can really be used as a proxy of neuronal activity patterns. Each voxel not only contains millions of neurons (Logothetis, 2008), but also it is not quite clear what information at the neuronal level is represented at the level of a voxel, given the hemodynamic transform required to infer the neuronal responses (Kriegeskorte et al., 2008). The RSA relies on the Johnson-Lindenstrauss lemma which states that if a projection from a neuron space to a voxel space is a random projection (i.e. each voxel is a randomly weighted combination of neurons of different selectivity), then the distances should be preserved (Johnson & Lindenstrauss, 1984). If neurons, and the capillaries linking their activity to the BOLD signal (Attwell et al., 2010), are not randomly distributed across voxels, we may not be able to make strong inferences about the underlying neural code based on the pattern similarity estimates. This simplification, however, is still useful for interpretation of the activity patterns in relation to computational models of cognition (Kriegeskorte & Diedrichsen, 2016).

7.9 Other limitations

Finally, the MTL is a particularly difficult region to image due to the BOLD signal dropout and low signal-to-noise ratio there (Carr, Rissman, & Wagner, 2010). This might have contributed to the inability to detect any memory-related activity in this region in the early days of fMRI research (Schacter & Wagner, 1999). The scanning sequences used in the two imaging studies presented in this thesis were optimized for reduction of signal dropout in the MTL, however with a method as sensitive to noise as the RSA, any loss of meaningful signal may have a substantial effect on the results.

7.10 Future directions

Understanding of the memory system is extremely important due to the impact of memory loss on people's life. Impairment in spatial memory abilities and disorientation are one of the earliest signs of Alzheimer's disease (Allison, Fagan, Morris, & Head, 2016; Chan et al., 2016), while imbalance between pattern separation and completion is implicated in the failures of memory in old age (Stark, Stevenson, Wu, Rutledge, & Stark, 2015) and Mild Cognitive Impairment (Bakker et al., 2015; Stark et al., 2013), an early diagnostic of cognitive decline which may result in Alzheimer's (Petersen, 2004). Expanding our knowledge on the workings of the memory system in health will eventually allow us to better understand and potentially predict memory failures in old age and disease. Diagnosis of pre-dementia stages allows for administration of therapy in early stages (Cummings, Doody, & Clark, 2007) and could help to evaluate the effects of potential future treatments. Due to the increasing number and the characteristics of the affected population, there is a high need for diagnostic tests that are non-invasive, as well as cheap and easy to administer in a standard clinical setting. The task presented in Chapter 5 and 6 offers a novel way of measuring pattern separation. More research is needed in order to fully understand the precise mechanisms captured by this measure, however it may prove to be a suitable alternative to tasks such as the Mnemonic Similarity Task (for fuller discussion of this topic, see Chapter 1.2.1 and Introduction to Chapter 4). Additionally, the task can be further altered to investigate other aspects of this mechanism. A few examples of potential modifications are proposed below.

7.10.1 Manipulation of similarity

The need for pattern separation is related to the similarity of input, and so it is recommended that the level of similarity (and so the level of interference) should be parametrically modulated (Liu et al., 2016). In the current paradigm, this could be achieved by manipulating the number of overlapping elements between pairs of events to see the effect of the degree of overlap on the dependency measure. It would also be possible to test whether the effects of perceptual and/or semantic similarity are the same as the effects of element-overlap by using either picture stimuli of varying degrees of similarity or manipulating the semantic similarity of items.

7.10.2 Pattern separation over time

Following earlier discussion about whether the effects of pattern separation are detectable immediately at encoding or whether distinct representations of complex events develop over time (see Discussion section of Chapter 6 as well as Chanales, Oza, Favila, & Kuhl, 2017 and Milivojevic et al., 2016), it would be interesting to look at gradual changes to event representations, either over the course of learning within the same session or over extended delays.

7.10.3 Use of populations with reduced pattern separation

Another way of testing the validity of the task presented here would be to include special populations known to show reduced pattern separation abilities on the Mnemonic Similarity Task: older participants (Stark et al., 2015, 2013), patients undergoing chemotherapy (Christie et al., 2012), patients with Mild-Cognitive Impairment (Bakker et al., 2015; Yassa et al., 2010), schizophrenia (Das et al., 2014) and depression (Déry et al., 2013). The inclusion of these groups could further strengthen the claim that the reduced dependency in the retrieval of overlapping memories is related to pattern separation.

7.10.4 Combining spatial and episodic approaches

Lastly, the two themes of this thesis could be combined to look at the pattern separation of memory episodes in an immersive spatial environment like the one used in Chapter 3. Both episodic and spatial memory are thought to rely on largely the similar neural systems and mechanisms (see Chapter 1), and yet they are most often investigated separately. There is plenty of evidence showing behavioural pattern separation in the spatial domain (Bonnici et al., 2012; Chanales et al., 2017; Kyle, Stokes, et al., 2015; Stokes et al., 2013) and in memory for complex episodes (Chadwick, Hassabis, & Maguire, 2011; Milivojevic et al., 2016), however so far there are is a lack of research looking at how the two domains interact together during exploration of the environment, despite the important role the spatial context plays in episodic memory (Byrne et al., 2007; Schiller et al., 2015; Tulving, 1972). Both Chadwick et al. (2011) and Milivojevic et al. (2016) used the 'background' of their video-stimuli to determine the spatial context of a scene - similarly to the behavioural task from Chapters 5 and 6, where spatial location was a 'static' contextual element. However, so far evidence is lacking for pattern separation of representations of complex episodes in a continuous VR environment. The paradigm used in Chapters 5 and 6 was adapted from a task designed to look at pattern completion (Horner & Burgess, 2013), which in turn was a variation of a VR task where the participants navigated around a virtual town, encountering avatars that presented them with objects (Trinkler et al., 2006), creating similar person-location-object memory engrams to the ones used here. Although that study did not find any evidence of pattern completion, this might have been due to a low number of trials, low performance and slightly different analysis methods. Using the present paradigm and measure of dependency, this VR paradigm could be easily adapted to look at pattern separation of overlapping episodic memories in a more immersive environment.

7.11 Conclusions

This thesis used a novel behavioural, univariate and multivariate (RSA) fMRI techniques to investigate how spatial and episodic memories are represented in the human brain. The findings extend the literature on the mechanisms related to memory for complex, associative information, and the role of the memory system in encoding, retrieval and imagery. The thesis presents novel findings as well as replications of effects previously observed using univariate methods. It extensively discusses the potential reasons for absence of some of the predicted effects (e.g. lack of evidence for grid-like representation of space), particularly in relation to the use of RSA and its related constraints, and the importance of a careful design of tasks optimised for this technique. Understanding of how the patterns of activity in the brain relate to complex cognitive functions is a fundamental challenge for the field of memory research, and this work contributes to our knowledge of how we can relate computational models of cognition to brain representations.

