

# **The Neural Processes Underpinning Episodic Memory**

**Demis Hassabis**

**Submitted for PhD in Cognitive Neuroscience**

**February 2009**



**University College London**

**Supervisor: Eleanor A. Maguire**

**Declaration:**

I, Demis Hassabis, 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:****Date:**

## **Abstract**

Episodic memory is the memory for our personal past experiences. Although numerous functional magnetic resonance imaging (fMRI) studies investigating its neural basis have revealed a consistent and distributed network of associated brain regions, surprisingly little is known about the contributions individual brain areas make to the recollective experience. In this thesis I address this fundamental issue by employing a range of different experimental techniques including neuropsychological testing, virtual reality environments, whole brain and high spatial resolution fMRI, and multivariate pattern analysis.

Episodic memory recall is widely agreed to be a reconstructive process, one that is known to be critically reliant on the hippocampus. I therefore hypothesised that the same neural machinery responsible for reconstruction might also support 'constructive' cognitive functions such as imagination. To test this proposal, patients with focal damage to the hippocampus bilaterally were asked to imagine new experiences and were found to be impaired relative to matched control participants. Moreover, driving this deficit was a lack of spatial coherence in their imagined experiences, pointing to a role for the hippocampus in binding together the disparate elements of a scene.

A subsequent fMRI study involving healthy participants compared the recall of real memories with the construction of imaginary memories. This revealed a fronto-temporo-parietal network in common to both tasks that included the hippocampus, ventromedial prefrontal, retrosplenial and parietal cortices. Based on these results I advanced the notion that this network might support the process of 'scene construction', defined as the generation and maintenance of a complex and coherent spatial context. Furthermore, I argued that this scene construction network might underpin other important cognitive functions besides episodic memory and imagination, such as navigation and thinking about the future.

It has been proposed that spatial context may act as the scaffold around which episodic memories are built. Given the hippocampus appears to play a critical role in imagination by supporting the creation of a rich coherent spatial scene, I sought to explore the nature of this hippocampal spatial code in a novel way. By combining high

spatial resolution fMRI with multivariate pattern analysis techniques it proved possible to accurately determine where a subject was located in a virtual reality environment based solely on the pattern of activity across hippocampal voxels. For this to have been possible, the hippocampal population code must be large and non-uniform. I then extended these techniques to the domain of episodic memory by showing that individual memories could be accurately decoded from the pattern of activity across hippocampal voxels, thus identifying individual memory traces.

I consider these findings together with other recent advances in the episodic memory field, and present a new perspective on the role of the hippocampus in episodic recollection. I discuss how this new (and preliminary) framework compares with current prevailing theories of hippocampal function, and suggest how it might account for some previously contradictory data.

**To**  
**Alexander and Arthur**

**The best in me**

'It's a poor sort of memory that only works backwards'

(Lewis Carroll, the White Queen to Alice in "Through the Looking-Glass")

## **Acknowledgements**

The work in this thesis could not have been accomplished without the generous help of a number of people. I wish to thank Eleanor Maguire, my principal supervisor, for her invaluable support and advice throughout my PhD, and for being so encouraging of new ideas. I want to also thank Geraint Rees, my second supervisor, for his guidance and infectious enthusiasm.

Then there are my friends and colleagues to thank from UCL: my partner in crime Dharshan Kumaran; Dean Mobbs for opening the door to exciting collaborations; Hugo Spiers and Benedetto de Martino for interesting science discussions; Carlton Chu and Martin Chadwick for work on multivariate methods; Jennifer Summerfield and Katherine Woollett for help on scene construction; Neil Burgess and Caswell Barry for useful theoretical discussions; Nikolaus Weiskopf and John Ashburner for their technical advice; and the support staff at the FIL, particularly Peter Aston, David Bradbury, Janice Glensman and Ric Davis for help with scanning and computers. Thanks also to the Brain Research Trust and the Wellcome Trust for funding the work in this thesis.

Finally, I would like to thank my mother and father for their spiritual and intellectual encouragement during my formative years without which I would not have been here today doing this PhD. And of course, most importantly, thanks to Teresa Niccoli who puts up with my craziness somehow and without whose love and support none of this could have happened.

# Contents

<b>1. Episodic Memory and the Hippocampus</b> .....	<b>15</b>
1.1 Introduction.....	16
1.2 Episodic memory.....	18
1.3 Reconstructive theories of episodic memory recall.....	21
1.4 The brain network underpinning episodic memory.....	23
1.4.1 The hippocampus.....	23
1.4.2 The episodic memory network.....	27
1.5 Hippocampal population codes.....	29
1.5.1 Information in the brain.....	30
1.5.2 The cell assembly hypothesis.....	31
1.5.3 Invariant representations in the hippocampus.....	33
1.5.4 Properties of the hippocampal population code.....	34
1.5.4.1 <i>Sparse codes</i> .....	34
1.5.4.2 <i>Topographical organisation</i> .....	35
1.6 Thesis overview.....	36
<b>2. Theories of Hippocampal Function</b> .....	<b>39</b>
2.1 Introduction.....	40
2.2 The Marr model.....	41
2.3 Declarative theory.....	42
2.4 Multiple trace theory.....	46
2.5 Cognitive map theory.....	50
2.6 Relational theory.....	55
2.7 Summary.....	60
<b>3. Methods</b> .....	<b>62</b>
3.1 Introduction.....	63
3.2 The biophysics of magnetic resonance imaging (MRI).....	63
3.2.1 Magnetic fields.....	63
3.2.2 Generating an MR signal.....	65
3.2.3 Tomographic image formation.....	67
3.2.4 Scanning parameters.....	67
3.3 The BOLD signal.....	69
3.3.1 The BOLD effect.....	69

3.3.2 The neurophysiology of the BOLD signal .....	70
3.4 Resolution of fMRI .....	72
3.5 Analysis of fMRI data.....	73
3.5.1 Spatial preprocessing .....	73
3.5.2 Processing of structural images .....	74
3.5.3 Statistical analysis and model specification .....	75
3.5.4 Model estimation and statistical inference.....	77
3.5.5 Random effects group analysis .....	78
3.5.6 Functional MRI experimental design and efficiency.....	78
3.6 Multivariate pattern classification techniques .....	79
3.6.1 The advantages of multivariate pattern analysis .....	80
3.6.2 General overview of classification methods .....	80
3.6.3 Decoding fine-grained patterns of representation .....	83
<b>4. Experiment 1: Patients with hippocampal amnesia can't imagine new experiences .....</b>	<b>85</b>
4.1 Introduction.....	86
4.2 Methods.....	88
4.2.1 Participants.....	88
4.2.2 Patients.....	89
4.2.2.1 P01 .....	91
4.2.2.2 P02 .....	91
4.2.2.3 P03 .....	92
4.2.2.4 P04 .....	92
4.2.2.5 P05 .....	93
4.2.3 Task and procedure.....	93
4.2.4 Scoring system .....	95
4.2.4.1 Content .....	95
4.2.4.2 Participant ratings.....	96
4.2.4.3 Spatial coherence index .....	96
4.2.4.4 Quality judgement.....	96
4.2.5 Calculating the experiential index.....	97
4.3 Results.....	98
4.3.1 Experiential index .....	98
4.3.2 Sub-components .....	101
4.3.2.1 Content .....	101
4.3.2.2 Participant Ratings .....	102
4.3.2.3 Spatial Coherence Index .....	103

4.3.2.4 <i>Quality Judgement</i> .....	105
4.3.3 <i>Outlier: P01</i> .....	106
4.4 <i>Discussion</i> .....	107
4.5 <i>Conclusions</i> .....	110
<b>5. Experiment 2: Using imagination to understand the neural basis of episodic memory</b> .....	<b>111</b>
5.1 <i>Introduction</i> .....	112
5.2 <i>Methods</i> .....	114
5.2.1 <i>Participants</i> .....	114
5.2.2 <i>Pre-scan interview</i> .....	114
5.2.3 <i>Stimuli</i> .....	117
5.2.4 <i>Task</i> .....	118
5.2.5 <i>Debriefing and scoring</i> .....	119
5.2.6 <i>Scanning parameters</i> .....	119
5.2.7 <i>Data analysis</i> .....	120
5.3 <i>Results</i> .....	121
5.3.1 <i>Behavioural data</i> .....	121
5.3.1.1 <i>Ratings summary</i> .....	121
5.3.1.2 <i>Formal ratings comparisons</i> .....	122
5.3.1.3 <i>Debriefing</i> .....	124
5.3.2 <i>Neuroimaging Data</i> .....	125
5.3.2.1 <i>Overall brain networks</i> .....	125
5.3.2.2 <i>Commonalities: the core scene construction network</i> .....	129
5.3.2.3 <i>Differences: additional requirements for real episodic memories</i> .....	131
5.3.2.4 <i>Additional analyses</i> .....	133
5.4 <i>Discussion</i> .....	133
5.5 <i>Conclusions</i> .....	137
<b>6. Experiment 3: Decoding neuronal ensembles in the human hippocampus</b> .....	<b>138</b>
6.1 <i>Introduction</i> .....	139
6.2 <i>Methods</i> .....	143
6.2.1 <i>Participants</i> .....	143
6.2.2 <i>Task and stimuli</i> .....	143
6.2.3 <i>Pre-scan training</i> .....	145
6.2.4 <i>Post-scan debriefing</i> .....	145
6.2.5 <i>Image acquisition</i> .....	146
6.2.6 <i>Imaging data pre-processing</i> .....	146

6.2.7 Univariate data analysis .....	146
6.2.8 Multivariate pattern classification.....	147
6.2.9 Multiclass classification .....	151
6.2.10 Reprojection and thresholding.....	152
6.2.11 Additional multivariate control analyses .....	153
<b>6.3 Results.....</b>	<b>155</b>
6.3.1 Behavioural data.....	155
6.3.2 Univariate neuroimaging data.....	157
6.3.3 Discriminating between two positions .....	157
6.3.4 Discriminating between four positions.....	159
6.3.5 Discriminating between the two environments.....	162
6.3.6 Sub-region dissociation .....	163
6.3.7 Permutation testing and thresholds.....	165
6.3.8 Additional multivariate control analyses .....	168
6.4 Discussion .....	169
6.5 Conclusions .....	173
<b>7. Experiment 4: Finding the Engram? .....</b>	<b>175</b>
7.1 Introduction.....	176
7.2 Methods.....	178
7.2.1 Participants .....	178
7.2.2 Pre-scan interview .....	178
7.2.3 Task and procedure.....	179
7.2.4 Debriefing .....	179
7.2.5 Image acquisition.....	180
7.2.6 Image data pre-processing.....	180
7.2.7 Multivariate pattern classification.....	181
7.3 Results.....	182
7.3.1 Behavioural data.....	182
7.3.2 Multivariate fMRI data.....	183
7.4 Discussion .....	185
7.5 Conclusions .....	189
<b>8. General Discussion.....</b>	<b>190</b>
8.1 Overview.....	191
8.2 Breaking down episodic memory using imagination.....	191
8.3 Scene construction .....	192
8.4 The core network.....	196

8.5 Other component processes of episodic memory .....	200
8.6 Downgrading the importance of time .....	202
8.7 Evolutionary function of construction .....	204
8.8 Interim summary .....	206
8.9 The dual function theory .....	207
8.9.1 Comparisons to the standard model and multiple trace theory .....	207
8.9.2 Evidence from patients .....	211
8.9.3 Evidence from neuroimaging .....	214
8.9.4 Compatibility with the cognitive map theory .....	215
8.9.5 Conclusion .....	216
<b>References .....</b>	<b>219</b>

## List of Figures

Figure 1. Multiple memory systems in the brain.....	17
Figure 2. Nissl-stained coronal section through human hippocampal formation.....	24
Figure 3. The episodic memory network.....	28
Figure 4. The cell assembly hypothesis.....	32
Figure 5. Schematic diagram of a simple model of the declarative theory.....	45
Figure 6. Hippocampal place cells.....	52
Figure 7. Firing fields of grid cells have a repetitive triangular structure.....	53
Figure 8. Schematic diagram of a simple relational network.....	57
Figure 9. Alignment of hydrogen atoms in a magnetic field.....	64
Figure 10. Generating an MR signal.....	66
Figure 11. Schematic of the magnetic properties of blood.....	69
Figure 12. The haemodynamic response function.....	71
Figure 13. Creation of condition-specific regressors in GLM.....	76
Figure 14. Illustration of a hypothetical experiment analysed using MPA.....	82
Figure 15. Orientation decoding from fMRI activity in visual cortex.....	84
Figure 16. Patient structural MRI scans.....	90
Figure 17. Examples of imagined experiences.....	99
Figure 18. Scores on the experiential index.....	100
Figure 19. Scores on the Spatial Coherence Index.....	105
Figure 20. Residual function of the right hippocampus in P01.....	107
Figure 21. Experimental design.....	117
Figure 22. Main scene conditions versus their respective object baselines.....	126
Figure 23. Main effect of Objects > Scenes.....	129
Figure 24. Brain areas in common to the three scene conditions.....	130
Figure 25. Comparison of real and imagined memories.....	132
Figure 26. The Experimental Task.....	142
Figure 27. Multivariate Pattern Analysis.....	149
Figure 28. Pairwise Position Classification.....	159
Figure 29. Four-way Position Classification.....	162
Figure 30. Environment Classification.....	163

Figure 31. 95th percentile threshold for each voxel individually.....	167
Figure 32. Example trial .....	182
Figure 33. Two-volume sliding time window classification .....	185
Figure 34. Constructing a scene from its elements.....	195
Figure 35. The core network .....	197
Figure 36. Mapping of component processes to cognitive functions .....	198

## List of Tables

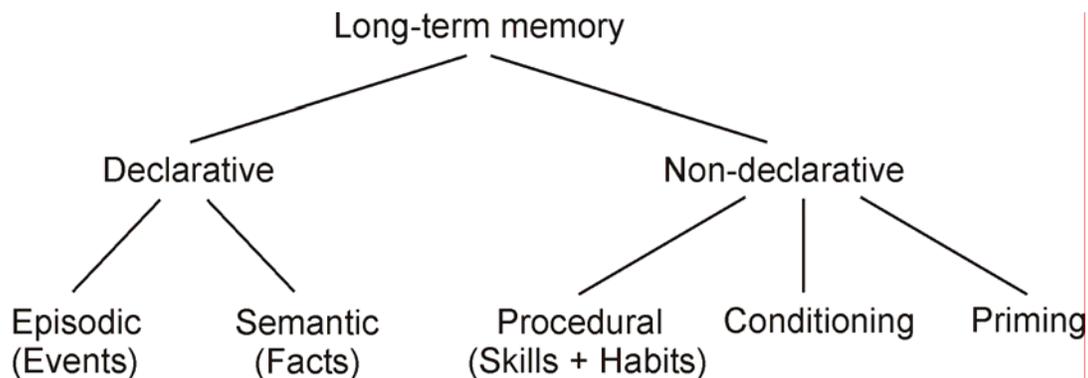
Table 1. Performance on the construction task.....	98
Table 2. Example of statement categorisation.....	102
Table 3. Spatial coherence index questionnaire.....	104
Table 4. Summary of conditions.....	115
Table 5. Behavioural ratings.....	122
Table 6. Recall of real episodic memories > recall of previously seen objects.....	127
Table 7. Newly imagined fictitious experiences > newly imagined objects.....	128
Table 8. Recall of previously imagined experiences > recall of previously imagined objects.....	128
Table 9. Conjunction analysis.....	131
Table 10. Recall of real episodic memories > recall of previously imagined experiences.....	133
Table 11. Example description of a fictitious scene.....	135
Table 12. 6-bit 'class codewords' for 4-way position classification.....	152
Table 13. Behavioural measures.....	156
Table 14. 95 <sup>th</sup> percentile accuracy value for prediction map thresholding*.....	165
Table 15. 95 <sup>th</sup> percentile averaged accuracy values for random runs*.....	166
Table 16. Control analyses.....	169

# Chapter 1

*Episodic Memory and the Hippocampus*

## 1.1 Introduction

Episodic memory (Tulving 2002b), the memory for our everyday personal experiences, is an essential ingredient in shaping how we perceive ourselves (Conway & Pleydell-Pearce 2000; Gallagher 2000). The hippocampus has been acknowledged to play a critical role in episodic memory ever since the famous case of patient HM who became profoundly amnesic following lesions of the medial temporal lobe (MTL) (Scoville & Milner 1957). The pattern of spared and impaired function exhibited by HM following his MTL lesions helped provide some of the impetus for the view, expounded by early theorists (Hirsh 1974; Nadel & O'Keefe 1974; Tulving 1972), that memory is not in fact a unitary system. His data added to a body of empirical evidence suggesting that there are in fact multiple memory systems in the brain (Hirsh 1974; Nadel & O'Keefe 1974; Tulving 1972), each mediated by a set of dissociable processes and regions (Figure 1)(Cohen & Squire 1980; Corkin 2002; Squire 1992; Squire et al. 2004). It was therefore proposed that the declarative memory system which encompasses consciously expressible memories for facts and events, corresponding to semantic and episodic memory respectively (Tulving 2002b), is critically dependent on the hippocampus and MTL (Cohen & Squire 1980; Corkin 2002; Squire 1992; Squire et al. 2004). By contrast, the nondeclarative memory system which includes priming, conditioning, and procedural memory such as skills and habits, is argued to be MTL-independent and thus spared in patients such as HM (Spiers et al. 2001; Squire et al. 2004).



**Figure 1. Multiple memory systems in the brain**

The traditional hierarchical classification of long-term memory. Declarative memory is critically dependent on the hippocampus and MTL whereas non-declarative memory is dependent on regions such as the striatum and cerebellum (Squire et al. 2004).

In this introductory chapter, I present a broad overview of what is now an extensive literature on episodic memory (Tulving 2002b), starting with Tulving's influential ideas (Tulving 1983; Tulving 2002b). I then outline the processes thought to support episodic memory recall (Schacter et al. 1998; Tulving 2002b), in particular reconstruction. Bartlett (Bartlett 1932) was among the first to pioneer ideas about the reconstructive nature of the memory system when he rejected the notion that memory was a process that simply involved the passive replay of a past experience via the awakening of a literal copy of experience. He famously stated that: "...the first notion to get rid of is that memory is primarily or literally reduplicative, or reproductive. In a world of constantly changing environment, literal recall is extraordinarily unimportant...if we consider evidence rather than supposition, memory appears to be far more decisively an affair of construction rather than one of mere reproduction..." (Bartlett 1932, pp. 204–205). The view that episodic memory is reconstructive is now widely accepted although the nature of the underlying processes involved remains poorly understood.

Next I survey briefly the patient studies that have helped to characterise the amnesic syndrome and originate the multiple memory systems view, with the hippocampus

playing a central role in declarative memory. This patient work has more recently been complemented by neuroimaging studies that inform about the wider brain network that, in conjunction with the hippocampus, underpins the episodic memory system. Finally, no comprehensive account of the memory system would be complete without an understanding of the mechanics of memory at the neuronal level. In the last section I briefly review the latest theories on how information is thought to be coded throughout the brain in general, and more specifically the fundamental properties of hippocampal population codes and the information they might represent. I conclude this introductory chapter with a summary of the aims of this thesis.

## **1.2 Episodic memory**

Declarative memory (Cohen & Squire 1980; Scoville & Milner 1957; Squire 1992; Squire et al. 2004), also referred to as propositional or explicit memory (Tulving 1983); Schacter 1996), is the memory for consciously expressible facts and events. In contrast, nondeclarative memory, also referred to as procedural or implicit memory, encompasses nonconscious memory such as that for skills, habits, conditioning and priming (Cohen & Squire 1980; Corkin 2002; Squire 1992; Squire et al. 2004). In a seminal essay published in 1972 (Tulving 1972), Tulving presciently proposed that declarative memory be further split into two distinct subtypes, episodic and semantic memory, and in so doing profoundly affected the direction of memory research (Tulving 2002b).

He defined episodic memory as the memory for the autobiographical episodes or specific events in our lives, which necessarily includes information about both the content of the experience and the spatial-temporal context in which it occurred - the “what, when, where” in Tulving’s classic taxonomy (Tulving 1983; Tulving 2002b). Recollection of such a memory entails a rich re-experiencing of the past event, effectively allowing one to mentally travel back in time (Tulving 1985). Tulving also proposed that recollection was accompanied by a specific kind of consciousness which he termed autonoetic consciousness, that enabled one to be aware of the self in subject time (Tulving 1985; Tulving 2002b). He went on to posit that the subjective awareness of time may be akin to a special, and separable, sense which he dubbed ‘chronesthesia’ (Buckner & Carroll 2007; Tulving 2002a; Tulving 2002b).

Semantic memory by contrast refers to acontextual factual knowledge about the world acquired during an experience, or across experiences, which then becomes separated from the specific context of the learning event itself (Tulving 2002b). This type of memory contributes to the formation and long-term representation of abstracted knowledge such as concepts, categories, facts, and word meanings. It also includes knowledge about ourselves (e.g. where we were born, where we lived, who our friends were), which some have referred to as personal semantics (Cermak & O'Connor 1983; Kopelman et al. 1989) in order to distinguish this type of memory from autobiographical episodes.

Tulving initially developed the episodic-semantic memory distinction as a useful psychological heuristic (Tulving 1972; Tulving 2002b). It was only later (Tulving 1983) that it became apparent that the distinction might also correspond to biological reality (Tulving 2002b). He therefore, controversially (see (Squire et al. 2004)), proposed that that episodic and semantic memory represented two functionally separable memory systems (Tulving 1983; Tulving 2002b). Episodic memory is thought to be a recently-evolved, late-developing, and early-deteriorating memory system (Tulving 2002b). It was suggested to be more vulnerable than other memory systems to neuronal dysfunction because of its complexity and, somewhat contentiously, Tulving claimed it is probably unique to humans (Tulving 2002b). Regarding this latter point, it should be noted that a series of elegant studies in scrub-jays (Clayton et al. 2003; Raby et al. 2007) have shown that non-humans can at least have 'episodic-like' memories, with knowledge of the what, when, and where of an event, although possibly without auto-noetic consciousness, an issue difficult to address in the absence of verbal abilities (Clayton et al. 2003; Raby et al. 2007; Suddendorf & Busby 2003; Suddendorf & Corballis 1997).

In terms of operationalising episodic memory, a wide range of tasks have been used in memory research that have claimed to be episodic, for example simple word-pair association tasks and personal semantic recall tasks (see (Steinvorth et al. 2005)). It is not clear however, if these types of tasks are truly episodic and tap into all of the processes involved in autobiographical memory (Cabeza & St Jacques 2007; Gilboa 2004; Maguire 2001a). Therefore, in the work undertaken for this thesis, I have largely employed rich and complex ecologically valid stimuli. Thus I shall refer to episodic

memory and autobiographical memory interchangeably in the text (although see (Gilboa 2004) for a review of the differences).

Of course the use of rich autobiographical memories as complex experimental stimuli presents numerous challenges. The difficulties in quantifying personal memories and assessing their retrieval are well-documented (Hodges 1995; Maguire 2001a; Warrington 1996), not least of which is the fact that by definition autobiographical memories are unique to an individual subject and therefore not easily generalisable. Controls need to be put in place for a host of often overlooked but potentially confounding factors including vividness, detail, recency, emotional salience, self-relevance and accuracy of the memories retrieved (Maguire 2001a). One commonly used testing method is the Galton-Crovitz word cue procedure whereby autobiographical memories are elicited from subjects in response to single word cues (Crovitz & Schiffman 1974). However, this relatively unconstrained approach leads to a high degree of variability between subjects across all the factors listed above (Hodges 1995; Maguire 2001a).

The Autobiographical Memory Interview (AMI) (Kopelman et al. 1989) was developed to standardise assessment in clinical contexts where subjects are questioned about memories from specific time periods, as well as personally relevant factual or semantic memories. To quantify the episodic qualities of a memory the AMI relies on a simple 3-point scoring system which has been criticised in some quarters (Levine et al. 2002) for its lack of granularity and its inability to measure the constituent factors that underpin episodic recollection. More sensitive scoring systems such as the Autobiographical Interview (Levine et al. 2002) have recently been developed that try and capture some of the subtle nuisances of episodic memory recall, but as yet there is no widely agreed standard adopted across all research groups (Bayley et al. 2005; Bayley et al. 2003; Squire et al. 2004).

For neuroimaging studies of episodic memory there are additional hurdles to be overcome. The effect of interview protocols prior to scanning on the subsequent imaging results has been questioned (Gilboa et al. 2004). Protocols and procedures vary widely between studies in terms of the amount of experimenter-subject interaction, the precise task instructions during scanning, and the amount of time elapsed between the interview

and scanning session (Maguire 2001a). Moreover, the status of the memory elicited via the interview session has to be considered. Does questioning subjects before scanning change the essential nature of the memory one wishes to examine? Is the memory now recent as opposed to what might have been a remote memory? Has the event been re-encoded and thus in some sense no longer the original event when recalled during scanning? Elegant (although labour intensive) procedures have been developed to mitigate against some of these issues like the use of confederates such as spouses to glean the information required for the scanning cues, and getting participants to record events for later experimental use (Gilboa et al. 2004; Ryan et al. 2001).

Then there are experimental design issues such as how long to give subjects for full recollection, and what comparison task to use - for example rest is known to activate many of the same regions as episodic memory recall including the hippocampus, so is not an appropriate baseline task (Maguire 2001a; Raichle et al. 2001). Thus the substantial challenges of using autobiographical memories as experimental stimuli was one of the many considerations during this thesis that led to the development of novel paradigms that tapped into autobiographical memory-like processes but were more constrained and easier to manipulate (see Chapters 4 and 5).

### **1.3 Reconstructive theories of episodic memory recall**

A rich recollective experience is a defining characteristic of episodic memory recall (Tulving 2002b). Recollection of this type of complex memory is widely accepted to be a (re)constructive process as opposed to the simple all-or-nothing retrieval of a perfect 'holistic' record (Bartlett 1932; Conway & Pleydell-Pearce 2000; Payne et al. 2002; Rubin et al. 2003; Schacter et al. 1998). Errors and inconsistencies such as intentional forgetting, confabulation (Burgess & Shallice 1996), misattribution and suggestibility are known to distort memories and can lead to incidents such as false witness statements (Jacobs & Nadel 1998; Schacter 2001; Weingardt et al. 1995). Such memory imperfections are important because they provide evidence that memory is not a literal reproduction of the past, but rather is a constructive process in which elements of information from various sources are pulled together (Bartlett 1932; Schacter et al. 1998). Episodic memory recall therefore can be conceptually divided into a number of component processes including a sense of subjective time (Tulving 2002b), connection

to the self (Conway & Pleydell-Pearce 2000; Gallagher 2000), narrative structure (Rubin et al. 2003), retrieval of relevant semantic information (Gottfried et al. 2004; Wheeler et al. 1997; Wheeler et al. 2000), feelings of familiarity (Wagner et al. 2005), and rich multimodal re-experiencing of the event (Tulving 2002b).

Constituent features of a memory, stored in their modality-specific cortical areas (Deichmann et al. 2003; Wheeler et al. 2000), are distributed widely across different parts of the brain, such that no single brain region contains a literal trace or engram that corresponds to the whole of a specific experience (Squire et al. 2004). Note, though, that this view does not preclude the idea of a 'simple' hippocampal memory trace or index that stores a distilled memory representation that connects to the wider cortical elements of the full memory (see Section 2.2)(Marr 1971; Squire et al. 2004). Retrieval of a past experience then involves the phenomenon of pattern completion (Marr 1971; McClelland et al. 1995), whereby a memory is pieced together from a set of distributed features that comprise a particular past experience (Schacter et al. 1998). Pattern completion describes the process by which a partial input (e.g. the sound of the sea) is transformed into the recall of the entire stored experience (e.g. an ocean voyage). This process is thought to be mediated by the intrinsic properties of auto-associative memory representations stored in the hippocampus (see Section 2.2), with perhaps the most famous example of pattern completion being Proust's remembrance of a childhood memory triggered by the taste of a Madeleine cake (Proust 1927).

Since a constructive memory system is prone to error, it must solve many problems to produce sufficiently accurate representations of a past experience. For example, the disparate features that constitute an episode must be linked or bound together at encoding. Failure to adequately bind together appropriate features can result in the common phenomenon of source memory failure, where people retrieve fragments of an episode but do not recollect, or misrecollect, how or when the fragments were acquired, resulting in various kinds of memory illusions and distortions (Schacter 2001). Furthermore, bound episodes must be kept separate from one another in memory (McClelland et al. 1995) otherwise if episodes overlap extensively, an individual may recall the general similarities or gist (Schacter & Slotnick 2004) common to episodes, but fail to distinguish one episode from another (Bartlett 1932; Schacter et al. 1998). Similarly, retrieval cues can potentially match with a number of stored experiences and

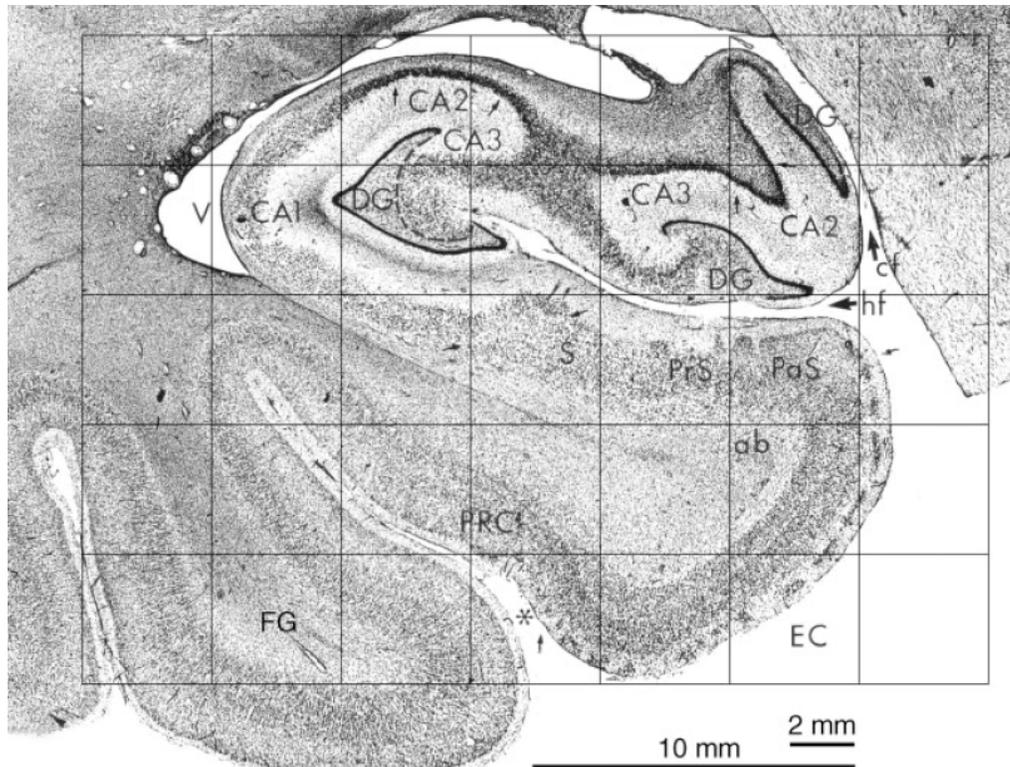
so retrieval often involves a preliminary stage, sometimes called retrieval mode (Rugg & Wilding 2000), in which guidance and orientation over the retrieval process is thought to be exercised by prefrontal cortex (Burgess & Shallice 1996; Schacter et al. 1998). Breakdowns in this process of formulating a retrieval description as a result of damage to the frontal cortex and other regions can sometimes produce striking memory errors, including confabulations regarding events that never happened e.g. (Burgess & Shallice 1996). All this serves to underscore just how complex it is trying to understand episodic memory recall.

From a computational perspective, reconstructing a memory from its components is more efficient in terms of storage capacity than the alternative of storing each memory separately as an intact record (Schacter 2001). Furthermore, this kind of storage structure lends itself conveniently to making abstract inferences and generalising across distinct experiences (McClelland et al. 1995; Norman & O'Reilly 2003). From an adaptive point of view, a flexible constructive memory system may confer other important advantages that I will discuss later. Episodic memory recall then, seems to be the result of a constructive process, but how exactly this is manifested cognitively and neurally is still not well understood.

## **1.4 The brain network underpinning episodic memory**

### **1.4.1 The hippocampus**

Throughout this thesis when referring to the hippocampus and MTL I will use the anatomical terminology of Amaral (1999). In this definition the hippocampus includes the hippocampus proper (fields CA1-CA4) and the dentate gyrus; the hippocampal formation includes the hippocampus, the entorhinal cortex and the subicular complex; and the medial temporal lobe includes the hippocampal formation, the perirhinal cortex (anterior parahippocampal gyrus) and the parahippocampal cortex (posterior parahippocampal gyrus) (Figure 2) (Amaral 1999).



**Figure 2. Nissl-stained coronal section through human hippocampal formation**

Abbreviations: CA1, CA2, CA3 - fields of the hippocampus. DG, dentate gyrus. PRC, perirhinal cortex. EC, entorhinal cortex. FG, fusiform gyrus. S, subiculum. PaS, parasubiculum. PrS, presubiculum. V, ventricle. cf, choroidal fissure. hf, hippocampal fissure. From (Amaral 1999).

The cytoarchitecture of the hippocampus is not suggestive of any overall anatomical functional structure beyond the broad subfield boundaries (Suzuki & Amaral 1994; Witter & Amaral 1991; Witter et al. 1989). However, at least in monkeys, there does appear to be some systematic structure in the pattern of connections from entorhinal cortex, the main input into the hippocampus, to the hippocampus, with local subdivisions of the entorhinal projecting to transverse sections of the hippocampus (Witter & Amaral 1991). Some have even suggested that the hippocampal cortex may be organised in parallel lamellae, both with regard to the neuronal and the vascular system. By means of this lamellar organization, small strips of the hippocampal cortex may operate as independent functional units, although excitatory and inhibitory transverse connections may modify the behaviour of the neighbouring lamellae (Andersen et al. 1971).

In humans, functional connectivity analysis on neuroimaging data acquired during the resting state (Buckner et al. 2008; Raichle et al. 2001) have shown that the anterior hippocampus is functionally highly correlated with entorhinal cortex, perirhinal cortex and lateral temporal cortex extending into the temporal pole. By contrast the body of the hippocampus is functionally highly correlated to midline posterior regions such as retrosplenial cortex, posterior cingulate cortex, as well as lateral parietal cortex and ventromedial prefrontal cortex (Kahn et al. 2008; Vincent et al. 2006).

The breakthrough in discovering the critical role played by the hippocampus and MTL in memory was made in 1957 with the famous case of patient HM described by Scoville and Milner (Scoville & Milner 1957). HM was rendered densely amnesic as a result of bilateral surgical resection of his MTL to treat medically intractable epilepsy (Corkin 2002; Scoville & Milner 1957). It was immediately evident following surgery that HM had a very profound impairment of recent memory in the apparent absence of other intellectual loss such as executive function, perception, language comprehension and lexical knowledge. For example he could not remember what he had for breakfast, and he could not find his way around the hospital or recognise members of hospital staff (Corkin 2002). Although HM's lesions encompassed the wider MTL, Scoville and Milner suggested that it was damage to the hippocampus, as opposed to the other structures that were excised (although see (Aggleton & Brown 1999)), that was the primary source of HM's amnesia, and they concluded that the severity of amnesia might be related to the size of the hippocampal removal (Corkin 2002; Nadel & Moscovitch 1997; Scoville & Milner 1957). This view was derived, in part, from converging evidence from two other cases that were described by Penfield and Milner (Penfield & Milner 1958). Both of their patients became amnesic after a left temporal lobectomy that included a large hippocampal removal, and both showed concomitant electrographic abnormality on the right side, indicating that, like HM's, their amnesia possibly resulted from a bilateral lesion. This hypothesis was confirmed in one of the patients who later came to autopsy (Penfield & Mathieson 1974).

HM's anterograde amnesia was near total, leaving him unable to acquire any new information that could be later recalled, although he could hold immediate impressions in his mind provided interfering activity did not distract his attention, leading to conclusions

that short-term memory is intact in hippocampal amnesia (Squire et al. 2004)(but see (Hannula et al. 2006)). In contrast, old memories from his childhood seemed to be intact (Scoville & Milner 1957; Squire 1992) although this finding was later disputed when HM was retested under more stringent conditions (Steinvorth et al. 2005) that showed his retrograde amnesia for episodic memories was much more extensive. The issue of temporally-graded retrograde amnesia is highly controversial (Bayley et al. 2003; Moscovitch et al. 2005; Squire et al. 2004) and although not the main focus of this thesis, is a subject I return to in more depth in later sections (see Sections 2.3, 2.4 and 8.9).

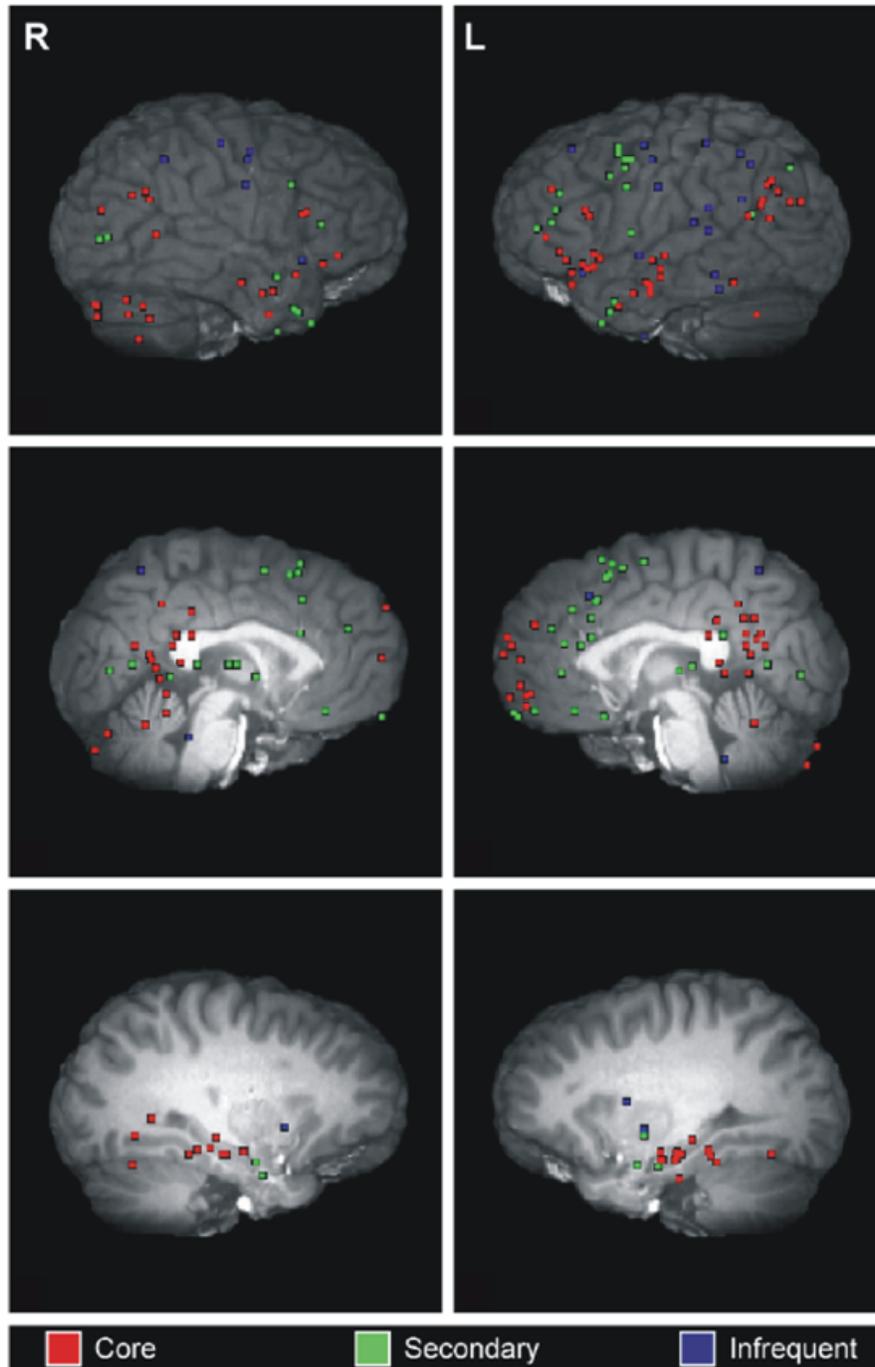
Further neuropsychological characterisation of HM revealed striking dissociations in the memory patterns arising from his MTL lesions. Despite his severe amnesia HM performed well on certain types of tasks where learning of new procedural information was required. The first experimental demonstration of this preserved learning was Milner's report that HM's performance on a mirror-tracing task improved over the course of three days (Corkin 2002). HM learned this new skill despite having no recollection at the end of learning of ever having performed the task (Corkin 2002). Subsequent work showed that motor skills were merely a subset of the learning and memory abilities spared following hippocampal lesions (Cohen & Eichenbaum 1993; Cohen & Squire 1980; Squire 1992). For instance, patients with hippocampal amnesia generally exhibit relatively preserved performance on tasks tapping into perceptual learning, classical conditioning, and repetition priming (Cohen & Eichenbaum 1993; Squire 1992). These findings paved the way for ideas about multiple memory systems (Hirsh 1974; Nadel & O'Keefe 1974; Tulving 1972) (e.g. procedural versus declarative) supported by separable neural systems only some of which were dependent on the hippocampal complex (Cohen & Eichenbaum 1993; Cohen & Squire 1980; Squire 1992).