References

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*(7), 1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Aggleton, J., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal, anterior thalamic axis. *Behavioral and Brain Sciences*, *22*(03), 425–444. https://doi.org/doi:10.1017/S0140525X99002034

Aggleton, J., Vann, S. D., Denby, C., Dix, S., Mayes, A. R., Roberts, N., & Yonelinas, A. P. (2005). Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2005.01.019

Aguirre, G. K., Detre, J. A., Alsop, D. C., & D'Esposito, M. (1996). The Parahippocampus Subserves Topographical Learning in Man. *Cerebral Cortex*, *6*(6), 823–829. Retrieved from http://dx.doi.org/10.1093/cercor/6.6.823

Aimone, J. B., Wiles, J., & Gage, F. H. (2006). Potential role for adult neurogenesis in the encoding of time in new memories. *Nature Neuroscience*, *9*(6), 723–727. https://doi.org/10.1038/nn1707

Allison, S. L., Fagan, A. M., Morris, J. C., & Head, D. (2016). Spatial navigation in preclinical Alzheimer's disease. *Journal of Alzheimer's Disease*, *52*(1), 77–90.

Aly, M., & Turk-Browne, N. B. (2015). Attention Stabilizes Representations in the Human Hippocampus. *Cerebral Cortex*, (February), 1–14. https://doi.org/10.1093/cercor/bhv041

Anacker, C., & Hen, R. (2017). Adult hippocampal neurogenesis and cognitive flexibility — linking memory and mood. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn.2017.45

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, *543*(7647), 719–722. https://doi.org/10.1038/nature21692

Attwell, D., Buchan, A. M., Charpak, S., Lauritzen, M., MacVicar, B. A., & Newman, E. A. (2010). Glial and neuronal control of brain blood flow. *Nature*, *468*(7321), 232.

Baker, S., Vieweg, P., Gao, F., Gilboa, A., Wolbers, T., Black, S. E., & Rosenbaum, R. S. (2016). The Human Dentate Gyrus Plays a Necessary Role in Discriminating New Memories. *Current Biology*, *26*(19), 2629–2634. https://doi.org/10.1016/j.cub.2016.07.081

Bakker, A., Albert, M. S., Krauss, G., Speck, C. L., & Gallagher, M. (2015). Response of the medial temporal lobe network in amnestic mild cognitive impairment to therapeutic intervention assessed by fMRI and memory task performance. *NeuroImage*, *7*, 688–698. https://doi.org/10.1016/j.nicl.2015.02.009

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

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Barense, M. D., Henson, R. N. A., Lee, A. C. H., & Graham, K. S. (2010). Medial temporal lobe activity during complex discrimination of faces, objects, and scenes: Effects of viewpoint. *Hippocampus*, *20*(3), 389–401. https://doi.org/10.1002/hipo.20641

Barron, H. C., Dolan, R. J., & Behrens, T. E. J. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience*, *16*(10), 1492.

Barry, C., Hayman, R., Burgess, N., & Jeffery, K. J. (2007). Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*, *10*(6), 682–684. https://doi.org/10.1038/nn1905

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., … Burgess, N. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, *17*(1–2), 71–97. https://doi.org/10.1515/REVNEURO.2006.17.1-2.71

Bastin, J., Vidal, J. R., Bouvier, S., Perrone-Bertolotti, M., Bénis, D., Kahane, P., … Epstein, R. A. (2013). Temporal components in the parahippocampal place area revealed by human intracerebral recordings. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(24), 10123–10131. https://doi.org/10.1523/JNEUROSCI.4646-12.2013

Bellmund, J. L. S., Deuker, L., Schröder, T. N., & Doeller, C. F. (2016). Grid-cell representations in mental simulation. *ELife*, *5*(AUGUST), 1–21. https://doi.org/10.7554/eLife.17089

Benoit, R. G., & Schacter, D. L. (2015). Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation. *Neuropsychologia*, *75*, 450–457.

Berron, D., Schütze, H., Maass, A., Cardenas-Blanco, A., Kuijf, H. J., Kumaran, D., & Düzel, E. (2016). Strong Evidence for Pattern Separation in Human Dentate Gyrus. *The Journal of Neuroscience*, *36*(29), 7569–7579. https://doi.org/10.1523/JNEUROSCI.0518-16.2016

Bhandari, A., Gagne, C., & Badre, D. (2018). Just above Chance: Is It Harder to Decode Information from Human Prefrontal Cortex Blood Oxygenation Level-dependent Signals? *Journal of Cognitive Neuroscience*, 1–26. https://doi.org/10.1162/jocn_a_01291

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: insights from spatial processing. *Nature Reviews. Neuroscience*, *9*(3), 182–194. https://doi.org/10.1038/nrn2335

Bird, C. M., Capponi, C., King, J. A., Doeller, C. F., & Burgess, N. (2010). Establishing the Boundaries: The Hippocampal Contribution to Imagining Scenes. *Journal of Neuroscience*, *30*(35), 11688–11695. https://doi.org/10.1523/JNEUROSCI.0723-10.2010

Bird, C. M., Keidel, J. L., Ing, X. L. P., Horner, X. A. J., & Burgess, N. (2015). Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex, *35*(43), 14426–14434. https://doi.org/10.1523/JNEUROSCI.1774-15.2015

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology: General*, *147*(2), 243–256. https://doi.org/10.1037/xge0000356

Bisby, J. A., Horner, A. J., Hørlyck, L. D., & Burgess, N. (2016). Opposing effects of negative emotion on amygdalar and hippocampal memory for items and associations. *Social Cognitive and Affective Neuroscience*, *11*(6), 981–990. https://doi.org/10.1093/scan/nsw028

Boccara, C. N., Sargolini, F., Thoresen, V. H., Solstad, T., Witter, M. P., Moser, E. I., & Moser, M. B. (2010). Grid cells in pre-and parasubiculum. *Nature Neuroscience*, *13*(8), 987–994. https://doi.org/10.1038/nn.2602

Boldrini, M., Fulmore, C. A., Tartt, A. N., Simeon, L. R., Pavlova, I., Poposka, V., … Mann, J. J. (2018). Human Hippocampal Neurogenesis Persists throughout Aging. *Cell Stem Cell*, *22*(4), 589–599.e5. https://doi.org/10.1016/j.stem.2018.03.015

Bonnici, H. M., Kumaran, D., Chadwick, M. J., Weiskopf, N., Hassabis, D., & Maguire, E. A. (2012). Decoding representations of scenes in the medial temporal lobes. *Hippocampus*, *22*(5), 1143–1153. https://doi.org/10.1002/hipo.20960

Borders, A. A., Aly, M., Parks, C. M., & Yonelinas, A. P. (2017). The hippocampus is particularly important for building associations across stimulus domains. *Neuropsychologia*, *99*(11), 335–342. https://doi.org/10.1016/j.neuropsychologia.2017.03.032

Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, (June), 1–11. https://doi.org/10.1038/nn.4573

Bosch, S. E., Jehee, J. F. M., Fernández, G., & Doeller, C. F. (2014). Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *The Journal of Neuroscience*, *34*(22), 7493–7500. https://doi.org/10.1523/JNEUROSCI.0805-14.2014

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(44), 13992–14003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009

Brown, M. B. (1975). A Method for Combining Non-Independent , One-Sided Tests of Significance., *31*(4), 987–992.

Buckley, M. J., & Gaffan, D. (1997). Impairment of visual object-discrimination learning after perirhinal cortex ablation. *Behavioral Neuroscience*, *111*(3), 467–475. https://doi.org/10.1037/0735-7044.111.3.467

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Burgess, N., Barry, C., & O'Keefe, J. (2007). An oscillatory interference model of grid cell firing. *Hippocampus*, *17*(9), 801–812. https://doi.org/10.1002/hipo.20327

Burgess, N., Becker, S., King, J. A., Keefe, J. O., & O'Keefe, J. (2001). Memory for events and their spatial context: models and experiments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *356*(1413), 1493–1503. https://doi.org/10.1098/rstb.2001.0948

Burgess, N., Jeffery, K. J., & O'Keefe, J. (Eds.). (1999). *The hippocampal and parietal foundations of spatial cognition. The hippocampal and parietal foundations of spatial cognition.* New York, NY, US: Oxford University Press.

Burgess, N., Maguire, E. a, Spiers, H. J., & O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *NeuroImage*, *14*(2), 439–453. https://doi.org/10.1006/nimg.2001.0806

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375. https://doi.org/10.1037/0033-295X.114.2.340.Remembering

Carr, V. A., Rissman, J., & Wagner, A. D. (2010). Imaging the human medial temporal lobe with high-resolution fMRI. *Neuron*, *65*(3), 298–308. https://doi.org/10.1016/j.neuron.2009.12.022

Chadwick, M. J., Hassabis, D., & Maguire, E. A. (2011). Decoding overlapping memories in the medial temporal lobes using high-resolution fMRI. *Learning & Memory*, *8*, 742–746. https://doi.org/10.1101/lm.023671.111

Chadwick, M. J., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2010). Decoding Individual Episodic Memory Traces in the Human Hippocampus. *Current Biology*, *20*(6), 544–547. https://doi.org/10.1016/j.cub.2010.01.053

Chadwick, M. J., Jolly, A. E. J., Amos, D. P., Hassabis, D., & Spiers, H. J. (2015). A Goal Direction Signal in the Human Entorhinal/Subicular Region. *Current Biology*, *25*(1), 87–92. https://doi.org/10.1016/j.cub.2014.11.001

Chan, D., Gallaher, L. M., Moodley, K., Minati, L., Burgess, N., & Hartley, T. (2016). The 4 Mountains Test: A Short Test of Spatial Memory with High Sensitivity for the Diagnosis of Pre-dementia Alzheimer's Disease. *Journal of Visualized Experiments*, (116), 1–11. https://doi.org/10.3791/54454

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers divergence of hippocampal representations. *Curr Biol*, *27*(15), 1–47. https://doi.org/10.1101/099226

Chavlis, S., & Poirazi, P. (2017). Pattern separation in the Hippocampus through the eyes of computational modeling. *Synapse*, (September 2016). https://doi.org/10.1002/syn.21972

Chen, G., King, J. A., Lu, Y., Cacucci, F., & Burgess, N. (2018). Spatial cell firing during virtual navigation of open arenas by head-restrained mice. *ELife*, *7*, e34789. https://doi.org/10.7554/eLife.34789

Christie, L. A., Acharya, M. M., Parihar, V. K., Nguyen, A., Martirosian, V., & Limoli, C. L. (2012). Impaired cognitive function and hippocampal neurogenesis following cancer chemotherapy. *Clinical Cancer Research*, *18*(7), 1954–1965. https://doi.org/10.1158/1078-0432.CCR-11-2000

Clelland, C. D., Choi, M., Romberg, C., Clemenson, G. D., Fragniere, A., Tyers, P., … Bussey, T. J. (2009). A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science (New York, N.Y.)*, *325*(5937), 210–213. https://doi.org/10.1126/science.1173215

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, Amnesia, and The Hippocampal System.* Cambridge, MA: MIT Press.

Colgin, L. L., Moser, E. I., & Moser, M. B. (2008). Understanding memory through hippocampal remapping. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2008.06.008

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468. https://doi.org/10.1126/science.aaf0941

Cummings, J. L., Doody, R., & Clark, C. (2007). Disease-modifying therapies for Alzheimer disease: challenges to early intervention. *Neurology*, *69*(16), 1622–1634.

Das, T., Ivleva, E. I., Wagner, A. D., Stark, C. E. L., & Tamminga, C. A. (2014). Loss of pattern separation performance in schizophrenia suggests dentate gyrus dysfunction. *Schizophrenia Research*, *159*(1), 193–197. https://doi.org/10.1016/j.schres.2014.05.006

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

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*, *100*(4), 2157–2162. https://doi.org/10.1073/pnas.0337195100

Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: Practices and pitfalls. *Annals of the New York Academy of Sciences*, *1296*(1), 108–134. https://doi.org/10.1111/nyas.12156

Day, M., Langston, R., & Morris, R. G. M. (2003). Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature*, *424*(6945), 205–209. https://doi.org/10.1038/nature01769

De Shetler, N. G., & Rissman, J. (2016). Dissociable profiles of generalization/discrimination in the human hippocampus during associative retrieval. *Hippocampus*, *121*(310), 1–22. https://doi.org/10.1002/hipo.22684

Decharms, R. C., & Zador, A. (2000). Neural representation and the cortical code. *Annual Review of Neuroscience*, *23*(1), 613–647.

Dees, R. L., & Kesner, R. P. (2013). The role of the dorsal dentate gyrus in object and object-context recognition. *Neurobiology of Learning and Memory*, *106*, 112–117. https://doi.org/10.1016/j.nlm.2013.07.013

Deng, W., Aimone, J. B., & Gage, F. H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nature Reviews Neuroscience*, *11*(5), 339–350. https://doi.org/10.1038/nrn2822

Déry, N., Pilgrim, M., Gibala, M., Gillen, J., Martin Wojtowicz, J., MacQueen, G., & Becker, S. (2013). Adult hippocampal neurogenesis reduces memory interference in humans: Opposing effects of aerobic exercise and depression. *Frontiers in Neuroscience*, *7*(7 APR), 1–15. https://doi.org/10.3389/fnins.2013.00066

Deuker, L., Bellmund, J. L. S., Schröder, T. N., & Doeller, C. F. (2016). Tracking the emergence of an event map in the hippocampus. *ELife*, 1–49. https://doi.org/10.7554/eLife.16534

Diedrichsen, J., & Kriegeskorte, N. (2016). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Computational Biology*, *13*(4), 1–35. https://doi.org/10.1101/071472

Diedrichsen, J., Ridgway, G. R., Friston, K. J., & Wiestler, T. (2011). Comparing the similarity and spatial structure of neural representations: A pattern-component model. *NeuroImage*, *55*(4), 1665–

1678. https://doi.org/10.1016/j.neuroimage.2011.01.044

Diedrichsen, J., & Shadmehr, R. (2005). Detecting and adjusting for artifacts in fMRI time series data. *NeuroImage*, *27*, 624–634. https://doi.org/10.1016/j.neuroimage.2005.04.039

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area (OPA) is causally and selectively involved in scene perception. *Journal of Neuroscience*, *33*(4). https://doi.org/10.1523/JNEUROSCI.4081-12.2013.