Since HM there have been many cases documented in the literature of amnesia occurring in the presence of hippocampal damage (for a comprehensive review see (Spiers et al. 2001)). Sensitive structural imaging techniques now allow for better characterisation of patient lesions and seem to confirm that selective damage to the hippocampus bilaterally is sufficient to cause profound amnesia (Spiers et al. 2001)(although it is still not possible to rule out the possibility of 'hidden damage' and disconnection – (Gaffan 2002; Squire et al. 2004; Squire & Zola 1996)). The question of

lateralisation is still unresolved with some evidence that damage to the left hippocampus may be enough to cause amnesia (Spiers et al. 2001). Functional neuroimaging in healthy control subjects have also hinted at a left lateralised bias in the role of the hippocampus in episodic memory, although the evidence is far from conclusive (Maguire 2001a; Svoboda et al. 2006). The large number of amnesic cases described in the literature with damage including the hippocampus reflects the susceptibility of the hippocampus to a range of pathological factors. Hippocampal cell damage can occur in a number of ways: loss of oxygen to the cells (anoxia, ischemia), physical damage from surgery or head injury, viral attack (meningitis, encephalitis), and autoimmune responses (e.g. systemic lupus erythematosus). Of these aetiologies, encephalitis, ischemia/anoxia and surgery are the most common. Surgery and encephalitis account for many of the cases with additional temporal lobe lesions, whereas ischemia/anoxia accounts for more of the selective cases (Spiers et al. 2001).

#### **1.4.2 The episodic memory network**

Whilst critical, the hippocampus does not support episodic memory on its own. With the advent of neuroimaging technology, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), it has become possible to observe the entire brain network underpinning episodic memory (Maguire 2001a; Svoboda et al. 2006). Neuroimaging studies of autobiographical memory have revealed a highly consistent and distributed network of brain regions, including the hippocampus, involved in supporting recollection (Maguire 2001a). In addition to the hippocampus, this network comprises dorsal and ventromedial prefrontal cortex (PFC), lateral prefrontal cortex, parahippocampal gyrus, lateral temporal cortices, temporo-parietal junction, thalamus, retrosplenial cortex (RSC), posterior cingulate cortex (PCC), precuneus, and cerebellum (Figure 3)(Maguire 2001a; Maguire & Frith 2003; Svoboda et al. 2006). Despite over a decade of activating this robust network, however, surprisingly little is understood about the contributions individual brain areas make to the overall recollective experience.



**Figure 3. The episodic memory network**

Significant peaks of activity from a meta-analysis of 24 neuroimaging studies of autobiographical memory (from (Svoboda et al. 2006)). The classic core episodic memory network can be seen in red and includes the hippocampus bilaterally, parahippocampal gyrus, retrosplenial, posterior cingulate and posterior parietal cortices and medial prefrontal cortex. Activations in core, secondary and infrequently reported regions are depicted across right and left, lateral, medial and subcortical planes.

For example, numerous patient and neuroimaging studies have demonstrated the importance of prefrontal cortex in episodic memory (Simons & Spiers 2003; Wheeler et al. 1997) but there is little consensus as to what processes the subregions of the prefrontal cortex specifically support in episodic memory, with ideas ranging from self-awareness to autonoetic consciousness to control processes (Rugg & Wilding 2000; Wheeler et al. 1997). The retrosplenial cortex too is almost always implicated in neuroimaging studies of autobiographical memory and focal damage to this area can cause amnesia (Maguire 2001b). Because of the retrosplenial cortex's critical role in spatial navigation (Maguire 2001b) and the fact that it is anatomically on the critical temporo-parietal pathway it has been speculated that it might play a role in viewpoint transformation from allocentric to egocentric spatial representations mediated by the MTL and parietal cortex respectively (Burgess 2002; Burgess et al. 2001). These examples serve to highlight the tantalising first steps being made towards understanding the complexities of episodic memory. As yet, however, a complete account of the episodic memory system encompassing all facets of the component cognitive processes and numerous brain regions involved and how they are so seamlessly integrated has so far proven elusive.

## **1.5 Hippocampal population codes**

In the previous section I reviewed briefly the patient and neuroimaging studies that have shown that the hippocampus is critical to episodic memory (Maguire 2001a; Spiers et al. 2001). Any comprehensive account of the hippocampal function in memory must necessarily tie the high level processes discussed previously to the neurobiological workings of the brain at a neuronal level. It has been suggested that when an event is experienced, a memory trace is established in the hippocampus that points to the cortical representations that contain the attended information comprising the event (see Chapter 2) (Marr 1971; Moscovitch et al. 2005; Squire et al. 2004). It is this memory trace that is then re-invoked during episodic memory recall which in turn reactivates the cortical representations via reciprocal connections (see Chapter 2). But what are the properties of these hippocampal memory traces at the neuronal level? What information does a single hippocampal neuron in a memory trace represent and how does this relate to the kinds of information encoded at a population level? How are these hippocampal

population codes organised, for instance how large are they and how robustly do they co-activate?

In this section I will briefly examine the literature covering some of these important theoretical questions starting with a summary of how information is thought to be coded in the brain in general before concentrating on the types of representations thought to be processed by the hippocampus specifically.

### **1.5.1 Information in the brain**

The brain comprises billions of neurons, which together form the world's most powerful information-processing machine. Information is encoded in the brain by populations or clusters of cells, rather than by single cells (Buzsaki 2004; Harris 2005; Pouget et al. 2000). Despite decades of research into the nature of the neural code, the fundamental principle by which these cells work together is still unknown (Harris 2005). There are two main schools of thought on the unit of computation used by the brain to transmit and store information. In a rate code, the only variable that neurons use to convey information is instantaneous firing rate, which is typically characterised by spike rate within a certain 'encoding time window' (deCharms & Zador 2000). Alternatively, in a temporal code, the exact timing of spike sequences also plays a part in information transmission. From a theoretical perspective, temporal coding has advantages. The set of all spike sequences is much larger than the set of instantaneous rates, and a neuron that distinguishes between sequences could therefore transmit a larger number of possible signals.

It is less clear, however, how a temporal code could be interpreted by downstream neurons. Neurons can be described, at least to first approximation, as 'leaky integrator' devices, in which the arrival of sufficient excitatory afferents within a time window - set by the width of an excitatory postsynaptic potential (EPSP) - leads to output spikes (Magee 2000). This scheme therefore seems better adapted to read a rate code, with the encoding time window equal to the EPSP width. Direct measurement of EPSP width is difficult in behaving animals. However, in vivo studies using anaesthetised animals and in vitro measurements of related parameters such as the membrane time constant indicate an approximate range of 10–30 milliseconds in adult pyramidal cells (Harris

2005), with the precise value depending on the level of inhibition and synaptic background activity (Harris 2005).

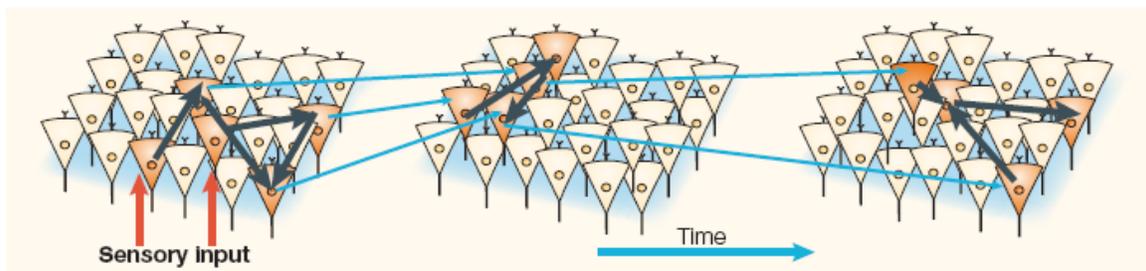
In sensory experiments it is often assumed that the primary function of a neural population code is to convey information about a stimulus. However, there is considerable evidence that the firing pattern of neurons reflects not just the physical nature of the stimulus, but also complex internal neural dynamics (Kosslyn et al. 1995; Kreiman et al. 2000). Indeed cortical circuits exhibit complex dynamics even in the absence of sensory stimuli (Kenet et al. 2003; Tsodyks et al. 1999). Ongoing activity that precedes sensory stimulation plays an important part in shaping neural activity during stimulus presentation, which indicates that it might be more accurate to regard sensory stimuli as modulating ongoing neural dynamics, rather than deterministically controlling firing patterns (Harris 2005).

Observations that psychological processes can be dependent on the internal state have recently led to a resurgence of interest in the cell assembly hypothesis (Harris 2005), first pioneered by Hebb in the 1950s (Hebb 1949). This proposes that the elemental currency of information is the firing of an organised cell assembly, not the behaviour of a single neuron (Hebb 1949). It has been recently suggested, somewhat controversially, that many findings that have been interpreted as evidence for temporal coding, such as spike train synchronicity and variability (Harris et al. 2002)(but also see (Huxter et al. 2003; Mehta et al. 2002; O'Keefe & Recce 1993), might in fact reflect the temporal dynamics of an underlying organisation of cell assemblies that use rate coding as their primary mode of communication (Harris 2005; Harris et al. 2003).

### **1.5.2 The cell assembly hypothesis**

The cell assembly hypothesis proposes that the key to internal cognitive processes lies in the recurrent nature of neuronal circuits. The theory is based on the premise, known as Hebb's rule, that synaptic connections become strengthened by synchronous activity of pre-synaptic and post-synaptic neurons (Hebb 1949). Repeated co-activation of a group of neurons during behaviour was proposed to cause the formation of a cell assembly – an anatomically dispersed set of neurons among which excitatory connections have been potentiated. Mutual excitation allows the cell assembly to later

maintain its activity, without requiring continuous sensory stimulation. Consequently, the activity of assemblies can become decoupled from external events, and can be initiated by internal factors such as the activity of other assemblies. A chain of assemblies, each one triggered by the last, is termed a ‘phase sequence’ (Harris 2005). The phase sequence allows for complex computations, which are only partially controlled by external input, and is the proposed substrate of an internal cognitive process (see Figure 4). A crucial feature of this theory is that the same assembly might be triggered by either sensory or internal factors (Harris 2005).



**Figure 4. The cell assembly hypothesis**

Sensory input (red arrows) drives certain neurons to fire. Thereafter, activity evolves owing to intrinsic cortical dynamics. Strengthened recurrent connections between members of a single assembly (black arrows) transiently stabilise assembly firing through mutual excitation. As the excitability of this assembly fades, inter-assembly connections (blue arrows) lead to subsequent activation of a new assembly. The resulting ‘phase sequence’ evolves through network dynamics and is not strictly determined by the time series of sensory inputs. The evolution of this phase sequence is the hypothesised substrate of internal cognitive processes. From (Harris 2005).

Although a phase sequence of cell assemblies might result in spike trains with properties superficially similar to those outputted by a temporal coding system, they are markedly different conceptually. In a temporal coding scheme, a local neural population uses precise patterns of spike times to transmit information to downstream targets. The downstream targets must then decode this temporal pattern to extract the information. This is in contrast to an assembly organisation, where sequences of assemblies will occur thus potentially creating what looks like a temporal structure in firing rates when local populations of cells are considered. However, in this system the fundamental currency of information processing is the firing of a single assembly, not the sequence or the time between spikes. The progression of assemblies in the phase sequence represents successive steps in a serial computation that is carried out by a larger

network, including both the local population and its targets. Critically, post-synaptic neurons do not need to 'decode' the phase sequence thus resolving a key problem with the temporal coding scheme (Harris 2005).

### **1.5.3 Invariant representations in the hippocampus**

Given the focus in this thesis on understanding episodic memories and the hippocampal memory traces underpin them, it is therefore pertinent to ask what kinds of information are represented by population codes in the hippocampus specifically. The best characterised example of high-level information stored by hippocampal neurons came with the seminal discovery of 'place cells' in the rat hippocampus (O'Keefe & Dostrovsky 1971; O'Keefe & Nadel 1978). These are neurons that encode the location of a freely moving rat within its environment, firing whenever it enters a restricted portion of the environment (the 'place field') independently of local sensory cues (O'Keefe & Dostrovsky 1971; O'Keefe & Nadel 1978). In open environments the firing of these place cells is independent of the orientation of the rat (Muller et al. 1994).

Place fields within the hippocampus are usually distributed in such a way as to cover all locations in the environment, but with considerable spatial overlap between the fields. As a result, a large population of cells will respond to any given location (O'Keefe & Nadel 1978). Immediate early gene imaging studies have confirmed electrophysiological findings (O'Keefe & Nadel 1978) that ~35% of neurons in the CA1 hippocampal subfield are active in a given environment (Guzowski et al. 1999; Redish et al. 2001; Vazdarjanova & Guzowski 2004). This is quite a high proportion and likely means that the population codes representing spatial memories are relatively large (Pouget et al.). Spatial memory is not just important for tasks such as navigation but also closely entwined with episodic recollection through the provision of the spatial context for a memory (Burgess 2002; Burgess et al. 2001). I will return to place cells and spatial memory in more detail in Section 2.5 and in Chapters 4 and 8.

Single cell recording from the hippocampus and MTL in humans have revealed cells that code for a range of different stimuli from animals to faces to objects (Kreiman et al. 2000). These neurons appear to be highly selective, activating in response to only a small percentage of the stimuli presented (Kreiman et al. 2000). Interestingly, the same

cells that responded to a visually presented stimulus also responded when patients imagined the previously viewed image in the absence of any sensory input (Kreiman et al. 2000). Thus it would appear that these neurons code for an abstract representation of the stimuli not directly linked to sensory perception (Kreiman et al. 2000) thus echoing some of the invariant properties of place cells (O'Keefe & Nadel 1978). More evidence for high-level abstract representations in the MTL came in a recent follow-up study where neurons were found to respond to multiple views of the same stimulus and in some cases even when that stimulus was presented in a different modality such as a letter string containing its name (Quiroga et al. 2005). These results therefore suggest an invariant, sparse and explicit code in the hippocampus that represents highly processed conceptual information in an abstracted manner.

#### **1.5.4 Properties of the hippocampal population code**

The convergent findings from single cell recording in humans and rats of invariant and abstract representations of spatial and visual information in the hippocampus and MTL are critical for our understanding of memory but what does it mean for the properties of the hippocampal population code itself?

##### *1.5.4.1 Sparse codes*

The consensus view is that the representations coded by hippocampal neuronal populations are sparse (e.g. (Redish et al. 2001)). Sparse codes have several theoretical advantages from a computational point of view (Waydo et al. 2006). Firstly, they mitigate against catastrophic interference between representations of similar memories (see Section 2.2) and secondly they allow auto-associative memory networks to store large numbers of representations (Treves & Rolls 1991), both desirable properties of a memory system.

However, sparseness has been used to describe a wide range of different representational scales from rather biologically implausible single 'grandmother' cells (Gross 2002)(but see (Pouget et al. 2000; Quiroga et al. 2005) for arguments against this view) to more than two million cells in other accounts (Waydo et al. 2006). In fact, recently detailed calculations were carried out by Waydo et al (Waydo et al. 2006) on the single-cell recoding data from the Quiroga et al study (Quiroga et al. 2005) to try quantify these rather vaguely defined terms. The Quiroga et al study (Quiroga et al. 2005)

revealed the existence of highly selective cells that, for example, respond strongly to different images of a single celebrity but not to 100 pictures of other people or objects (Quiroga et al. 2005). By applying probabilistic reasoning, they estimated that around 0.54% of neurons in the MTL were activated by a typical stimulus (Waydo et al. 2006). Furthermore, if we assume that a typical adult recognises between 10,000 and 30,000 discrete objects (Biederman 1987), a sparseness of 0.54% implies that each neuron fires in response to 50–150 distinct representations (Waydo et al. 2006). This is indeed a 'sparse' representation. However, assuming there are the order of  $10^9$  neurons in the human MTL (Andersen et al. 2007; Henze et al. 2000), this still corresponds to a large number of neurons (~5 million) being activated by a typical stimulus (Waydo et al. 2006).

#### *1.5.4.2 Topographical organisation*

The hippocampus exhibits a coarse topographical organisation in its intrinsic and extrinsic connections (Amaral 1993), and broad septotemporal gradients of encoding have also been documented (Jung et al. 1994). However, the firing patterns of neighbouring cells tend to be uncorrelated under a variety of conditions (Guzowski et al. 1999; O'Keefe & Nadel 1978; Redish et al. 2001). In fact, it has generally proven impossible to predict the firing correlates of a given hippocampal cell in one context from its properties in other contexts with dissimilar spatial cues (O'Keefe & Nadel 1978; Redish et al. 2001). This independence of firing correlates across contexts is thought to be incompatible with a large-scale tendency for anatomically neighbouring neurons to exhibit correlated firing (Redish et al. 2001).

Two studies, however, have suggested that there might be some degree of clustering of firing rate correlations within the hippocampus (Eichenbaum et al. 1989; Hampson et al. 1999). Hampson et al (Hampson et al. 1999) showed a particularly striking periodic laminar topography in the hippocampus when recording from rats performing a spatial delayed-nonmatch-to-sample (DNMS) short-term memory task. Separate segments of the hippocampus were found to have clusters of cells that appeared to code for the spatial and non-spatial aspects of the task (Hampson et al. 1999). Results from these studies (Eichenbaum et al. 1989; Hampson et al. 1999) reporting the clustering of cells with overlapping spatial selectivities into wide bands has been interpreted as evidence for a global topographical organisation of the neuronal code across the hippocampus (Hampson et al. 1999).

However, the Hampson et al (1999) result has since largely been discredited due to issues with the tetrode registration techniques used between animals (Redish et al. 2001). A rebuttal study found that there was no indication that the firing correlates of hippocampal cells located anatomically near each other were any more correlated than cells that were situated far apart, or indeed any more correlated than predicted by chance (Redish et al. 2001). It is thus unlikely that the inputs to neighbouring hippocampal cells are significantly correlated or that the presence of low-resistance gap junctions in those same cells leads to a significant increase in correlated firing, at least during active behaviour (Redish et al. 2001). Therefore the current consensus is that hippocampal representations do not have an explicit topographical organisation and instead these population codes consist of neurons that are randomly and uniformly distributed (O'Keefe et al. 1998; Redish et al. 2001).

## **1.6 Thesis overview**

The primary aim of this thesis is to advance the conceptual understanding of the cognitive processes underpinning episodic memory and to establish the neural bases that support them, with a particular emphasis on the hippocampus.

I sought to characterise the mechanisms underpinning episodic memory by adopting a new approach to deconstructing episodic memory into its constituent processes. Episodic memory recall is thought to be a reconstructive process rather than a literal reproduction of the past. Therefore, reasoning that constructive and reconstructive tasks might share processes in common, I devised a novel 'construction' task, namely richly imagining new experiences, as a well-matched comparison task for episodic memory recall.

In Experiment 1 I tested patients with focal damage to the hippocampus bilaterally on imagining new experiences. This study revealed for the first time a surprising deficit in hippocampal patients on this primarily non-mnemonic task. Furthermore, this deficit appeared to be driven by an inability to construct and maintain a spatially coherent scene, a process I called 'scene construction'. Experiment 2 built on these patient findings with an fMRI study involving healthy subjects engaged in imagination and

memory recall tasks. By comparing and contrasting the recall of real and imaginary memories I established the brain network that underpinned scene construction, confirming the involvement of the hippocampus, and demonstrated that it could account for a large portion of the episodic memory network. Moreover, I identified two brain regions, the anterior medial prefrontal cortex and posterior cingulate cortex, which distinguished real from imaginary memories and thus might be responsible for giving one the phenomenological feeling that an event really happened. Recently it has been suggested that a 'core' network may support a common process involved in many disparate cognitive functions including navigation, future thinking, and mind wandering. In the general discussion, I present evidence to suggest that scene construction may be a good candidate for that common core process.

No account of episodic memory could be complete without an understanding of the mechanics of memory at the neuronal level. One of the best characterised neural correlates of high level behaviour comes from the single cell recording of place cells in the rat hippocampus. In order to look at memory representations in healthy humans, a method is required that permits fine-grained information to be examined non-invasively using fMRI. In Experiment 3 I attempted to address this issue by combining high spatial resolution fMRI with sensitive multivariate pattern classification techniques whilst subjects were engaged in a simple navigation task in a virtual reality environment. Using multivariate classification methods I was able to accurately predict where a subject was standing in the environment solely from the activity patterns of voxels in the hippocampus. This result demonstrated that, contrary to current consensus, the hippocampal population codes for spatial representations must necessarily be large and have anisotropic structure. Experiment 4 successfully extended these novel analysis techniques to episodic memories and the hippocampal memory traces that underpin them. In this preliminary study I was able to accurately predict which of two real memories a subject was recalling at any one time in the scanner solely from patterns of activity in hippocampal voxels, perhaps pinpointing an individual memory trace for the first time.

In conclusion I discuss the work in this thesis in the context of other recent advances in the memory field. I believe that taken together, recent empirical data point to the need for a revision of the accepted role of the hippocampus in episodic memory. In the last

chapter I tentatively outline a new preliminary framework for the hippocampal role in episodic recollection and suggest how it may be possible to reconcile some of the more controversial findings in the memory literature.

This thesis contains work reported in the following peer-reviewed publications:

- **“Patients with hippocampal amnesia cannot imagine new experiences”** Hassabis D, Kumaran D, Vann SD, Maguire EA (2007). *Proceedings of the National Academy of Sciences USA*. 104, 1726-31.
- **“Using imagination to understand the neural basis of episodic memory”** Hassabis D, Kumaran D, Maguire EA (2007). *Journal of Neuroscience*. 27, 14365-74.
- **“Deconstructing episodic memory with construction”** Hassabis D, Maguire EA (2007). *Trends in Cognitive Sciences*. 11, 299-306.
- **“Decoding neuronal ensembles in the human hippocampus”** Hassabis D, Chu C, Rees G, Weiskopf N, Molyneux PD, Maguire EA (In press). *Current Biology*.
- **“The construction system of the brain”** Hassabis D, Maguire EA (In press). *Philosophical Transactions of the Royal Society, London: Series B*.
- **“Cortical midline involvement in autobiographical memory”** Summerfield JJ, Hassabis D, Maguire EA (2009). *Neuroimage*. 44, 1188-200.
- **“How does the brain construct scenes?”** Summerfield JJ, Hassabis D, Maguire EA (In preparation).

Work also undertaken during my PhD but not reported here:

- **“When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans”** Mobbs D, Petrovic P, Marchant JL, Hassabis D, Weiskopf N, Seymour B, Dolan RJ, Frith CD (2007). *Science*. 317, 1079-83.
- **“Impaired spatial and non-spatial configural learning in patients with hippocampal pathology”** Kumaran D, Hassabis D, Spiers HJ, Vann SD, Vargha-Khadem F, Maguire EA (2007). *Neuropsychologia*. 45, 2699-711.
- **“Choking on the money: Reward based performance decrements are associated with midbrain activity”** Mobbs D, Hassabis D, Marchant, J, Seymour B, Weiskopf N, Dolan RJ, Frith CD (In press). *Psychological Science*.
- **“Tracking the Emergence of Conceptual Knowledge during Human Decision Making”** Kumaran D, Summerfield JJ, Hassabis D, Maguire EA (submitted).
- **“Autobiographical memory in semantic dementia: a longitudinal fMRI study”** Maguire EA, Kumaran D, Hassabis D, Kopelman MD (submitted).
- **“From fear to panic: The neural organization of defensive systems in humans”** Mobbs D, Marchant JL, Hassabis D, Seymour B, Tan G, Gray M, Petrovic P, Dolan RJ, Frith CD (submitted).

# Chapter 2

*Theories of Hippocampal Function*

## 2.1 Introduction

As alluded to in the previous chapter, the hippocampus is essential to episodic memory, and therefore deserves a pivotal place in any discussion of memory function. In the 50 years since the discovery of the critical importance of the hippocampus (Scoville & Milner 1957), many theories have been proposed regarding its role in supporting memory; indeed it is one of the most intensively researched areas of cognitive neuroscience. In this chapter I will briefly summarise five of the most influential and prominent theories of hippocampal function that have proven useful in accounting for experimental phenomena and have been widely adopted and supported. This is not intended to be an exhaustive review of the hippocampal theory literature (for a more comprehensive summary and details of other theories, the reader is directed to: (Morris 2006)), but instead here I concentrate on the themes and the testable predictions made by these five major theories that relate most directly to the work undertaken in this thesis.

The work of Marr (Marr 1971) created a framework for many subsequent computational models and theories of hippocampal-neocortical function. The declarative theory, or standard model, of memory (Alvarez & Squire 1994; Squire 1992; Squire et al. 2004) built upon these ideas and created a distinction between the types of memory impaired (declarative) and spared (procedural) in patients with amnesia due to MTL damage. A key, and contentious, aspect of this theory was the temporally graded nature of retrograde amnesia, with the view that episodic memory gradually becomes independent of the hippocampus over time through consolidation. The declarative theory still dominates the memory field today.

The multiple trace theory (Moscovitch et al. 2006; Moscovitch et al. 2005; Nadel & Moscovitch 1997) was developed to challenge some of the basic tenets of the standard model that were suggested to be at odds with much of the empirical evidence. For instance, it was argued that two types of declarative memory, episodic and semantic, were differentially supported by the MTL and that in the case of episodic memory, retrograde amnesia was temporally ungraded and global in nature. The idea of multiple memory traces was invoked to reconcile these perceived discrepancies with a new hippocampally-mediated trace created each time an old memory was retrieved.

Other prominent accounts of hippocampal function have focused on the kinds of representations supported by the hippocampus. The cognitive map theory, on the basis of the discovery of place cells (O'Keefe & Dostrovsky 1971), argued for the primacy of space and the creation of allocentric maps as the primary function of the hippocampus. The relational theory (Cohen & Eichenbaum 1993) on the other hand concentrated on the key role of the hippocampus in creating associations in general and building these associations into flexible relational networks. By this view then spatial context is just a special instance of a relational network, albeit an important one.

Finally, I conclude with some ideas about how these seemingly opposing points of view might be reconciled that led to the studies undertaken in this thesis.

## **2.2 The Marr model**

It is hard to overestimate the influence Marr's work (Marr 1971) had on subsequent ideas, models and theories of hippocampal function. In laying down some of the fundamental principles of associative recall, Marr proposed that the hippocampus indexed an event through the formation of a 'simple representation' which was synaptically linked to the full representation in the neocortex. Thus, the simple representation itself was not viewed to contain the full details of the event but was nonetheless capable of reinstating it through the reactivation of neocortical representations. The rapid formation and storage of a memory index was achieved via Hebbian learning, a process whereby connections between coactive neurons strengthen (Hebb 1949) and are thought to be facilitated by synaptic long-term potentiation (Bliss 2006). This results in patterns of neuronal activity that are auto-associative, and it is precisely these kinds of patterns that Hebb referred to as an engram or memory trace (Hebb 1949).

The intrinsic circuitry of the hippocampus is ideally suited to supporting associative memories. The CA3 region in particular, with its high density of recurrent collaterals, is often modelled as an auto-associative network allowing the elements of a pattern to be associated together (Hasselmo & Schnell 1994; Marr 1971; Treves & Rolls 1992; Wills et al. 2005). Another benefit of an auto-associative system is that it is conducive to pattern

completion, a powerful process that allows the recovery of the entire hippocampal representation from just a partial cue (McClelland et al. 1995). An elegant study by Nakazawa et al (Nakazawa et al. 2002) provided empirical evidence that the CA3 region supports pattern completion. Mutant mice with highly selective loss of NMDA receptors in the CA3 subregion showed normal acquisition in the Morris water maze task but performed poorly on the spatial task under partial cue conditions when intact pattern completion processes were needed.

Other important features of Marr's model included the idea that the hippocampus was in fact a temporary store with the day's experiences transferred, via the process of consolidation, to the neocortex for permanent storage during sleep. Marr also regarded the simple representations to be sparsely encoded in the hippocampus, i.e. involving only a small, distributed population of neurons, in order to mitigate against the potential for interference between similar memories. The hippocampal indexing system as proposed by Marr (Marr 1971) is therefore highly efficient from a computational point of view and is strikingly reminiscent of the way hash table look-ups (Aho et al. 1983) are used to efficiently store and retrieve large amounts of data in computer science applications, an influence that is perhaps not altogether surprising given Marr's mathematical background.

## **2.3 Declarative theory**

The declarative theory (Cohen & Squire 1980; Squire 1992), or standard model, of memory takes many of the concepts in Marr's model and extends them to account for the pattern of impaired and spared performance in MTL-induced amnesia in humans and animals. The declarative theory splits memories into two broad types: declarative and nondeclarative. Declarative memory, thought to be critically dependent on the MTL and hippocampus, includes the capacity to recollect facts and events, and can be classed as the set of memories that are consciously accessible and expressible. In contrast, nondeclarative memory encompasses skills and habits, conditioning and priming, and is argued to be MTL-independent, supported instead by structures such as the cerebellum and striatum (Squire et al. 2004). Given that only the declarative memory system was disrupted in the amnesic syndrome, it was therefore concluded that only declarative

memory was dependent on the MTL/hippocampus (Cohen & Squire 1980; Squire 1992; Squire et al. 2004).

The declarative theory makes three principle claims about the MTL system: (1) it is primarily concerned with memory, (2) it operates with neocortex to establish and maintain long-term memory, and (3) ultimately, through a process of consolidation, memory becomes independent of the MTL. Another important feature of the declarative theory is its rejection of proposals that advocate a simple division of labour between the functions of the hippocampus and adjacent medial temporal cortex structures, such as the perirhinal cortex (Murray & Bussey 2001).

The declarative theory has its origins in the characterisation of the amnesic syndrome, beginning with HM in 1957 (Corkin 2002; Scoville & Milner 1957). The hallmark of the impairment following MTL lesions appeared to be profound multi-modal forgetfulness against a background of spared intellectual, perceptual and short-term memory abilities (Squire et al. 2004). However, each of these claims of spared function has been disputed with recent evidence of some impairment in perceptual abilities (Graham & Gaffan 2005; Lee et al. 2005b) and even short-term working memory deficits (Hannula et al. 2006) in hippocampal patients.

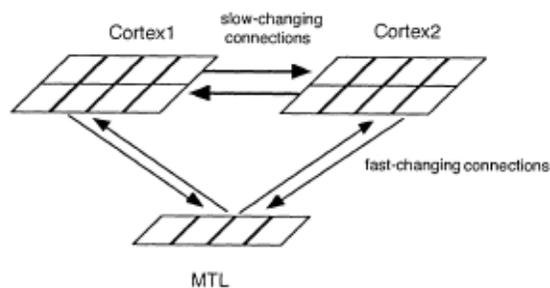
The subsequent memory impairment is widely regarded as primarily one of storage not retrieval with the hippocampus/MTL needed to establish representations in long-term memory (Squire 1992; Squire et al. 2004). However, the declarative theory, largely on the basis of claims of spared remote memory in MTL patients (Manns et al. 2003), takes the view that the MTL is not the permanent repository of memory but rather that this is elsewhere in the neocortex. Numerous animal studies have unequivocally shown the process of consolidation with typical time course of around 30 days. For example, in a study of trace eyeblink conditioning in the rat (Takehara et al. 2003), hippocampal lesions 1 day after learning nearly abolished the conditioned response, but the lesion had no effect after 4 weeks. In contrast, lesions of medial prefrontal cortex affected the conditioned response only marginally 1 day after learning but severely affected performance after 4 weeks and had an intermediate effect after 2 weeks (Squire et al. 2004).

Temporal gradients of retrograde amnesia have also been well described in patients with focal hippocampal damage (Bayley et al. 2003; Manns et al. 2003). Here, amnesia extends across a period of years rather than weeks and there is evidence for sparing of remote memory for facts (semantic memory) as well as remote memory for autobiographical events (Bayley et al. 2003). In one study, 8 patients and 25 age-matched controls responded to 24 cue words and produced the same number of remote memory recollections with a comparable number of details (Bayley et al. 2003). On the other hand several patient studies have found amnesia in hippocampal patients to extend globally in time, with an inability to recall events even from their early life (Cipolotti et al. 2001; Maguire et al. 2006a; Moscovitch et al. 2005; Steinworth et al. 2005). However the proponents of the standard model have largely dismissed these findings as due to hitherto undocumented damage outside the MTL (Squire et al. 2004). The large amount of neuroimaging evidence demonstrating that the hippocampus is involved in episodic retrieval irrespective of the age of the memory (Addis et al. 2004; Maguire 2001a; Maguire & Frith 2003; Svoboda et al. 2006) is similarly dismissed as most likely due to the encoding of new information connected with the task of remembering the event in the scanner rather than the retrieval of the old experience (Squire et al. 2004), despite the fact that re-encoding factors were controlled for in many of these studies e. g. (Gilboa et al. 2004; Maguire & Frith 2003).

Spatial memory too is regarded as only temporarily dependent on the hippocampus. Whilst advocates of the standard model, in the face of overwhelming empirical evidence (Hartley et al. 2003; Maguire et al. 1998; O'Keefe & Dostrovsky 1971), would accept that the hippocampus and parahippocampal gyrus have an important role to play in spatial memory, they view spatial memory as a subset, albeit a good example, of declarative memory. As such they consider spatial memories to be consolidated over time out of the hippocampus to neocortex just as with other types of declarative memory. This goes against the idea that the hippocampus is always critical for spatial memory as expressed by the cognitive map theory (see Section 2.5) and supported by findings in rats which require their hippocampus for navigation no matter how long ago the spatial memories were acquired (Clark et al. 2005; Mumby et al. 1999; Squire et al. 2004; Sutherland et al. 2001). The standard model maintains that this is not the case at least for humans (Squire et al. 2004). In support of this, EP, a patient with bilateral hippocampal damage, was shown to be able to recall the spatial layout of the region where he grew up and

from which he moved away as a young adult more than 50 years earlier (Teng & Squire 1999). However, the status of remote spatial memory in the context of bilateral hippocampal lesions may be more nuanced than suggested by the findings from EP (Rosenbaum et al. 2005), as sparing of coarse but not detailed remote spatial memory was reported in another hippocampal-damaged patient TT (Maguire et al. 2006a).

Another key aspect of the standard model is that it maintains there are no clear subdivisions of labour between different structures of the MTL. Greater activation in hippocampus has been reported in classic remember/know paradigms during fMRI for 'remember' responses (Yonelinas et al. 2002). This has led to the suggestion that the hippocampus codes for recollective aspects of declarative memory. However, this could also be explained by simple differences in the amount of information retrieved or the confidence in the response itself (Squire et al. 2004). Similarly, dissociations between associative memory and single-item memory have not provided sharp anatomical dichotomies, with fMRI signals often correlated in both the hippocampal region and the parahippocampal gyrus and in some cases entorhinal cortex also (see p.293 of (Squire et al. 2004) for more). It is also worth noting that the standard model's unitary view of MTL function is at odds with other theories such as the relational theory (see Section 2.6) which attributes the formation of associations and relational networks to the hippocampus only (Cohen & Eichenbaum 1993).



**Figure 5. Schematic diagram of a simple model of the declarative theory**

A simple computational model of the declarative theory taken from (Alvarez & Squire 1994). Areas cortex1 and cortex2 represent association neocortex. Each unit in each of the areas is reciprocally connected to each unit in the other areas. There are no connections within areas, only a form of winner-takes-all inhibition. A key feature of the model is that connections to and from the MTL area (thin lines) change much faster than the connections between two cortical areas (thick lines) (Alvarez & Squire 1994)(also see (McClelland et al. 1995; Norman & O'Reilly 2003) for the evolutionary function of having a fast and a slow learning system in the brain).

The declarative theory, or standard model, is still predominant in the field of memory today. The idea of a hippocampal memory trace indexing stored representations in neocortex has its origins in Marr's model and the process of consolidation would appear to be substantially proven and widely accepted, although there remains the complex outstanding issue of reconsolidation (Nadel & Land 2000)(Dudai 2004). A simple computational model of the main features of the declarative theory is shown in Figure 5 (Alvarez & Squire 1994). The standard model as with any incumbent theory has had its fair share of criticism from accusations of circular definition (any task that an amnesic patient is able to do is by definition not a declarative task), to experimental ambiguity in animal models (i.e. the difficulties associated with showing an animal is consciously expressing a memory), to the existence of the temporal gradient, and MTL subregion non-specificity, yet it continues to show its considerable durability in the face of concerted opposition.

## **2.4 Multiple trace theory**

The multiple trace theory (MTT) was primarily developed to address problems perceived in the standard model (Moscovitch et al. 2006; Moscovitch et al. 2005; Nadel & Moscovitch 1997). Its criticisms revolve around two central aspects of the standard model (Nadel & Moscovitch 1997), firstly that episodic (Tulving 2002b) and semantic memory, key subdivisions of declarative memory, might be dissociable with respect to hippocampal dependence, and secondly that the pattern and extent of the retrograde amnesia observed after MTL lesions is not temporally graded.

According to multiple trace theory, the hippocampal complex rapidly and obligatorily encodes all information that is attended (i.e. consciously apprehended) and links the neocortical neurons that represent that experience into a memory trace (Moscovitch et al. 2005). Echoing Marr's model (Marr 1971) this information is proposed to be sparsely encoded in a distributed network of hippocampal complex neurons that act as a pointer, or index, to the neurons representing the attended information (Marr 1971; Teyler & DiScenna 1986). Where the MTT departs from the standard model is with the idea that for episodic memory there is no prolonged consolidation process that slowly strengthens the neocortical components of a memory to the point where the memory becomes independent of the hippocampus/MTL. Instead, formation and consolidation of memory

traces for an episodic memory is viewed to be relatively rapid, lasting on the order of seconds or at most days (Moscovitch et al. 2005). The MTT moreover posits that each time an old memory is retrieved, a new hippocampally-mediated trace is created so that old memories are represented by more or stronger hippocampal/MTL-neocortical traces than are new ones and, therefore it is argued, are less susceptible to disruption from brain damage than are more recent memories. If the memory trace for autobiographical episodes is distributed in the hippocampus, the extent and severity of retrograde amnesia, and perhaps the slope of the gradient, will be related to the extent and location of the damage to the extended hippocampal system (Nadel & Moscovitch 1997) (Moscovitch et al. 2005).