Dimsdale-Zucker, H. R., & Ranganath, C. (2019). Representational Similarity Analyses: A Practical Guide for Functional MRI Applications. In *Handbook of Behavioral Neuroscience* (Vol. 28, pp. 509–525). Elsevier. https://doi.org/10.1016/B978-0-12-812028-6.00027-6

Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nature Communications*, *9*(1), 294. https://doi.org/10.1038/s41467-017-02752-1

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657–661. https://doi.org/10.1038/nature08704

Doeller, C. F., Barry, C., & Burgess, N. (2012). From cells to systems: grids and boundaries in spatial memory. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *18*(6), 556–566. https://doi.org/10.1177/1073858411422115

Duncan, K., Sadanand, A., & Davachi, L. (2012). Memory's penumbra: episodic memory decisions induce lingering mnemonic biases. *Science*, *337*(6093), 485–487.

Eichenbaum, H., & Cohen, N. J. (2014). Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? *Neuron*, *83*(4), 764–770. https://doi.org/10.1016/j.neuron.2014.07.032

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*. https://doi.org/10.1016/S0896-6273(00)80773-4

Eichenbaum, H., & Lipton, P. a. (2008). Towards a functional organization of the meidal temporal lobe memory system: Role of the parahippocampal and medial entorhinal areas. *Hippocampus*, *18*(12), 1314–1324. https://doi.org/10.1002/hipo.20500.Towards

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *23*(1), 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Ekstrom, A. D., Kahana, M., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184–188. https://doi.org/10.1038/nature01955.1.

Ekstrom, A. D., & Ranganath, C. (2017). Space , Time and Episodic Memory : the Hippocampus is all over the Cognitive Map, *093052*, 1–16.

Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation.

Trends in Cognitive Sciences, *12*(10), 388–396. https://doi.org/10.1016/j.tics.2008.07.004

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-Specific Scene Representations in Human Parahippocampal Cortex, *37*, 865–876.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Epstein, R., Parker, W. E., & Feiler, A. M. (2007). Where Am I Now? Distinct Roles for Parahippocampal and Retrosplenial Cortices in Place Recognition. *Journal of Neuroscience*, *27*(23), 6141–6149. https://doi.org/10.1523/JNEUROSCI.0799-07.2007

Epstein, R., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*(11), 1504–1513. https://doi.org/10.1038/nn.4656

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*, 1–10. https://doi.org/10.1038/ncomms11066

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*(7132), 190–194. https://doi.org/10.1038/nature05601

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, *20*(2), 226–241. https://doi.org/10.1016/j.cogbrainres.2004.02.012

Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., & Denis, M. (1997). Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. *Neuroreport*, *8*(3), 739–744.

Gilbert, P. E., Kesner, R. P., & Lee, I. (2001). Dissociating hippocampal subregions: A double dissociation between dentate gyrus and CA1. *Hippocampus*, *11*(6), 626–636. https://doi.org/10.1002/hipo.1077

Gilmore, A. W., Nelson, S. M., & McDermott, K. B. (2015). A parietal memory network revealed by multiple MRI methods. *Trends in Cognitive Sciences*, *19*(9), 534–543. https://doi.org/https://doi.org/10.1016/j.tics.2015.07.004

Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*, *3*(4), 491–516. https://doi.org/10.1002/hipo.450030410

Gluth, S., Sommer, T., Rieskamp, J., & Büchel, C. (2015). Effective connectivity between hippocampus and ventromedial prefrontal cortex controls preferential choices from memory. *Neuron*, *86*(4), 1078–1090.

Goldenberg, G., Schuri, U., Grömminger, O., & Arnold, U. (1999). Basal forebrain amnesia: does the nucleus accumbens contribute to human memory? *Journal of Neurology, Neurosurgery &Amp; Psychiatry*, *67*(2), 163 LP-168. Retrieved from http://jnnp.bmj.com/content/67/2/163.abstract

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*. https://doi.org/10.1002/wcs.26

Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical Reinstatement Mediates the Relationship Between Content-Specific Encoding Activity and Subsequent Recollection Decisions. *Cerebral Cortex*, *24*(12), 3350–3364. Retrieved from http://dx.doi.org/10.1093/cercor/bht194

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. https://doi.org/10.1038/nature03721

Hartley, T., Burgess, N., Lever, C., Cacucci, F., & O'Keefe, J. (2000). Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus*, *10*(4), 369–379. https://doi.org/10.1002/1098-1063(2000)10:4<369::AID-HIPO3>3.0.CO;2-0

Harvey, C. D., Collman, F., Dombeck, D. A., & Tank, D. W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, *461*(7266), 941–946. https://doi.org/10.1038/nature08499

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using Imagination to Understand the Neural Basis of Episodic Memory. *Journal of Neuroscience*, *27*(52), 14365–14374. https://doi.org/10.1523/JNEUROSCI.4549-07.2007

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*(5), 1726–1731. https://doi.org/10.1073/pnas.0610561104

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11*(7), 299–306. https://doi.org/10.1016/j.tics.2007.05.001

Hasselmo, M. E. (2009). A model of episodic memory: Mental time travel along encoded trajectories using grid cells. *Neurobiology of Learning and Memory*, *92*(4), 559–573. https://doi.org/10.1016/j.nlm.2009.07.005

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, *37*(6), 1027–1041. https://doi.org/10.1016/S0896-6273(03)00144-2

Henson, R., Hornberger, M., & Rugg, M. D. (2005). Further dissociating the processes involved in recognition memory: An fMRI study. *Journal of Cognitive Neuroscience*, *17*(7).

https://doi.org/10.1162/0898929054475208

Henson, R., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and Familiarity in Recognition Memory : An Event- Related Functional Magnetic Resonance Imaging Study. *Journal of Neuroscience*, *19*(10), 3962–3972.

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

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

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*(5), 7462. https://doi.org/10.1038/ncomms8462

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like processing of imagined navigation. *Current Biology*, *26*(6), 842–847. https://doi.org/10.1016/j.cub.2016.01.042

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology. General*, *142*(4), 1370–1383. https://doi.org/10.1037/a0033626

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, *24*(9), 988–992. https://doi.org/10.1016/j.cub.2014.03.012

Horner, A. J., Gadian, D. G., Fuentemilla, L., Jentschke, S., Vargha-Khadem, F., & Duzel, E. (2012). A Rapid, Hippocampus-Dependent, Item-Memory Signal that Initiates Context Memory in Humans. *Current Biology*, *22*(24), 2369–2374. https://doi.org/10.1016/j.cub.2012.10.055

Howard, M. W. (2017). Temporal and spatial context in the mind and brain. *Current Opinion in Behavioral Sciences*, *17*, 14–19. https://doi.org/10.1016/j.cobeha.2017.05.022

Huffman, D. J., & Stark, C. E. L. (2014). Multivariate pattern analysis of the human medial temporal lobe revealed representationally categorical cortex and representationally agnostic hippocampus. *Hippocampus*, *10*, 1–10. https://doi.org/10.1002/hipo.22321

Hunsaker, M. R., & Kesner, R. P. (2013). The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neuroscience and Biobehavioral Reviews*, *37*(1), 36–58. https://doi.org/10.1016/j.neubiorev.2012.09.014

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior. *Brain Research Reviews*, *31*, 6–41.

Ino, T., Inoue, Y., Kage, M., Hirose, S., Kimura, T., & Fukuyama, H. (2002). Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. *Neuroscience Letters*, *322*(3), 182–186. https://doi.org/10.1016/S0304-3940(02)00019-8

Ison, M. J., Quian Quiroga, R., & Fried, I. (2015). Rapid Encoding of New Memories by Individual Neurons in the Human Brain. *Neuron*, *87*(1), 220–230.

https://doi.org/10.1016/j.neuron.2015.06.016

Jacob, P.-Y., Casali, G., Spieser, L., Page, H., Overington, D., & Jeffery, K. (2016). An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. *Nature Neuroscience*, *20*(2). https://doi.org/10.1038/nn.4465

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., … Kahana, M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience*, *16*(9), 1188–1190. https://doi.org/10.1038/nn.3466

Jeneson, A., Kirwan, C. B., Hopkins, R. O., Wixted, J. T., & Squire, L. R. (2010). Recognition memory and the hippocampus: A test of the hippocampal contribution to recollection and familiarity. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *17*(1), 63–70. https://doi.org/10.1101/lm.1546110

Johnson, W. B., & Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, *26*(189–206), 1.

Johnston, S. T., Shtrahman, M., Parylak, S., Gonçalves, J. T., & Gage, F. H. (2016). Paradox of pattern separation and adult neurogenesis: A dual role for new neurons balancing memory resolution and robustness. *Neurobiology of Learning and Memory*, *129*, 60–68. https://doi.org/10.1016/j.nlm.2015.10.013

Julian, J. B., Keinath, A. T., Frazzetta, G., & Epstein, R. A. (2018). Human entorhinal cortex represents visual space using a boundary-anchored grid. *Nature Neuroscience*, *21*(2), 191–194. https://doi.org/10.1038/s41593-017-0049-1

Kellenbach, M. L., Brett, M., & Patterson, K. (1984). Actions Speak Louder Than Functions : The Importance of Manipulability and Action in Tool Representation. *Journal of Cognitive Neuroscience*, 30–46.

Kempermann, G., Gage, F. H., Aigner, L., Song, H., Curtis, M. A., Thuret, S., … Frisén, J. (2018). Human Adult Neurogenesis: Evidence and Remaining Questions. *Cell Stem Cell*, *23*, 1–6. https://doi.org/10.1016/j.stem.2018.04.004

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

Kim, M., Jeffery, K. J., & Maguire, E. A. (2017). Multivoxel pattern analysis reveals 3D place information in the human hippocampus. *The Journal of Neuroscience*, 2703–2716. https://doi.org/10.1523/JNEUROSCI.2703-16.2017

Kim, M., & Maguire, E. A. (2018). 3D grid cells in human entorhinal cortex: Theoretical and methodological considerations and fMRI findings. *BioRxiv*, 282327. https://doi.org/10.1101/282327

Kirschen, G. W., Shen, J., Tian, M., Schroeder, B., Wang, J., Man, G., … Ge, S. (2017). Active dentate granule cells encode experience to promote the addition of adult- born hippocampal neurons. *The Journal of Neuroscience*, *1*(631). https://doi.org/10.1523/JNEUROSCI.3417-16.2017

Kirwan, B. C., Hartshorn, A., Stark, S. M., Goodrich-Hunsaker, N. J., Hopkins, R. O., & Stark, C. E. L. (2012). Pattern separation deficits following damage to the hippocampus. *Neuropsychologia*, *50*(10), 2408–2414. https://doi.org/10.1016/j.neuropsychologia.2012.06.011

Kirwan, B. C., & Stark, C. E. L. (2004). Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus*, *14*(7), 919–930. https://doi.org/10.1002/hipo.20014

Kirwan, B. C., & Stark, C. E. L. (2007). Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *14*(9), 625–633. https://doi.org/10.1101/lm.663507

Knierim, J. J., & Zhang, K. (2012). Attractor Dynamics of Spatially Correlated Neural Activity in the Limbic System. *Annual Review of Neuroscience*, *35*(1), 267–285. https://doi.org/10.1146/annurev-neuro-062111-150351

Kragel, J. E., Morton, N. W., & Polyn, S. M. (2015). Neural activity in the medial temporal lobe reveals the fidelity of mental time travel. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*(7), 2914–2926. https://doi.org/10.1523/JNEUROSCI.3378-14.2015

Kriegeskorte, N., & Diedrichsen, J. (2016). Inferring brain-computational mechanisms with models of activity measurements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1705), 1–25. https://doi.org/10.1098/rstb.2016.0278

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(11), 4. https://doi.org/10.3389/neuro.06.004.2008

Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience*, *10*(7), 908–914. https://doi.org/10.1038/nn1918

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 LP-5908. Retrieved from http://www.pnas.org/content/108/14/5903.abstract

Kumaran, D., & Maguire, E. A. (2007). Match Mismatch Processes Underlie Human Hippocampal Responses to Associative Novelty. *Journal of Neuroscience*, *27*(32), 8517–8524. https://doi.org/10.1523/JNEUROSCI.1677-07.2007

Kunz, L., Schröder, T. N., Lee, H., Montag, C., Lachmann, B., Sariyska, R., … Axmacher, N. (2015). Reduced grid-cell-like representations in adults at genetic risk for Alzheimer’s disease. *Science*, *350*(6259), 430–433. https://doi.org/10.1126/science.aac8128

Kyle, C. T., Smuda, D. N., Hassan, A. S., & Ekstrom, A. D. (2015). Roles of human hippocampal subfields in retrieval of spatial and temporal context. *Behavioural Brain Research*, *278*, 549–558. https://doi.org/10.1016/j.bbr.2014.10.034

Kyle, C. T., Stokes, J. D., Lieberman, J. S., Hassan, A. S., & Ekstrom, A. D. (2015). Successful retrieval of competing spatial environments in humans involves hippocampal pattern separation mechanisms. *ELife*, *4*, e10499. https://doi.org/10.7554/eLife.10499

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. L. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *18*(1), 15–18. https://doi.org/10.1101/lm.1971111

Lambrey, S., Doeller, C., Berthoz, A., & Burgess, N. (2012). Imagining Being Somewhere Else: Neural Basis of Changing Perspective in Space. *Cerebral Cortex*, *22*(1), 166–174. https://doi.org/10.1093/cercor/bhr101

LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *The Journal of Neuroscience*, *33*(13), 5466–5474. https://doi.org/10.1523/JNEUROSCI.4293-12.2013

Lassalle, J. M., Bataille, T., & Halley, H. (2000). Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task. *Neurobiology of Learning and Memory*, *73*(3). https://doi.org/10.1006/nlme.1999.3931

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science (New York, N.Y.)*, *315*(5814), 961–966. https://doi.org/10.1126/science.1135801

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science (New York, N.Y.)*, *305*(5688), 1295–1298. https://doi.org/10.1126/science.1100265

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, *29*(31), 9771–9777.