There is overwhelming evidence (Kapur 1999; Moscovitch et al. 2005; Nadel & Moscovitch 1997; Steinvorth et al. 2005; Warrington 1996) suggesting that episodic and semantic memories are affected differentially by retrograde amnesia with the severity of the deficit in episodic memory in general more pronounced. Retrograde amnesia for episodic memory after large MTL lesions is usually very extensive and severe often extending for decades, or even a lifetime, far longer than it would be biologically plausible for a consolidation process to last (Moscovitch et al. 2005; Nadel & Moscovitch 1997). By contrast, retrograde amnesia for semantic memory is much less extensive and is often temporally graded, as is the case also for memory of public events and personalities, and even more so for 'personal semantics' i.e. facts about oneself (Kapur 1999; Moscovitch et al. 2005). This poses a significant problem to the standard model which treats all types of declarative memory as a unified whole, equivalently supported by the MTL. Further evidence comes from a study of patients with developmental amnesia (Vargha-Khadem et al. 1997) in which it was demonstrated that semantic learning was spared in the context of focal hippocampal damage and episodic amnesia. Proponents of the standard model, however, have argued that in this instance findings might be due to the special case of developmental amnesia leading to compensatory functional reorganisation (Squire et al. 2004).

The issue of temporally graded amnesia in hippocampal patients has also proven to be highly contentious, with claims made by the standard model of spared remote memory. Patient studies on this issue have been inconclusive with claims (Bayley et al. 2005; Bayley et al. 2003; Squire et al. 2004) and counterclaims (Cipolotti et al. 2001; Maguire

et al.; Moscovitch et al. 2005; Steinvorth et al. 2005; Viskontas et al. 2000) being made in support of both sides of the argument leading to something of an impasse. On some points, however, both sides would agree, that episodic memory is densely impaired in retrograde amnesia often lasting 25-40 years, and the severity of the deficit can correlate with the amount of MTL tissue damaged (Moscovitch et al. 2005). Nadel and Moscovitch, in their original formulation of the multiple trace theory (Nadel & Moscovitch 1997), pointed out that even some results of the original proponents of the standard model (Rempel-Clower et al. 1996) indicate that patients with large hippocampal lesions have no discernible gradient for the most recent three decades and have very little or no autobiographical memory except for the most remote time period tested. They also argued that it was difficult to conceive of an adaptive basis for a biologically plausible consolidation process that is almost as long as the average human life span. Thus, the estimate of the time needed to consolidate autobiographical memories made by the standard model is so long, then for all practical purposes the hippocampal complex would need to be involved in recovering most of them throughout one's life (Nadel & Moscovitch 1997).

Indeed Nadel and Moscovitch (Nadel & Moscovitch 1997) point out that the originators of the standard model (Nadel was a co-author of one of the earliest articles on the standard model (Squire et al. 1984)) did not have such long intervals in mind for the consolidation process when they proposed that memory could be subserved by extra-hippocampal regions alone after consolidation was complete (Nadel & Moscovitch 1997; Squire et al. 1984). Data from the majority of animal studies appear to support the same conclusions, although the data are not entirely consistent (see Table 2 in (Nadel & Moscovitch 1997)). When the task is one for which the hippocampus is absolutely essential during learning (e.g. spatial tasks, such as the water maze, the radial maze or scene discrimination), then no matter when the lesions are made in the hippocampus, deficits ensue. This is comparable to the result seen in humans for autobiographical memory, and suggests that in some domains at least, the hippocampus is involved at all stages of learning and remembering.

The neuroimaging evidence on the other hand has been largely unequivocal with the majority of studies of autobiographical memory retrieval (Maguire 2001a; Maguire & Frith 2003; Svoboda et al. 2006) finding hippocampal involvement irrespective of memory

age. It should, however, be noted that no reverse gradient, with greater hippocampal activation for older memories as predicted by the multiple trace theory (Moscovitch et al. 2005), has been clearly demonstrated to date either (Addis et al. 2004; Gilboa et al. 2004; Maguire & Frith 2003; Steinworth et al. 2006). Several studies (Gilboa et al. 2004; Maguire 2001a; Maguire & Frith 2003) controlled for effects of re-encoding thus rebuffing one of the main criticisms of neuroimaging studies by advocates of the standard model (Squire et al. 2004) that any hippocampal activations found simply reflect encoding of the re-experienced event in the scanner (Squire et al. 2004). For example Gilboa et al (Gilboa et al. 2004) used picture stimuli as cues for memory and in the critical contrast compared the recall of one's own personal memories with the encoding of another participant's picture cues. A further interesting result from this study was the finding that left hippocampal activation was correlated with the richness and vividness of the recalled memory (Gilboa et al. 2004). This was confirmed in another study (Addis et al. 2004) that also showed that recency did not have a significant effect on hippocampal activation when recollective qualities such as vividness were included as covariates. Taken together, the implications of these results is that the hippocampus is always required for episodic memory retrieval so long as the memory retrieved is vivid and detailed enough. Furthermore, recent memories are in general more vivid than remote memories and hence may also explain some temporal gradient findings (Maguire & Frith 2003; Moscovitch et al. 2005). In fact it has recently been suggested that discrepancies between studies of remote memory in hippocampal patients (Bayley et al. 2005; Bayley et al. 2003) might be accounted for by differences in the quality or richness of the recollective experience, a feature that is not always captured by existing scoring systems (Addis et al. 2004; Gilboa et al. 2004; Kopelman et al. 1989; Levine et al. 2002; Moscovitch et al. 2005).

Similarly, with respect to spatial memory specifically, and in contrast to the standard model (Squire et al. 2004), MTT suggests that detailed spatial representations that can support the rich re-experiencing of the environment are always mediated by the hippocampus (Moscovitch et al. 2005). By contrast generic allocentric spatial information, dubbed "schematic or semantic spatial memory" (Moscovitch et al. 2005) and often sufficient for navigation (Teng & Squire 1999; Winocur et al. 2005), is mediated initially by the hippocampus, but like other forms of semantic memory, can exist independently of it once the memory has been consolidated (Moscovitch et al.

2005; Squire et al. 2004). Interestingly this pattern of spatial memory function was recently found in a study that tested hippocampal-damaged patient TT on a navigation task (Maguire et al. 2006a). He was found to have spared performance on routes that involved the use of major artery roads, spatial information that presumably was well-schematised, but impaired on routes that involved the use of minor roads and thus required access to detailed spatial representations mediated by the hippocampus (Maguire et al. 2006a). This view (Moscovitch et al. 2005), and pattern of impairment (Maguire et al. 2006a), also potentially poses a problem for the cognitive map theory in its original formulation (O'Keefe & Nadel 1978) which predicts that flexible navigation in environments learned recently or remotely would be impaired following hippocampal damage (see Section 2.5) (O'Keefe & Nadel 1978).

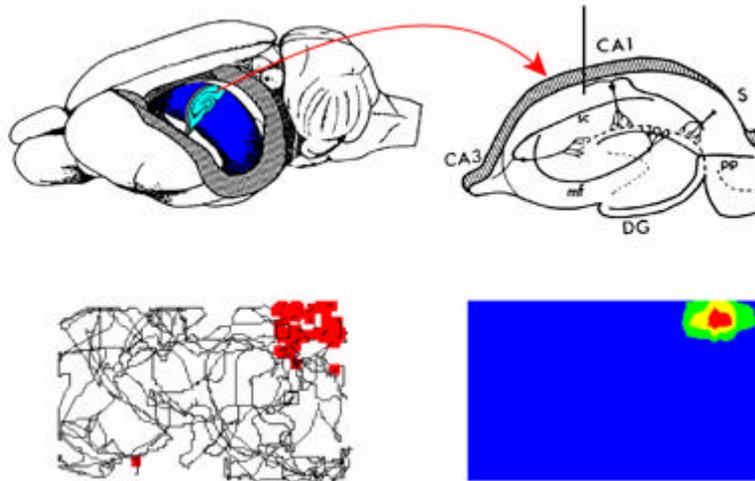
In summary, according to MTT the temporal gradient observed in some patient studies may be related either to stronger and more distributed traces in the hippocampus, making them more resilient than recent memories to partial damage of the hippocampus, or to the greater semanticisation of remote, as compared to recent, memories, making the former less dependent on the hippocampus. Neuroimaging evidence points to hippocampal activation being related to the recollective qualities of autobiographical memory, such as details, vividness, emotion, and personal significance rather than the age of the memory per se (Addis et al. 2004; Gilboa et al. 2004). Given that these factors can covary with memory age, one may erroneously interpret the changes in hippocampal activation with memory age as suggesting that its involvement in autobiographical memory is temporally limited when in fact once all factors are taken into account, memory age has been demonstrated to have no influence on hippocampal activation (Gilboa et al. 2004). Overall these data has been interpreted by advocates of MTT to be supportive of MTT over the standard model (Moscovitch et al. 2005) but even the originators admit that MTT as initially formulated (Nadel & Moscovitch 1997) would need to be modified to account for some of these more recent findings (Gilboa et al. 2004; Moscovitch et al. 2005) (see Section 8.9 for an alternative account).

## **2.5 Cognitive map theory**

It was Kant who first postulated that space was a fundamental organising principle of the mind (O'Keefe & Nadel 1978). He advanced the notion that the mind contained an a

priori system for organising sensations into a spatial framework and that this framework was absolute and not relative in nature. Empirical evidence for this kind of spatial mental construct came with the finding of place cells in the rat hippocampus (O'Keefe & Dostrovsky 1971) (see Section 1.5.3), arguably still the most impressive neural correlate of a high level behaviour discovered to date. This then led to the formulation of the cognitive map theory of hippocampal function which proposed that the function of the hippocampus was to construct and maintain allocentric (i.e. in world-centred coordinates) maps of the environment (O'Keefe & Nadel 1978).

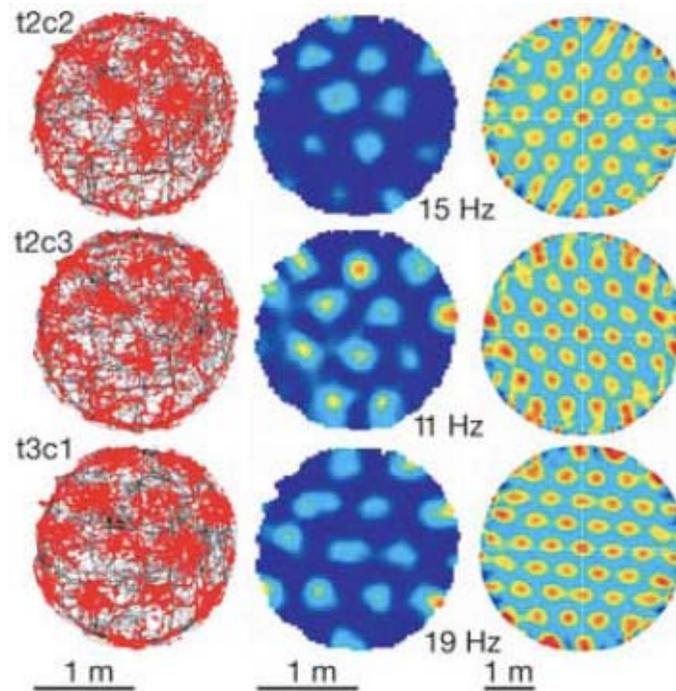
Place cells are neurons that fire invariantly when an animal is at a particular location in an environment irrespective of other factors such as what the animal might be doing, the direction it may be heading, or current visual input (Figure 6)(O'Keefe 2006), hence the notion that place cells code for location in an allocentric manner. More recent work has confirmed the existence of neurons with similar properties to place cells in both monkeys (Colombo et al. 1998)(but see (Rolls et al. 1998) for an alternative view) and humans (Ekstrom et al. 2003). Place cells exhibit striking location-selective responses, increasing their firing rate by orders of magnitude once the animal enters their receptive field (O'Keefe & Nadel 1978). In a given environment, perhaps around 20-30% of hippocampal CA1 neurons have robust place fields (Guzowski et al. 1999; O'Keefe 2006). Interestingly, however, the current consensus is that place cells are not topographically organised in the hippocampus, with cells representing adjacent locations in the environment not anatomically closer than would be expected by chance (Redish et al. 2001)(but see (Eichenbaum et al. 1989; Hampson et al. 1999) and Chapter 6). Place selective responses appear rapidly upon exposure to a new environment (i.e. within minutes: (Wilson & McNaughton 1993)), and once formed tend to remain stable over a long period (up to months), although this is reduced in aged animals (Barnes et al. 1987).



**Figure 6. Hippocampal place cells**

Rodent brain exposed (upper left panel) with schematic diagram of electrode within CA1 region (upper right panel). Traversals by rat in rectangular environment (lower left panel) with rate map of a place cell firing in top right corner of enclosure (lower right panel). Adapted from O'Keefe Lab web page ([www.anat.ucl.ac.uk/research/okeefe/okeefe\\_research.html](http://www.anat.ucl.ac.uk/research/okeefe/okeefe_research.html)).

Recently the operating mechanisms of the spatial system have been further illuminated by the discovery of grid cells in the medial entorhinal cortex (MEC) of the rat (Figure 7)(Hafting et al. 2005), a region that lies one synapse upstream from hippocampal place cells. Grid cells show multiple peaks of firing that are regularly spaced throughout the environment, suggesting the presence of an underlying metric such as a path integration system driven by self-motion cues. Grids are expressed immediately in a novel environment, and a given MEC cell fires with constant spacing regardless of the sensory environment (Moser et al. 2008). However, this spacing distance varies between individual cells (in a systematic way across the caudal-ventral axis (Moser et al. 2008) allowing groups of grid cells to cover an entire environment, mimicking the overlap of multiple sheets of graph paper with varying orientations and scales. Place cells, therefore, could be driven partly by multiple inputs from grid cells that have peaks in a given location. In addition, hippocampal place ensembles may incorporate information concerning landmarks, perhaps arriving via the perforant path from the lateral entorhinal cortex (LEC), to form unique representations of individual environments (Jeffery 2007; McNaughton et al. 2006).



**Figure 7. Firing fields of grid cells have a repetitive triangular structure**

Three simultaneously recorded grid cells during 30 minutes of running in a large circular enclosure. Left: trajectory of rat (black) with superimposed spike locations (red). Middle: colour-coded rate map with the peak rate indicated. Red is maximum and dark blue zero. Right: spatial autocorrelation for each rate map. From (Hafting et al. 2005).

Lesions studies, particularly in the setting of the Morris water maze, provide further compelling evidence for the critical role of the hippocampus in spatial tasks that require an allocentric representation of the environment (Morris 2006; Morris et al. 1982). In the Morris water maze task, rats learn to escape from the pool by swimming to a hidden platform. Following exploration, rats are rapidly able to locate the unmarked platform after only a few visits, and take novel shortcuts in the face of obstructions or novel starting positions (Morris et al. 1982). Rats with hippocampal lesions are severely impaired when they are required to locate the platform starting from a different position on each trial. In contrast, they successfully learn to swim to the platform if the navigation task is made egocentric by maintaining the starting position across trials (Eichenbaum et al. 1990). Thus the Morris water maze task demonstrates that the hippocampus supports

efficient spatial navigation through the mediation of an allocentric representation of the environment.

Evidence in humans also points to a role for the hippocampus in spatial memory (Burgess et al. 2002). Patients with damage to the hippocampus bilaterally exhibit marked impairments in learning to navigate around new environments (Burgess et al. 2002; Maguire et al. 2006b). Furthermore, the deficit in spatial memory in some patients has been linked to their inability to utilise allocentric, as opposed to egocentric representations (Hartley et al. 2007). Functional neuroimaging studies have also implicated the hippocampus in spatial navigation (Hartley et al. 2003; Maguire et al. 1998; Spiers & Maguire). For example Hartley et al. (2003) showed that when subjects were required to navigate within a virtual reality environment, hippocampal activation correlated with wayfinding accuracy (i.e. finding novel routes between destinations). Kumaran and Maguire (2005) also showed that simply navigating a complex associational network (e.g. one's social network) was not enough to activate the hippocampus, the network had to be spatial in nature (but see (Cohen & Eichenbaum 1993) and see Section 2.6). Structural MRI studies have provided further evidence for the critical role of the hippocampus in spatial memory, where a study conducted on a group of expert human navigators showed a marked increase in grey matter in the posterior hippocampus seemingly at the cost of a decrease in grey matter in the anterior hippocampus, both changes correlating with the degree of navigational experience accrued (Maguire et al. 2006b).

Taken together then, convergent physiological, lesion and neuroimaging evidence in both rats and humans broadly supports the hypothesis that the hippocampus plays a critical role in spatial memory through the provision of allocentric representations of the environment. The original formulation of the cognitive map theory stated that the involvement of the hippocampus in supporting these kinds of representations was not time-limited but rather always required for the maintenance of allocentric spatial information (O'Keefe & Nadel 1978; Nadel & Moscovitch 1997). Although the majority of the available animal data supports this view (Clark et al. 2005; Morris et al. 1982; Mumby et al. 1999; Squire et al. 2004; Sutherland et al. 2001) the spared navigational function of hippocampal patient EP (Teng & Squire 1999) and complex pattern of

impairment of patient TT (Maguire et al. 2006a) potentially pose a problem for cognitive map theory (Moscovitch et al. 2005; Squire et al. 2004).

Another outstanding critical question is how this spatial role for the hippocampus can account for autobiographical memory deficits following damage to the hippocampal system. This question was considered in the original formulation of the cognitive map theory (O'Keefe & Nadel 1978) where it was suggested that the human hippocampus, although retaining an important role in spatial processing, also incorporates both a linear sense of time and language (O'Keefe & Nadel 1978). A lateralisation of function was proposed with the right hippocampus supporting spatial and episodic functions, and the left hippocampus performing linguistic functions involved in coding of aspects of the structure of language and verbal memory (O'Keefe & Nadel 1978). In this way, the pivotal role of the human hippocampus in autobiographical memory was accounted for by its ability to represent the spatiotemporal context of events, a defining characteristic of autobiographical memory (Burgess 2002; Burgess et al. 2001; Tulving 2002b). Critically, therefore, it was argued that amnesic patients with hippocampal damage should show "a deficit in those learning situations which are dependent upon the retention of *contextual information*" (O'Keefe & Nadel 1978). The hippocampus, therefore was viewed to mediate the binding of the spatial and temporal context (i.e. the ongoing state of the world) of events, to the content (i.e. what happened). Hence the cognitive map theory proposes that the hippocampus, through the provision of a spatiotemporal context, provides the scaffold on top of which autobiographical memories are built (Burgess 2002; Burgess et al. 2001; Byrne et al.; O'Keefe & Nadel 1978).

## **2.6 Relational theory**

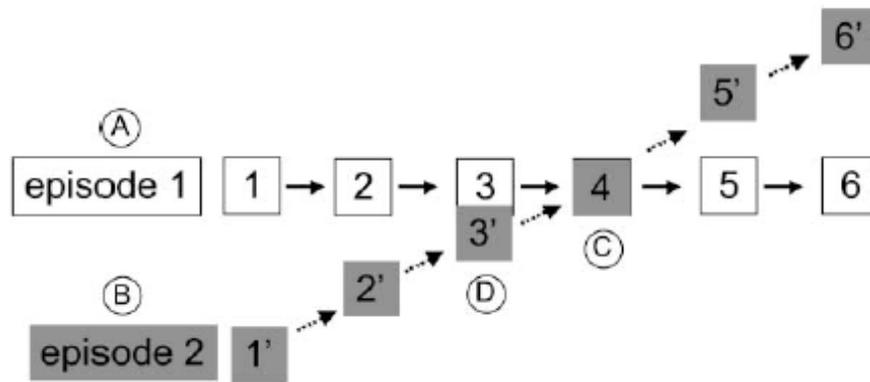
The relational theory (Cohen & Eichenbaum 1993), like the declarative theory, can be considered to have its origins in the characterisation of the amnesic syndrome beginning with HM in 1957 (Corkin 2002; Scoville & Milner 1957). In common with the declarative theory (Squire 1992), the involvement of the hippocampus is considered to be time-limited. It accounts for the pattern of impaired and spared performance of amnesic patients by postulating the existence of at least two independent memory systems (relational vs procedural) with distinct neural substrates, only one of which (the MTL/relational system) is damaged in amnesia. Critically, however, the relational theory

proposes that the MTL system plays an important role in some tasks but not others, due to the unique nature of representations it supports (Cohen & Eichenbaum 1993). In contrast, the declarative theory holds that other factors, namely the ability to consciously access propositional (i.e. either true or false) information, constitutes the crucial distinction between MTL-dependent and independent tasks (Squire 1992). Given this representational emphasis, the relational theory is proposed to account for more empirical data than the declarative theory, which by emphasising the importance of conscious access to stored memories has less clear-cut applicability to research in animals.

Drawing on the privileged anatomical position of the hippocampus in the sensory hierarchy, the relational theory argues that the MTL and hippocampus in particular, represent the “outcomes of processing” of different neocortical modules that constitute a given learning experience or event (Cohen & Eichenbaum 1993). The hippocampus is viewed to represent these outcomes in a highly interconnected multidimensional space where all possible relations between multiple elements are captured. This property of hippocampal representations is argued to allow relational memories stored in this network to be accessed and flexibly utilised/recombined in numerous ways depending on task requirements. The hippocampus, therefore, is held to be critical for the flexible expression of relational memories, encompassing memory for facts and personal experiences, as well as finding novel detours, and transitive inference. Procedural memories, in contrast, which support performance on tasks such as priming and skill learning, are mediated by the tuning and modification of localised neocortical processors during learning, rather than through the representation of outcomes of processing (Cohen & Eichenbaum 1993). As a result, procedural memories are proposed to be relatively inflexible, with applicability limited to specific tasks and circumstances that are effectively repetitions of the original learning experience.

According to the relational theory the elemental building blocks of hippocampal representations are viewed to be (1) associative (2) sequential and (3) relational networks (Figure 8)(Eichenbaum 2004). As such, the notion that space or allocentric spatial representations constitute the organising principle of hippocampal function, as postulated by the cognitive map theory, is rejected. Instead, place cells and the capacity for spatial navigation, are viewed to be an emergent property of the hippocampal

relational networks that equally represent spatial and non-spatial information. The relational theory, therefore, can be considered to retain the essence of traditional animal learning theories which argue that spatial learning is entirely explicable through the formation of stimulus response associations (see (Morris 2006) for discussion).



**Figure 8. Schematic diagram of a simple relational network**

The network is composed of two episodic memories (A and B). Each memory is construed as a sequence of events, which each element representing the conjunction of an event and the place where it occurred. C is an element that contains the same features in both episodes. D is an element that contains only some of the common information. From (Eichenbaum 2004).

The hippocampus is proposed to play a key role in supporting both spatial and non-spatial associative codings that mediate the rapid binding of the disparate elements of events (e.g. (Marr 1971; McClelland et al. 1995)), in particular between perceptually distinct arbitrarily paired items (e.g. face-voice association). A fundamental feature of hippocampal associative representations is held to be their compositionality, whereby both the links between individual elements and the elements themselves are encoded. In contrast, configural representations (Mayes et al. 2007), for example, can be considered to disregard information concerning the elements themselves (Mayes et al. 2007) and according to the relational theory are supported by regions outside the hippocampus (e.g. PHG) (Moses & Ryan 2006).

Proponents of the relational theory have emphasised empirical evidence that the firing of hippocampal pyramidal cells is governed not only by the location occupied by the animal, but also by other factors such as the demands of the behavioural task and the presence of non-spatial cues (e.g. olfactory stimuli) (Eichenbaum et al. 1999; Otto & Eichenbaum 1992). A study by Wood et al (Wood et al. 1999) that specifically examined the extent to which hippocampal neurons coded spatial as compared to non-spatial variables during an odour DNMS task yielded findings consistent with this notion. Rats were trained to approach a small cup of sand scented with one of nine odours. On any given trial, the cup was placed in one of nine locations. If the odour was different from the odour sampled on the last trial (i.e. non-match), the cup contained food which could be obtained by digging. The rationale for the experimental design was that non-spatial variables (e.g. odour; match/non-match judgement etc.) were uncorrelated with the location occupied by the animal, allowing the spatial and non-spatial coding properties of hippocampal neurons to be distinguished. Given that hippocampal neurons recorded during the 1 second prior to arrival at the cup were found to code both spatial and non-spatial variables, it was argued that this study provides evidence of the general associative/relational coding function of the hippocampus. As such, it has been suggested that the hippocampus automatically represents the attended features of the environment, irrespective of whether it is necessary for task performance, effectively providing a record of episodic experiences involving places, stimuli and actions taken within the environment.

According to the relational theory, sequence representations constitute a fundamental aspect of hippocampal representations and act to link temporally punctuate event codings into extended episodic memories (Eichenbaum 2004). In this way, they are viewed to support the encoding of the flow of episodic experiences over time, allowing the mental replay of autobiographical events. Further, it is argued that journeys taken in the environment are coded as sequences of events, linked through overlapping elements into networks that afford the capacity to take novel detours (Kumaran & Maguire 2006). The relational theory also postulates that episodic memories do not exist as isolated sequential representations, but are linked to other experiences through shared elements to create relational networks (Eichenbaum 2004). These shared representational elements encode features (e.g. stimuli, places) common to different

experiences, allowing memories to be indirectly related to one another (e.g. in making novel inferences). Hence, hippocampal representations are viewed to capture regularities that cut across different experiences. In this way, it argued that episodic memories are linked through their common features into relational networks within a memory space, supporting flexible memory expression.

The study by Wood et al. described above (Wood et al. 1999), also provides evidence that hippocampal neurons encode both the unique and common aspects of experiences. A few hippocampal neurons were found to be selective for unique experiences, for example the execution of a specific behaviour (i.e. match response) in response to the presence of a particular odour at a given place. In contrast, a greater proportion of cells reflected commonalities across different experiences, for example the occurrence of a specific odour ('odour cell') or the presence of the rat at a specific spatial location ('place cell'). According to this view, therefore, place cells, do not constitute components of a cognitive map, but rather reflect "nodal" codings that link different episodes which occurred in the same location (Eichenbaum et al. 1999). Similarly, odour cells link different episodes in which the same odour was experienced.

In summary then, the relational theory offers a representational account of the hippocampal contribution to memory designed to explain the pattern of spared and impaired performance on tasks across species (Cohen & Eichenbaum 1993). As such, the hippocampus is viewed to support relational representations, built from a combination of associative and sequential codings which are linked through common elements to form a memory space. In particular, the notion that the hippocampus plays an exclusive or even predominant role in performing spatial computations that underlie a spatial map of the environment is rejected. Rather, place cells, like odour cells, are proposed to be examples of nodal codings that represent features that are shared across different experiences. Spatial navigation, therefore, is mediated not by a cognitive map as such, but instead by the flexible use of networks of memory representations encoding previous routes traversed in the environment. Moreover, the critical role of the hippocampus in spatial tasks such as the water maze is argued to represent merely a subset of relational capabilities mediated by the hippocampus, that also includes performance on transitive inference tasks (Eichenbaum 2004).

## 2.7 Summary

The influential theories reviewed in this chapter represent over 50 years of research and thinking into the role of the hippocampus in memory (Andersen et al. 2007; Scoville & Milner). Each of the theories in their own way makes a valuable contribution to our understanding of the hippocampus and memory function but none deals with the whole spectrum of empirical data and behaviour documented in the voluminous literature in the field.

Nevertheless, there are a number of points of general agreement. It is clear that memory is not a unitary function and that there are multiple dissociable memory systems in the brain with declarative memory critically reliant on the hippocampal complex (Cohen & Eichenbaum 1993; Scoville & Milner 1957; Squire 1992). The idea of a hippocampal memory index (Marr 1971) that reinstates the full representation of a memory by reactivating its cortical components is widely accepted (Moscovitch et al. 2005; Squire et al. 2004). As is the fact that anterograde memory loss is total in hippocampal amnesia (Moscovitch et al. 2005; Squire et al. 2004). There is also broad agreement that informational index stored in the hippocampus can eventually over time consolidate to neocortex and become hippocampal independent (Squire et al. 2004), at least for some types of memory such as semantic memory (Moscovitch et al. 2005; Squire et al. 2004).

However, there are numerous areas of disagreement. For instance there is continued controversy as to the temporal extent of the episodic memory loss exhibited in hippocampal amnesia, with the standard model claiming the sparing of remote episodic memories (Bayley et al. 2003; Squire et al. 2004) whilst the multiple trace theory argues that the deficit is global (Moscovitch et al. 2005; Nadel & Moscovitch 1997). The fate of episodic versus semantic memories and whether they are similarly reliant on the hippocampus over time is also disputed (Moscovitch et al. 2005; Squire et al. 2004). Another unresolved issue is the debate about whether, at the representational level, space is the primary organising principle of hippocampal function (O'Keefe & Nadel 1978) or whether space should be viewed as a special form of a more general associative capability (Cohen & Eichenbaum 1993).

There are also a host of other issues on the periphery of the mainstream theoretical debates that are largely only indirectly addressed by the theories considered in this

chapter. This includes controversial evidence that the hippocampal complex might be involved in the perception of certain types of complex stimuli (Graham & Gaffan 2005; Lee et al. 2005b; Murray & Bussey 2001), and that short-term memory might not be intact in hippocampal patients for certain types of tasks (Hannula et al. 2006)(but see (Squire et al. 2004)). Additionally there is the issue of tasks that appear to be critically reliant on the hippocampus but that are not directly related to autobiographical memories such as thinking about the future (Klein et al. 2002), called episodic future thinking (Atance & O'Neill 2001), and navigation (Burgess et al. 2002).

Finally, these theories of hippocampal function, whilst dealing with memory traces, consolidation and representations, do not for the most part, attempt to relate these mechanistic concepts with the widely accepted high level psychological concepts such as how they might fit with the reconstructive nature of memory (Bartlett 1932; Schacter et al. 1998) or facilitate the rich recollective experience that accompanies and defines episodic memory recall (Tulving 2002b). Nor do they consider the component processes that make up episodic memory recall and the functions the hippocampus might help support in the context of the wider brain network that supports episodic memory recall and is consistently activated in neuroimaging studies (Maguire 2001a; Svoboda et al. 2006).

To conclude, whilst all the theories presented in this chapter likely capture important aspects of hippocampal function in episodic memory, none of them offer a completely satisfactory account that can explain all extant data. With the help of some novel paradigms, I sought to formulate a new perspective on the role of the hippocampus in episodic recollection that could help to reconcile some of the seemingly incompatible views and findings discussed in this chapter.

# Chapter 3

*Methods*

## **3.1 Introduction**

Three methodological approaches were employed in this thesis. Experiment 1 involved the neuropsychological testing of patients with amnesia associated with primary damage to the hippocampus bilaterally. Experiment 2 used whole-brain fMRI to investigate brain activity while healthy volunteers performed different cognitive tasks. These data were analysed in a standard univariate manner using a general linear model. In Experiments 3 and 4, high spatial resolution fMRI was employed and the data were analysed using multivariate pattern analysis techniques.

In this chapter, I provide an overview of the general principles of fMRI. I also outline the basic rationale behind univariate and multivariate analysis techniques for fMRI. A description of the neuropsychological methods used in Experiment 1 are given in Chapter 4, and more specific methodological details pertaining to multivariate pattern analysis are provided in Chapters 6 and 7.

## **3.2 The biophysics of magnetic resonance imaging (MRI)**

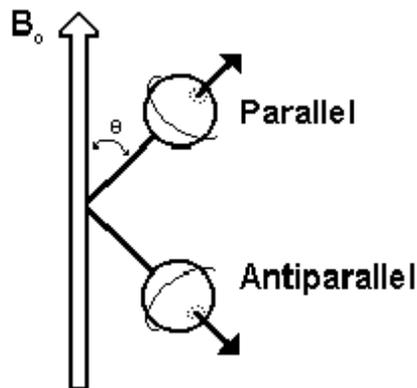
Functional MRI allows neural activity to be indirectly measured whilst human subjects perform cognitive tasks. Given that fMRI does not involve radiation and has excellent spatial resolution and adequate temporal resolution, it has become a widely used tool. During fMRI scanning, the blood-oxygen-level-dependent (BOLD) mechanism is used to infer neural activity from changes in the  $T_2^*$  signal that occur as a result of alterations in intracerebral blood flow. This chapter reviews the basis of fMRI and the BOLD signal (Logothetis 2003; Logothetis 2008), starting with a conceptual overview of MRI upon which fMRI relies (Frackowiak et al. 2004). The essential parameters of acquisition of fMRI images are then outlined, followed by the principles governing the analysis of fMRI data and fMRI design.

### **3.2.1 Magnetic fields**

MRI relies on the principle of nuclear magnetic resonance (NMR) discovered by Purcell and Bloch in the 1940s (Bloch et al. 1946; Purcell et al. 1946). fMRI is a special form of magnetic resonance imaging (MRI), used widely in clinical practice, and is based on the

same basic principles as MRI. All atomic nuclei possess the quantum property of spin. MRI techniques measure the effects of changing the spin of particular atomic nuclei, such as  $^1\text{H}$  and  $^{13}\text{C}$ , which have an odd number of nucleons (protons plus neutrons).

In living tissue such as the brain, the most abundant source of protons is the hydrogen atom in the form of water. The hydrogen nucleus is positively charged and the spinning motion of this charge induces a local magnetic field. These hydrogen nuclei (or protons) therefore behave like small magnets, i.e. they have a magnetic moment. In the absence of a magnetic field, these individual spins are randomly orientated and the bulk material has no net magnetisation. However, in an externally applied magnetic field,  $B_0$ , the individual spins align with the external magnetic field. If the spin is not completely aligned with the direction of the magnetic field  $B_0$ , this causes the proton to revolve, or precess, around the field direction (Figure 9). The frequency with which the axis rotates around the field direction is called the resonance or 'Larmor' frequency, and is directly proportional to the field strength  $B_0$ .



**Figure 9. Alignment of hydrogen atoms in a magnetic field**

Diagram illustrating alignment of hydrogen atoms either parallel or anti-parallel to external magnetic field ( $B_0$ ).

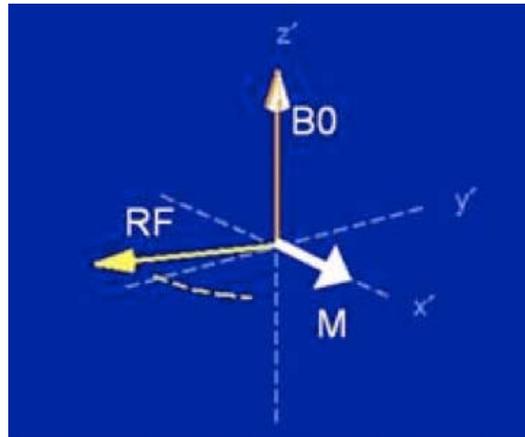
Quantum mechanics dictates that a spin can have different energies depending on the orientation of its magnetic moment with respect to the applied magnetic field: when the magnetic moment is aligned with the field, its energy will be lower than when it opposes the field. For the simple spin system of  $^1\text{H}$ , the magnetic moment can have two

orientations with respect the magnetic field, either against it (high energy state) or along it (low, ground energy state). The amount of energy required to flip orientations is so small that the normal thermal energy available at room temperature is enough to flip spins. All of the signals generated are, therefore, based on small differences between these energy states. The fact that the energy differences are small is one reason why MRI techniques tend to be safe but also why they are typically limited by signal strength.

The sum over all the nuclei in an object volume gives the net magnetisation for the body of tissue. Magnetic fields are described in a Cartesian co-ordinate system, with the z-axis being in the direction of the applied magnetic field. The resting net magnetisation is called the equilibrium magnetisation. In equilibrium, more spins are in the low than in the high energy state. Summing the contributions of individual magnetic moments will, therefore, give a net magnetic moment along the direction of the applied magnetic field. Even though a small part of the rotating magnetisation of each nucleus has a component projecting into the xy-plane, there is no equilibrium net 'transverse magnetisation' because the average magnetisation in this xy-plane over all nuclei cancels to zero.

### **3.2.2 Generating an MR signal**

Spins can be excited from the low to the high energy state by the use of an oscillating radiofrequency (RF) pulse. This RF pulse is generated by applying an oscillating radiofrequency electromagnetic field ( $B_1$ ) perpendicular to the main magnetic field ( $B_0$ ). To achieve the most efficient transfer of energy, the oscillation frequency of the  $B_1$  field should be the same as the spin Larmor frequency (i.e. 64 MHz for a proton in a 1.5 Tesla field and 128 MHz for a proton in a 3 Tesla field - the field strengths of the two types of scanners used in this thesis). The application of the RF pulse at the Larmor frequency therefore causes some of the protons in the low-energy state to absorb energy and transition into the high-energy state. This tips the net magnetisation vector ( $M$ ) of the tissue towards the transverse xy-plane (Figure 10). When sufficient energy is provided to produce a  $90^\circ$  change in the net magnetisation, the 'flip angle' of the RF pulse is said to be set at  $90^\circ$ , as was the case in all experiments in this thesis. Oriented perpendicular to  $B_0$  is a receiver coil which measures, through current induction, the transverse magnetisation vector.



**Figure 10. Generating an MR signal**

A radiofrequency pulse (RF) is applied perpendicular to the external magnetic field,  $B_0$  and tips the net magnetisation vector ( $M$ ) of the tissue towards the transverse  $xy$ -plane.

Given that the RF pulse is only transient, the spin orientations will tend to rebound to the direction of  $B_0$ . This occurs through the protons losing the energy gained from the RF pulse in one of two ways. Firstly, to neighbouring tissue as heat (spin-lattice or  $T_1$  relaxation), for a given medium, the rate of spin-lattice relaxation is determined by a constant called  $T_1$ .  $T_1$  varies for different elements: for example, the protons in water have a longer  $T_1$  than those in fat because the carbon bonds in fat resonate near the Larmor frequency, which facilitates the transfer of energy to the lattice. In the human brain, the different water content of grey and white matter (71% and 84%, respectively) means that  $T_1$  contrast can be used to provide contrast between these two tissues. White matter has a shorter  $T_1$  relaxation time than grey matter and hence appears brighter on  $T_1$ -weighted images.

Second, protons can lose energy to other spinning nuclei (spin-spin or  $T_2$  relaxation). In a perfectly uniform magnetic field the  $T_2$  relaxation or spin-spin relaxation describes the disappearance of coherence of the magnetic moment in the  $xy$ -plane. As neighbouring spins pass energy from one to another, their rotations become desynchronised. However, magnetic fields are typically not perfectly uniform, but locally inhomogeneous. Importantly, the inhomogeneity of the magnetic fields affects the  $T_2$  relaxation. Hence the parameter  $T_2^*$  (i.e. apparent  $T_2$ ) captures the combined effect of perfect  $T_2$

relaxation and the magnetic field inhomogeneity and forms the basis of the measured MRI signal used in fMRI.

### **3.2.3 Tomographic image formation**

Simply placing sample tissue (e.g. the brain) within a homogenous  $B_0$  field will not produce tomographic MR images because all protons will experience (roughly) the same magnetic field and, hence, the frequencies of their emitted signal will all be identical. Additional manipulations are therefore needed to encode spatial information. In brief, this involves the application of a second magnetic field (“gradient field”) that allows slice selective excitation. During imaging acquisition, spatial locations are encoded by both the frequency and phase of the detected signal. Frequency encoding is determined by applying the frequency encoding gradient in the x-axis (conventionally) during acquisition. This is followed by phase encoding along a second orthogonal axis (y-axis).