Lever, C., Wills, T., Cacucci, F., Burgess, N., & O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature*, *416*(6876), 90–94. https://doi.org/10.1038/416090a

Libby, L. A., Hannula, D. E., & Ranganath, C. (2014). Medial Temporal Lobe Coding of Item and Spatial Information during Relational Binding in Working Memory. *Journal of Neuroscience*, *34*(43), 14233–14242. https://doi.org/10.1523/JNEUROSCI.0655-14.2014

Liu, K. Y., Gould, R. L., Coulson, M. C., Ward, E. V, & Howard, R. J. (2016). Tests of pattern separation and pattern completion in humans - a systematic review. *Hippocampus*, *26*(6), 207–217. https://doi.org/10.1002/hipo.22561

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. https://doi.org/10.1038/nature06976

Long, N. M., Sperling, M. R., Worrell, G. A., Davis, K. A., Gross, R. E., Lega, B. C., … Kahana, M. J. (2017). Contextually Mediated Spontaneous Retrieval Is Specific to the Hippocampus. *Current Biology*, *27*(7), 1074–1079. https://doi.org/10.1016/j.cub.2017.02.054

Mack, M. L., & Preston, A. R. (2016). Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. *NeuroImage*, *127*, 144–157. https://doi.org/10.1016/j.neuroimage.2015.12.015

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, *104*(3), 667–676.

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. https://doi.org/10.1098/rstb.1971.0078

McAvoy, K. M., & Sahay, A. (2017). Targeting Adult Neurogenesis to Optimize Hippocampal Circuits in Aging. *Neurotherapeutics*, *14*(3). https://doi.org/10.1007/s13311-017-0539-6

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), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McDonald, R. J., & White, N. M. (1995). Hippocampal and nonhippocampal contributions to place learning in rats. *Behavioral Neuroscience*, *109*(4), 579–593. https://doi.org/10.1037/0735-7044.109.4.579

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, *83*(1), 202–215. https://doi.org/10.1016/j.neuron.2014.05.019

McKenzie, S., Keene, C., Farovik, A., Blandon, J., Place, R., Komorowski, R., & Eichenbaum, H. (2015). Representation of memories in the cortical–hippocampal system: Results from the application of population similarity analyses. *Neurobiology of Learning and Memory*, *134*, 178–191. https://doi.org/10.1016/j.nlm.2015.12.008

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the "cognitive map". *Nature Reviews Neuroscience Neuroscience*, *7*(8), 663–678. https://doi.org/10.1038/nrn1932

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

Milivojevic, B., Varadinov, M., Vicente Grabovetsky, A., Collin, S. H. P., Doeller, C. F. F., Grabovetsky, A. V., … Doeller, C. F. F. (2016). Coding of Event Nodes and Narrative Context in the Hippocampus. *Journal of Neuroscience*, *36*(49), 12412–12424. https://doi.org/10.1523/JNEUROSCI.2889-15.2016

Miller, J. F., Fried, I., Suthana, N., & Jacobs, J. (2015). Repeating Spatial Activations in Human Entorhinal Cortex. *Current Biology*, *25*(8), 1080–1085. https://doi.org/10.1016/j.cub.2015.02.045

Milner, B., Corkin, S., & Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM. *Neuropsychologia*, *6*(3), 215–234.

Mizumori, S. J., & Williams, J. D. (1993). Directionally selective mnemonic properties of neurons in the lateral dorsal nucleus of the thalamus of rats. *The Journal of Neuroscience*, *13*(9), 4015–4028. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8366357

Morgan, L. K., Macevoy, S. P., Aguirre, G. K., & Epstein, R. A. (2011). Distances between real-world locations are represented in the human hippocampus. *Journal of Neuroscience*, *31*(4), 1238–1245. https://doi.org/10.1523/JNEUROSCI.4667-10.2011.Distances

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., & O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, *297*(5868), 681–683. https://doi.org/10.1038/297681a0

Moser, E. I., Roudi, Y., Witter, M. P., Kentros, C., Bonhoeffer, T., & Moser, M.-B. (2014). Grid cells and cortical representation. *Nature Reviews. Neuroscience*, *15*(7), 466–481. https://doi.org/10.1038/nrn3766

Motley, S. E., & Kirwan, C. B. (2012). A Parametric Investigation of Pattern Separation Processes in the Medial Temporal Lobe. *Journal of Neuroscience*, *32*(38), 13076–13084. https://doi.org/10.1523/JNEUROSCI.5920-11.2012

Muller, R. U., & Kubie, J. L. (1987). The Effects of Changes in the Environment Hippocampal Cells on the Spatial Firing of. *Journal of Neuroscience*, *7*(7), 1951–1968.

Muller, R. U., Kubie, J. L., & Saypoff, R. (1991). The hippocampus as a cognitive graph. *Hippocampus*, *1*(3), 243–246. https://doi.org/10.1002/hipo.450010306

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-Selective Cortical Regions in Human and Nonhuman Primates. *Journal of Neuroscience*, *31*(39), 13771–13785. https://doi.org/10.1523/JNEUROSCI.2792-11.2011

Nau, M., Navarro Schröder, T., Bellmund, J. L. S., & Doeller, C. F. (2018). Hexadirectional coding of visual space in human entorhinal cortex. *Nature Neuroscience*, *21*(188). https://doi.org/10.1038/s41593-017-0050-8

Neunuebel, J. P., & Knierim, J. J. (2012). Spatial Firing Correlates of Physiologically Distinct Cell Types of the Rat Dentate Gyrus. *Journal of Neuroscience*, *32*(11), 3848–3858. https://doi.org/10.1523/JNEUROSCI.6038-11.2012

Neunuebel, J. P., & Knierim, J. J. (2014). CA3 retrieves coherent representations from degraded input: Direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*. https://doi.org/10.1016/j.neuron.2013.11.017

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003553

Norman, K. A., & O'Reilly, R. C. (2003). Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences*, *97*(20), 11120–11124. Retrieved from http://www.pnas.org/content/97/20/11120.abstract

O' Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, *381*(6581), 425–428. https://doi.org/10.1038/381425a0

O'Craven, K. M., & Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023. https://doi.org/10.1162/08989290051137549

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175. https://doi.org/10.1016/0006-8993(71)90358-1

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.

O'Keefe, J., Nadel, L., Keightley, S., & Kill, D. (1975). Fornix lesions selectively abolish place learning in the rat. *Experimental Neurology*, *48*(1), 152–166.

O'Keefe, J., & Speakman, A. (1987). Single unit activity in the rat hippocampus during a spatial memory task. *Experimental Brain Research*, *68*(1), 1–27. https://doi.org/10.1007/BF00255230

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal Conjunctive Encoding , Storage , and Recall : Avoiding a Trade-off, *4*(6), 661–682.

Okada, K., Vilberg, K. L., & Rugg, M. D. (2012). Comparison of the neural correlates of retrieval success in tests of cued recall and recognition memory. *Human Brain Mapping*, *33*(3), 523–533. https://doi.org/10.1002/hbm.21229

Olson, J. M., Tongprasearth, K., & Nitz, D. A. (2016). Subiculum neurons map the current axis of travel. *Nature Neuroscience*, *20*(2), 170. https://doi.org/http://dx.doi.org/10.1101/050641

Pajkert, A., Finke, C., Shing, Y. L., Hoffmann, M., Sommer, W., Heekeren, H. R., & Ploner, C. J. (2017). Memory integration in humans with hippocampal lesions. *Hippocampus*, *27*(12), 1230–1238. https://doi.org/10.1002/hipo.22766

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage*, *47*(4), 1747–1756. https://doi.org/10.1016/j.neuroimage.2009.04.058

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., & Vetter, T. (2009). A 3D Face Model for Pose and Illumination Invariant Face Recognition. *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference On*, 296–301. https://doi.org/10.1109/AVSS.2009.58

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science (New York, N.Y.)*, *310*(5756), 1963–1966. https://doi.org/10.1126/science.1117645

Poole, W., Gibbs, D. L., Shmulevich, I., Bernard, B., & Knijnenburg, T. A. (2016). Combining dependent P-values with an empirical adaptation of Brown's method. *Bioinformatics*, *32*(17), i430–i436. https://doi.org/10.1093/bioinformatics/btw438

Pourtois, G., Schwartz, S., Spiridon, M., Martuzzi, R., & Vuilleumier, P. (2009). Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. *Cerebral Cortex*, *19*(8), 1806–1819. https://doi.org/10.1093/cercor/bhn210

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*(1), 2–13. https://doi.org/10.1016/j.neuropsychologia.2003.07.006

Rebola, N., Carta, M., & Mulle, C. (2017). Operation and plasticity of hippocampal CA3 circuits: implications for memory encoding. *Nature Reviews Neuroscience*, *18*(4), 208. https://doi.org/10.1038/nrn.2017.10

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. https://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*, 101–128. https://doi.org/10.1146/annurev-psych-120710-100344

Ritchey, M., Wing, E. A., Labar, K. S., & Cabeza, R. (2013). Neural Similarity Between Encoding and Retrieval is Related to Memory Via Hippocampal Interactions. *Cerebral Cortex*, (23), 2818–2828. https://doi.org/10.1093/cercor/bhs258

Robin, J., Wynn, J., & Moscovitch, M. (2016). The spatial scaffold: The effects of spatial context on memory for events. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(2), 308–315. https://doi.org/10.1037/xlm0000167

Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, *7*, 74. https://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*, *129*, 4–28. https://doi.org/10.1016/j.nlm.2015.07.008

Rolls, E. T., Stringer, S. M., & Trappenberg, T. P. (2002). A unified model of spatial and episodic memory. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*(1496), 1087–1093. https://doi.org/10.1098/rspb.2002.2009

Rosenbaum, R. S., McKinnon, M. C., Levine, B., & Moscovitch, M. (2004). Visual imagery deficits, impaired strategic retrieval, or memory loss: disentangling the nature of an amnesic person's autobiographical memory deficit. *Neuropsychologia*, *42*(12), 1619–1635.

Rugg, M. D., Johnson, J. D., Park, H., & Uncapher, M. R. (2008). Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspective. *Progress in Brain Research*, *169*, 339–352.

Rugg, M. D., Vilberg, K. L., Mattson, J. T., Yu, S. S., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: FMRI evidence. *Neuropsychologia*, *50*(13), 3070–3079. https://doi.org/10.1016/j.neuropsychologia.2012.06.004

Rumiati, R. I., Weiss, P. H., Shallice, T., Ottoboni, G., Noth, J., Zilles, K., & Fink, G. R. (2004). Neural basis of pantomiming the use of visually presented objects, *21*, 1224–1231. https://doi.org/10.1016/j.neuroimage.2003.11.017

Sack, A. T. (2009). Parietal cortex and spatial cognition. *Behavioural Brain Research*, *202*, 153–161. https://doi.org/10.1016/j.bbr.2009.03.012

Sahay, A., Scobie, K. N., Hill, A. S., O'Carroll, C. M., Kheirbek, M. A., Burghardt, N. S., … Hen, R. (2011). Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature*, *472*(7344), 466–470. https://doi.org/10.1038/nature09817

Sahay, A., Wilson, D. A., & Hen, R. (2011). Pattern separation: a common function for new neurons in hippocampus and olfactory bulb. *Neuron*, *70*(4), 582–588. https://doi.org/10.1016/j.neuron.2011.05.012

Samsonovich, A., & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *The Journal of Neuroscience*, *17*(15), 5900–5920. https://doi.org/10.1523/JNEUROSCI.17-15-05900.1997

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M.-B., & Moser, E. I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science (New York, N.Y.)*, *312*(5774), 758–762. https://doi.org/10.1126/science.1125572

Schacter, D. L., & Addis, D. R. (2007). Constructive memory: The ghosts of past and future. *Nature*. https://doi.org/10.1038/445027a

Schacter, D. L., & Addis, D. R. (2009). On the nature of medial temporal lobe contributions to the constructive simulation of future events. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1245–1253. https://doi.org/10.1098/rstb.2008.0308

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661. https://doi.org/10.1080/08995600802554748

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*, *76*(4), 677–694. https://doi.org/10.1016/j.neuron.2012.11.001

Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: mechanisms and functions. *Current Opinion in Behavioral Sciences*, *17*, 41–50. https://doi.org/10.1016/j.cobeha.2017.06.002

Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus*, *9*(1), 7–24.

Schiller, D., Eichenbaum, H., Buffalo, E. A., Davachi, L., Foster, D. J., Leutgeb, S., & Ranganath, C. (2015). Memory and Space: Towards an Understanding of the Cognitive Map. *Journal of Neuroscience*, *35*(41), 13904–13911. https://doi.org/10.1523/JNEUROSCI.2618-15.2015

Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *12*(1), 103–113. https://doi.org/10.1136/jnnp.20.1.11

Setlow, B. (1997). The nucleus accumbens and learning and memory. *Journal of Neuroscience Research*, *49*(5), 515–521. https://doi.org/10.1002/(SICI)1097-4547(19970901)49:5<515::AID-JNR1>3.0.CO;2-E

Solstad, T., Solstad, T., Boccara, C. N., Boccara, C. N., Kropff, E., Kropff, E., … Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*(12), 1865–1868. https://doi.org/10.1126/science.1166466

Sorrells, S. F., Paredes, M. F., Cebrian-Silla, A., Sandoval, K., Qi, D., Kelley, K. W., … Alvarez-Buylla, A. (2018). Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature*, *555*(7696), 377–381. https://doi.org/10.1038/nature25975

Soto, F. A., & Ashby, F. G. (2015). Categorization training increases the perceptual separability of novel dimensions. *Cognition*, *139*, 105–129. https://doi.org/10.1016/j.cognition.2015.02.006

Squire, L. R. (1986). Mechanisms of memory. *Science (New York, N.Y.)*, *232*(4758), 1612–1619. https://doi.org/10.1176/appi.ajp.162.11.2201

Staresina, B. P., & Davachi, L. (2009). Selective and Shared Contributions of the Hippocampus and Perirhinal Cortex to Episodic Item and Associative Encoding Bernhard. *Journal of Cognitive Neuroscience*, *20*(8), 1478–1489. https://doi.org/10.1162/jocn.2008.20104.Selective

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic Reinstatement in the Medial Temporal Lobe. *Journal of Neuroscience*, *32*(50), 18150–18156. https://doi.org/10.1523/JNEUROSCI.4156-12.2012

Stark, S. M., Stevenson, R., Wu, C., Rutledge, S., & Stark, C. E. L. (2015). Stability of age-related deficits in the mnemonic similarity task across task variations. *Behavioral Neuroscience*, *129*(3), 257–268. https://doi.org/10.1037/bne0000055

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Stark, S. M., Yassa, M. A., & Stark, C. E. L. (2010). Individual differences in spatial pattern separation performance associated with healthy aging in humans. *Learning & Memory (Cold Spring Harbor,* *N.Y.)*, *17*(6), 284–288. https://doi.org/10.1101/lm.1768110