The characteristics of the applied gradients determine the resolution of MRI images acquired. Step-wise increases in gradients partition the sample into small cubes, termed voxels (volume-elements). Protons within a given voxel behave as a single entity given that they are subjected to the same frequency and phase encoding. The voxel size, therefore, determines the resolution of the MRI image and was set at 3x3x3 cubic mm for Experiment 2 and at a higher resolution of 1.5x1.5x1.5 cubic mm for Experiments 3 and 4. Note that larger voxel sizes increase signal-to-noise ratios (SNR), and therefore sensitivity, but have the disadvantage of lumping protons with very different spin behaviour together which can produce a misleading signal.

### **3.2.4 Scanning parameters**

Various scanner parameters can be chosen in order to optimise for particular aspects of an experimental design including, for example, the signal-to-noise ratio, brain coverage, distortion, BOLD sensitivity, or voxel resolution. Some of the main parameters are listed below:

1) *TR*: The repetition time is the interval between two consecutive 90° RF pulses and controls the effect of T1 relaxation. For example, a short TR (and TE-see below) will

emphasise the T1 characteristics of tissue and produce a “T1-weighted” image, whereas a long TR and TE produce a “T2-weighted image”.

2) *TE*: The “echo time” is the time between initial RF excitation and readout of the MR signal which occurs following additional gradients (or RF pulses) that are applied to produce echoes and maximise the MR signal. In echo-planar imaging (EPI-see below), echoes are formed after the initial 90° RF pulse by using gradient fields to dephase and then rephase spins (rather than via echo RF pulses used in “spin-echo” imaging used in clinical settings). The purpose of these echoes is to resynchronise the protons which have started losing coherence following the initial RF pulse, and therefore enhance the measured MR signal.

3) *Image contrast*: This is based on the difference in signal intensity between different areas. This is determined by a number of factors, including T1 and T2 relaxation times and spin density (see above). There are three main types of contrasts: T1-weighted, T2-weighted and spin density-weighted.

4) *Echo-planar Imaging (EPI)*: EPI was used in all studies in this thesis. EPI was first introduced by Mansfield in 1977 (Mansfield 1977). It allows the fast-acquisition of MR images (i.e. with a TR of a few seconds) necessary for fMRI. Other MRI sequences used in clinical settings sample data at a considerably slower rate.

All experiments in this thesis were conducted at the Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London. Experiment 2 was conducted on a Siemens SONATA system (Siemens, Erlangen, Germany), operating at 1.5T. A TR of 4.05 seconds was used (45 slices with a slice acquisition time of 90ms) allowing for whole brain coverage at 3mm resolution.

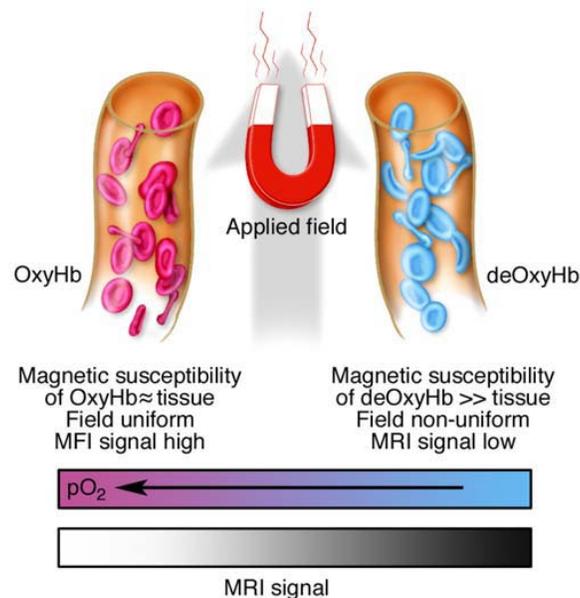
Experiments 3 and 4 were conducted on a head-only Siemens Magnetom ALLEGRA system (Siemens, Erlangen, Germany), operating at 3T, affording the opportunity to obtain data at a higher spatial resolution of 1.5 mm. Given that in the latter two studies I was primarily interested in the medial temporal lobe regions I used a 35-slice sequence that yielded the required partial volume brain coverage with a fast TR of 3.57s (i.e. 35 slices with a slice acquisition time of 102ms) in order to maximise the number of training

volumes acquired for the training of classifiers. Further scanning parameter details can be found in the Methods sections of the individual experimental chapters (see Chapters 5, 6 and 7).

### 3.3 The BOLD signal

#### 3.3.1 The BOLD effect

The BOLD effect relies upon magnetisation differences between oxygenated (oxyHb) and deoxygenated haemoglobin (deoxyHb) and was first described by Ogawa and colleagues in 1990 (Ogawa & Lee 1990; Ogawa et al. 1990a; Ogawa et al. 1990b). DeoxyHb is paramagnetic in nature (Pauling & Coryell 1936), meaning that it alters the magnetic field into which it is introduced. This is the case because unbound iron containing haem groups are present in deoxyHb. The magnetic state of blood, therefore, reflects its level of oxygenation (Figure 11).



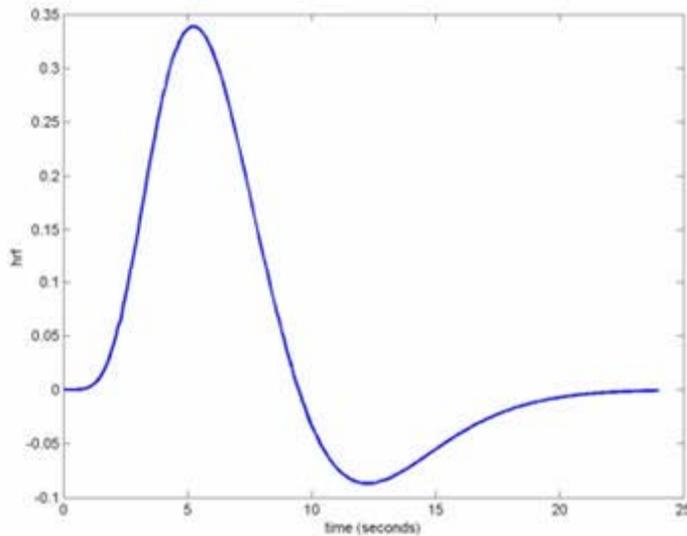
**Figure 11. Schematic of the magnetic properties of blood**

Deoxygenated blood has a higher magnetic susceptibility than tissue (OxyHb – oxyhemoglobin; DeOxyHb – deoxyhemoglobin).

As discussed previously, the  $T2^*$  of hydrogen protons is influenced by interactions between the protons themselves and also by local  $B_0$  inhomogeneities caused by different magnetic properties of various molecules. Paramagnetic molecules, such as deoxyhaemoglobin induce local magnetic field inhomogeneities and therefore contribute to the decay of transverse magnetisation and consequently shorten the  $T2^*$  decay time. Therefore changes in the ratio of deoxyhaemoglobin to oxyhaemoglobin results in changes to the  $T2^*$  parameter. For example, an increase in dexoxyHb causes the measured  $T2^*$  to decrease. This effect was demonstrated empirically *in vivo* in animal work carried out by Ogawa and colleagues (Ogawa et al. 1990b) and Turner and colleagues (Turner et al. 1991). Ogawa and colleagues used strong magnetic fields ( $\sim 7T$ ) to study the rat brain. In 1992, three groups took the further step of demonstrating that the difference in  $T2^*$  signal produced by deoxyhaemoglobin concentrations *in vivo* in humans is sufficient to act as a contrast source (Kwong et al. 1992; Ogawa et al. 1992). This signal source was termed the BOLD contrast.

### **3.3.2 The neurophysiology of the BOLD signal**

The BOLD contrast mechanism, therefore, links the oxygenation state of blood to the measured MR signal. Although the relationship between neural activity in a given region and changes in blood oxygenation is still not fully understood, a consensus has been reached on several main points which also accounts for the basic shape of the BOLD response (Figure 12)(see for review: (Logothetis 2003)).



**Figure 12. The haemodynamic response function**

x-axis: Time (seconds). Y-axis: Amplitude of response (arbitrary units). Response peaks at around 6 seconds after initial stimulation.

At the initial phase of neural activity in a given region, oxygen demand increases causing a small increase in the proportion of deoxyHb (initial dip in BOLD response). Consequently, capillary vasodilatation is initiated with the resulting oversupply of oxygenated blood to the neurally active region that occurs around 1 second after the initial stimulation (Heeger & Ress 2002; Ress et al. 2000). This phase causes a large decrease in the proportion of deoxyHb (peak of BOLD response). The BOLD contrast increase evoked by this deoxyhaemoglobin decrease is the signal that is typically measured in fMRI experiments.

Recent empirical work provides evidence that the measured BOLD signal does indeed reliably reflect neural activity. A study by Logothetis et al (Logothetis et al. 2001) (see also: (Ress et al. 2000)) used simultaneous intracortical recordings and fMRI measurements from macaque visual cortex to demonstrate that the BOLD response directly reflects increase in neuronal activity, correlating in a roughly linear fashion. However, the BOLD signal was found to correlate better with local field potentials (LFPs), rather than neuronal spiking activity. LFPs, and therefore the BOLD signal, mainly reflect the incoming inputs into an area as well as the processing of this input information by local (cortical) circuitry (including both excitatory and inhibitory

interneurons). That single-cell recording experiments in animals generally measure neuronal spiking activity which represents the output of given brain region should be borne in mind when interpreting data from animal electrophysiological studies and human fMRI studies (Logothetis 2003). In general it is reasonable to assume that output activity will usually correlate with inputs to a given area (i.e. LFPs) (Logothetis 2003).

### **3.4 Resolution of fMRI**

Whereas neuronal dynamics occur in the millisecond time frame, the BOLD response takes a number of seconds to evolve. The dynamics of the neurovascular response cause the BOLD signal to be delayed in time, peaking at around 6-10 seconds (Logothetis 2003). This means that although several fMRI images can theoretically be acquired in a second, the temporal smoothing of the underlying neuronal signal, effected by the BOLD response, ultimately dictates fMRI's effective temporal resolution (Frackowiak et al. 2004). Of note, other techniques such as EEG and MEG allow far superior temporal resolution, in the order of milliseconds, although at the cost of inferior spatial resolution. The temporal resolution of PET is far worse than fMRI, in the order of minutes.

With regard to spatial resolution, the images acquired in Experiment 2 had a resolution of 3x3x3 cubic mm voxels and through the use of interleaved slices (see Section 6.2) the images acquired in Experiments 3 and 4 had a resolution of 1.5x1.5x1.5 cubic mm voxels. However, the actual spatial resolution achieved is less than the voxel size and is more difficult to quantify for a number of reasons. Firstly, given that the BOLD contrast mechanism is used to infer neural activity from changes in local blood flow, it is difficult to determine the precise location of neural activity itself. Secondly, variations in individual anatomy may reduce the actual spatial resolution when a group analysis is performed, although this is not an issue for single subject analysis as performed in Experiments 3 and 4. Thirdly, fMRI data are preprocessed prior to statistical analysis, a procedure that involves smoothing of the images which reduces spatial specificity. Nevertheless, the spatial resolution of fMRI is far superior to that of the other techniques used to study human brain function such as PET, MEG and EEG (Frackowiak et al. 2004).

## 3.5 Analysis of fMRI data

fMRI allows inferences to be made about differences in regional brain activity between different conditions or states. In this section I outline the standard steps by which fMRI data is analysed in a classic univariate manner such as that used in Experiment 2 (see Frackowiak et al., 2004 for review of whole analysis procedure). Note that the multivariate analysis techniques used in Experiments 3 and 4 whilst following some of the preprocessing steps below, such as realignment and smoothing, use a largely bespoke set of tools that are described in detail in Chapters 6 and 7.

As a first step, the initial “dummy” volumes acquired in each session (normally numbering ~six) are discarded to allow for T1-equilibration effects. The following steps are aimed to make the data conform to a known anatomical space. This is necessary to assign an observed response to a particular brain structure or cortical area. Hence, the analysis of fMRI data starts with a series of spatial transformations of brain images aimed at reducing artifactual variance components in each voxel time series. Following this preprocessing, a statistical model must be created in order to draw inferences about differences in regional brain activity between different experimental conditions (Friston et al. 1995). Analyses are typically performed on the time-series data for each voxel in the brain, as though they are independent of each other. All analyses of fMRI data were performed using Statistical Parametric Mapping (SPM5 [www.fil.ion.ucl.ac.uk/spm/software/spm5/](http://www.fil.ion.ucl.ac.uk/spm/software/spm5/)) software implemented in Matlab 7.1 (The Mathworks, USA).

### 3.5.1 Spatial preprocessing

The imaging time series is first realigned to a common reference frame to correct for subject movement during scanning. This realignment is carried out since even small degrees of head movement (i.e. in the order of millimetres) can cause considerable distortions of signal intensities. The aim of the realignment procedure is to ensure that the same voxel coordinate refers to the same location in the brain in every EPI image in the data series. Realignment involves the estimation of six parameters (three translation and three rotation parameters) of the affine rigid body transformation that minimises the differences between each successive scan and the first. These parameters are used to

align each EPI image with the common reference image, often the first image in the time-series.

Following realignment, all scans must be transformed into a standard stereotactic space, in this case the Montreal Neurological Institute (MNI) space. This ensures that the same voxel coordinate refers to the same location across different subjects within the same experiment, and across experiments. This is normally achieved by geometrically distorting images into a standard shape through the process of spatial normalisation. Normalisation is performed efficiently using a 12-parameter affine transformation along with a nonlinear transformation, and relies upon a standard template image (in this case the standard T1 template based on the MNI reference brain). This template conforms to a standard geometric space and coordinate system such that  $x$  refers to the laterality of the location (positive is right),  $y$  refers to the anterior-posterior dimension (positive is anterior), and  $z$  refers to the dorsal-ventral dimension (positive is dorsal) and the origin (i.e. coordinate  $[0, 0, 0]$ ) refers to the anterior commissure (AC), using millimetre (mm) as unit.

Following normalisation, the EPI images are spatially smoothed by applying a Gaussian filter to each voxel. The extent of the smoothing is determined by the full-width at half-maximum (FWHM) of the kernel, and is typically between 6 and 10mm in fMRI studies. Data are smoothed for several reasons. For example, in a group analysis, data are averaged across many subjects, it is necessary to smooth the data so that regional effects are expressed at a spatial scale where homologies in functional anatomy exist over subjects. Furthermore, Gaussian field theory, which underpins statistical inference (see below) requires that the data be smoothed.

### **3.5.2 Processing of structural images**

For each participant, a T1-weighted anatomical image was acquired. These structural images were coregistered to, and normalised using parameters generated from, processing of the functional images. A mean image was generated by averaging together the structural images of all participants in each experiment. The results of the fMRI statistical group analyses (i.e. SPMs: see below) were overlaid onto this mean

structural image so as to better define the anatomical regions in which activations occur at the group level.

### 3.5.3 Statistical analysis and model specification

Statistical analysis involves the General Linear Model (GLM) which is also the model for a variety of standard statistical analysis e.g. linear regression, ANOVAs. This model explains variation in the data,  $Y$ , in terms of a linear combination of the explanatory variables ( $X$ ), plus an error term:

$$Y_j = X_{j1}b_1 + \dots + X_{jn}b_n + \dots + X_{jN}b_N + e_j$$

The  $b_n$  are unknown parameters (betas), corresponding to each of the  $N$  explanatory variables for the  $j$ th observation of  $Y$ . The errors  $e$  are assumed to be identically and normally distributed (see below). A mass univariate approach is employed in SPM, whereby the time-series for each voxel are treated as independent from one another, with the same regression model applied to each voxel. This is in contrast to the multivariate approach adopted for Experiments 3 and 4 which looks for informative patterns of activity across multiple voxels (see Section 3.6). The regression model is often termed the design matrix ( $X$ ) in SPM, so as to reflect that its construction is based on the experimental design. Each column of the design matrix is formed by a column vector and called a regressor. The same is true for the time-series data for each individual voxel ( $Y$ ). For  $J$  observations of  $Y$ , the general linear model can be expressed in matrix formulation:

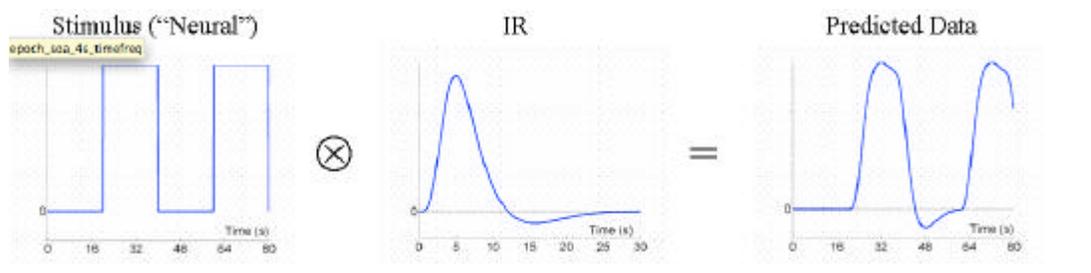
$$Y = Xb + e$$

for the column vector of observations  $Y$ , the column vector of error terms  $e$  and the column vector of parameters  $b$ ;  $b = [b_1 \dots b_j \dots b_J]^T$ . Matrix  $X$ , of size  $J \times L$ , is the design matrix.

The implementation of the GLM relies upon the assumption that the errors (or residuals) are identically and normally distributed. However, this not the case when time-series data are concerned. The haemodynamic response is of longer duration than the typical

scan acquisition time, which leads to serial correlations among the error terms. The general linear model accounts for these autocorrelations by imposing a known temporal smoothing function on the time-series and adjusting the estimators and degrees of freedom accordingly (Frackowiak et al. 2004).

Condition-specific regressors specify the onset of trials, for each experimental session separately. These trial onsets are first convolved with either a boxcar (i.e. square wave) function or a stick (delta) function depending on whether the experiment is event-related or blocked (see below). Next, these functions are then convolved with the chosen haemodynamic response function (HRF) to take into account the temporal characteristics of the BOLD response (Figure 13). In this thesis, the canonical HRF was used exclusively (Friston et al. 1998).



**Figure 13. Creation of condition-specific regressors in GLM**

Schematic diagram illustrating how condition specific regressors that specify the onset and duration of trials (left panel) are convolved with the haemodynamic response function (middle panel) to yield regressors that model predicted neural activity (right panel).

Condition-specific regressors, therefore, capture the temporal characteristic and intensity of predicted BOLD responses associated with a given condition. Hence, each column in the design matrix corresponds to some effect that one has built into the experiment (i.e. experimental manipulation), or effects that may confound the results. The latter include a series of terms that are designed to remove or model low-frequency variations in signal due to artifacts such as aliased biorhythms or scanner drift. Regressors designed to covary out movement related effects and generated by the realignment process (above) are also included in the composition of the design matrix.

### 3.5.4 Model estimation and statistical inference

The goal of the model estimation process is to estimate the best linear combination of regressors to provide an optimal fit with the time series data. Often this proceeds using the least squares estimation approach and results in the generation of parameter estimates, or betas. Betas reflect the intensity of the haemodynamic response caused by that task condition in a given voxel. For example, if a given condition elicits little haemodynamic response in a given voxel, then model fitting would be best achieved by assigning a low beta to the relevant regressor.

Statistical inferences about the parameter estimates are made using their estimated variances. This allows for two types of statistical test. One can test the null hypothesis that all the estimates are zero using the  $F$  statistic to give an SPM $\{F\}$  or, alternatively, that some particular linear combination or a “contrast” (e.g. a subtraction) of the estimates is zero using a SPM $\{T\}$ . The T statistic is calculated by dividing the contrast (specified by contrast weights) of parameter estimates by the standard error of that contrast. This error term is estimated using the variance of the residuals about the least squares fit. An example of contrast weights could be  $[1 \ -1 \ 0 \ 0 \ \dots]$  to compare the differential responses evoked by two conditions that have been modeled by the first two condition-specific regressors (columns) in the design matrix. This procedure results in an ‘image’ of statistics (i.e. Statistical Parametric Map - SPM). Drawing statistical inferences from this SPM relies upon reliably locating voxels where an effect is present whilst controlling against the probability of false positives (i.e. the multiple comparisons problem).

The problem of multiple comparisons arises when we try to reject the null hypothesis by using more than one statistical test. Given that there are hundreds of thousands of voxels in the brain, the problem of multiple comparisons (i.e. the chance of finding false positives) is particularly severe. In classical statistics, an alpha level of 0.05 is traditionally used. In functional MRI, there is a general consensus that a more conservative threshold of 0.001 uncorrected should be employed to take account of the issue of multiple comparisons. This alpha value is derived from approaches based on Gaussian random field theory (Frackowiak et al. 2004). Given that fMRI images are usually smoothed before statistical analysis, one can conceive of entities that consist of clusters of voxels (“resels”). Since there are fewer resels, as compared to the number of

voxels, whole brain correction based on Gaussian random field theory typically gives a higher voxel-wise alpha value, compared for example to an overly stringent Bonferroni correction. It also follows that using a larger smoothing kernel (FWHM) increases statistical power. However, this arises at the cost of reduced spatial specificity.

### **3.5.5 Random effects group analysis**

Random effects analyses are employed to allow inferences to be made about the population from which the sample of subjects was drawn. In contrast, fixed effects analyses are less generalisable given that a significant effect may be driven by a small proportion of the subjects. Typically, a contrast of parameter estimates from a 1st level analysis is entered into a random effects analysis. This type of analysis, therefore, is not at risk of being biased by strong effects in a subset of subjects, given that the effect size is set against the between subject variability. It follows that more subjects are required to achieve a significant result with random effects analyses, as the degrees of freedom depend on the number of subjects scanned, a suitable minimum number of subjects being 12. Random effects analyses are typically a one sample t-test testing whether the estimated effect size (i.e. contrast) is significantly greater than zero across all subjects. An alternative approach is to conduct a repeated measures ANOVA which may be useful in analyses of covariance (i.e. ANCOVA). Random effects analyses were adopted for the analysis of Experiment 2. However, multivariate analysis for Experiments 3 and 4 was done on a single subject basis (see Chapters 6 and 7).

### **3.5.6 Functional MRI experimental design and efficiency**

Two basic types of experimental design are used in fMRI studies: blocked designs and event-related designs. In block designs, haemodynamic responses to changes in brain state are measured whilst subjects engage in a task for extended periods of time (e.g. an epoch of 30 seconds). During event-related fMRI, in contrast, the haemodynamic response to a single event can be examined in a context (i.e. state) independent fashion. Event-related and block designs are typically modelled using stick or boxcar functions, respectively. Block designs are highly efficient in statistical terms (Friston et al. 1998) and were used exclusively in this thesis as the paradigms employed were set in ecologically valid contexts and thus naturally tended towards longer trial epochs. In

blocked designs, the length of epochs is an important factor affecting efficiency with maximum power achieved with a design involving 16 seconds of task alternating with a similar period of time off task. This was borne in mind during the design of each study along with considerations about inter-trial time intervals which were generally set long enough to allow the haemodynamic response to fall back to baseline.

It is worth stressing that considerations relating to design efficiency need to be balanced with psychological constraints and a compromise reached. Indeed, throughout the experiments presented in this thesis, a strong emphasis was placed upon ensuring psychological validity of the relevant paradigms established by extensive behavioural piloting. Although care was taken to maximise design efficiency as far as possible, precedence was given to psychological constraints. For example, in Experiment 2 where subjects were required to richly imagine new experiences the length of the task epochs was determined following pilot studies that established the optimal block duration. In this epoch durations happened to match with the length thought to be optimal (i.e. around 16 seconds) purely on statistical grounds. Specific details pertaining to individual experiments are described in the relevant chapters.

### **3.6 Multivariate pattern classification techniques**

To date most fMRI studies have been analysed using conventional mass univariate methods such as the general linear model described above (Frackowiak et al. 2004). However, recently there has been a surge of interest in the application of multivariate pattern classification techniques to neuroimaging analysis (Cox & Savoy 2003; Haxby et al. 2001; Haynes & Rees 2005a; Haynes & Rees 2006; Kamitani & Tong 2005; Norman et al. 2006; Polyn et al. 2005). Univariate analysis considers the activity of each voxel in isolation. By contrast multivariate pattern analysis (MPA) can extract information distributed in the pattern of activity of multiple voxels. A major development in the last few years has been the realisation that fMRI data analysis can be construed, at a high level, as a pattern-classification problem (i.e. how we can recognise a pattern of brain activity as being associated with one cognitive state versus another). As such, techniques that have been developed for pattern classification and data mining in other domains (e.g. image recognition) can be productively adapted and applied to fMRI data analysis.

### **3.6.1 The advantages of multivariate pattern analysis**

There are several advantages of MPA over conventional univariate approaches. Firstly, the weak information available at each location can be accumulated in an efficient way across many spatial locations (Haynes & Rees 2005a). Secondly, even if two single brain regions do not individually carry information about a cognitive state, they might nonetheless do so when jointly analysed (Sidtis et al. 2003). Thirdly, most conventional studies employ processing steps (such as spatial smoothing) that remove fine-grained spatial information that might carry information about perceptual or cognitive states of an individual. This information is discarded in conventional analyses, but can be revealed using methods that simultaneously analyse the pattern of brain activity across multiple locations. Fourthly, conventional studies usually probe whether the average activity across all task trials during one condition is significantly different from the average activity across all time points during a second condition. Typically these studies acquire a large number of samples of brain activity to maximise statistical sensitivity. However, by computing the average activity, information about the functional state of the brain at any given time point is lost. By contrast, the increased sensitivity of decoding-based approaches potentially allows even online estimates of a person's perceptual or cognitive state (Haynes & Rees 2005b; Polyn et al. 2005). In summary, conventional univariate analyses may be inadvertently throwing away signal that is only present in the pattern of activity of multiple voxels.

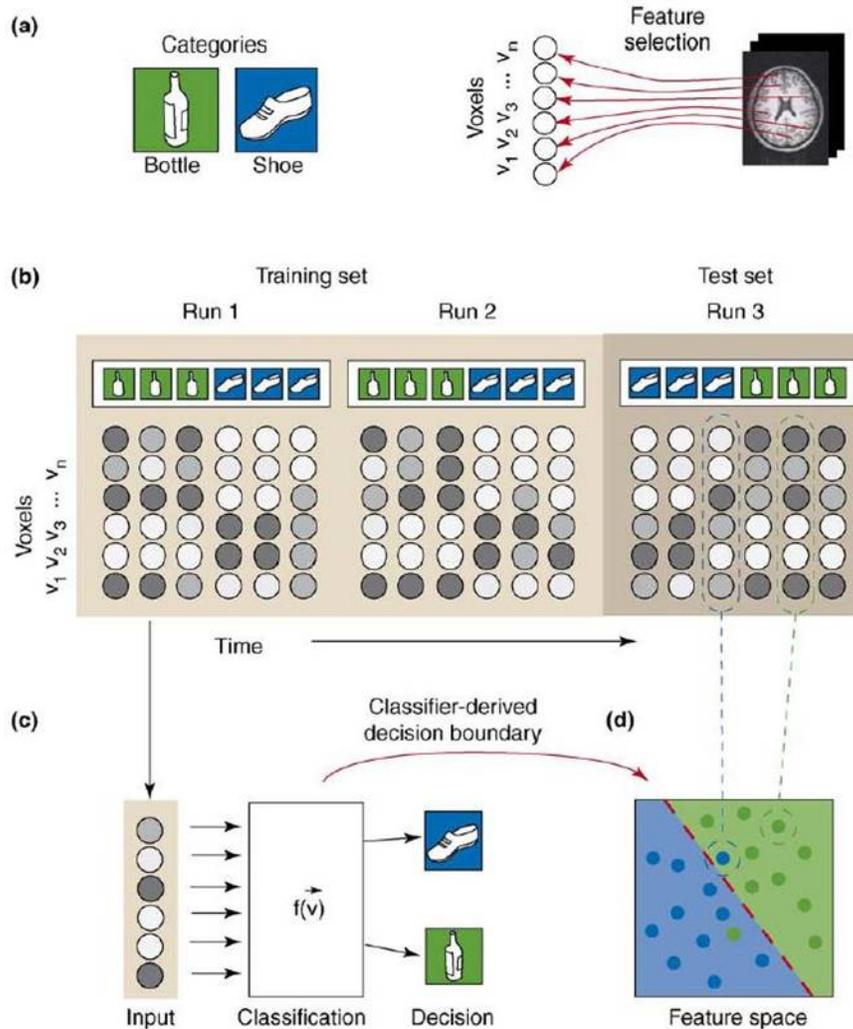
### **3.6.2 General overview of classification methods**

Here I will give a brief and general overview of MPA methods. More details and study-specific experimental procedures can be found in the methods sections of Chapters 6 and 7. The basic MPA method is a straightforward application of pattern classification techniques, where the patterns to be classified are typically vectors of voxel activity values. Figure 14 illustrates the four basic steps in an MVPA analysis. The first step, feature selection, involves deciding which voxels will be included in the classification analysis (Figure 14a). There are number of ways of doing this and the best way depends on the specific paradigm and question being addressed. The simplest method involves selecting regions of voxels based on a priori anatomical interest (Haxby et al. 2001).

Other approaches taken have involved the first pass use of univariate statistics such as t-tests on individual voxel activities between conditions (Haynes & Rees 2005a) to maximise the chance of finding a discriminable set of voxels for pattern classification.

The second step, pattern assembly, involves sorting the data into discrete 'brain patterns' corresponding to the pattern of activity across the selected voxels at a particular time in the experiment (Figure 14b). Brain patterns are labelled according to which experimental condition generated the pattern; this labelling procedure needs to account for the fact that the haemodynamic response measured by the scanner is delayed and spread over time, relative to the instigating neural event. The third step, classifier training, involves feeding a subset of these labelled patterns into a multivariate pattern classification algorithm. Based on these patterns, the classification algorithm learns a function that maps between voxel activity patterns and experimental conditions (Figure 14c). The fourth step is generalisation testing: given a new pattern of brain activity (not previously presented to the classifier), can the trained classifier correctly determine the experimental condition associated with that pattern (Figure 14d).

Machine learning researchers have developed an enormous range of classification algorithms that can potentially be used in MPA studies (see (Duda et al. 2001) for details) including correlation-based classifiers (Haxby et al. 2001), neural networks without a hidden layer (Polyn et al. 2005) and linear discriminant analysis (Haynes & Rees 2005a; Haynes & Rees 2005b). In this thesis (Experiments 3 and 4) I used linear support vector machines (SVMs) (Cortes & Vapnik 1995) as they are computationally efficient and are well-suited to problems, such as neuroimaging, where the feature space (in this case voxels) is relatively large compared to the number of observations (in this case brain images) (Cortes & Vapnik 1995).



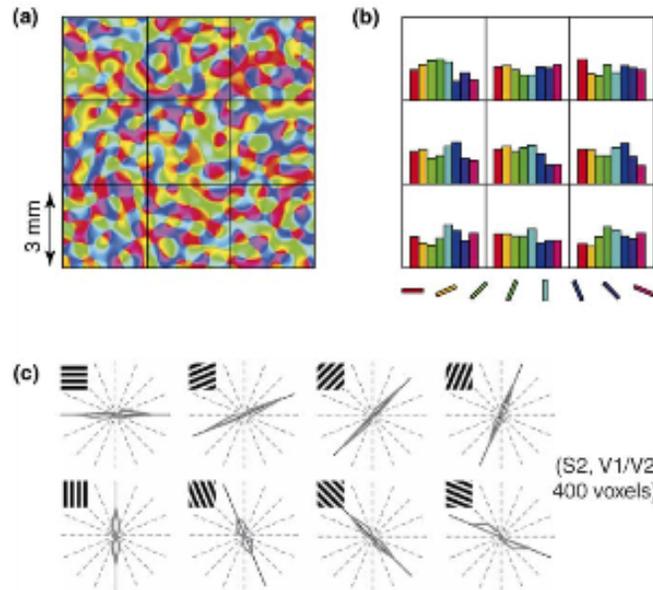
**Figure 14. Illustration of a hypothetical experiment analysed using MPA**

(a) Subjects view stimuli from two object categories (bottles and shoes). A 'feature selection' procedure, such as the application of an anatomical mask, is used to determine which voxels will be included in the classification analysis. (b) The fMRI time series is decomposed into discrete brain patterns that correspond to the pattern of activity across the selected voxels at a particular point in time. Each brain pattern is labelled according to the corresponding experimental condition (bottle versus shoe). The patterns are divided into a training set and a testing set. (c) Patterns from the training set are used to train a classifier function that maps between brain patterns and experimental conditions. (d) The trained classifier function defines a decision boundary (red dashed line, right) in the high-dimensional space of voxel patterns (collapsed here to 2-D for illustrative purposes). Each dot corresponds to a pattern and the colour of the dot indicates its category. The background colour of the figure corresponds to the guess the classifier makes for patterns in that region. The trained classifier is used to predict category membership for patterns from the test set. The figure shows one example of the classifier correctly identifying a bottle pattern (green dot) as a bottle, and one example of the classifier misidentifying a shoe pattern (blue dot) as a bottle. From (Norman et al. 2006).

### **3.6.3 Decoding fine-grained patterns of representation**

The experimental opportunities afforded by MPA go beyond the extraction of extra information from patterns of activity in multiple voxels. Using MPA, fine-grained neural patterns below the resolution of a voxel can be detected via the phenomenon of voxel sampling bias (Haynes & Rees 2005a). Many detailed neural features are represented at a much finer spatial scale in the cortex than the resolution of fMRI. For example, neurons coding for high-level visual features in the inferior temporal cortex are organised in columnar representations that have a finer spatial scale than the resolution of conventional human neuroimaging techniques (Wang et al. 1996). Similarly, low-level visual features, such as the orientation of a particular edge, are encoded in the early visual cortex at a spatial scale of a few hundred micrometres (Obermayer & Blasdel 1993).

Nevertheless, recent work demonstrates that pattern-based decoding of BOLD contrast fMRI signals acquired at relatively low spatial resolution can successfully predict the perception of such low-level perceptual features (an example is shown in Figure 15). For example, the orientation (Haynes & Rees 2005a; Kamitani & Tong 2005), and perceived colour (Haynes & Rees 2005b) of a visual stimulus presented to an individual can be predicted by decoding spatially distributed patterns of signals from local regions of the early visual cortex. These spatially distributed response patterns reflect biased low-resolution sampling by fMRI voxels of slight irregularities in such high resolution feature maps (Haynes & Rees 2005a; Kamitani & Tong 2005) (Figure 15a). In other words as long as the low resolution neural features being investigated are not uniformly distributed across voxels there will be some asymmetry between conditions across patterns of multiple voxels that could potentially be picked up by a sensitive classifier.



**Figure 15. Orientation decoding from fMRI activity in visual cortex**

Parts (a) and (b) illustrate how voxels acquire weak sensitivity for line orientation. (a) This shows a simulated orientation tuning map for a patch of visual cortex (different colors indicate different orientations), with a voxel-sized (3 mm) grid superimposed on the map. (b) This shows the distribution of orientation selectivity values for each of the nine 'voxels' shown in (a). Although all of the orientations are represented inside each voxel, the distribution of selectivity values is slightly different for each voxel. The classifier is able to exploit these small per-voxel irregularities in selectivity to decode orientation from multi-voxel patterns. (c) This illustrates the performance of the classifier in (Kamitani & Tong 2005) for a single subject. The polar plots show the classifier's orientation predictions for eight different (actual) line orientations; predictions were based on 400 voxels in V1 and V2. For these voxels, most of the classifier's predictions exactly matched the correct orientation, and the classifier's (rare) mistakes were all tightly clustered around the correct orientation. From (Kamitani & Tong 2005).

# Chapter 4

*Experiment 1: Patients with hippocampal amnesia can't imagine new experiences\**

\*This chapter derives in part from: "Patients with hippocampal amnesia cannot imagine new experiences" Hassabis D, Kumaran D, Vann SD, Maguire EA (2007). *Proceedings of the National Academy of Sciences USA*. 104, 1726-31.

## **Precis**

The current general consensus (as outlined in Section 1.3) is that episodic memory recall is a reconstructive process rather than a literal reproduction of the past (Bartlett 1932; Schacter et al. 1998). If this is the case then I wondered if a purely constructive task, such as imagination, would rely on some of the same processes and therefore be supported by similar brain regions to reconstruction. The hippocampus has long-been known to play a critical role in episodic memory (see Section 1.4.1) so I tested patients with hippocampal amnesia on their imagination abilities. I hoped not only to reveal and characterise a novel deficit, but also to provide new insights into the role of the hippocampus in episodic memory.

## **4.1 Introduction**

Each of us has our own unique personal past, comprising a myriad of autobiographical experiences accrued over a lifetime. Recollection of these rich autobiographical or episodic memories has been likened to mentally travelling back in time and re-experiencing one's past (Tulving 2002b). It has long been known that the hippocampus and related MTL structures play a critical role in supporting episodic memory (Scoville & Milner 1957), and damage to the hippocampus alone is sufficient to cause amnesia (Rempel-Clower et al. 1996; Spiers et al. 2001; Zola-Morgan et al. 1986). How exactly the hippocampus supports episodic memory (Eichenbaum 2004; Moscovitch et al. 2005; Squire et al. 2004), or indeed whether its involvement is time-limited (Bayley et al. 2005; Bayley et al. 2003; Squire 1992; Squire et al. 2004) or permanent (Moscovitch et al. 2006; Moscovitch et al. 2005; Nadel & Moscovitch 1997) is uncertain, however. Numerous studies have attempted to settle this debate by ascertaining the status of remote episodic memory in patients with hippocampal amnesia (Spiers et al. 2001) but without resolution thus far. This is not altogether surprising as studying memory for personal experiences is fraught with methodological issues (Bayley et al. 2003; Levine et al. 2002; Maguire 2001a), not least of which is how to generalise across individuals when autobiographical memories are unique to each person (Burgess 2002; Moscovitch et al. 2006).

In this study, I sought to further the understanding of the role of the hippocampus in episodic memory by adopting a novel approach. If patients with hippocampal damage are impaired at recollecting past events, can they imagine new experiences? While there have been some suggestions that amnesic patients have difficulty envisioning themselves in the future (Atance & O'Neill 2001; Klein et al. 2002; Rosenbaum et al. 2005; Tulving 1983), surprisingly, the more general question of whether imagining new experiences depends upon a functioning hippocampus has not been formally addressed. In fact, episodic memory and imagining or constructing events share striking similarities in terms of the psychological processes engaged (Conway & Pleydell-Pearce 2000; Greenberg & Rubin 2003; Schacter 1996). These include imagery (Rubin et al. 2003), sense of presence (Tulving 2002b), retrieval of semantic information and multi-modal details (Wheeler et al. 1997), and narrative structure (Rubin et al. 2003). Moreover, both episodic memory and construction involve the salient visualisation of an experience within a rich spatial setting or context (Burgess 2002; Burgess et al. 2001), and therefore differ markedly from 'simple' visual imagery (e.g. for faces or single objects) (Kosslyn et al. 2001) which is thought not to depend the hippocampus (Rosenbaum et al. 2004). Constructions, then, have much in common with episodic memories but have the advantage of being easier to systematise and experimentally manipulate. For example, all patients can be asked to construct the same fictitious situations, and their performances can be compared and contrasted more directly than would be possible in a standard episodic memory recall paradigm.

I therefore tested whether a group of patients (n=5) with amnesia associated with bilateral hippocampal damage (see Section 4.2.2) and a group of matched control subjects (n=10) could construct new imagined experiences in response to short verbal cues which outlined a range of simple commonplace scenarios (see Section 4.2.3). When imagining a new experience participants were explicitly told not to describe a remembered event or any part of one but rather to give free reign to their imaginations and create something new. They were also encouraged to 'see the situation and setting in their mind's eye' as if they themselves were physically present, and to describe as many sensory and introspective details about the situation as they could. These descriptions were then scored along a range of parameters in order to address two questions: (i) is the hippocampus critical for imagining experiences, paralleling its vital

role in recollecting the past; and if so (ii) is there a specific hippocampal mechanism underpinning imagining that might also bear on its role in episodic memory?