Steemers, B., Vicente-Grabovetsky, A., Barry, C., Smulders, P., Burgess, N., & Doeller, C. F. (2016). Hippocampal Attractor Dynamics Predict Memory- Report Hippocampal Attractor Dynamics Predict Memory-Based Decision Making. *Current Biology*, *26*, 1–8. https://doi.org/10.1016/j.cub.2016.04.063

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, *492*(7427), 72–78. https://doi.org/10.1038/nature11649

Stern, C. E., Corkin, S., Gonzalez, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., … Rosen, B. R. (1996). The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *93*(16), 8660–8665. https://doi.org/10.1073/pnas.93.16.8660

Stokes, J., Kyle, C., & Ekstrom, A. (2013). Complementary Roles of Human Hippocampal Subfields in Differentiation and Integration of Spatial Context. *Journal of Cognitive Neuroscience*, *27*(3), 546–559. https://doi.org/10.1162/jocn

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(2), 642–647. https://doi.org/10.1073/pnas.0610082104

Taube, J. S. (1998). Head direction cells and the neurophysiological basis for a sense of direction. *Progress in Neurobiology*, *55*(3), 225–256.

Taube, J. S. (2007). The Head Direction Signal: Origins and Sensory-Motor Integration. *Annual Review of Neuroscience*, *30*(1), 181–207. https://doi.org/10.1146/annurev.neuro.29.051605.112854

Taube, J. S. (2017). New building blocks for navigation. *Nature Neuroscience*, *20*(2), 131–133. https://doi.org/10.1038/nn.4480

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *The Journal of Neuroscience*, *10*(2), 420–435. https://doi.org/10.1212/01.wnl.0000299117.48935.2e

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A Map for Social Navigation in the Human Brain. *Neuron*, *87*(1), 231–243. https://doi.org/10.1016/j.neuron.2015.06.011

Thakral, P. P., Benoit, R. G., & Schacter, D. L. (2017). Characterizing the role of the hippocampus during episodic simulation and encoding. *Hippocampus*, *27*(12), 1275–1284. https://doi.org/10.1002/hipo.22796

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Towse, B. W., Barry, C., Bush, D., & Burgess, N. (2014). Optimal configurations of spatial scale for grid cell firing under noise and uncertainty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635). https://doi.org/10.1098/rstb.2013.0290

Treves, A., & Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, *2*(2), 189–199. https://doi.org/10.1002/hipo.450020209

Treves, A., Tashiro, A., Witter, M. E., & Moser, E. I. (2008). What is the mammalian dentate gyrus good for? *Neuroscience*, *154*(4), 1155–1172. https://doi.org/10.1016/j.neuroscience.2008.04.073

Trinkler, I., King, J. A., Spiers, H. J., & Burgess, N. (2006). Part or parcel? Contextual binding of events in episodic memory. In *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience* (pp. 53–83). https://doi.org/10.1093/acprof:oso/9780198529675.003.0003

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, *1*, 381–403.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12.

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

Tulving, E., Markowitsch, H. J., Kapur, S., Habib, R., & Houle, S. (1994). Novelty encoding networks in the human brain: positron emission tomography data. *Neuroreport*, *5*(18), 2525–2528. https://doi.org/10.1097/00001756-199412000-00030

Uncapher, M. R., & Rugg, M. D. (2005). Encoding and the Durability of Episodic Memory: A Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, *25*(31), 7260–7267. https://doi.org/10.1523/JNEUROSCI.1641-05.2005

Valyear, K. F., Cavina-Pratesi, C., Stiglick, A. J., & Culham, J. C. (2007). Does tool-related fMRI activity within the intraparietal sulcus reflect the plan to grasp? *Human Brain Mapping Journal*, *36*(3), 94–108. https://doi.org/10.1016/j.neuroimage.2007.03.031

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*(5324), 376–380.

Vazdarjanova, A., & Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *24*(29), 6489–6496. https://doi.org/10.1523/JNEUROSCI.0350-04.2004

Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nature Neuroscience*, *16*(4), 388–390. https://doi.org/10.1038/nn.3345

Visser, R. M., Scholte, H. S., & Kindt, M. (2011). Associative learning increases trial-by-trial similarity of BOLD-MRI patterns. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(33), 12021–12028. https://doi.org/10.1523/JNEUROSCI.2178-11.2011

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2015). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wheeler, M. E., & Buckner, R. L. (2004). Functional-anatomic correlates of remembering and knowing.

Neuroimage, *21*(4), 1337–1349.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N., & Keefe, J. O. (2005). Attractor Dynamics in the Hippocampal Representation of the Local Environment. *Science*, *308*(5723), 873–876. https://doi.org/10.1126/science.1108905.Attractor

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

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Michael, C. (2015). Retrieval Induces Adaptive Forgetting of Competing Memories via Cortical Pattern Suppression. *Nature Neuroscience*, *18*(4), 582–589. https://doi.org/10.1038/nn.3973.Retrieval

Wimber, M., Bauml, K.-H., Bergstrom, Z., Markopoulos, G., Heinze, H.-J., & Richardson-Klavehn, A. (2008). Neural Markers of Inhibition in Human Memory Retrieval. *Journal of Neuroscience*, *28*(50), 13419–13427. https://doi.org/10.1523/JNEUROSCI.1916-08.2008

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273.

Wing, E. A., Ritchey, M., & Cabeza, R. (2014). Reinstatement of Individual Past Events Revealed by the Similarity of Distributed Activation Patterns during Encoding and Retrieval. *Journal of Cognitive Neuroscience*, *27*(4), 679–691. https://doi.org/10.1162/jocn_a_00740

Wolbers, T., Wiener, J. M., Mallot, H. A., & Buchel, C. (2007). Differential Recruitment of the Hippocampus, Medial Prefrontal Cortex, and the Human Motion Complex during Path Integration in Humans. *Journal of Neuroscience*, *27*(35), 9408–9416. https://doi.org/10.1523/JNEUROSCI.2146-07.2007

Woodruff, C. C., Johnson, J. D., Uncapher, M. R., & Rugg, M. D. (2005). Content-specificity of the neural correlates of recollection. *Neuropsychologia*, *43*(7), 1022–1032.

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, *21*(9), 968–979. https://doi.org/10.1002/hipo.20808

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

Yassa, M. A., Stark, S. M., Bakker, A., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2010). High-resolution structural and functional MRI of hippocampal CA3 and dentate gyrus in patients with amnestic Mild Cognitive Impairment. *NeuroImage*, *51*(3), 1242–1252. https://doi.org/10.1016/j.neuroimage.2010.03.040

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*(11), 3002–3008.

Zhang, T., Heuer, H. W., & Britten, K. H. (2004). Parietal area VIP neuronal responses to heading stimuli are encoded in head-centered coordinates. *Neuron*, *42*(6), 993–1001.

Zola-Morgan, S., & Squire, L. R. (1993). Neuroanatomy of memory. *Annual Review of Neuroscience*, *16*(1), 547–563.

Zugaro, M. B., Arleo, A., Berthoz, A., & Wiener, S. I. (2003). Rapid spatial reorientation and head direction cells. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(8), 3478–3482. https://doi.org/23/8/3478 [pii]