## **4.2 Methods**

### **4.2.1 Participants**

Five patients took part (all male, one left-handed) each with primary damage to the hippocampus bilaterally, and concomitant amnesia. All but one patient have been reported previously (Aggleton et al. 2005; Hartley et al. 2007; Maguire et al. 2006a; Samarasekera et al. 2007) (see Section 4.2.2 for summaries of each patient case). Hippocampal damage in three of the cases resulted from limbic encephalitis associated with voltage gated potassium channel antibodies (VGKC-Ab), meningoencephalitis and then recurrent meningitis in one case, and limbic encephalitis (not associated with VGKC-Ab) in one case. The mean age of the patients was 52.8 years (SD 18.5, range 24-70), years of education 14.0 years (SD 3.7, range 11-19) and verbal IQ was 103.2 (SD 11.7, range 90-116). All of the patients had significant impairment of anterograde memory, some deficient on both recognition and recall tests, others on recall tests alone. Retrieval of pre-morbid semantic memory was intact in all cases, while retrograde memory for episodic experiences was impaired, with the amnesic period ranging from 10 years to a complete lifetime. Language, perceptual, verbal fluency, and executive functioning were within the normal range in all cases.

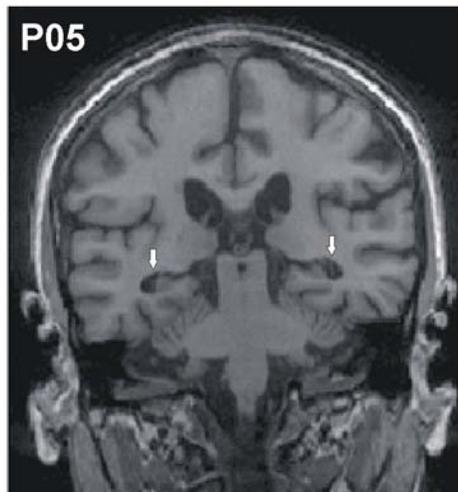
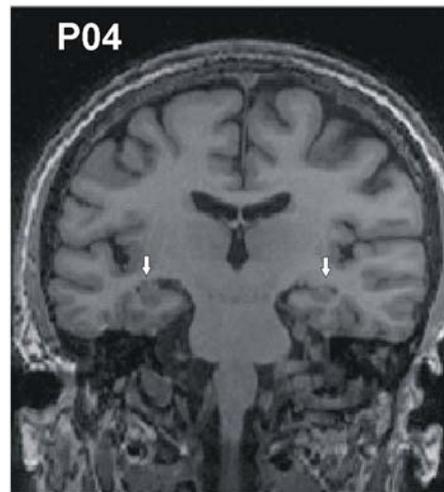
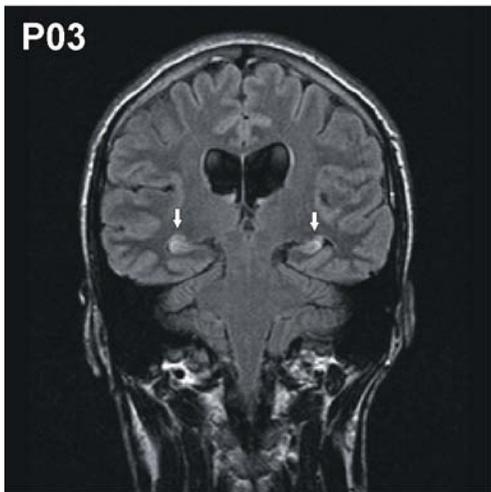
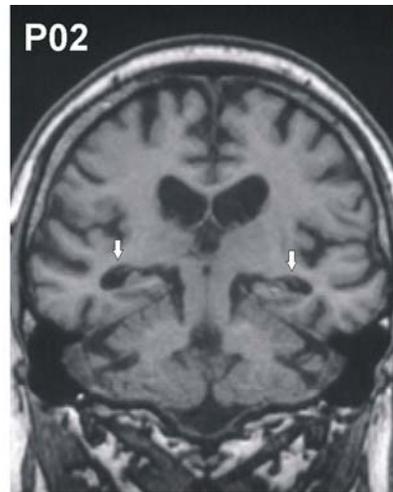
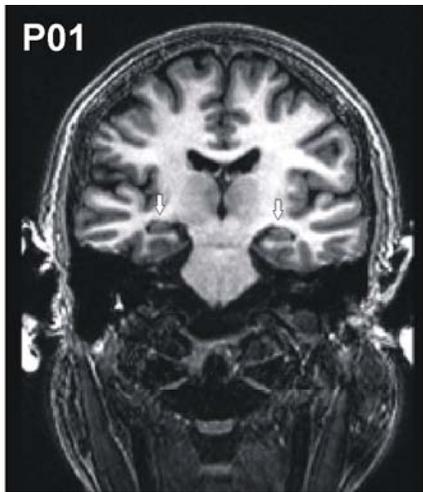
Lesions were confirmed by structural MRI scans, and appeared to implicate the hippocampi, with no evidence of damage in adjacent medial temporal areas. It is notoriously difficult, if not impossible, to be certain in vivo that lesions are selective to a particular brain region. Even scrupulous measurements or ratings of tissue volume from MRI scans cannot provide a definitive indication as to whether the tissue is functioning or not (Maguire et al. 2005; Maguire et al. 2006a; Maguire et al. 2001b) (see Section 4.2.2), or the functional effect of a lesion on wider brain systems. In the patients tested, the primary area of damage in every case seemed restricted to the hippocampi. This was the only area of overlap. Their neuropsychological profiles suggested an isolated memory impairment, and their performance on the experimental task was remarkably homogenous (excluding P01 – see Section 4.3.3). Notwithstanding the difficulties

inherent in the field, I therefore feel that the present data permit conclusions to be drawn regarding the hippocampus.

Ten healthy control participants also took part (all male, one left-handed). The mean age of the control subjects was 52.2 years (SD 16.9, range 25-76), years of education was 14.1 years (SD 2.8, range 11-17), and verbal IQ was 104.3 (SD 6.3, range 94-112). There was no significant difference between the patients and controls on these background characteristics (age  $p=0.95$ ; education  $p=0.95$ ; IQ  $p=0.81$ ). As well as comparing the two groups, I ensured that each patient was matched exactly to two of the control subjects on age, education and IQ. All participants gave informed written consent to participation in the study in accordance with the local research ethics committee.

#### **4.2.2 Patients**

In this section I provide basic summaries of each patient's case. Full details of neuropsychological tests and scores are available in other published reports of these cases (references provided below). Where information relevant to this study is not provided elsewhere, I include it here in the summary.



**Figure 16. Patient structural MRI scans**

Coronal sections through the MRI brain scans of each patient, where the damaged hippocampi are indicated by white arrows. Note that P03's scan is a FLAIR image, while the other scans are T1 images.

#### *4.2.2.1 P01*

This patient was male, right-handed, and aged 46 at the time of testing. His case has been described in detail elsewhere (patient KN (Aggleton & Brown 1999; McKenna & Gerhand 2002)). To summarise, this university-educated former industrial biochemist contracted meningoencephalitis at the age of 34 and then recurrent meningitis. He was left without useful motor function below T12, and amnesia. Recent MRI scanning showed bilateral abnormalities in the occipital lobes, with the other main location of volume reduction in the hippocampi (reduced by 48.8% in the left and 46.2% in the right) (see Figure 16). In terms of his neuropsychological profile, his IQ was in the high average range (113) (WTAR (Wechsler 2001)). He performed normally on tests of language, executive function, and perception. His anterograde memory for episodic information was grossly impaired, while he retained some ability to acquire new semantic information. He performed within normal limits on some tests of recognition, but was very impaired on tests of recall. His retrograde memory for autobiographical events was grossly impaired across four decades, with virtually no reliable recollections (as verified by his spouse), while his retrograde memory for personal and general semantics was intact.

#### *4.2.2.2 P02*

This patient was male, right-handed, and aged 66 at the time of testing. His case has been described in detail elsewhere (patient TT (Maguire et al. 2006a)). To summarise, prior to his illness, he worked for 37 years as a licensed London taxi driver. He presented with a 6-week history of rapidly progressive amnesia and confusion, 2 weeks after a diarrhoeal illness. Four months later, he had a series of complex partial and tonic-clonic seizures. After extensive clinical investigations, it was established that he had limbic encephalitis associated with voltage gated potassium channel antibodies (VGKC-Ab (Vincent et al. 2004)). His most recent MRI brain scan showed damage throughout the length of both hippocampi. There was some generalised atrophy, however in this context entorhinal, perirhinal and parahippocampal cortices appeared intact, as did the mamillary bodies, fornix and thalamic nuclei (see Figure 16). In terms of his neuropsychological profile, his verbal IQ was average (90). He performed normally on a wide range of tests including those assessing executive, perceptual and language functions. By contrast, he had a severe impairment of anterograde memory for episodic and semantic information, encompassing recognition and recall, in the visual and verbal

domains. He had profound retrograde amnesia for autobiographical event memories extending back over sixty years.

#### 4.2.2.3 P03

This patient was male, left-handed, and aged 24 at the time of testing. His case has been described elsewhere (patient 3 (Samarasekera et al. 2007)). To summarise, three years earlier he had been at university studying for a PhD. He presented following two generalised seizures which followed a flu-like illness lasting three days. He was admitted to the intensive care unit because of persistent seizures that were brought under control after treatment. At that time he reported symptomatic problems with memory. MRI scan early in the course showed high FLAIR signal in both hippocampi (see Figure 16). The diagnosis was limbic encephalitis, in this instance not associated with VGKC-Ab. On follow up at 27 months he had been forced to abandon his higher degree because of a persisting memory impairment. In terms of his neuropsychological profile, his verbal IQ was in the high average range (116). At the time of this study, language, perceptual and executive functions (verbal fluency scaled scored 9) were within the normal range. He was impaired on anterograde memory tests in the visual domain in particular. He had a significant retrograde amnesia for autobiographical events – Autobiographical Memory Interview: *Childhood* semantic 13.5/21; autobiographical 3/9; *Early Adult Life* semantic 18.5/21; autobiographical 4/9; *Recent Life* semantic 17/21; autobiographical 6/9).

#### 4.2.2.4 P04

This patient was male, right-handed, and aged 58 at the time of testing. His case has been described elsewhere (patient KC3 (Hartley et al. 2007)). He left school aged 11 but later returned to complete a university degree. Prior to his illness he had worked as a market trader for more than 20 years. He presented with sudden onset memory difficulties associated with what he described as ‘panic attacks’ but which were subsequently confirmed as seizures. MRI scan revealed abnormal high signal restricted to the hippocampi and amygdalae. After extensive clinical investigations, it was established that he had limbic encephalitis associated with VGKC-Ab. Scanned 18 months later, the abnormalities that were evident on MRI had resolved, although his hippocampi were noted to be small (Figure 16). He is not able to return to work because of persisting memory problems. In terms of his neuropsychological profile, his verbal IQ was in the average range (99). Although English was not his first language, limiting the

range of tests that could be administered, at the time of this study he performed within normal limits on tests of language, and perception, and scored just below average on a test of verbal fluency (scaled score 7). On anterograde memory tests, his performance was low average-average on tests of verbal and visual recognition. In contrast, his verbal recall was impaired. Testing of retrograde memory (verified by his spouse) revealed amnesia dating back approximately 10 years for autobiographical events.

#### *4.2.2.5 P05*

This patient was male, right-handed and aged 70 at the time of testing. He was a retired painter and decorator who had left school aged 15. He presented with a persistent memory impairment. Clinical investigations revealed that he had limbic encephalitis associated with VGKC-Ab. His most recent MRI scan showed apparently selective atrophy of both hippocampi along their entire length (see Figure 16). In terms of his neuropsychological profile, his IQ was within the average range (VIQ 103, PIQ 90, FS IQ 96). At the time of this study, his language skills were intact (WAIS-R vocabulary high average; Graded Naming Test low average), while perceptual (VOSP cube analysis 8/10) and executive abilities (Weigl Sorting Test passed, Trail Making Test passed, verbal fluency scaled score 8) were also average. In contrast, his anterograde memory was significantly impaired at recognition (Warrington Recognition Memory Test words and faces both less than 5<sup>th</sup> percentile) and recall (WMS-R Logical Memory delayed recall less than 5<sup>th</sup> percentile). His retrograde amnesia for autobiographical events stretched back over approximately 20 years.

#### **4.2.3 Task and procedure**

Each participant was tested individually and sat facing the examiner. Testing sessions were digitally recorded to enable transcription and later scoring of participants' responses. The requirements of the imagining new experiences task were explained, and several examples provided. During this practice phase it was also established that patients could remember the instructions and the cues throughout a construction trial. Commonplace ordinary settings were chosen as scenarios to minimise the difficulty level, and be as independent from a participant's innate creative ability as possible. The scenarios also purposely encompassed a wide variety of different subject matter from

the man-made to the natural, and the busy to the isolated to ensure there were no content biases.

Each participant was tested on 10 scenarios covering a range of themes. Seven were standard commonplace scenarios (involving a beach, museum, pub, port, market, forest, and castle setting). I also examined the effect of scenarios that were explicitly self-relevant and potentially plausible in the future, so-called 'episodic future thinking' (Atance & O'Neill 2001; Klein et al. 2002) (possible Christmas event, possible event over next weekend, possible future meeting with a friend). Performances on the two scenario types were initially analysed separately. However, both had identical patterns of results, and so for clarity I present the results collapsed across scenarios (see Section 4.3).

For each scenario a short description was read out loud by the interviewer from a prepared script (e.g. 'Imagine you're lying on a white sandy beach in a beautiful tropical bay'; see also Figure 17) and the participant was instructed to vividly imagine the situation from the cue and describe it in as much (multi-modal) detail as possible. Participants were explicitly told not to recount an actual memory or any part of one but rather create something new. A printed text card was placed on the desk in front of the participant summarising the main concept of the scenario to act as a reminder if needed. Participants were allowed to continue with their descriptions until they came to a natural end or they felt nothing else could be added. Descriptions were recorded and later scored across a number of ratings using a similar method to the Autobiographical Interview (Levine et al. 2002) that has been used to assess the richness of episodic memories, but with additional ratings and measures to fit the requirements of our task (see Section 4.2.4).

A probing protocol dictated the appropriate use of statements used by the examiner during the session. These mostly took the form of general probes encouraging further description (e.g. 'can you see anything else in the scene?'), or asking for further elaboration on a theme introduced by the subject (e.g. 'can you describe the fishing boat in more detail?' in response to the subject saying 'I can see a small fishing boat gently rocking out in the sea'). It was strictly prohibited for the examiner to introduce any concept, idea, detail or entity that had not already previously been mentioned by the subject. After each scenario, participants were asked to rate their constructions across a

number of different categories (see Sections 4.2.3 and 4.3). At various points during a trial, and prior to the post-scenario ratings, the examiner verified that the participant still recalled the task instructions, the scenario in question, and the scenario he had created.

In a second version of the task, the 'assisted' version (see Section 4.3), the scenarios were similar in nature and the task identical except that before commencing imagination, four-colour photographs of objects relevant to the scenario, a relevant sound was played, and a relevant smell was provided. These were available to the subject throughout the imagination task, with the sound periodically played by the examiner, and the subject invited to smell the relevant odour.

#### **4.2.4 Scoring system**

After extensive pilot testing, I identified the key elements that captured the essence of detailed descriptions, as agreed across the experimenters. Each of these elements (content, participant ratings, spatial coherence index, quality judgements) was regarded as important, but I reasoned that the most directly observable measure, information content across the 4 categories, should receive the most weight (almost 50%). Participant ratings and spatial coherence received roughly similar proportions of the remainder, and quality judgments slightly more. However, as stated in Section 4.3, even when the quality judgement score was omitted, the findings were unchanged.

A composite score, the Experiential Index, ranging from 0-60, measuring the overall richness of the imagined experience, was calculated from the four sub-components described below.

##### *4.2.4.1 Content*

Each scenario description was segmented into a set of statements (see Table 2). Every statement was then classified as belonging to one of 4 main categories: spatial reference, entity presence, sensory description, or thought/emotion/action. Repeated statements, irrelevant details and other tangential information that could not be classified into one of these 4 categories were discarded. Extensive pilot studies indicated that the production of 7 details per category was an optimal reflection of performance whilst ensuring that those with more circuitous descriptions were not unfairly advantaged.

Thus, the score for each details category was capped at a maximum of 7, and the total possible content score for an imagined experience was out of 28.

#### *4.2.4.2 Participant ratings*

Two subjective self-ratings contributed to the Experiential Index, each varying on a scale from 1 to 5: sense of presence (1 - 'did not feel like I was there at all'; 5 - 'felt strongly like I was really there') and perceived salience (1 - 'couldn't really see anything'; 5 - 'extremely salient').

#### *4.2.4.3 Spatial coherence index*

As part of the feedback on each scenario participants were presented with a set of 12 statements each providing a possible qualitative description of the newly constructed experience (see Table 3). Participants were instructed to indicate the statements they felt accurately described their construction. They were free to identify as many or as few as they thought appropriate. Of the 12 statements, 8 were 'integrated' and indicated that aspects of the scene were contiguous (e.g. 'I could see the whole scene in my mind's eye') and 4 were 'fragmented' and indicated that aspects of the scene were not contiguous (e.g. 'It was a collection of separate images'). One point was awarded for each integrated statement selected and one point taken away for each fragmented statement. This yielded a score between -4 and +8 that was then normalised around zero to give final Spatial Coherence Index score ranging between -6 (totally fragmented) and +6 (completely integrated). Any construction with a negative Spatial Coherence Index was considered to be incoherent and fragmented. Only positive spatial coherence ratings contributed to the overall Experiential Index (see below) so as not to over-penalise fragmented descriptions.

#### *4.2.4.4 Quality judgement*

The final scoring component was the scorer's assessment of the overall quality of the construction. Scorers were instructed to rate on a scale of 0-10 how well they felt the description evoked a detailed 'picture' of the experience in their own mind's eye (0 - no picture at all... 10 - vivid, extremely rich picture). Quality ratings could range from 0 (indicating the construction was completely devoid of details and with no sense of experiencing) to 10 (indicating an extremely rich and highly evocative construction that appeared to emerge from an extremely vivid imagining).

#### 4.2.5 Calculating the experiential index

All the scoring elements were combined to create the overall Experiential Index, using the following formula:

$$\text{Experiential Index} = T + P + S + C + Q$$

Where:

T – Content score 0-28

P – Sense of presence 0-4 (rescaled from 1-5; see below)

S – Perceived salience 0-4 (rescaled from 1-5; see below)

C – Spatial Coherence 0-6

Q – Quality rating 0-18 (rescaled from 0-10; see below)

This yielded final Experiential Index scores ranging from 0 (not experienced at all) to a maximum of 60 (extremely richly experienced). Specifically, the various scoring components to the Experiential Index were: 28 from the categorised content statements, 8 from the participant self-ratings, P and S (which were rescaled for the purposes of contributing to the index to allow for a theoretical Experiential Index minimum of 0 to be possible), 6 from the spatial coherence index, and 18 from the scorer quality judgement. Note, for quality judgement, instead of requiring scorers to give a rather cumbersome rating between 0-18, it was deemed simpler for scorers to assign a more natural rating between 0-10 and for this score to later be multiplied by a factor of 1.8 to achieve the same effect.

Similar to the approach used in studies of autobiographical memory (Levine et al. 2002), the primary scorer was not blind to subject status. Therefore in order to assess inter-rater reliability and scoring system robustness, 2 subjects (1 patient and 1 control) were selected at random and their constructions (representing 13.3% of all constructions) were assessed by a second trained scorer who was blind to subject identity. A two-way ANOVA was performed with scorer and subject as factors. There was no significant main effect of scorer on Experiential Index ratings ( $F(1,9) = 0.117, p=0.74$ ) indicating a high degree of scorer reliability.

## 4.3 Results

### 4.3.1 Experiential index

Table 1 shows the mean scores on these measures for patients and controls. Examples of scenario cues, excerpts from patients' constructions, and those of matched control subjects are shown in Figure 17. The patient group scored significantly lower on the overall Experiential Index than the control group ( $p=0.002$ ) (Figure 18), thus revealing that the ability to richly imagine new experiences is compromised in the context of bilateral hippocampal damage.

**Table 1. Performance on the construction task**

	Mean (SD)	
	Patients (n=5)	Controls (n=10)
<b>Overall Richness:</b>		
Experiential Index	27.54 (13.12)	45.06 (4.02)
<b>Sub-components:</b>		
<i>Content</i>		
Spatial References	2.38 (1.82)	5.28 (1.15)
Entities Present	4.94 (1.26)	6.49 (0.42)
Sensory Descriptions	4.12 (1.03)	5.64 (0.72)
Thought/Emotion/Actions	2.76 (1.77)	5.52 (0.64)
<i>Participant Ratings</i>		
Sense of Presence	3.46 (1.15)	3.65 (0.49)
Perceived Salience	3.52 (1.19)	3.88 (0.48)
<i>Spatial Coherence</i>		
Spatial Coherence Index	0.10 (3.21)	3.68 (1.30)
<i>Scorer Rating</i>		
Quality Judgement	3.88 (2.70)	7.13 (0.96)

**Cue: *Imagine you are lying on a white sandy beach in a beautiful tropical bay***

**P03:** As for seeing I can't really, apart from just sky. I can hear the sound of seagulls and of the sea... um... I can feel the grains of sand between my fingers... um... I can hear one of those ship's hooters [laughter]... um... that's about it. *Are you're actually seeing this in your mind's eye?* No, the only thing I can see is blue. *So if you look around what can you see?* Really all I can see is the colour of the blue sky and the white sand, the rest of it, the sounds and things, obviously I'm just hearing. *Can you see anything else?* No, it's like I'm kind of floating...

**CON:** It's very hot and the sun is beating down on me. The sand underneath me is almost unbearably hot. I can hear the sounds of small wavelets lapping on the beach. The sea is a gorgeous aquamarine colour. Behind me is a row of palm trees and I can hear rustling every so often in the slight breeze. To my left the beach curves round and becomes a point. And on the point there are a couple of buildings, wooden buildings, maybe someone's hut or a bar of some sort. The other end of the beach, looking the other way, ends in big brown rocks. There's no one else around. Out to sea is a fishing boat. It's quite an old creaking looking boat, chugging past on its small engine. It has a cabin in the middle and pile of nets in the back of the boat. There's a guy in the front and I wave at him and he waves back...[continues]...

**Cue: *Imagine that you are standing in the main hall of a museum containing many exhibits***

**P05:** [pause] There's not a lot as it happens. *So what does it look like in your imagined scene?* Well, there's big doors. The openings would be high, so the doors would be very big with brass handles, the ceiling would be made of glass, so there's plenty of light coming through. Huge room, exit on either side of the room, there's a pathway and map through the centre and on either side there'd be the exhibits [pause] I don't know what they are [pause]...there'd be people. [pause] To be honest there's not a lot coming. *Do you hear anything or smell anything?* No, it's not very real. It's just not happening. My imagination isn't... well, I'm not imagining it, let's put it that way. Normally you can picture it can't you? I'm not picturing anything at the moment. *So are you seeing anything at all?* No.

**CON:** I'm standing in the hallway of this museum I've never visited before. There's an atmosphere here of people moving in expectation towards some paintings and sculptures. As with many museums it's a beautiful place, it's architecturally well constructed and pleasant to the eye. *So what does it look like?* It's a pillared hall and the floor is marble, the ceiling is domed and sculpted. There is a buzz about the place, I think there must some special exhibition on which I had not expected, because I've come just to see the general exhibits. In this hallway - although there are paintings there - there's nothing I actually recognise even though I'm fairly interested in art. I'd imagine that at some time a lot of these paintings would probably have actually rested in churches rather than galleries or private homes. The paintings are all the way round and I'm looking at one in particular which is almost straight ahead of me...[continues]...

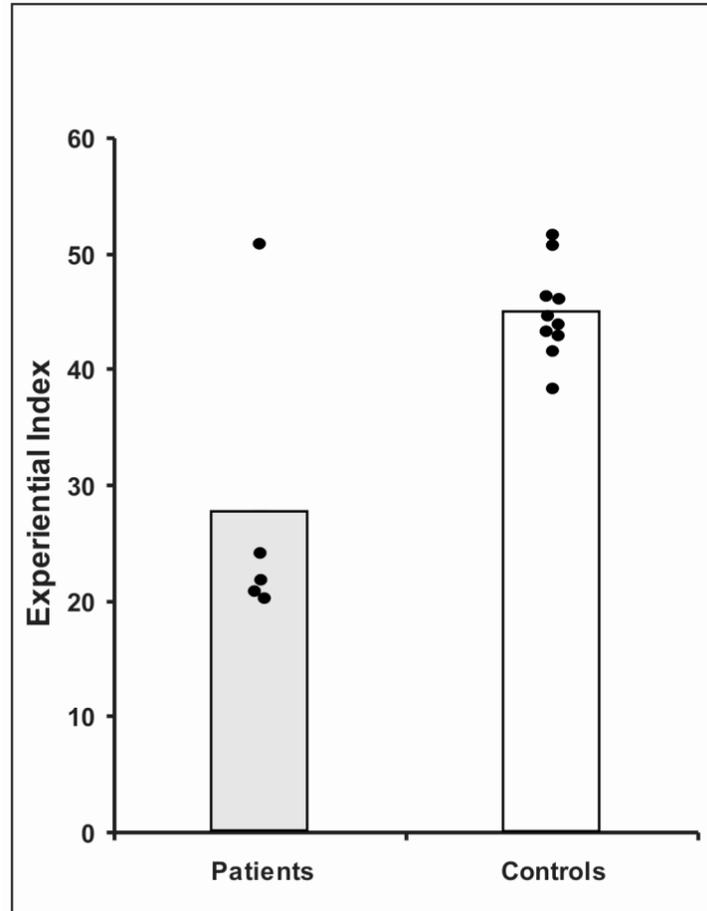
**Cue: *Imagine that you are standing in the middle of a bustling street market***

**P04:** Well, it shouldn't be difficult for someone who has done it for 24 and a half years [here P04 is referring to his previous job of being a market trader - see Supplementary Methods online]. The surrounding is very pleasant, it's a beautiful day. There's no bad smells, there is no dirt, which is quite unusual. Everything seems to be okay, everything is pleasant. People seem to be quite well disposed and basically there is a sense of [pause] happiness. *And if you look round what can you see?* When I look around I see people, very many people. Most of all... um... not many men, all I see are young ladies. And basically they are all in a hurry, clearly, because they would only have so much time. And... it's a precious moment. [pause] *And what does this market look like?* It could be almost any market, it could be St. Albans, it could be Petticoat Lane, it could be the Roman Road on Saturday, it could be Leather Lane on a Friday, it could be Charter Street when the days were good. *Okay, but what does it look like to you?* Well, it could be a market in Milan. Markets have something in common which is magical and that magic can only be experienced by someone who appreciates that sort of thing. A market is a wonderful place, maybe not for everyone, but on a lovely day I think you'd find as much happiness there as anywhere.

**CON:** Right, okay. So, on either side of me I've got stalls and it's noisy. We have a person on my right who is selling fruit and veg and they're telling us that the bananas are on special offer this week and they're shouting about that. On my left side I've got somebody selling china and he's trying to do the same thing. He's very keen you buy from him, 2 for 1 pound, that sort of thing. And further on you've got a guy selling suitcases and then somebody else selling some hardware stuff and a bit further on somebody selling fabric and there's somebody else on this side selling clothes. The people selling the fruit and veg are selling it off of something that looks like a cart and has got wheels, sort of parallel to where I'm looking. And everything has got an awning over it; they've all got their own individual covers. For some reason it feels quite narrow, but probably not as narrow as I think it is but more because of the number of people around and the sort of bustling that's going on. *So what does the market look like?* It's in a street and on this side there's a wall and on the other side a lot of the stalls are actually the fronts of shops, so they're actually part of the shop and it's got a sort of cobbled floor...[continues]...

### **Figure 17. Examples of imagined experiences**

Representative excerpts from transcriptions relating to three of the scenarios, with the cue at the top, followed by an excerpt from a patient's transcription, followed by that of a control subject who was age, education, and IQ matched to that patient. It is notable that P04, who was a market trader prior to his illness, scored poorly on the scenario based in a market (see third scenario above). This is despite having at his disposal a wealth of semantic knowledge accrued over 20 years of working in a market setting. Interviewer's probing comments in italics. Relevant background is information noted in square brackets.



**Figure 18. Scores on the experiential index**

The data points for every patient and control subject are shown. Vertical bars represent group means.

In the first instance, I determined if the impairment might be because patients found the task more difficult than the control subjects. After imagining each new experience, subjects rated how difficult they found this on a scale from 1-5 (1 - very easy...5 - very difficult). There was no significant difference between patients and controls in perceived difficulty of imagining scenarios (patient mean 2.20 (1.07), controls 2.13 (0.64),  $p=0.87$ ). I next considered the extent to which imagined experiences were derivatives of actual memories in the control subjects, despite the instruction to create something new. After each construction participants were asked to rate on a scale of 1-5 its similarity to an actual memory, in whole or in part (1 - nothing at all like any memories... 5 - exactly like a memory). The purpose of this measure was to verify that control subjects had adhered to the instructions to create something new. The low overall mean of the control

subjects (2.03 (0.62)) confirmed that this was the case, and therefore that successful construction of rich new experiences was not dependent on recalling real memories. The patient mean was 2.37 (0.90), and there was no significant difference between groups ( $p=0.40$ ).

I next sought to investigate the source of the patients' Experiential Index deficit further by analysing the sub-components that comprised this composite score.

### **4.3.2 Sub-components**

#### *4.3.2.1 Content*

Each scenario description given by the participant was segmented into a set of statements. Every statement was then classified as belonging to one of 4 main categories: spatial reference, entity presence, sensory description, or thought/emotion/action (TEA). The spatial reference category encompassed statements regarding the relative position of entities within the environment, directions relative to the participant's vantage point, or explicit measurements (e.g. 'behind the bar...' or 'to my left I can see...' or 'the ceiling is about 40ft high'). The entity category was a simple count of how many distinct entities (e.g. objects, people, animals) were mentioned (e.g. 'I can see some birds'). The sensory descriptions category consisted of any statements describing (in any modality) properties of an entity (e.g. the chair I'm sitting on is made of wood) as well as general weather and atmosphere descriptions (e.g. 'it is very hot' or 'the room is very smoky'). Finally the TEA category covered any introspective thoughts or emotional feelings (e.g. 'I have a sense of being alone') as well as the thoughts, intentions and actions of other entities in the scene (e.g. 'he seems to be in a hurry' or 'the barman is pouring a pint') (see Table 2 for an example categorisation).

**Table 2. Example of statement categorisation**

---

	SD	SPA	EP	TEA
The sea is very blue, if you look a bit further out you can see the waves breaking. There are				
palm trees behind me. There's some big shells in the area and a few people around but it's	EP	SPA	EP	SPA
not too busy. There are some people swimming... um... younger folk, and people out there	SD	TEA	SD	
sunbathing. I can also see a boat now, a native fishing boat, basically with an out-rigger, and			EP	SD
there are people with nets coming in. And there's somebody to my left with a barbecue	EP	EP	TEA	SPA
cooking the fish which I can smell as well...	TEA	SD		EP

---

SD = sensory descriptions (here n=6); SPA = spatial references (4); EP = entities present (distinct and unique, i.e. excluding repetitions of the same entity, 8 in this excerpt); TEA = thoughts/emotions/actions (5).

For each category, patients produced fewer details than controls: spatial references ( $p=0.002$ ), entities present ( $p=0.003$ ), sensory descriptions ( $p=0.005$ ), TEAs ( $p=0.001$ ). Although impaired, all patients performed above floor and were able to produce at least some associations and even substantial descriptions relevant to the scenario in question, as can be appreciated from Figure 17. None had deficits in semantic retrieval. Nevertheless, I next asked whether an inability to retrieve relevant semantic information might underpin their construction deficit. One patient was selected at random and tested in the same way on additional scenarios. However, on this occasion, before starting to imagine, the patient was provided with pictures of objects, as well as sounds and smells relevant to the scenario (see Section 4.2.3). This information, which remained present during the task, thus eschewed the need for retrieval from memory of relevant semantic information, and using the provided materials appropriately could produce a reasonable score. Despite this, however, performance did not improve, with no difference in Experiential Index scores between original and 'assisted' scenarios ( $p=0.96$ ), in the overall Content score ( $p=0.50$ ), or in the Spatial Coherence index ( $p=0.85$ ).

#### 4.3.2.2 Participant Ratings

Perhaps their imagined experiences lacked some key experiential qualities. Patients were tested for an explicit awareness of this by asking them to rate each construction on a scale of 1-5 for sense of presence (1 - 'did not feel like I was there at all'...5 - 'felt strongly like I was really there'). There was no significant difference between patients

and controls in perceived sense of presence (patient mean 3.46 (1.15), controls 3.65 (0.49),  $p=0.65$ ). Participants were also asked to rate the perceived salience of each imagined experience on a scale of 1-5 (1 - 'couldn't really see anything'...5 - 'extremely salient'). As with sense of presence, there was no significant difference between patients and controls in perceived salience (patient mean 3.52 (1.19), controls 3.88 (0.48),  $p=0.41$ ). Thus while patients were impaired in terms of the amount of content included in their imagined experiences, their subjective feelings of salience and sense of presence were no different to control subjects indicating that the deficit is subtle in nature. Although the feedback received from patients was confidently and promptly given, it cannot be completely excluded that these subjective ratings are somehow systematically different in patients compared with controls. This could be because they can less accurately remember their immediate task performance due to their amnesia, or the fact they don't have the same volume of rich experiences available to draw upon. However, if these participant ratings were somehow inaccurately scored by patients, and they should have scored themselves lower, then this would have only resulted in their Experiential Index being further decreased.

I next considered another factor fundamental to episodic experiences, namely the extent to which they felt like they were taking place in an integrated and coherent spatial context as opposed to merely being a fragmented collection of images.

#### *4.3.2.3 Spatial Coherence Index*

The Spatial Coherence Index is a measure of the contiguousness and spatial integrity of the imagined scene. After each scenario, participants were presented with a set of statements each providing a possible qualitative description of the imagined experience (see Table 3). They were instructed to indicate which statements they felt accurately described their construction. They were free to identify as many or as few as they thought appropriate. The Spatial Coherence Index was then derived from the statements identified (see Section 4.2.4.3). Some of the statements indicated that aspects of the scene were integrated (e.g. 'I could see the whole scene in my mind's eye') while others indicated that aspects of the scene were fragmented (e.g. 'It was a collection of separate images'). Participants were blind to the purpose of the statements and the concept of coherence.

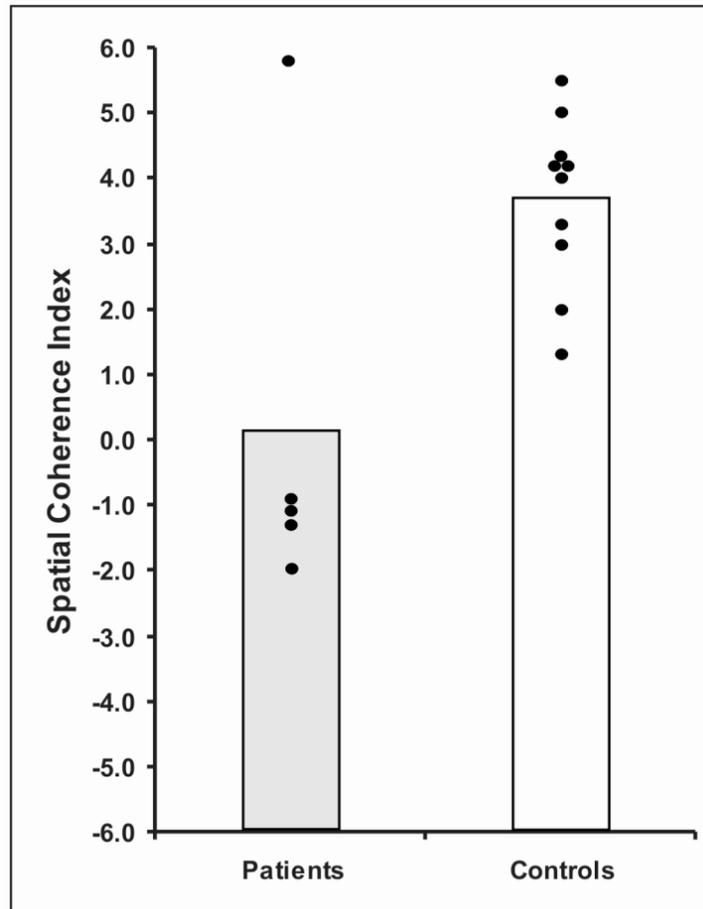
**Table 3. Spatial coherence index questionnaire**

---

1. It was quite fragmented
  2. I saw the scene in colour
  3. It was similar to looking at a picture or seeing it on TV
  4. I could see individual details, but it didn't all fit together as a whole scene
  5. I would find it easy to answer questions about the scene
  6. It wasn't so much a scene as a collection of images
  7. I was able to use some senses other than vision e.g. sound, smell
  8. I could see it as one whole scene in my mind's eye
  9. I was able to think of details associated with the general theme
  10. I would find it easy to give further details of the surroundings in the scene
  11. It wasn't a scene you could step into; it wasn't really joined-up
  12. I would find it easy to substitute an aspect of the scene for something else
- 

After each imagined scenario participants were instructed to indicate which of the above 12 statements they felt accurately described their construction. They were free to identify as many or as few as they thought appropriate. Of the 12 statements, 8 indicated that the imagined scene was 'integrated' and 4 that it was fragmented (statements 1, 4, 6, 11) (see Section 4.2.4.3).

Compared with controls, the feedback from the patients indicated that their imagined experiences were fragmentary and lacking in coherence ( $p=0.007$ ; see Figure 19). Although the Spatial Coherence Index encompasses a range of concepts, to ensure that the patient feedback was reliable I performed a correlation analysis between the spatial coherence scores and the data from the spatial reference content category, the most closely related (though narrower) objective measure. A strong trend was observed ( $p=0.07$ ) which, although was not significant, indicates that it is likely that something concrete and real was captured by the derived spatial coherence measure. Notably, when performing the spatial coherence task, several of the patients identified fragmentation as a relevant difficulty with the task at hand, but also pertinent to their memory problems in general.



**Figure 19. Scores on the Spatial Coherence Index**

The data points for every patient and control subject are shown. Vertical bars represent group means.

#### 4.3.2.4 Quality Judgement

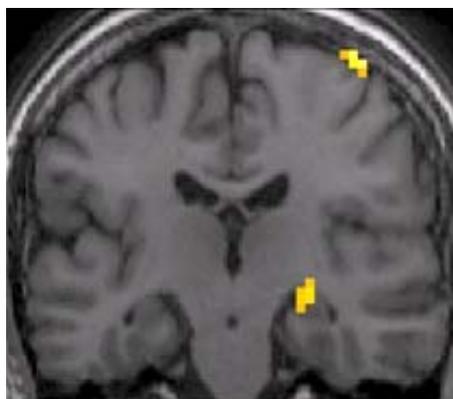
One final scoring component, the Quality Judgement, was the scorer's assessment of the overall quality of each construction (see Section 4.2.4.4). This was included in order to provide a measure of the range, diversity and quality of the details described and to ensure that short, succinct, but nonetheless salient descriptions were not unfairly penalised due to their brevity. Scorers were instructed to rate on a scale of 0-10 how well they felt the description evoked a detailed 'picture' of the experience in their own mind's eye (0 – no picture at all... 10 – vivid, extremely rich picture). The imagined experiences of the patients were judged as significantly poorer in quality than those of the controls ( $p=0.004$ ). The pattern of quality judgement scores matched the pattern of

the overall Experiential Indices very closely and in fact when an analysis was done without the contribution of the quality judgements to the Experiential Index, the patients still scored significantly lower than controls ( $p=0.001$ ).

#### **4.3.3 Outlier: P01**

While overall the patient group was impaired at imagining new experiences compared with the control group, examination of Figures 17 and 18 shows that one of the five patients (P01) (see Section 4.2.2) appeared to be unimpaired on the task. This was confirmed when his performance was compared with his two exactly matched control subjects. A separate analysis of the data excluding P01 and his two matched control subjects, however, made no difference to the statistical significance of any of the results reported above. The fact that the patient group were so impaired even when P01 was included illustrates the extent of the deficit in imagining new experiences. Furthermore, the other four patients were remarkably homogenous in their performances (see Figures 17 and 18), despite differences in age, IQ, and memory profiles (see Section 4.2.2).

P01 (patient KN (Aggleton & Brown 1999; McKenna & Gerhand 2002)) took part in a separate functional MRI (fMRI) experiment designed to ascertain if his residual hippocampal tissue was capable of activation (Maguire, Vann, Aggleton, unpublished observations). The paradigm and analysis parameters used were identical to that reported in control subjects by Maguire & Frith (Maguire & Frith 2004). Auditory stimuli and fMRI were used to assess brain activations associated with the incidental acquisition of facts about the world of the sort we are exposed to day to day. A control task was included where subjects heard novel sentences relating to unfamiliar people that did not convey general knowledge. In control subjects, the left hippocampus was active during knowledge acquisition compared with the control task. In P01 the hippocampus was also more active during knowledge acquisition, but this time on the right (Figure 20).



**Figure 20. Residual function of the right hippocampus in P01**

Activation, in patient P01, of the residual tissue of the right hippocampus during a semantic learning task.

This result suggests that the residual tissue in this patient's hippocampi, at least on the right, might be functionally active, and may contribute to his ability to acquire new semantic information, and imagine new experiences. He also retained the information presented during scanning at least several months after this initial exposure. Therefore whilst P01 was densely amnesic for episodic information (Aggleton & Brown 1999; McKenna & Gerhard 2002) his residual function of the right hippocampus might have been sufficient to support imagination and possibly even some semantic learning.

#### **4.4 Discussion**

In this experiment, I devised a novel paradigm where participants, rather than recollecting the past, had to imagine new experiences. The present results demonstrate that amnesic patients with bilateral hippocampal damage were significantly impaired on this task, and therefore enable several novel conclusions to be drawn. To begin with, this is the first systematic study to formally document that patients with hippocampal amnesia have a deficit in richly imagining new experiences. Secondly, in revealing this I show that the role of the hippocampus extends beyond reliving past experiences, encompassing not only imagining plausible self-relevant future events, but also more generally the construction of fictitious experiences. Thirdly, these findings offer some insight into a mechanism whose absence could underpin all of these deficits. The patients' imagined experiences were strikingly deficient in spatial coherence, resulting in

their constructions being fragmented and lacking in richness. The hippocampus, therefore, may make a critical contribution to the creation of new experiences by providing the spatial context or environmental setting into which details are bound (Burgess et al. 2001; Eichenbaum 2004; Moscovitch et al. 2005; O'Keefe & Nadel 1978). Given how closely imagined experiences match episodic memories, the absence of this function mediated by the hippocampus, may fundamentally affect the ability to re-experience or re-construct past events.

Whenever one examines patients with amnesia, and in particular using a novel task that takes more than a few seconds, it could be argued that the results merely reflect their anterograde memory impairment. However, patients had access to a reminder cue during each trial, and the examiner verified throughout that they had not forgotten the task, the instructions, or their own constructions. There were no instances of confusion, requests for clarification, and in particular no evidence of repetition, which one might have expected if patients were forgetting recently generated scene elements and then constructing them anew (see excerpts in Figure 17).

According to the standard model of memory (see Section 2.3) (Alvarez & Squire 1994; Squire 1992; Squire & Alvarez 1995; Squire et al. 2004) the role of the hippocampus in episodic memory is time-limited, with these memories consolidated to the neocortex over time. Within this framework it is held that the neocortex contains generalised representations for spatial contextual (e.g. a beach, a market) and non-spatial (e.g. object) memories, and therefore supports remote memory independently of any contribution from the hippocampus/MTL (Squire et al. 2004). According to this view then, successful imagination of experiences would be expected to occur in the presence of hippocampal lesions, by the coordination of activity in multiple neocortical areas, perhaps mediated by the goal-directed temporary or online binding capacities of the prefrontal cortex in working memory (Fuster 1997; Moses & Ryan 2006; Prabhakaran et al. 2000). The demonstration that patients with hippocampal amnesia are impaired at generating new imagined experiences poses a challenge to the standard model.

It could be argued, however, that the creation of imagined new experiences relies upon retrieval of recent episodic memories, a process severely disrupted in hippocampal amnesics. Although this possibility cannot be entirely excluded, it is probably unlikely for

several reasons. Firstly, commonplace everyday scenarios were specifically selected to increase the dependence of constructions on generalised semantic memory representations formed from numerous prior experiences, and thus minimise any possible contribution of recent (or even remote) episodic memories. Therefore, an inability to use recent episodic memories to aid their constructions is likely to have caused at most, a mild impairment on the construction task, rather than the devastated performance observed. Secondly, we explicitly instructed participants not to recount an actual memory or any part of one. That our instructions were adhered to is evidenced by control subjects reporting a low dependence on episodic memories in order to produce rich constructions. Finally, the fact that one densely amnesic patient was unimpaired on the construction task is further evidence against intact episodic retrieval capabilities being a necessary prerequisite for imagining new experiences.

Whilst the present findings are inconsistent with the standard model (Squire et al. 2004) they accord well with suggestions that the hippocampus plays a critical role in imagining experiences through the provision of spatial context, in perpetuity, either through its ability to process spatial information (Burgess et al. 2001; Moscovitch et al. 2005; O'Keefe & Nadel 1978), or bind together disparate elements of the imagined scene (Eichenbaum 2004; Graham & Gaffan 2005; Hannula et al. 2006; Lee et al. 2005b; McClelland et al. 1995). Given the striking similarities between the process of imagining new experiences and reliving past memories (Conway & Pleydell-Pearce 2000; Greenberg & Rubin 2003; Schacter 1996) the present findings may also have implications for the status of remote episodic memory following hippocampal damage. It has been suggested that discrepancies between studies of remote episodic memory in hippocampal patients (Bayley et al. 2003) might be accounted for by differences in the quality or richness of the recollective experience, a feature which is not always captured by existing scoring systems (Kopelman et al. 1989; Levine et al. 2002; Moscovitch et al. 2005). Indeed, recent evidence suggests that the hippocampus may be critical for recollecting vivid, detailed episodic memories, regardless of their age (Addis et al. 2004; Gilboa et al. 2004; Moscovitch et al. 2005; Steinworth et al. 2006). The present results are consistent with this perspective and moreover suggest the critical attribute determining whether internally generated experiences, either real or imaginary, are hippocampal-dependent may be the extent to which they are vividly (re)experienced.

One caveat to the findings is that P01 was unimpaired on the construction task. While the other four patients, despite variations in age, education, IQ, and memory profiles, were found to be strikingly homogenous in their performances, P01 is clearly distinct. Although the reason for his spared performance is uncertain, it may relate to a degree of residual hippocampal function. Indeed, this patient retains the ability to acquire new semantic information (Aggleton et al. 2005). Moreover, in a separate fMRI study, P01's spared capacity for encoding new semantic information (retained even when tested a few months later) was associated with significant activation in residual hippocampal tissue (see Section 4.3.3). This suggests, therefore, that this patient may have had sufficient functional hippocampal tissue to allow him to perform successfully on the construction task.

## **4.5 Conclusions**

In this study, I demonstrate that patients with hippocampal amnesia are impaired on a novel task that shares many similarities with episodic memory, namely the construction task. As such I was able to provide new insights into the nature of essential neural mechanisms carried out by the hippocampus. The current task was optimised to examine spatial deficits within the more general context of rich experiencing. In the future, it will be interesting to examine if the lack of coherence observed extends to other non-spatial aspects of an imagined event. Furthermore, it will be important to clarify the precise relationship between construction and episodic memory, and establish whether the hippocampus plays a similar role in mediating both these functions.

# Chapter 5

*Experiment 2: Using imagination to understand the neural basis of episodic memory\**

\*This chapter is derived in part from: "Using imagination to understand the neural basis of episodic memory" Hassabis D, Kumaran D, Maguire EA (2007). *Journal of Neuroscience*. 27, 14365-74.

## **Precis**

In light of the results from Experiment 1, specifically that the lack of spatial coherence in the imagined scenes appeared to be driving the patient deficit, I defined a new component process, namely 'scene construction', to encapsulate these findings (Hassabis & Maguire 2007). The process of scene construction involves the generation, maintenance and visualisation of a complex spatial setting in which an event (real or imaginary) can be mentally experienced (Burgess et al. 2001; Hassabis & Maguire 2007). Although we have seen that the hippocampus plays a critical integrational role in scene construction (Chapter 4) it is unlikely to supporting this process on its own (also see Section 8.3 for further discussion). In the present fMRI study I sought to establish the wider neural network underpinning scene construction by testing healthy subjects on the imagination task whilst being scanned. In addition, by comparing the retrieval of imagined memories to that of real memories I hoped to elucidate the mechanisms by which the brain can distinguish reality from fiction and in so doing perhaps better understand what it is about episodic memory that makes it so special (Tulving 2002b).

## **5.1 Introduction**

As outlined in Chapter 1, a rich recollective experience is a defining characteristic of episodic memory recall, the memory for our everyday personal experiences (Tulving 2002b). Recollection of this type of memory is widely accepted to be a reconstructive process (see Section 1.3) (Bartlett 1932; Conway & Pleydell-Pearce 2000; Hassabis & Maguire 2007; Rubin et al. 2003; Schacter & Addis 2007; Schacter et al. 1998). Numerous fMRI studies investigating the neural basis of episodic memory recall (Cabeza & St Jacques 2007; Maguire 2001a; Svoboda et al. 2006), and more recently the closely related task of future thinking (Addis et al. 2007; Atance & O'Neill 2001; Szpunar et al. 2007), have revealed a consistent and distributed network of associated brain regions. This comprises dorsal and ventromedial prefrontal cortex (PFC), lateral prefrontal cortex, the hippocampus, parahippocampal gyrus, lateral temporal cortices, temporo-parietal junction, thalamus, retrosplenial cortex (RSC), posterior cingulate cortex (PCC), precuneus, and cerebellum (see Figure 3) (Cabeza & St Jacques 2007; Maguire 2001a; Maguire & Frith 2003; Svoboda et al. 2006). However, surprisingly little

is understood about the contributions individual areas make to the overall recollective experience.

In order to progress this issue further I employed a novel paradigm where healthy subjects, as well as recollecting past memories, had to imagine new fictitious experiences (see Chapter 4) during fMRI scanning. Episodic memory and imagining fictitious experiences share striking similarities in terms of the psychological processes engaged (Conway & Pleydell-Pearce 2000; Greenberg & Rubin 2003; Schacter & Addis 2007) including imagery (Rubin et al. 2003) and the retrieval of relevant semantic information (Wheeler et al. 1997). Moreover, both tasks involve the process of scene construction (Hassabis & Maguire 2007). Scene or event construction involves the mental generation and maintenance of a complex and coherent scene or event. This is achieved by the reactivation, retrieval and integration of relevant semantic, contextual and sensory components, stored in their modality specific cortical areas (Wheeler et al. 2000), the product of which has a coherent spatial context (see Chapter 4), and can then later be manipulated and visualised. Note that this kind of complex 'scene' imagery differs markedly from 'simple' imagery (Kosslyn et al. 2001) (e.g. for faces or single objects) in that it requires the binding of disparate (possibly multi-modal) elements of a scene into a coherent whole. As such scene construction likely recruits a set of additional processes underpinned by regions such as the hippocampus (see Chapter 4) not thought to be involved in simple imagery (Rosenbaum et al. 2004). Critically, imagining fictitious scenarios, in contrast to tasks that require thinking about plausible personal future events such as those used in several recent episodic memory studies (Addis et al. 2007; Szpunar et al. 2007), results in experiences that are not explicitly temporal in nature. Furthermore, a purely created imagined experience does not have the same reliance or effect on the imaginer's self-schema compared with a real episodic memory (Conway & Pleydell-Pearce 2000; Gallagher 2000) although both likely involve the adoption of an egocentric viewpoint (Burgess 2006) on the part of the imaginer. By using imagining fictitious experiences as a comparison task for episodic memory, it is possible to address some key questions not amenable to previous studies (Addis et al. 2007; Szpunar et al. 2007).

Specifically, in the present study subjects were asked to recall very recent episodic memories, retrieve fictitious experiences previously constructed in a pre-scan interview

one week before, and to construct new fictitious experiences during fMRI scanning. I reasoned that brain regions engaged in common during episodic memory retrieval and imagined experiences would likely support scene construction irrespective of whether experiences were real or fictitious (see Chapter 4). I further hypothesised that brain areas selectively active during episodic memory recall would mediate specific aspects of remembering one's own past experiences including those related to self-schema (Conway & Pleydell-Pearce 2000; Gallagher 2000) and mental time travel (Tulving 2002b). In this way, I hoped not only to characterise the neural circuitry underlying the (re)construction of complex scenes, but also to gain new insights into the component processes supported by different brain regions in the episodic memory network (Maguire 2001a; Svoboda et al. 2006).

## **5.2 Methods**

### **5.2.1 Participants**

Twenty one healthy, right-handed, native English speakers participated in the experiment (10 males, mean age = 24.8 years (SD = 3.8), age range 18-31). All subjects gave informed written consent to participation in accordance with the local research ethics committee.

### **5.2.2 Pre-scan interview**

Interview sessions were conducted with subjects approximately one week prior to scanning (mean=6.24 days, SD=1.45). Subjects sat facing the experimenter and engaged in four tasks (see Table 4 for conditions summary). They were told at the outset to try and remember as much as possible about the things they were going to do during the interview session as their memory would be tested in the subsequent scanning session.

**Table 4. Summary of conditions**

Abbreviation	Description
Main conditions – scenes	
RM	Recall of a recent real episodic memory elicited in the pre-scan interview
IS	Recall of an imagined fictitious scene previously constructed in the pre-scan interview
NS	Construction of a novel fictitious scene for the first time during scanning
Control conditions – objects	
RO	Recall of an acontextual object visually presented during the pre-scan interview
IO	Recall of an imagined acontextual object previously created in the pre-scan interview
NO	Construction of a novel acontextual object for the first time during scanning

First they were shown 10 pictures of everyday objects on a computer screen (e.g. a red fire extinguisher) (condition: 'RO' – real objects). Each object was shown centrally in isolation for 20 seconds set against a plain white background along with an accompanying line of text (in black along the bottom of the screen) describing its key features (e.g. 'a red fire extinguisher with a silver label and black nozzle'). Subjects were instructed to use the time to try and remember as many details about the object as possible and also to form a strong mental image of the object. Then short descriptions of 10 further everyday objects (e.g. 'a fancy gold-plated pen with a silver nib and the initials JT engraved along the casing') were read out by the experimenter (condition: 'IO' – imagined objects). Subjects closed their eyes and were given 20 seconds to form a vivid and detailed mental picture of the object. The importance of visualising the object either against a plain background or simply floating in mid-air, i.e. in the absence of any background context, was stressed and this was verified by the experimenter after each object was imagined.

Next, as in Experiment 1, subjects were asked to describe 10 fictitious scenes in response to short verbal cues read out by the experimenter (condition: 'IS' – imagined scene) outlining a range of commonplace scenarios (e.g. 'Imagine you are lying on a

sandy beach in a tropical bay. Describe what you can see, hear, smell and feel in as much detail as possible'). Commonplace, ordinary settings were chosen to minimize the difficulty level and to be as independent from a subject's innate creative ability as possible. The scenarios also purposely encompassed a wide variety of different subject matters from the manmade to the natural and the busy to the isolated to ensure there were no content biases. Thus the imagined scenarios could be considered as 'episodic-like' experiences in that they involved the salient experiencing of complex scenes that included people and actions, with the implicit constraint that any event imagined to be taking place necessarily had a relatively short duration and occurred in one context.

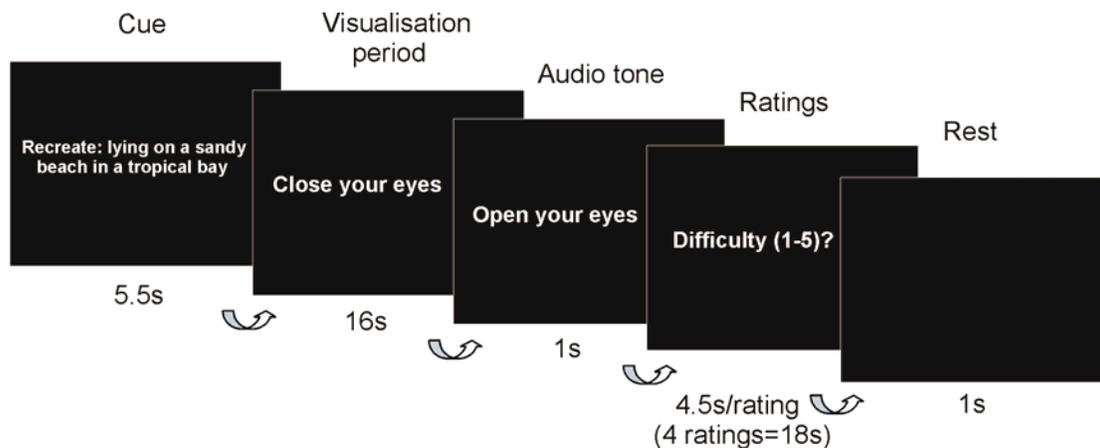
Subjects were asked to close their eyes and vividly imagine the scene and then given 2-3 minutes to describe it in as much (multi-modal) detail as possible. Importantly, subjects were explicitly told not to recount an actual memory or any part of one or something they planned to do but rather to create something completely new. After each scenario, participants were asked to rate their imagined scenes for vividness (1 – not vivid...5 – very vivid) and for similarity to a real memory (1 – nothing like any memory...5 – exactly like a memory). All descriptions were digitally recorded for later scoring purposes.

Finally, 10 recent real episodic memories were elicited from the subjects (condition: 'RM' – real memory). Memories had to be emotionally neutral, specific in time and place, and vividly recalled to be accepted as a stimulus. If a memory involved a long sequence of events (e.g. going to the cinema one evening) subjects were instructed to focus on a single shorter temporal element of that event and describe just it and its immediate location (e.g. buying the ticket at the ticket booth in the cinema lobby) so as to qualitatively match the imagined scene conditions as closely as possible. Subjects were encouraged to think of memories that were set in different distinct environments.

Overall it is important to note that because tasks were carried out in a single pre-scan session in one sitting and one context, all conditions were matched for effects related to the memory of the interview itself. In addition, only very recent autobiographical memories (mean age 12.82 days old (SD = 2.77)) were used as stimuli in order to be temporally comparable to the imagined memories.

### 5.2.3 Stimuli

Stimuli for the scanning session consisted of simple one-line text cues presented centrally on the screen in white on a plain black background (Figure 21). Each cue started with a keyword indicating the trial type followed by a very brief description of the scene or object to be visualised. Four keywords were used: 'Recall' indicated that the description that followed was either a real autobiographical memory (different for each subject) or an object that had been seen in the interview session (RM and RO conditions); 'Recreate' indicated trials in which an imagined scene or object previously created in the interview session was to be reconstructed (IS and IO conditions); 'Imagine' prefixed trials in which a new fictitious scene or object was to be imagined for the first time in the scanner ('NS', new scene, and NO, new object, conditions). 'Focus' was used to indicate a low imagery baseline task of where subjects had to imagine and then focus on a white crosshair on a black background.



**Figure 21. Experimental design**

A text cue, prefixed by an instruction keyword denoting trial type ('Recall', 'Recreate', 'Imagine'), was presented for 5.5s describing the scene or object to be visualised. Subjects were then instructed to close their eyes and begin visualising the scene or object in as much detail as possible for the entire 16s duration. A simple audio tone, played through headphones and lasting 1s, indicated the end of the visualisation period at which point the subject opened their eyes. Subjects then used a 5-button MR-compatible box to rate their just visualised scene or object across 4 ratings on five-point scales: difficulty, vividness, coherence and memory. Subjects were given 4.5s to respond per rating, resulting in an overall rating period of 18 seconds. This was followed by a 1s rest period where a blank screen was presented prior to the start of the next trial.

#### **5.2.4 Task**

Outside the scanner, the instructions and keywords were explained to subjects prior to scanning, and they received extensive training to ensure they were thoroughly familiarized with all aspects of the task. Subjects also had a practice session in the scanner prior to the main scanning sessions consisting of one trial per condition. There were 7 conditions (3 scene conditions: RM, IS, NS; 3 object conditions: RO, IO, NO; and the low imagery baseline condition) with 10 trials per condition, yielding 70 trials in total. As in the pre-scan interview session, subjects were explicitly told that for the scenes and objects they would newly imagine in the scanner (NS, NO) they should not give an actual memory or any part of one but rather to create something completely new.

Scanning consisted of four main sessions lasting around 11 minutes each during which 17 or 18 trials were pseudorandomly presented such that 2 or 3 trials were presented per condition, and never the same condition twice in a row. Text cues remained on the screen for 5.5s and were then replaced by a 'close your eyes' text instruction (see Figure 21). At this point subjects were instructed to close their eyes immediately and begin visualising the scene or object (the 'visualisation period') as vividly and, in the case of the 'memory' conditions (RM, RO, IS, IO), as accurately as possible.

Extensive piloting in the scanning environment revealed that subjects required ~16s to comfortably perform the visualisation tasks. Subjects were required to focus on the scene or object they were imagining, adding more details if necessary, for the entire 16s duration of the visualisation period. A simple 1s audio tone signalled the end of the visualisation period (at which point subjects opened their eyes) and the start of the ratings phase. Using an MR-compatible 5-button key pad subjects scored their just visualised scene or object across 4 ratings: difficulty (how hard was the trial: 1 – easy...5 – hard), vividness (salience of the imagery: 1 – not vivid...5 – very vivid), coherence (contiguosness of the spatial context: 1 – fragmented...5 – completely contiguous), and memory (how much like a memory the visualised scene or object was: 1 – nothing like a memory...5 – exactly like a memory). For each rating subjects were given 4.5s to respond. This was followed by a 1s period of rest before the next cue was presented. The scenes and objects used for the new imagination conditions (NS, NO) and those imagined in the pre-scan interview (IS, IO) were counter-balanced across subjects.

### **5.2.5 Debriefing and scoring**

After scanning, subjects were thoroughly debriefed. Subjects were first tested on their memory for all the seen (RO) and imagined objects (IO) from the pre-scan interview. The cues for the objects were read out and subjects provided as many details as they could remember. An object was regarded as accurately remembered if the subject was able to correctly name at least 2 salient features of the object correctly with no incorrect details.

Then two real memories (RM), two previously imagined scenes (IS), and two newly imagined scenes (NS) were randomly selected and subjects were asked to describe in detail what they had visualised in the scanner. These descriptions were digitally recorded for later transcription and scoring. The number of distinct details was calculated, and for the previously imagined scene condition (IS) the details were also compared to those described when that same scene was initially created in the pre-scan interview, thus providing an objective measure of how accurately the scene had been recreated in the scanner.

Subjects were then asked how emotional the three main scene conditions made them feel overall on a scale from -3 (negative) to +3 (positive). Finally, they were also asked to rate on a five-point scale how much they felt themselves to be an actual part of the visualised scenes ('sense of presence') (1 – not at all...5 – felt like I was really there).

### **5.2.6 Scanning parameters**

T2\*-weighted EPI images with BOLD contrast were acquired on a 1.5 tesla Siemens AG (Erlangen, Germany) Sonata MRI scanner. Standard scanning parameters to achieve whole-brain coverage were used: 45 oblique axial slices angled at 30 degrees in the anterior-posterior axis, 2mm thickness (1mm gap), repetition time of 4.05s. The first 6 'dummy' volumes from each session were discarded to allow for T1 equilibration effects. A T1-weighted structural MRI scan was acquired for each subject after the four main scanning sessions. Images were analysed in a standard manner using the statistical parametric mapping software SPM5 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). Spatial preprocessing consisted of realignment and normalisation to a standard EPI template in Montreal Neurological Institute (MNI) space with a resampled voxel size of 3x3x3mm, and

smoothing using a Gaussian kernel with full width at half maximum of 8mm (see Chapter 3).

### **5.2.7 Data analysis**

After preprocessing, statistical analysis was performed using the general linear model (see Section 3.5). The experiment had three main 'scene' conditions and three baseline 'object' conditions. The main interest was in the 16 second visualisation period when subjects were vividly recalling or imagining the scenes or objects. This period was modelled as a boxcar function (of 16 seconds duration) and convolved with the canonical haemodynamic response function to create regressors of interest. Subject-specific movement parameters were included as regressors of no interest. Subject-specific parameter estimates pertaining to each regressor (betas) were calculated for each voxel. These parameter estimates (collapsed across sessions) were entered into a second level random-effects analysis using a one-way ANOVA.

Results are reported in a priori regions of interest (previously identified in numerous neuroimaging studies of episodic memory (Maguire 2001a; Rugg et al. 2002; Svoboda et al. 2006) at  $p < 0.001$ , uncorrected for multiple comparisons, with an extent threshold of 5 or more contiguous voxels. Activations in other brain regions are reported for completeness at a threshold of  $p < 0.001$  uncorrected for multiple comparisons, but were only considered significant if they survive whole brain correction for multiple comparisons at  $p < 0.05$  (in line with established procedures (Frackowiak et al. 2004)).

A standard conjunction analysis (Friston et al. 2005), as implemented in SPM5 and reported at  $p < 0.001$ , was performed in order to reveal the regions activated in common between the main scene conditions as compared to their respective individual baseline object conditions. Two additional analyses were also carried out in which effects due to difficulty, vividness, coherence, and memory were either factored out by including these ratings as covariates of no interest in a second level ANOVA, or treated as covariates of interest. All activations are displayed on sections of the average structural image of all the participants. Reported voxels conform to MNI coordinate space. Right side of the brain is displayed on the right side.

## 5.3 Results

### 5.3.1 Behavioural data

#### 5.3.1.1 Ratings summary

During scanning, after each visualisation period (Figure 21) subjects were immediately asked to rate on five-point scales the just (re)constructed scene or object across four measures (see Section 5.2.4). Overall the conditions were generally very well-matched (see Table 5). All conditions were rated as low in difficulty (all conditions had a mean less than 1.9) and the images visualised as vivid (all conditions had a mean greater than 3.9). The scenes (re)created as part of the three main conditions (RM, IS, NS) were all rated to be coherent and contiguous (all had a mean greater than 3.9), conversely, the background contexts in the three object control conditions (RO, IO, NO) were rated as fragmentary (all object conditions had a mean coherence less than 1.8), indicating that instructions to minimise context and focus solely on visualising the required object had been adhered to. Finally, the four 'memory' conditions (both real and previously imagined: RM, RO, IS, IO) that required recall of a memory previously elicited/established in the prior interview session were well remembered (all had a mean greater than 3.6). The two 'new' conditions (NS, NO) did not rely heavily on explicit episodic memories (each had a mean memory rating less than 2.0), thus confirming that instructions not to use a memory or any part of one or something they planned to do but instead to create something new had been followed.

**Table 5. Behavioural ratings**

Conditions	Ratings (1 – low ... 5 – high): mean (SD)			
	Difficulty	Vividness	Coherence	Memory
Real recent memories (RM)	1.36(0.37)	4.41(0.44)	4.25(0.76)	4.56(0.50)
Previously imagined scenes (IS)	1.69(0.41)	4.07(0.41)	4.06(0.58)	3.97(0.68)
Newly imagined scenes (NS)	1.85(0.45)	3.91(0.64)	3.91(0.58)	1.94(0.33)
Previously seen objects (RO)	1.80(0.42)	4.04(0.41)	1.49(0.48)	3.89(0.61)
Previously imagined objects (IO)	1.84(0.46)	3.91(0.44)	1.49(0.54)	3.66(0.75)
Newly imagined objects (NO)	1.65(0.44)	4.09(0.43)	1.67(0.56)	1.74(0.54)

### 5.3.1.2 Formal ratings comparisons

As expected, all three main scene conditions were rated as more coherent than their respective object baselines (RM > RO,  $t(20) = 13.52$ ,  $p < 0.001$ ; IS > IO,  $t(20) = 13.60$ ,  $p < 0.001$ ; NS > NO,  $t(20) = 11.03$ ,  $p < 0.001$ ) thus confirming that objects had been visualised acontextually as instructed. RM was rated as easier, more vivid and more like a real memory than the pictures of objects seen in the interview session (RO) (RM > RO; difficulty:  $t(20) = -4.89$ ,  $p < 0.001$ ; vividness:  $t(20) = 4.36$ ,  $p < 0.001$ ; memory:  $t(20) = 4.41$ ,  $p < 0.001$ ). Also, imagining new scenes (NS) was rated as significantly harder than imagining new objects (NO) ( $t(20) = 2.88$ ,  $p < 0.01$ ). There were no other significant differences.

Direct comparison of the three main scene conditions with each other revealed no significant difference in difficulty, vividness or coherence between any of the conditions except for recalling real autobiographical memories (RM) which was rated as significantly easier and more vivid than the other two scene conditions (difficulty: RM > IS,  $t(20) = -4.77$ ,  $p < 0.001$ , RM > NS,  $t(20) = -6.67$ ,  $p < 0.001$ ; vividness: RM > IS,  $t(20)$

= 5.51,  $p < 0.001$ , RM > NS,  $t(20) = 4.96$ ,  $p < 0.001$ ). Real memories were also better remembered (i.e. had a higher memory rating) than imaginary memories (IS) ( $t(20) = 6.17$ ,  $p < 0.001$ ). Of course new imagined experiences (NS) relied on actual memories significantly less than the recall of real (RM) or imaginary memories (IS) (RM > NS,  $t(20) = 17.85$ ,  $p < 0.001$ ; IS > NS,  $t(20) = 11.18$ ,  $p < 0.001$ ) thus confirming that subjects had adhered to instructions to create something completely new and fictitious for the NS condition.

Direct comparison of the three object control conditions (RO, IO, NO) with each other revealed no significant difference in difficulty, vividness, coherence or memory except for recalling previously imagined objects (IO) which was rated as harder than imagining new objects (NO) ( $t(20) = 2.42$ ,  $p < 0.03$ ). Again, as with the scenes, new imagined objects (NO) relied on actual memories significantly less than the recall of real (RO) or imaginary objects (IO) (RO > NO,  $t(20) = 12.99$ ,  $p < 0.001$ ; IO > NO,  $t(20) = 8.67$ ,  $p < 0.001$ ) thus confirming that subjects had adhered to instructions to create new fictitious objects for the NO condition.

The pattern of performance observed with the real memory (RM) condition was expected, as retrieving a recent real salient personal memory and vividly reliving it is not a difficult task for a healthy subject. It should be noted that although RM condition was rated as easier, more vivid and better remembered, this was in the context of all other tasks being rated as easy, salient and well-remembered also.

At the interview session one week prior to scanning, subjects were asked to rate the vividness and the use of memory after each imagined scene they constructed. There was no significant difference in vividness between the scenes constructed in the interview (mean = 3.98; SD = 0.46) when compared to those same scenes later recalled in the scanner (IS) ( $t(20) = -0.89$ ,  $p = 0.39$ ) nor compared to new scenes freshly created in the scanner for the first time (NS) ( $t(20) = 0.51$ ,  $p = 0.62$ ). There was also no significant difference between the (low) use of episodic memory in imagining a new scene for the first time whether done in the interview session (mean = 1.75; SD = 0.31) or in the scanner (mean = 1.94; SD=0.33) ( $t(20) = -1.97$ ,  $p = 0.06$ ) again confirming that in both cases something new had been created.

Overly positive or negative real autobiographical memories were discarded at the interview stage, as were memories that the subject could not visualise saliently or pinpoint accurately in time. All real memories taken forward to scanning were therefore recent (mean = 12.82 days old; SD = 2.77), emotionally neutral, specific in time and vividly recollected. For the imagination conditions scenes and objects were counter-balanced across subjects for those that were presented during the pre-scan interview (IS, IO) and those that were presented for the first time in the scanner (NS, NO).

### *5.3.1.3 Debriefing*

After scanning, subjects were thoroughly debriefed. Subjects were asked how emotional the three main scene conditions made them feel overall on a scale from -3 (negative) to +3 (positive). As expected, the subjects rated all 3 conditions as being emotionally neutral (RM, mean = 0.24 (SD = 1.22); IS, mean = 0.62 (SD = 0.86); NS, mean = 0.38 (SD = 0.67), with no significant difference between any of the conditions (RM > IS,  $t(20) = -1.00$ ,  $p = 0.33$ ; RM > NS,  $t(20) = -0.47$ ,  $p = 0.64$ , IS > NS,  $t(20) = 1.75$ ,  $p = 0.10$ ).

Subjects were also asked to rate (scale of 1-5, where 1 is low) how much they felt themselves to be an actual part of the visualised scenes ('sense of presence'). As expected, real memories were associated with a stronger sense of presence than imagined scenes whether recalled or newly created in the scanner (RM, mean = 4.32 (SD = 0.66); IS, mean = 3.81 (SD = 0.73); NS, mean = 3.62 (SD = 0.79); RM > IS,  $t(20) = 3.86$ ,  $p < 0.001$ ; RM > NS,  $t(20) = 4.26$ ,  $p < 0.001$ ; IS > NS,  $t(20) = 1.09$ ,  $p = 0.29$ ).

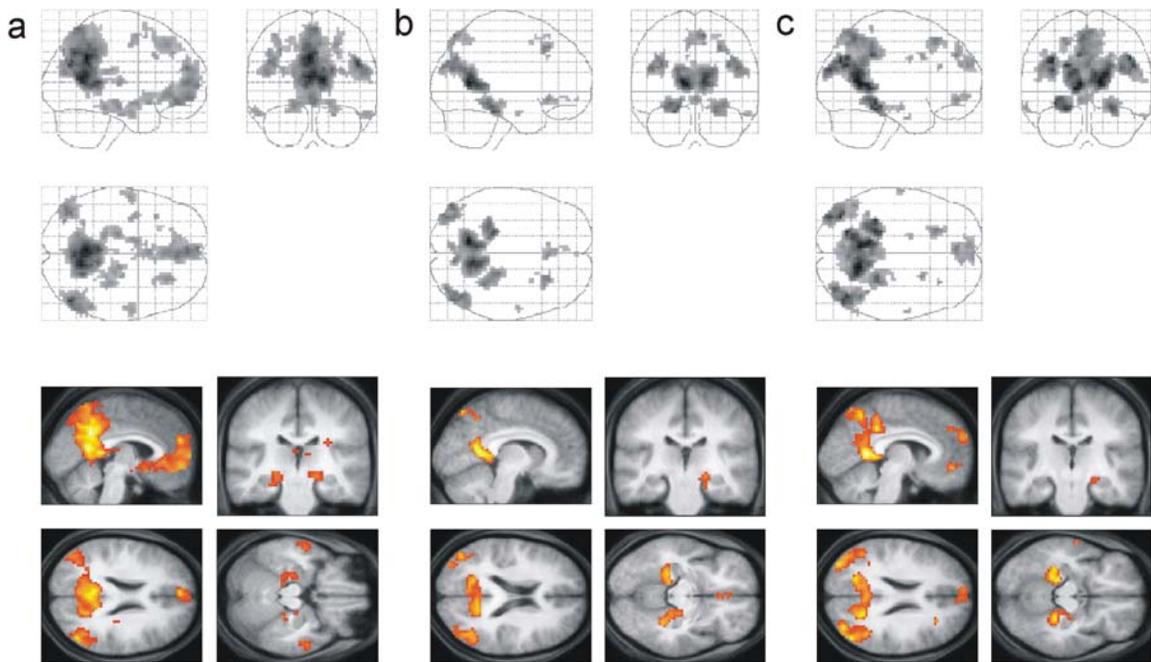
The scenes visualised in the three main conditions were well-matched for number of details, with no significant difference between any of the conditions (RM, mean = 7.12 (SD = 1.51); IS, mean = 7.36 (SD = 1.85); NS, mean = 7.29 (SD = 1.59); RM > IS,  $t(20) = -0.69$ ,  $p = 0.50$ ; RM > NS,  $t(20) = -0.59$ ,  $p = 0.56$ ; IS > NS,  $t(20) = 0.26$ ,  $p = 0.80$ ). Also, when describing details of the previously imagined scenes (IS), during debriefing, subjects mentioned an average of 89.7% of the same details they had produced when creating the scene originally in the pre-scan interview, thus indicating that they had recalled the previously imagined scenes (IS) with a high degree of accuracy and thereby confirming the behavioural rating (memory rating, mean = 3.97(0.68)) obtained in the scanner.

Memory for both the seen (RO) and imagined objects (IO) from the pre-scan interview was tested. All objects in the study had a number of salient features (e.g. colour, texture, adornments – see Section 5.2.2). When cued in the scanner only the basic name of the object was given. In the debriefing session an object was regarded as accurately remembered if the subject was able to name correctly and unprompted at least two salient features of the object. Objects were well remembered with a mean of 8.33 (SD=1.35) and 8.43 (SD=1.29) of the seen and imagined objects remembered respectively out of a possible 10, and no significant difference between the two conditions ( $t(20) = -0.30, p = 0.77$ ).

### **5.3.2 Neuroimaging Data**

#### *5.3.2.1 Overall brain networks*

In order to appreciate the overall network associated with each scene condition, I first contrasted each with its respective control condition. Contrasting the recall of recent episodic memories to the recall of acontextual objects previously seen in the interview session (RM > RO) replicated the well-established episodic memory retrieval network consistently activated in studies of autobiographical memory (Cabeza & St Jacques 2007; Maguire 2001a; Rugg et al. 2002; Svoboda et al. 2006) (see Table 6 and Figure 22a), consisting of the hippocampus bilaterally, parahippocampal gyrus (PHG), retrosplenial cortex, precuneus, posterior cingulate cortex (PCC), posterior parietal cortex, right thalamus, middle temporal cortex bilaterally, and medial prefrontal cortex.



**Figure 22. Main scene conditions versus their respective object baselines**

The top panels show sagittal, coronal and axial images from a 'glass brain', which enables one to appreciate activations in all locations and levels in the brain simultaneously. The bottom panels show activations on a selection of relevant sagittal, coronal and axial sections from the averaged structural MRI scan of the 21 subjects, at a threshold of  $p < 0.001$  uncorrected. **(a)** RM > RO: this contrast reveals the well-established network for episodic memory retrieval that includes bilateral hippocampi, parahippocampal gyrus, retrosplenial and posterior parietal cortices, right thalamus, middle temporal cortices and medial prefrontal cortex. Table 6 details the coordinates of all the activation peaks. **(b)** NS > NO: this contrast reveals a similar network for imagining new fictitious experiences that includes right hippocampus, parahippocampal gyrus, retrosplenial and posterior parietal cortices, and ventromedial prefrontal cortex. Table 7 details the coordinates of all the activation peaks. **(c)** IS > IO: this contrast also reveals a similar network for recalling imagined fictitious experiences previously created in a pre-scan interview that includes right hippocampus, parahippocampal gyrus, retrosplenial and posterior parietal cortices, and medial prefrontal cortex. Table 8 details the coordinates of all the activation peaks.

**Table 6. Recall of real episodic memories > recall of previously seen objects**

Region	Peak Coordinate (x, y, z)	Z
Medial prefrontal cortex	-3, 51, 24	5.00
Ventromedial prefrontal cortex	3, 42, -9	4.79
Right superior frontal sulcus	27, 36, 39	4.39
Left middle temporal cortex	-57, -9, -21	4.10
Right middle temporal cortex	57, -6, -24	3.96
Left hippocampus	-18, -21, -15	4.54
Right hippocampus	18, -24, -12	4.19
Right thalamus	15, -18, 9	4.30
Left parahippocampal gyrus	-18, -36, -15	4.28
Right parahippocampal gyrus	24, -33, -18	3.97
Left retrosplenial cortex	-3, -57, 15	6.04
Right retrosplenial cortex	12, -60, 12	6.11
Posterior cingulate cortex	-3, -54, 33	6.39
Precuneus	3, -66, 39	5.67
Left posterior parietal cortex	-48, -78, 18	4.46
Right posterior parietal cortex	48, -72, 27	5.71
Medial posterior parietal cortex	3, -63, 63	4.94

\*RM>RO

By contrasting the construction of imagined scenes to the imagination of acontextual objects, both newly imagined in the scanner (NS > NO), I identified brain regions involved in the process of creating and imagining vivid fictitious scenes (see Table 7 and Figure 22b) including the right hippocampus, PHG, retrosplenial and posterior parietal cortices. The involvement of the hippocampus provides further evidence that the process of scene construction involves the hippocampus, consistent with the findings from Experiment 1 (see Chapter 4).

**Table 7. Newly imagined fictitious experiences > newly imagined objects\***

Region	Peak Coordinate (x, y, z)	Z
Ventromedial prefrontal cortex	3, 24, -9	4.27
Right superior frontal sulcus	27, 27, 45	4.42
Right middle temporal cortex	57, -6, -24	3.70
Right hippocampus	21, -24, -12	3.86
Left parahippocampal gyrus	-18, -36, -15	4.28
Right parahippocampal gyrus	33, -42, -12	4.43
Left retrosplenial cortex	-12, -60, 9	6.08
Right retrosplenial cortex	12, -57, 15	5.52
Right precuneus	9, -57, 48	3.91
Left posterior parietal cortex	-48, -78, 24	4.75
Right posterior parietal cortex	45, -66, 24	4.75
Medial posterior parietal cortex	9, -75, 57	4.73

\*NS&gt;NO

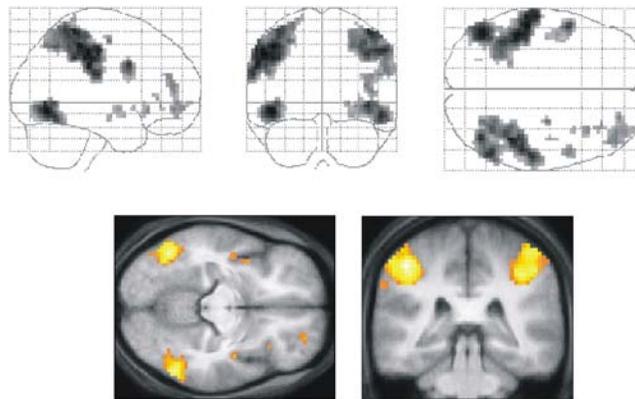
Contrasting the recall of imagined scenes to the recall of imagined acontextual objects, both previously constructed in the interview session prior to scanning (IS > IO), revealed a similar network as that engaged during the creation of new constructed scenes (i.e. NS > NO) (see Table 8, Figure 22c), but with additional activations in the medial superior frontal gyrus (BA9), and also precuneus and PCC.

**Table 8. Recall of previously imagined experiences > recall of previously imagined objects\***

Region	Peak Coordinate (x, y, z)	Z
Medial superior frontal gyrus	-3, 45, 36	4.53
Ventromedial prefrontal cortex	-6, 48, -9	4.35
Left superior frontal sulcus	-24, 24, 51	4.48
Left middle temporal cortex	-63, -12, -18	3.93
Right hippocampus	24, -21, -15	3.49
Left parahippocampal gyrus	-21, -39, -18	5.57
Right parahippocampal gyrus	27, -36, -18	5.17
Left retrosplenial cortex	-9, -48, 6	5.87
Right retrosplenial cortex	15, -57, 9	6.21
Left posterior cingulate cortex	-9, -42, 36	5.29
Precuneus	-3, -60, 54	4.88
Left posterior parietal cortex	-36, -75, 30	5.72
Right posterior parietal cortex	45, -69, 33	5.51
Medial posterior parietal cortex	3, -66, 60	4.71

\*IS&gt;IO

Although not the main focus, it is reported for completeness that the brain regions showing increased activity for objects compared with scenes ( $RO + IO + NO > RM + IS + NS$ ) (see Figure 23) included lateral occipital cortex (LOC) bilaterally (-45, -66, -9;  $z = 5.28$ ; 42, -69, -3;  $z = 4.90$ ), intra-parietal sulcus (IPS) bilaterally (-54, -30, 42;  $z = 5.24$ ; 51, -33, 45;  $z = 4.89$ ), and right lateral prefrontal cortex (39, 48, 15;  $z = 3.96$ ). This is strikingly consistent with other studies that have investigated objects as compared to places (e.g. (Sugiura et al. 2005)), and has been attributed to greater attention and focus being directed towards the object when present in isolation (Sugiura et al. 2005).



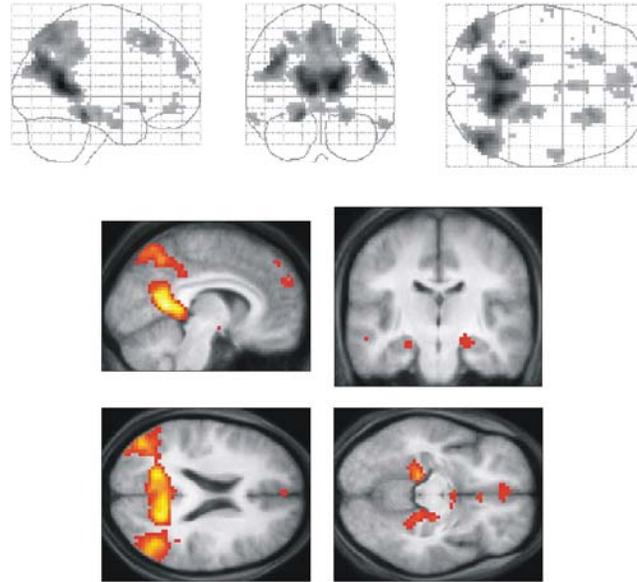
**Figure 23. Main effect of Objects > Scenes**

The top panels show sagittal, coronal and axial images from a 'glass brain', which enables one to appreciate activations in all locations and levels in the brain simultaneously. This contrast,  $[(RO+IO+NO) > (RM+IS+NS)]$ , reveals brain regions classically associated with supporting object representations and manipulation including the lateral occipital cortex bilaterally, intra-parietal sulcus bilaterally, and right lateral prefrontal cortex. Views of these brain regions are also shown in the lower panels on axial and coronal sections from the averaged structural MRI scan of the 21 subjects, at a threshold of  $p < 0.001$  uncorrected.

### 5.3.2.2 Commonalities: the core scene construction network

Having observed in the separate contrasts above similarities in the brain networks underpinning the three scene conditions, I next sought to formally examine commonalities between them. A conjunction analysis (Friston et al. 2005) of the main conditions (RM, IS, NS) against their respective controls (RO, IO, NO) allowed the common brain network involved in the underlying core process of constructing, maintaining and visualising complex scenes to be isolated. This network (see Table 9

and Figure 24) included bilateral hippocampi, parahippocampal gyrus, retrosplenial cortex, precuneus, posterior parietal cortex and ventromedial prefrontal cortex (vmPFC).



**Figure 24. Brain areas in common to the three scene conditions**

A conjunction analysis revealed the brain regions activated in common by the three scene conditions and therefore likely involved in 'scene construction', the primary process these three conditions have in common. This network included bilateral hippocampi, parahippocampal gyrus, retrosplenial and posterior parietal cortices, middle temporal cortices and medial prefrontal cortex. Table 9 details the coordinates of all the activation peaks. Views of this distributed brain network are also shown in the lower panels on a selection of relevant sagittal, coronal and axial sections from the averaged structural MRI scan of the 21 subjects, at a threshold of  $p < 0.001$  uncorrected.

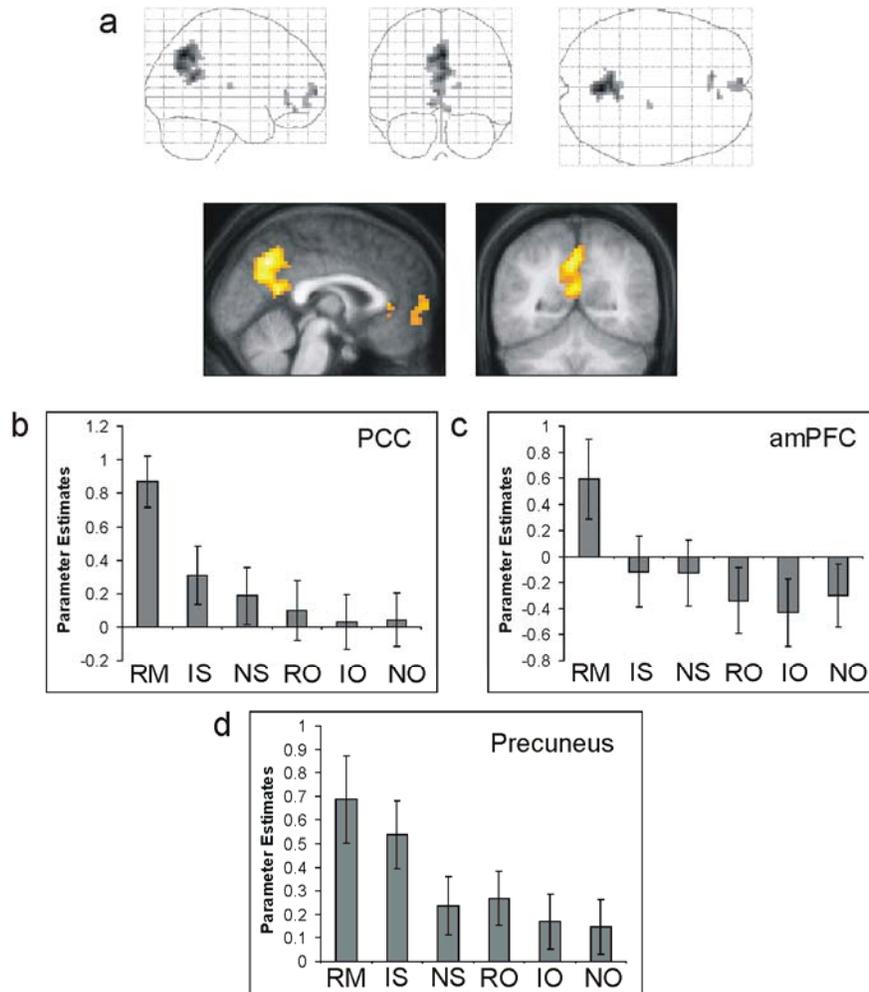
**Table 9. Conjunction analysis\***

Region	Peak Coordinate (x, y, z)	Z
Medial prefrontal cortex	-6, 54, 27	4.52
Ventromedial prefrontal cortex	-3, 48, -12	3.90
Left superior frontal sulcus	-24, 24, 51	5.37
Right superior frontal sulcus	24, 24, 48	5.54
Left middle temporal cortex	-63, -9, -18	4.29
Right middle temporal cortex	57, -3, -24	4.67
Left hippocampus	-21, -21, -21	3.90
Right hippocampus	24, -21, -15	5.04
Left parahippocampal gyrus	-18, -33, -15	7.52
Right parahippocampal gyrus	21, -33, -18	5.51
Left retrosplenial cortex	-6, -54, 9	>8
Right retrosplenial cortex	12, -54, 9	>8
Posterior cingulate cortex	-3, -39, 42	5.23
Left precuneus	-6, -63, 51	6.12
Right precuneus	6, -66, 60	5.71
Left posterior parietal cortex	-33, -78, 39	6.60
Right posterior parietal cortex	48, -72, 27	>8
Medial posterior parietal cortex	-6, -78, 54	4.94

\*The three contrasts entered into the conjunction analysis were: (RM-RO) (IS-IO) (NS-NO)

### 5.3.2.3 Differences: additional requirements for real episodic memories

Having established the brain areas the scene conditions share in common, I next examined differences between them, specifically which brain regions distinguished between the recall of real and fictitious events. This was addressed by contrasting the recall of real memories with the recall of previously imagined memories (minus their respective baseline conditions to control for any effects of external versus internally generated stimuli and creative processes:  $(RM - RO) > (IS - IO)$ ). The precuneus, PCC and anterior medial prefrontal (amPFC; BA 10) were preferentially engaged for real compared to imagined memories (see Table 10 and Figure 25a). The plots of the parameter estimates in these three areas (Figure 25b-d) clearly show that activation in these areas is most sensitive to the recall of real memories. No changes in activity were observed in the reverse contrast  $[(IS - IO) > (RM - RO)]$ .



**Figure 25. Comparison of real and imagined memories**

**(a)** Contrasting recall of real memories to the recall of imaginary memories whilst controlling for externally versus internally generated stimuli and creativity processes,  $(RM - RO) > (IS - IO)$ , revealed that the precuneus, PCC, and amPFC were preferentially engaged for real memories. Table 10 details the coordinates of all the activation peaks. Views of these brain regions are also shown in the lower panels on a sagittal and coronal section from the averaged structural MRI scan of the 21 subjects, at a threshold of  $p < 0.001$  uncorrected. **(b-d)** Condition specific parameter estimates (betas) in arbitrary units at the peak voxels. Bars represent the SE. RM = recall of real memories, IS = recall of previously imagined scenes, NS = constructing new fictitious scenes, RO = recall of previously seen objects, IO = recall of previously imagined objects, NO = imagining new objects. **(b)** At the peak voxel in PCC. **(c)** At the peak voxel in amPFC. **(d)** At the peak voxel in the precuneus.

**Table 10. Recall of real episodic memories > recall of previously imagined experiences\***

Region	Peak Coordinate (x, y, z)	Z
Anterior medial prefrontal cortex	-3, 63, 6	4.54
Posterior cingulate cortex	-3, -54, 36	4.61
Precuneus	3, -63, 39	5.21

\*(RM – RO) > (IS – IO)

I was also able to ask another question, namely which brain areas were differentially engaged by the retrieval, as compared to the encoding, of an (fictitious) episodic-like event. We contrasted the recall of previously imagined memories to the encoding of newly imagined events (IS > NS) which revealed increased activity in the precuneus bilaterally (-12, -63, 30;  $z = 4.21$ ; 9, -69, 36;  $z = 3.95$ ) for recall of previously imagined scenes. This effect is also apparent in the parameter estimates plot in Figure 25d. No changes in activity were observed in the reverse contrast (NS > IS).

#### 5.3.2.4 Additional analyses

Even though the ratings for the conditions were overall quite similar in terms of difficulty, vividness, coherence, and memory, I nevertheless included the ratings as covariates of no interest in a second level ANOVA in a separate analysis (see Section 5.2.7). However, this made no qualitative difference to the results, which was expected, due to the high degree of similarity between the ratings across conditions. In another analysis I treated the ratings as covariates of interest, asking whether activity in any brain areas correlated with the ratings. There were no significant findings, again most likely because the tasks were generally well matched.

## 5.4 Discussion

In this fMRI study I reveal the extended brain network, including the hippocampus, associated with scene construction (Hassabis & Maguire 2007), a key process in common to real and imaginary memories, and crucial to the recollective experience of recalling an episodic memory (Tulving 2002b). Moreover, by utilising rich imaginary memories as a well-matched comparison task we were able to isolate processes specific to episodic memory, such as the use of self-schema (Conway & Pleydell-Pearce 2000;

Gallagher 2000), a sense of familiarity and mental time travel (Tulving 2002b). Previous studies that have compared episodic memory and future thinking were unable to make this distinction because these features were present to a similar degree in both autobiographical memories and personal future scenarios (Addis et al. 2007; Atance & O'Neill 2001; Szpunar et al. 2007). Taken together, therefore, the present study advances efforts to dissociate the functions of the many brain regions consistently activated in studies of episodic memory recall (Cabeza & St Jacques 2007; Maguire 2001a; Svoboda et al. 2006) and episodic future thinking (Addis et al. 2007; Atance & O'Neill 2001; Szpunar et al. 2007).

Our results show that when subjects construct new fictitious scenes an extended brain network involving the hippocampus, parahippocampal gyrus, retrosplenial cortices, posterior parietal cortices, and ventromedial prefrontal cortex is active. Critically, I show using a conjunction analysis that this network is not specific to the construction of new fictitious experiences, but is also engaged when subjects remember both previously imagined experiences as well as their own real personal experiences. The implication is, therefore, that this distributed network supports cognitive operations engaged in common during the three main conditions in our experiment, most likely the (re)construction, maintenance and visualisation of complex scenes. Importantly, these effects cannot be attributed to differences in either difficulty or vividness between the conditions given that these variables were generally comparable.

A key aspect of the experimental design was the use of control conditions involving the vivid visualisation of acontextual objects which allowed the processes of “simple” imagery (Kosslyn et al. 2001) (e.g. of objects) to be distinguished from that of complex scene construction and visualisation, and attribute the latter to a distributed network including the hippocampus (see Chapter 4). Indeed, the finding of significantly greater activation in the object conditions (i.e. Objects > Scenes) in brain areas often associated with supporting object representations and manipulations (Sugiura et al. 2005) (i.e. LOC and IPS), lends further support to our assertion that scene construction represents a dissociable cognitive process with a distinct neural basis. As such, a large portion of the episodic memory recall network (Cabeza & St Jacques 2007; Maguire 2001a; Svoboda et al. 2006) would seem to be more accurately characterised as the network responsible for the construction of complex scenes. Given this, it would seem reasonable to propose

that real memories are reconstructed along very similar lines to the way imagined events are constructed, dovetailing with the idea that memory is (re)constructive in nature (Bartlett 1932; Conway & Pleydell-Pearce 2000; Hassabis & Maguire 2007; Rubin et al. 2003; Schacter & Addis 2007; Schacter et al. 1998).

Another aim of this study was the investigation of areas differentially engaged by episodic memories and (re)construction of imagined fictitious experiences. When constructing fictitious experiences subjects were instructed not to describe a memory or something they planned to do but instead create something completely new. That novel and original experiences were constructed is evidenced by the low scores obtained on the 'use of memory' behavioural measure. Furthermore, it is unlikely that subjects engaged in mental time travel when describing their fictitious scenes as comparisons of the fictitious scene descriptions obtained in this study (see Table 11) to typical descriptions obtained in future thinking studies (see Appendix A.2 (Addis et al. 2007)) show clear qualitative differences.

***Table 11. Example description of a fictitious scene***

---

Cue: Imagine standing by a small stream somewhere deep in a forest

"It's a pine forest. What I can see on the ground all around me are patches of pine needles and brown earth with nothing really growing. The tree trunks are quite narrow. Overhead are the spikes of the green pines and you can only just see the sky. There's a pine needle smell but down towards the stream there's a slightly rotting smell. It's quite a narrow stream with stones in it and dark water rushing round them causing little white water eddies. There's not much life around the stream and the banks are quite steep sloping down to the stream. It's peaceful and quiet..."

---

This is a representative example of the kinds of descriptions subjects gave in response to the scene construction cues. Note that no temporal statements, such as 'I will be', 'I plan to' or 'I'm going to', were used as compared to future event descriptions where they are commonplace (see Appendix A.2 in Addis et al., 2007).

In future event descriptions temporal phrases such as 'I will be', 'I plan to' and 'I'm going to', and familiar objects and people (e.g. 'my sister is there') are commonplace (Addis et al. 2007) in contrast to imagined fictitious scenes where generally none of these features are present. I therefore sought to capitalise on the absence of self-projection in time and reduced reliance on self-schema processes in the imagination task to identify brain regions that might support these specific qualities of episodic memory, both past and future (Schacter & Addis 2007). Contrasting the retrieval of recent real episodic

memories to the retrieval of recently constructed fictitious experiences, whilst controlling for external versus internal generation effects with the respective baselines, revealed increased activity in the amPFC, PCC and precuneus. It should be noted that this differential activity cannot be due to differences in retrieval effort, as recalling real memories was rated as easier than the other conditions. It is also unlikely to be explained by vividness; although real memories were rated as more vivid, this is set against an overall context of high vividness across all conditions. This was confirmed by the covariate analysis where covarying out the effect of vividness made no qualitative difference to the regions activated in the contrasts.

Interestingly, significantly greater activation was observed in precuneus when the recall of previously imagined experiences was compared to the creation of new imagined experiences. As such, activation in the precuneus may reflect the relative familiarity of the visualised experience (i.e. RM>IS>NS – see Figure 25d) with real memories, perhaps not surprisingly given their often highly familiar content, more familiar than imaginary memories, in line with proposals concerning the function of this region derived from studies of recognition memory (Hornberger et al. 2006; Rugg et al. 2002; Vincent et al. 2006; Wagner et al. 2005).

Activation in amPFC and PCC, on the other hand, was only observed during episodic memory recall, suggesting that these regions support functions that are specific to episodic memory over and above scene construction (Tulving 2002b). In fact the pattern of activation revealed by this contrast bears a striking resemblance to networks found to support self-reflection (Johnson et al. 2002), theory of mind (Amodio & Frith 2006; Kumaran & Maguire 2005) and episodic future thinking (Addis et al. 2007) suggesting that the amPFC and PCC might be supporting processes related to the self (Conway & Pleydell-Pearce 2000) and mental time travel (Tulving 2002b; Wheeler et al. 1997). Indeed, when viewed in terms of component processes, episodic future thinking (Atance & O'Neill 2001) and recalling past events share many of the same underlying processes. These include not only scene construction (which they have in common with imagining fictitious experiences) but also self-schema processing, self-projection in time (forwards instead of backwards in the case of future thinking), and a sense of familiarity – processes present in imagined fictitious experiences either to a much lesser extent or not at all.

Perhaps it is not altogether surprising therefore that recent fMRI studies have found comprehensive overlap in the activation patterns associated with recalling the past and thinking about plausible self-relevant future events (Addis et al. 2007; Szpunar et al. 2007). I suggest, therefore, that during episodic memory retrieval (and perhaps also episodic future thinking), the interaction or cooperation between the self-processing and familiarity functions performed by the amPFC/PCC and precuneus respectively may be sufficient to distinguish between real and fictitious memories and likely reflects the greater sense of presence reported by our subjects for real memories.

## **5.5 Conclusions**

To conclude, I have demonstrated that a distributed brain network, including the hippocampus, is recruited during both episodic memory recall and the visualisation of fictitious experiences. The present data provides evidence for the role of this brain network in scene construction, a critical process underpinning rich recollection, episodic future thinking and imagination. These findings also provide insights into the neural basis of processes specific to episodic memory, suggesting a model in which activation in amPFC, PCC and precuneus interact to allow real memories to be distinguished from imaginary ones. Moreover, the notion that scene construction is a key component of episodic memory recall lends further support to constructive theories of episodic memory (Bartlett 1932; Schacter et al. 1998) and the recently proposed constructive episodic simulation function of the episodic memory system (Schacter & Addis 2007; Schacter et al. 2007).

This study demonstrates that imagining new experiences is a manipulable and useful experimental tool which may prove fruitful in further advancing our understanding of episodic memory and the processes underpinning it. In the future, it will be important to define the exact function of individual regions within the scene construction network, explain the role of these functions in apparently unrelated tasks such as word-pair recognition tasks that also often involve this network, and discover the precise mechanisms which give rise to the phenomenological feeling one experiences during episodic memory recall that allows us to know that an event really happened.

# Chapter 6

*Experiment 3: Decoding neuronal ensembles in the human hippocampus\**

\*This chapter is derived in part from: "Decoding neuronal ensembles in the human hippocampus" Hassabis D, Chu C, Rees G, Weiskopf N, Molyneux PD, Maguire EA (In press). Current Biology.

## **Precis**

Experiments 1 and 2 showed that the process of scene construction is a key component process that underpins both episodic memory and imagination. The hippocampus plays an important role in scene construction by binding the disparate elements of a scene together into a coherent spatial context (see Chapter 4). But is it just about space or is this dynamic integrational role for the hippocampus (see Section 8.9) important for the binding of non-spatial elements? This question has its origins in the long-standing theoretical debate between the Cognitive Map Theory (see Section 2.5) and the Relational Theory (see Section 2.6) as to the special status or otherwise of space. In order to address this question in healthy humans it would be useful to devise a way of looking at fine-grained hippocampal representations non-invasively with fMRI. In this study I attempt to achieve this by combining the use of high spatial resolution fMRI with sensitive multivariate pattern analysis techniques (see Section 3.6). In so doing I illuminate some of the key properties of the population codes underlying spatial representations in the human hippocampus.

## **6.1 Introduction**

Information about the environment is thought to be encoded in the brain by activity in large populations of neurons (Buzsaki 2004; Harris 2005; Hebb 1949; Pouget et al. 2000). In order to understand the properties and dynamics of population codes, it is necessary to specify how they can be decoded in order to extract the precise information they represent (Harris 2005). This enterprise is at the heart of neuroscience, and provides a substantial challenge (Buzsaki 2004). Decoding the activity of single, or small numbers of, neurons has been highly successful, the best characterised example being the memory-related response of hippocampal place cells that fire invariantly when an animal is at a particular spatial location (Ekstrom et al. 2003; Moser et al. 2008; O'Keefe & Nadel 1978). It is not clear, however, what information such place cells represent at the population level, as recording *in vivo* from thousands of hippocampal neurons simultaneously is not currently possible (Buzsaki 2004; Dombek et al. 2007; Lin et al. 2005; Wilson & McNaughton 1993). Other techniques such as immediate early gene imaging have provided some insights into memory representations at the population level (Guzowski et al. 1999; Redish et al. 2001), but have limited temporal resolution (in

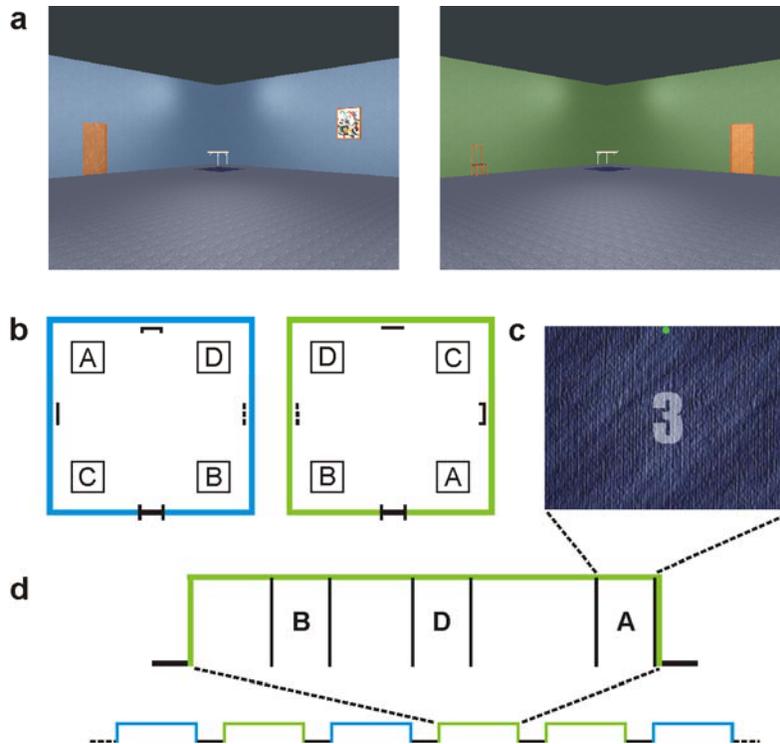
the order of minutes) and do not provide an *in vivo* measure, making it difficult to isolate with precision the specific feature of a stimulus, memory or behaviour associated with gene expression.

Recently, invasive approaches to examining how neurons encode information (Ekstrom et al. 2003; Kreiman et al. 2000; Quiroga et al. 2005) have been complemented by multivariate pattern analyses of non-invasive human functional MRI (fMRI) data (Haynes & Rees 2006; Norman et al. 2006). Functional MRI measures signals that are indirectly correlated with neuronal activity simultaneously in many individual voxels. Each voxel, depending on its size and location, contains thousands of neurons. Conventional univariate fMRI analysis methods focus on activity in each individual voxel in isolation. In contrast, multivariate pattern analyses harvest information from local patterns of activity expressed across multiple voxels and hence large neuronal populations (see Section 3.6). Not only can such novel analyses imply the presence of neuronal representations previously thought below the spatial resolution of fMRI (see Section 3.6.3) (Haynes & Rees 2005a; Kamitani & Tong 2005), but the ensemble activity of such distributed patterns can predict the perceptual state or intention of an individual with high accuracy (Haynes et al. 2007). However, to date there has been only limited application of this approach to memory (Polyn et al. 2005) and none that has focused specifically on decoding activity in the hippocampus, despite its critical mnemonic role (Andersen et al. 2007). This is perhaps not surprising as making discriminations on the basis of activity in the hippocampus and surrounding medial temporal lobe (MTL) regions only, presents a far more challenging classification problem than simply using whole brain information in a category-based design that results in large activity differences across multiple brain regions (Polyn et al. 2005).

However, successful decoding from focal hippocampal fMRI signals would have significant implications for understanding how information is represented within neuronal populations in the human hippocampus and for appreciating fundamental properties of the hippocampal population code. The current consensus from invasive animal studies (Guzowski et al. 1999; Redish et al. 2001) and computational models (Hartley et al. 2000; Samsonovich & McNaughton 1997) is that this population code is random and uniformly distributed, casting doubt on some earlier studies that suggested a potential functional structure in the hippocampus (Eichenbaum et al. 1989; Hampson et al. 1999).

However, if there is a functional organisation to the hippocampal population code then activity at the voxel level should also be non-uniform making classification possible using multivariate methods applied to human fMRI data (Haynes & Rees 2006; Norman et al. 2006).

I set out to test this hypothesis by combining fMRI at high spatial resolution with multivariate pattern analysis techniques (Haynes & Rees 2006; Kriegeskorte et al. 2006; Norman et al. 2006) to investigate if it was possible to accurately predict the precise position of an individual within an environment from patterns of activity across hippocampal voxels alone (see Section 3.6). I used an interactive virtual reality (VR) spatial navigation task (Figure 26) given that spatial navigation critically relies on the hippocampus (Andersen et al. 2007; O'Keefe & Nadel 1978). Importantly, by holding visual inputs and task constant after successful navigation to a position within the VR environment, it made it possible to isolate and characterise the 'abstract' (i.e. independent of current sensory inputs) internal representation of the environment's layout. Using this approach I show that non-invasive in vivo measurements of activity across the population of neurons in the human hippocampus can be used to precisely decode and accurately predict the position of an individual within their environment.



**Figure 26. The Experimental Task**

**(a)** The virtual reality environment comprised two separate and distinct environments, a blue room and a green room. Each room was 15mx15m and contained four ‘target’ positions which participants were instructed to navigate between as quickly and as accurately as possible following extensive pre-training. **(b)** Schematic of the room layouts with the four target positions, labelled A, B, C, and D. These targets were visually delineated by identical cloth rugs (i.e. not by letters which are depicted here only for ease of reference), placed on the floor at those positions and each 1.5mx1.5m. Single objects (door, chair, picture, and clock with different exemplars per room but of similar size and colour) were placed along the centre of each wall to act as orientation cues. Identical small tables were placed in all four corners of the rooms to help visually delineate the wall boundaries. Single trials involved participants being instructed to navigate to a given target position using a keypad. The trial order was designed to ensure that the number of times a target position was visited starting from another target position was matched across positions to control for goal and head direction. Once the intended destination was reached the participant pressed a trigger button causing the viewpoint to smoothly transition to look vertically downwards at the floor (as if bowing one’s head) to reveal the rug on the floor marking the target position – shown in **(c)**. At this point a 5-second countdown was given, denoted by numerals displayed in white text overlaid on the rug (the number ‘3’ is shown here as an example), and followed by the text label of the next target position (i.e. ‘A’, ‘B’, ‘C’ or ‘D’). The viewpoint then smoothly transitioned back to the horizontal, and navigation control was returned to the participant. **(d)** Environment blocks in each room consisted of 2-4 navigation trials and were counter-balanced across participants.

## 6.2 Methods

### 6.2.1 Participants

Four healthy right-handed males with prior experience of playing first-person videogames participated in the experiment (mean age 24.3 years, SD 3.2, age range 21-27). A case-plus-replication approach, standard in the literature (e.g. (Friston et al. 1999)), was adopted in which each of our four participants was treated as an independent case for the purposes of analysis, with detailed single case analyses followed by replication of the findings in each independent test case. All had normal or corrected-to-normal vision. All participants gave informed written consent to participate in accordance with the local research ethics committee.

### 6.2.2 Task and stimuli

During scanning participants were required to navigate as quickly as possible between four arbitrary target locations in two different virtual reality environments (see Figure 26). The virtual reality environment was implemented with a modified version of the graphics engine used in the videogame Fable (<http://www.lionhead.com/fable/index.html>). The room interiors were designed in the architectural package Sketch-up (<http://sketchup.google.com>) and imported into the graphics engine. The code for the environment, controls and scanner pulse synchronisation was written in C++ with Microsoft Visual Studio (<http://msdn.microsoft.com/en-gb/vstudio/products/default.aspx>).

Participants controlled their movement through the environment with a 4-button MRI-compatible control pad. The buttons were configured to move forward, rotate left, rotate right and to signal that a target destination had been reached. Each room was 15mx15m and perspective was set at the height of an average person around 1.8m above ground. The four target positions (A, B, C, and D) were situated 3m in from the corners and visually delineated by identical cloth rugs. Each rug (and hence each target area) was 1.5mx1.5m. Identical small square tables were placed in each corner to aid visibility and were irrelevant as cues for the navigation task.

The two rooms were matched in terms of size, shape, luminosity, emotional salience, contents and floor colour. The rooms were designed so that spatial relationships

between neighbouring object categories as well as the target position labels were orthogonal for each room. Participants navigated through the rooms at a fast walking speed of 1.9m/s. It was important for movement to be at a realistic speed and under participant control as self-motion is thought to play an important part in the spatial updating process (Bird & Burgess 2008; Byrne et al. 2007). Hence the use of interactive virtual reality was highly suited for extraction of position information that was as ecologically valid as possible.

Once a target location was reached the viewpoint transitioned downwards so the identical floor texture occupied the entire field of view thus ensuring visual input was matched perfectly across positions. At this point a 5s countdown was given followed by the letter of the next location, displayed for 2s, during which time the participant was stationary and viewing the floor ('stationary phase'). The viewpoint then transitioned back to the horizontal and the participant navigated to the next location as quickly and accurately as possible. Navigation blocks consisting of 2-4 individual trials were interspersed with a 13s period of rest during which a fixation cross was presented on a plain black screen. The label of the next target position was then displayed for 2s before the participant was placed anew in one of the rooms with his back facing the closed door as if he had just entered the room.

The trial and room orders were pseudorandomised and fully counter-balanced across participants. Each environment (i.e. blue or green room) was visited 20 times during the scanning session giving 40 environment blocks in total. Within each room every target position was visited 14 times giving 112 trials in total. In order to maintain attention during the stationary countdown period, catch trials were included that involved an incidental visual task. The countdown numbers were displayed in white text, but occasionally one would flash red for 200ms. Participants were instructed to press the trigger button as quickly as possible upon spotting a red number. There were 8 catch trials spread throughout the scanning session, one at each target position, and always at the end of a block. The volumes acquired during these catch trials were excluded from the analyses.

### **6.2.3 Pre-scan training**

Prior to scanning participants were pre-exposed to the two environments during an extensive training session where they were instructed to familiarise themselves with the keypad controls and were introduced to the navigation task. They were given no prior knowledge of the mapping of labels to target positions but were encouraged to learn through trial-and-error at their own pace. Feedback was given in the form of a red dot displayed centrally at the top of the screen when the trigger button was pressed over an incorrect target position. No suggestion was given to the participants to use any particular strategy for learning the two room layouts and they were instead encouraged to utilise whatever method worked for them. The training program periodically switched between the blue and green rooms to allow both layouts to be learned (training order was counter-balanced across participants).

Before proceeding to the scanner, participants had to reach criterion performance in a behavioural test where they had to correctly complete 10 navigation trials in a row in each room to ensure the two layouts had been well-learned. It took approximately 30 minutes of training for participants to reach this level of performance. The training protocol was carried out in order to minimise any learning or novelty effects during the scanning session and to allow time for the neural representations for each environment to form and stabilise (Wilson & McNaughton 1993).

### **6.2.4 Post-scan debriefing**

After scanning participants were debriefed and asked about the navigational strategies they used to learn the room layouts initially and to navigate between target positions during scanning. They were then asked to rate their emotional response to each room on a scale from -3 (really dislike) to +3 (really like) and also how difficult they found the task on a scale from 1 (easy) to 5 (hard). Finally, the participants were asked about any other observations they had pertaining to the layouts, the room representations, and their overall concentration during the task.

### **6.2.5 Image acquisition**

A 3T Magnetom Allegra head scanner (Siemens Medical Solutions, Erlangen, Germany) operated with the standard transmit-receive head coil was used to acquire functional data using a T2\*-weighted single-shot echo-planar imaging (EPI) sequence (in-plane resolution = 1.5x1.5 mm<sup>2</sup>; matrix = 128x128; field-of-view = 192x192 mm<sup>2</sup>; 35 slices acquired in an interleaved order; slice thickness = 1.5 mm with no gap between slices; echo time TE = 30 ms; asymmetric echo shifted forward by 26 phase-encoding (PE) lines; echo spacing = 560 μs; repetition time TR = 3.57 s; flip angle  $\alpha = 90^\circ$ ). All data were recorded in one single uninterrupted functional scanning session (total volumes acquired for each participant: s1 636 volumes; s2 640 volumes; s3 658 volumes; s4 670 volumes). An isotropic voxel size of 1.5 x 1.5 x 1.5 mm<sup>3</sup> was chosen for an optimal trade-off between BOLD sensitivity and spatial resolution. Further, the isotropic voxel dimension reduced resampling artifacts when applying motion correction.

In order to minimise repetition time whilst also optimising coverage of the regions of interest in the medial temporal lobe, partial functional volumes were captured at an angle of 5 degrees in the anterior-posterior axis (Deichmann et al. 2003)(see Figure 27b). Susceptibility-induced loss of BOLD sensitivity in the medial temporal lobe was intrinsically reduced by the high spatial resolution and adjusting the EPI parameters for the given slice tilt (z-shim gradient prepulse moment = 0 mT/m\*ms; positive PE polarity). A T1-weighted high-resolution whole-brain structural MRI scan was acquired for each participant after the main scanning session (1 mm isotropic resolution, 3D MDEFT).

### **6.2.6 Imaging data pre-processing**

This consisted of realignment to correct for motion effects and minimal spatial smoothing with a 3mm FWHM Gaussian kernel. The first six “dummy” volumes were discarded to allow for T1 equilibration effects (Frackowiak et al. 2004).

### **6.2.7 Univariate data analysis**

First, a standard univariate statistical analysis was performed with a general linear model implemented in SPM5 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). The paradigm consisted of 20 blocks of trials in each of the blue and the green rooms, pseudorandomly ordered, lasting an

average of 42.5s (SD 1.36s) resulting in the acquisition of approximately 12 volumes per block, and separated by rest periods of 15s (13s rest plus 2s next target indication – see above). Within an environment block the stationary period of each navigation trial was also separately modelled as a ‘mini-block’ consisting of the two volumes immediately following the start of the stationary period at the target position. There were 14 navigation trials for each of the four positions in the two rooms yielding a total of 112 position trials in total.

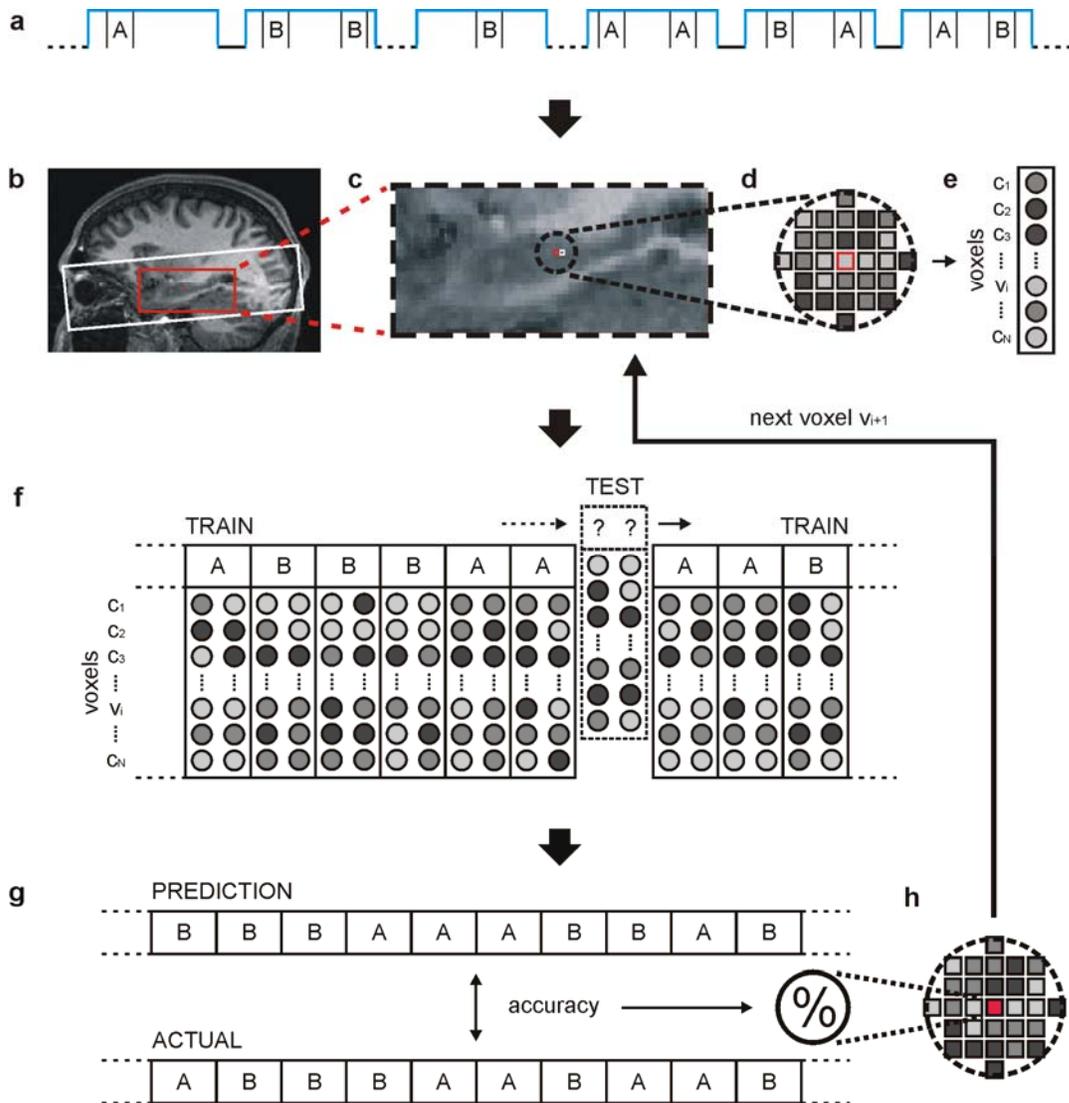
Each environment block and position mini-block was modelled as a boxcar function and convolved with the canonical hemodynamic response function to create regressors of interest and participant-specific movement parameters were included as regressors of no interest. Participant-specific parameter estimates pertaining to each regressor (betas) were then calculated for each voxel. Given the a priori interest in the MTL, a significance threshold of  $p < 0.001$ , uncorrected for multiple comparisons, was employed with an extent threshold of five or more contiguous voxels (Frackowiak et al. 2004). The data were also examined using a much more liberal threshold of  $p < 0.05$  uncorrected for multiple comparisons.

### **6.2.8 Multivariate pattern classification**

I then performed a multivariate pattern analysis (Haynes & Rees 2006; Norman et al. 2006) designed to identify brain regions where distributed fMRI activation patterns carried information about the environment a participant was in or his position in that environment. A linear detrend was run on the pre-processed images to remove any noise due to scanner drift or other possible background sources (LaConte et al. 2005). Next the image data was convolved with the canonical hemodynamic response function to increase the signal-to-noise ratio effectively acting as a low-pass filter (Frackowiak et al. 2004). BOLD signal has an inherent delay of around 6s to peak response relative to stimulus onset due to the hemodynamic response function (Frackowiak et al. 2004) and applying this convolution in effect delayed the peak by another 6s, giving a total delay of 12s. To best compensate for this delay all onset times were shifted forwards in time by three volumes yielding the best approximation to the 12s delay given a TR of 3.57s and rounding to the nearest volume (Haynes & Rees 2006; Norman et al. 2006).

The first volume and the last four volumes of each environmental block were discarded to allow for any orientation effects to settle (due to appearing suddenly in a room) and to exclude catch trials (always at the end of a block when present). Three separate multivariate classifications were carried out to (i) discriminate between which of two target positions in a single room the participant was standing ('pairwise'); (ii) discriminate between all four target positions in a single room ('4-way'); (iii) discriminate between which of the two room environments the participant was in ('environment'). The same technique, described next, was used in all three types of classification.

In order to search in an unbiased fashion for informative voxels and maximise sensitivity, I used a novel variant of the "searchlight" approach (Haynes et al. 2007; Kriegeskorte et al. 2006), a multivariate feature selection method (Kriegeskorte et al. 2006; Norman et al. 2006) that examines the information in the local spatial patterns surrounding each voxel  $v_i$  (Figure 27). This approach has another important advantage in that it results in statistical maps that allow for the anatomical mapping of the spatial pattern of informative voxels to be appreciated. Thus, for each  $v_i$ , I investigated whether its local environment contained information that would allow accurate decoding of the current position. For a given voxel  $v_i$ , I first defined a small spherical clique of  $N$  voxels  $c_{1,N}$  with radius of three voxels centered on  $v_i$ . A radius of three voxels was reported to be the optimal size for a clique by Kriegeskorte et al. (2006) although this may be partially dependent on the resolution of the acquired images. For each voxel  $c_{1,N}$  in the fixed local clique, the voxel intensity from each image was extracted yielding an  $N$ -dimensional pattern vector for each image.



**Figure 27. Multivariate Pattern Analysis**

An example multivariate analysis of a pairwise position classification, in this case discriminating between position A and position B in the blue room (see Figure 26). **(a)** Only volumes acquired while the participant was standing at these two blue room positions were entered into the analysis. **(b)** Coverage for functional scanning is shown as a white bounding box. The search space for the searchlight algorithm (Kriegeskorte et al. 2006; Norman et al. 2006), anatomically defined to encompass the entire hippocampus and wider MTL bilaterally, is shown as a red bounding box. **(c)** The search space was stepped through voxel by voxel. For each voxel  $v_i$  (example  $v_i$  outlined in red) a spherical clique (radius 3 voxels) of  $N$  voxels  $c_{1,N}$  was extracted with voxel  $v_i$  at its centre **(d)** to produce an  $N$ -dimensional pattern vector for each volume **(e)**. **(f)** Each pattern vector was labelled according to the corresponding experimental condition (position A versus position B) and then partitioned into a training set (solid lines) and an independent test set (dashed line and indented). Patterns of activity across the voxel

clique from the training set were used to train a linear SVM classifier which was then used to make predictions about the labels of the test set. A standard k-fold cross-validation testing regime was implemented ensuring that all pattern vectors were used once as the test data set. **(g)** This cross-validation step therefore yielded a predicted label for every pattern vector in the analysis that was then compared to the real labels to produce an overall prediction accuracy for that voxel clique. This accuracy value was stored with the voxel  $v_i$  for later thresholding and reprojection back into structural image space **(h)**. The whole procedure was then repeated for the next voxel  $v_{i+1}$  (outlined in white in **(c)**) along in the search space until all voxels in the search space had been considered.

Multivariate pattern recognition was then used to assess how much position and environment information was encoded in these local pattern vectors. This was achieved by splitting the image data (now in the form of pattern vectors) into two segments: a 'training' set used to train a linear support vector pattern classifier (with fixed regularisation hyperparameter  $C = 1$ ) to identify response patterns related to the two conditions being discriminated, and a 'test' set used to independently test the classification performance. The classification was performed with a support vector machine (SVM) (Duda et al. 2001) using the LIBSVM implementation (<http://www.csie.ntu.edu.tw/wcjlin/libsvm>).

A standard k-fold cross-validation testing regime was used (Duda et al. 2001; Haynes & Rees 2005a; Haynes & Rees 2006), where k equalled the number of blocks, with each block set aside in turn as the test data and the rest of the blocks used to train the classifier (see Figure 27f). This procedure was then repeated until all blocks had been assigned once as the test data (the cross-validation step). Thus the pairwise position classification involved a 28-fold cross-validation step (14 position X and 14 position Y mini-blocks of two volumes each - as the stationary phase lasted 7s and the scanning repetition time was 3.57s this consisted of the two volumes immediately following the onset of the stationary period); the 4-way position classification involved a 56-fold cross-validation step (14 mini-blocks of length two volumes, for each of 4 positions); and the environment classification involved a 40-fold cross-validation step (20 blue room and 20 green room blocks with an average of seven volumes per block). Every volume within a test block was individually classified following the cross-validation step, thus yielding an overall percentage accuracy for the clique centred around voxel  $v_i$  for all the volumes in the entire experimental session (see Figure 27g).

This decoding accuracy was stored with voxel  $v_i$  for subsequent reprojection as a 'prediction map' (see Section 6.2.10 below) and the entire procedure was repeated on a voxel-by-voxel basis until all voxels in a previously defined region of interest had been considered. In this case the search regions were anatomically defined with two large rectangular bounding boxes (each composed of 6750 voxels – see Figure 27b) covering both the right and left medial temporal lobes and thus encompassing our apriori regions of interest, i.e. the hippocampus and parahippocampal gyrus. Good overall classification accuracy for a voxel  $v_i$  implies that patterns in the surrounding local clique of voxels encode information about the current position/environment of the participant.

### **6.2.9 Multiclass classification**

Standard SVMs are binary classifiers that operate on two-class problems. However, they can be arbitrarily extended to work in cases where there are more than two classes (Allwein et al. 2001) as with the 4-way position classification performed here. Typically, this is done by reducing the single multiclass problem into multiple binary classification problems that can be solved separately and then recombined to provide the final class prediction (Allwein et al. 2001; Mourao-Miranda et al. 2006). For the 4-way position classification problem in the present study I used the well-established approach of Error Correcting Output Codes (Dietterich & Bakiri 1995; Mourao-Miranda et al. 2006) to assign a unique binary string of length  $n$  to each of the 4 classes (class "codewords" representing positions A, B, C, D respectively) where  $n$  corresponds to the number of binary classifiers performed. There are 6 possible pairwise comparisons that can be made between 4 positions, therefore all 6 pairwise binary classifications were performed and combined their outputs to produce codewords 6-bits in length (see Table 12 for the codeword matrix) with each bit representing the output from a single binary classifier.

**Table 12. 6-bit 'class codewords' for 4-way position classification**

Position	Binary Classifier Output					
	(A v B)	(A v C)	(A v D)	(B v C)	(B v D)	(C v D)
A	1	1	1	0	0	0
B	-1	0	0	1	1	0
C	0	-1	0	-1	0	1
D	0	0	-1	0	-1	-1

'1' or '-1' indicates output from that binary classifier is expected to be class 1 (i.e. class listed first) or class 2 respectively. '0' indicates output from that binary classifier is considered neutral for that label assignment.

These output strings were then compared against all 4 of the pre-assigned class codewords to determine the final predicted class. This was achieved by computing the Hamming distance (i.e. the number of bits which differ between two binary strings) (Hamming 1950) between the output string and the class codewords to find the closest class codeword and thus by association the predicted class. Apart from this extra multiclass procedure, the 4-way position classification was performed with the same methods as the other binary classifications.

### **6.2.10 Reprojection and thresholding**

Once the classifications were completed and decoding accuracies stored for each voxel in the search region, I proceeded to reproject these values back into structural brain image space to allow the resultant prediction maps to be visually inspected. These prediction maps were then thresholded at a percentage accuracy value that was significantly above that expected by chance. This significance threshold was determined using the classical method of nonparametric permutation testing (Fisher 1935; Nichols & Holmes 2002), requiring minimal assumptions (for example about the shape of the population distribution) for validity. The entire classification procedure outlined above was performed 100 times with a different random permutation of the training labels for each classification type for each participant. The individual voxel accuracy values from each of these 100 random runs were then concatenated into one population and the accuracy value at the 95th percentile of this aggregated distribution was calculated. This

procedure therefore yielded a percentage accuracy value for each individual participant above which a voxel's accuracy was considered significant, equating to a confidence level of  $p < 0.05$  uncorrected in a standard t-test.

The multiple comparisons problem was accounted for by performing a standard test for the difference between two population proportions (Daniel & Terrell 1995). If significant voxels were false positives due to random variation, the proportion of significant voxels should be uniform over the entire search space (see Figure 27). To test this null hypothesis, two anatomical masks, one covering the hippocampus bilaterally and the other the parahippocampal gyrus bilaterally, were created for each individual participant by hand with MRICro (<http://www.sph.sc.edu/comd/rorden/micro.html>), using each participant's structural MRI scan for guidance. The proportion of significant voxels for each region was determined (i.e. active voxels/total voxels) for each prediction map (i.e. pairwise position, 4-way position in blue room, 4-way position in green room, and environment) for each participant. A two-tailed test for difference between proportions was performed for each of the prediction maps to determine if the proportion of active voxels in the hippocampus was significantly different from that in the parahippocampal gyrus.

In addition to the voxel count difference of proportions test described above, I used a second analytic approach to test for a region x classification type interaction between the hippocampus and parahippocampal gyrus. The informational content of BOLD signals for each type of classification was assessed (i.e. pairwise position and context) for both regions. A classifier trained on A v B in the blue room (comprising primarily hippocampal voxels – see Figure 28) was tested on discriminating between the blue room and green room contexts. Conversely a classifier trained on blue versus green room context (comprising primarily parahippocampal voxels – see Figure 30) was tested on discriminating between positions A v B in the blue room. To enable comparison, a single accuracy value was generated for each classifier (see Section 6.2.11).

### **6.2.11 Additional multivariate control analyses**

In order for a classifier to successfully decode brain activity, the difference between two conditions must be systematic and consistent across the majority of the training

examples. For example this is not the case for the navigation periods between locations which were self-paced and thus exhibited considerable trial-by-trial variability in terms of the paths taken and the objects viewed en route and thus cannot have been a contributing factor to successful classification at the destination locations. In order to empirically confirm this hypothesis, a further binary classification analysis was performed this time only using the volumes acquired during the navigation periods prior to reaching a destination location. This analysis was carried out for all four participants in the blue room on the navigation periods prior to reaching position A versus the navigation periods prior to reaching position B.

Whilst the task was optimally designed to test the hypothesis that discrimination between within-environment locations reflects neural signals related to space, it is necessary to consider alternative, possibly confounding, factors. I performed a series of new analyses, following the same procedure in each case. The subset of hippocampal voxels that could discriminate between within-environment locations was taken (i.e. those from the pairwise position classification) and I now tested whether signals from these voxels could instead discriminate the alternative factors. If they could, then any conclusion that signals from these voxels encoded within-environment locations would be called into question.

Three new classification analyses were undertaken: (1) Destination labels: If a classifier learns to discriminate patterns of activity that represent the verbal label "A" in environment 1, does this classifier perform above chance in discriminating label "A" in environment 2? (2) Nearby objects: looking at Figure 26, the landmarks near locations B and A in the blue room resemble those of locations B and C in the green room. If a classifier learns to distinguish locations B and A in the blue room, does this classifier perform above chance in discriminating locations B and C in the green room? (3) Position relative to global 'room-centred' coordinates: if a classifier learns locations in the blue room can it discriminate corresponding locations in the green room? In this analysis 'locations' refer to spatial, room-centred coordinates, i.e. relative to the entry point (the door, which was the most-used orienting object).

For each of the three analyses, I took the subset of voxels from our original searchlight analyses (see Figure 28) that had accurately discriminated within-room position, and

now trained a new classifier on all the examples of the particular alternative factor selected for analysis in one of the rooms (e.g. A v B in the blue room). This new classifier was then tested on the discrimination between the corresponding confounding factor in the other room (e.g. A v B in the green room). Critically, this ensures that training and test data are independent and dissociates testing of these confounding factors (which are common to each room) from neural signals related to space. To enable comparison, a single accuracy value was generated for each participant and confounding factor (and for the original within-environment position discrimination using the same methodology).

For the destination label factor, a classifier trained on A v B in the blue room was tested on discriminating A v B in the green room. For the nearby objects factor, a classifier trained on A v B in the blue room was tested on discriminating B v C in the green room, as those pairs of positions had the same neighbouring object exemplars in both rooms. Finally, for the relative position factor using global room-centred coordinates as anchored by the door, a classifier trained on C v D in the green room was tested on discriminating A v D in the blue room, as those pairs of positions had the same relative position in each room as viewed from the door. For the two participants who had counter-balanced trial positions in the blue and the green rooms, these were swapped for the purposes of training and testing the new classifiers.

## **6.3 Results**

### **6.3.1 Behavioural data**

These two environments were designed to be austere to minimise the impact of extraneous sensory inputs. Apart from colour, which acted as a simple unambiguous retrieval cue for each room and is processed in extra-striate cortex (Gegenfurtner & Kiper 2003), the two environments were well-matched in terms of design and layout (see Figure 26). Within each room target positions and trial types were also well-matched. This was confirmed behaviourally with no significant difference between time spent in each of the rooms ( $p = 0.38$ ), or average times to navigate to each target position ( $p > 0.48$  for all pairwise comparisons; see Table 13).

**Table 13. Behavioural measures**

Condition	Time (s) for each participant: mean (SD)				
	s1	s2	s3	s4	Avg
<b>Blue Room</b>					
Overall*	20.36(13.20)	20.65(13.47)	23.08(14.42)	23.73(13.86)	21.96(1.70)
Navigate to A	6.37(1.93)	5.92(1.05)	6.71(2.32)	7.02(1.83)	6.51(0.47)
Navigate to B	5.43(1.23)	5.95(1.01)	6.97(4.36)	6.97(2.89)	6.33(0.77)
Navigate to C	6.30(1.49)	6.08(1.68)	5.83(1.19)	7.12(1.98)	6.33(0.56)
Navigate to D	5.48(1.00)	6.22(1.19)	8.05(5.83)	7.45(3.95)	6.80(1.16)
<b>Green Room</b>					
Overall*	20.59(13.31)	20.19(13.23)	21.01(14.23)	21.34(12.97)	21.03(0.93)
Navigate to A	6.15(2.49)	5.92(0.83)	6.25(1.72)	6.77(1.32)	6.27(0.36)
Navigate to B	5.71(0.96)	5.79(0.96)	6.73(2.89)	6.18(1.05)	6.10(0.47)
Navigate to C	5.34(1.63)	6.07(1.07)	5.55(1.21)	7.24(1.47)	6.05(0.85)
Navigate to D	6.95(2.87)	5.75(0.76)	6.13(1.08)	6.42(2.52)	6.31(0.51)

\*Overall times exclude the stationary phases

After the scan, a debriefing interview was conducted. Participants were encouraged to explain the strategies they used for learning the room layouts initially and how they performed the navigation task during scanning. All reported forming a 'mental map' of the each room's layout and only occasionally using one of the wall objects (three used the door, one used the clock) as an anchor point to maintain their bearings. All reported viewing their mental map from an overhead aerial viewpoint but two also reported being aware of the angle of the next target position from a first-person perspective during the navigation task. None reported using a verbal strategy or mentally verbalising the destination location labels during the task.

The task was generally found to be quite easy ( ratings from 1(easy) to 5 (hard), mean rating 2.0 (0.82) ), and both rooms were equivalent in emotional salience with slightly positive feelings reported for both ( -3 (really dislike) to +3 (really like); green room +0.75 (1.50), blue room +0.25 (0.96);  $p=0.59$  ). All participants noticed the different object configurations in each room, treated each room as a separate environment (not spatially

connected in any way), and were able to maintain concentration on the navigation task throughout.

### **6.3.2 Univariate neuroimaging data**

Prior to our main multivariate pattern analysis, a conventional univariate analysis (Frackowiak et al. 2004) was performed with a general linear model. Any activity differences associated with the two environments was characterised by contrasting the blocks of trials in the blue room with those in the green room and, as expected, no significant differences in activity were detected anywhere in the brain for any of the participants.

Similarly when contrasting between pairs of target position mini-blocks, no significant differences in activity were detected anywhere in the brain for any of the participants for any of the pairwise comparisons. This was the case even at an extremely liberal statistical threshold of  $p < 0.05$  uncorrected for multiple comparisons. These null univariate results were expected because conventional univariate analysis works by measuring the difference in average voxel activity between conditions (Frackowiak et al. 2004). With conditions almost identically matched in terms of stimuli and task, it is no surprise that this method did not reveal any significant differences in average univariate activity, hence the advantage of using a multivariate approach sensitive to patterns of information across groups of voxels (Haynes & Rees 2006).

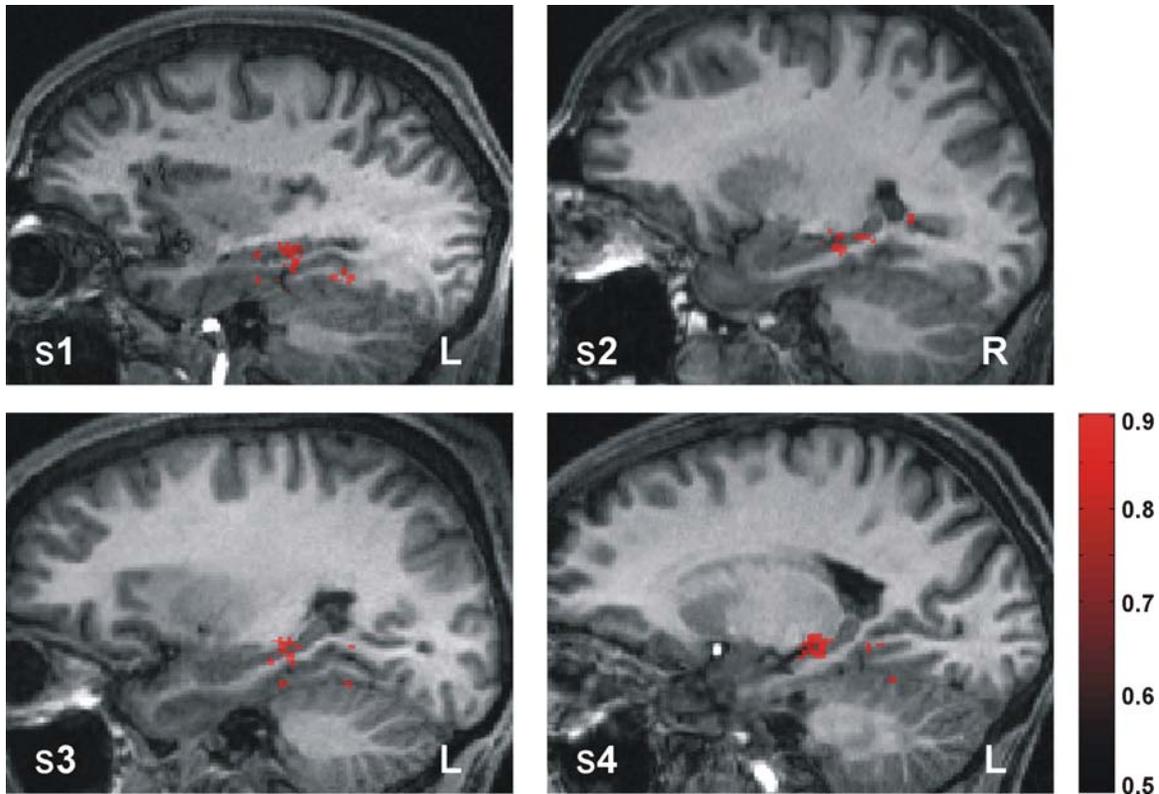
### **6.3.3 Discriminating between two positions**

I first investigated if it was possible to accurately predict where a participant was located within a room solely from the pattern of fMRI BOLD responses across multiple voxels in the hippocampus and MTL. To address this question, comparisons were made between arbitrarily selected pairs of positions (A v B and C v D) in both rooms. Importantly, after navigation when participants reached a target position, the default horizontal viewpoint transitioned smoothly downwards by 90 degrees so that the entire visual display was occupied solely by an identical view of the floor (see Figure 26c). Critically, only volumes capturing fMRI activity during this stationary phase (see Figure 26d) at the target positions when the participant was viewing the floor were entered into the analysis. This

is a key aspect of the study design, because visual stimuli such as objects and boundaries are known to be processed by the MTL (Bar 2004; Bird & Burgess 2008; Epstein & Kanwisher 1998; Janzen & van Turenout 2004; Kreiman et al. 2000; Quiroga et al. 2005).

By removing visual input as a confounding factor it was therefore possible to isolate the internal representation of spatial location as the only difference between conditions. Moreover, the task design minimised or controlled for any other potential confounding psychological factors during this period, as also confirmed in the debriefing (see Section 6.3.8). The imaging data were then divided into independent training and test sets (see Figure 27) with the former used to train a linear support vector machine (SVM) classifier. The performance of this classifier was evaluated by running it on the independent test data and obtaining a percentage prediction accuracy value.

Using a multivariate 'searchlight' approach to feature selection (Haynes et al. 2007; Kriegeskorte et al. 2006; Norman et al. 2006), we stepped through a large search space encompassing the MTL (see Figure 27) and identified spherical cliques of voxels whose spatial patterns of activity enabled the classifier to correctly discriminate between two positions significantly above chance ( $p < 0.05$  uncorrected, using the statistically conservative approach of nonparametric permutation testing and accounting for the multiple comparisons problem ((Fisher 1935; Nichols & Holmes 2002) - see Sections 6.2.10 and 6.3.6). Voxels at the centre of cliques whose accuracies survived this thresholding, and were therefore important for accurately distinguishing between the two experimental conditions (e.g. position A versus position B), were then reprojected back onto the structural brain image of the participant to produce 'prediction maps'. Remarkably, this process revealed large numbers of voxels in the body-posterior of the hippocampus bilaterally that accurately discriminated the position of the participant (Figure 28).



**Figure 28. Pairwise Position Classification.**

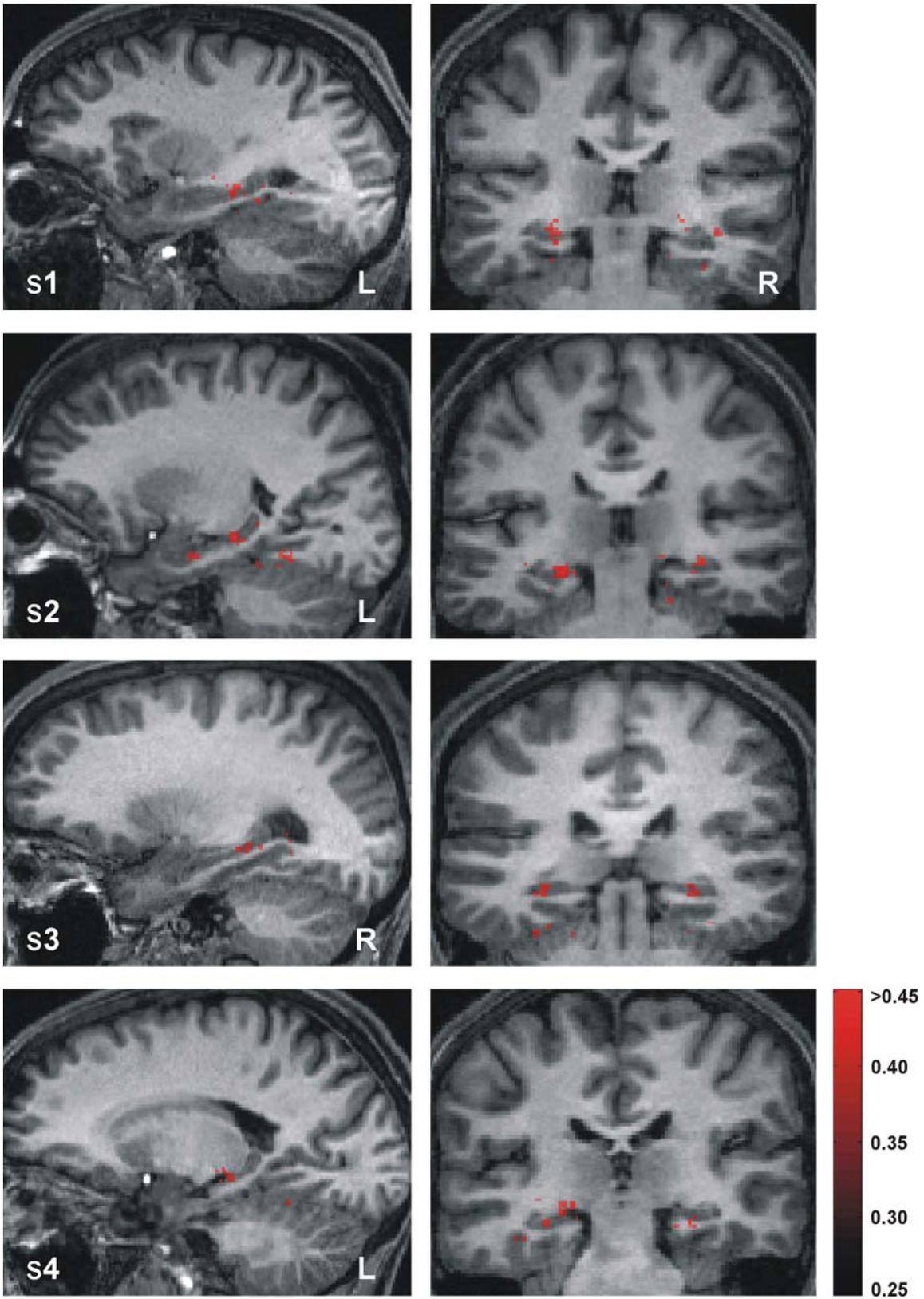
Prediction maps showing the accuracies of the voxels at the centre of searchlight cliques that discriminate between two arbitrarily chosen target positions in a room (apriori selected to be A v B and C v D) significantly above chance (50%). The resultant prediction map for a participant, bounded by the search space (indicated by the red box in Figure 27b), is projected onto their structural brain image. A sagittal section for each participant is displayed showing that voxels in the body-posterior of the hippocampus bilaterally are crucial for accurate position discrimination by the classifier. The findings are highly consistent across participants. The red bar indicates percentage accuracy values as a fraction (significance threshold set at 66.07% for all participants - see Tables 14 and 15 for thresholding and comparison pair details). ‘R’ and ‘L’ are right and left sides of the brain respectively.

### 6.3.4 Discriminating between four positions

I next investigated whether there were voxels in the hippocampus capable of discriminating simultaneously between all four target positions in a room. Using the same protocol as above all six possible pairwise classifiers were performed for each room (comparing positions A v B, A v C, A v D, B v C, B v D, and C v D against each other – see Figure 26) and combined their results into error correcting output codes from

which resultant predictions were determined by computing the nearest Hamming distance to a real label code (see Section 6.2.9).

Although these four-way classifications are dependent on a linear combination of the pairwise classifications above, they provide distinct information about the data because significant voxel accuracy in pairwise classification does not necessitate significant accuracy in four-way classification. Significant voxels were again reprojected back onto the structural brain image of a participant to produce prediction maps. This revealed a focal cluster of voxels in the body-posterior of the hippocampus bilaterally that allowed for accurate differentiation between all four positions in a room, again independent of visual input (Figure 29), a result that was markedly consistent across participants. There were very few discriminating voxels elsewhere in the MTL, thus demonstrating the specific involvement of the hippocampus in representing spatial positions.

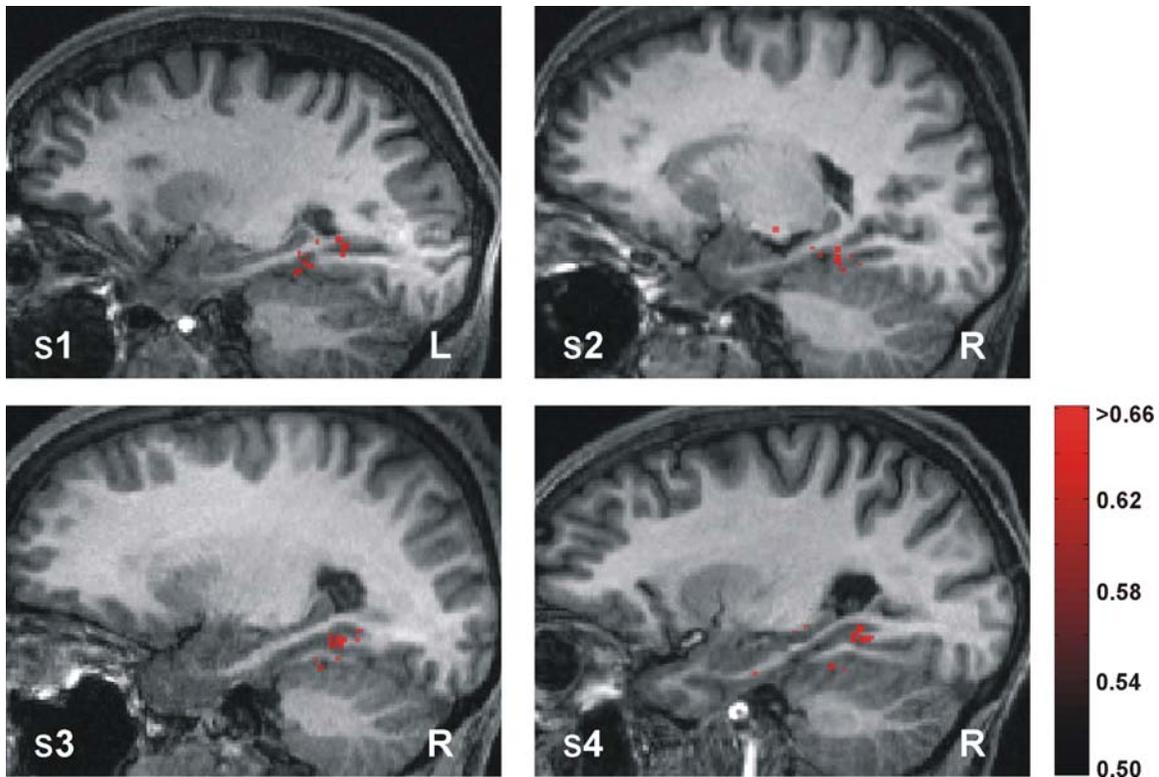


### **Figure 29. Four-way Position Classification**

Prediction maps, bounded by the search space (indicated by the red box in Figure 27b) and projected onto each participant's structural brain image, showing the accuracies of the voxels at the centre of searchlight cliques that discriminate between all 4 target positions in the same room significantly above chance (25%). Sagittal and coronal sections for each participant are displayed on left and right panels respectively, showing that voxels in the body-posterior of the hippocampus bilaterally are crucial for accurate 4-way position discrimination by the classifier. The findings are highly consistent across participants. The red bar indicates percentage accuracy values as a fraction (significance threshold set at 33.04% for all participants - see Tables 14 and 15 for thresholding details). Four-way position discrimination in the green room is shown for participants 1 and 2, and in the blue room for participants 3 and 4. 'R' and 'L' are right and left sides of the brain respectively.

### **6.3.5 Discriminating between the two environments**

While spatial positions of the participant within the environment were represented almost exclusively in the hippocampus, our findings also highlighted an interesting dissociation between the hippocampus and parahippocampal gyrus. In a separate multivariate analysis I tested whether it was possible to accurately predict which environment, the blue or green room, a participant was in during navigation. The prediction maps obtained revealed voxels in the parahippocampal gyrus bilaterally that allowed for differentiation between environments (Figure 30). In contrast to the position analysis, minimal numbers of voxels were found in the hippocampus that accurately discriminated between the two environments.



**Figure 30. Environment Classification**

Prediction maps, bounded by the search space (indicated by the red box in Figure 27b) and projected onto each participant's structural brain image, showing the accuracies of the voxels at the centre of searchlight cliques that discriminate between the blue room and the green room significantly above chance. A representative sagittal section for each participant is displayed showing that voxels in the posterior parahippocampal gyrus bilaterally are crucial for accurate discrimination between the two environments by the classifier. The result is consistent across participants. Note the dissociation between the parahippocampal gyrus prediction maps here and the hippocampus prediction maps observed for position discrimination (see Figures 28 and 29). The red bar indicates percentage accuracy values as a fraction (significance thresholds were set for each participant between 57.45% and 58.00% – see Tables 14 and 15). 'R' and 'L' are right and left sides of the brain respectively.

### 6.3.6 Sub-region dissociation

For each classification type the differences in numbers of discriminating voxels present in the hippocampus and parahippocampal gyrus respectively was formally quantified by performing a difference of population proportions (Daniel & Terrell 1995) significance test on the two anatomically defined regions (see Section 6.2.10). For the pairwise (see Figure 28) and 4-way (see Figure 29) position prediction maps for both the blue and green rooms the proportion of active voxels in the hippocampus was significantly higher

than for the parahippocampal gyrus in all participants (s1: pairwise, HC=218/3027 voxels, PHG=133/2461 voxels,  $p=0.007$ ; 4-way blue, HC=215/3027, PHG=101/2461  $p<0.001$ ; 4-way green, HC=175/3027, PHG=80/2461,  $p<0.001$ ) (s2: pairwise, HC=118/4032, PHG=70/3822,  $p=0.002$ ; 4-way blue, HC=190/4032, PHG=140/3822,  $p=0.02$ ; 4-way green, HC=367/4032, PHG=133/3822,  $p<0.001$ ) (s3: pairwise, HC=261/3665, PHG=69/3888,  $p<0.001$ ; 4-way blue, HC=245/3665, PHG=164/3888,  $p<0.001$ ; 4-way green, HC=316/3665, PHG=118/3888,  $p<0.001$ ) (s4: pairwise, HC=254/3738, PHG=219/3863,  $p=0.042$ ; 4-way blue, HC=182/3728, PHG=140/3863,  $p=0.007$ ; 4-way green, HC=128/3738, PHG=86/3863,  $p=0.002$ ).

Conversely for the environment discrimination (see Figure 30) the proportion of active voxels in the parahippocampal gyrus was significantly higher than for the hippocampus in all participants (s1: HC=30/3027, PHG=61/2461,  $p<0.001$ ; s2: HC=117/4032, PHG=146/3822,  $p=0.024$ ; s3: HC=32/3665, PHG=133/3888,  $p<0.001$ ; s4: HC=49/3738, PHG=78/3863,  $p=0.016$ ). These findings therefore quantify the dissociation observed in the activity patterns of the hippocampus and parahippocampal gyrus (see Figures 28, 29 and 30) and allow the null hypothesis to be rejected by demonstrating that there is a significant difference in the proportions of active voxels between the hippocampus and parahippocampal gyrus. These findings also mitigate against the multiple comparisons problem as the data cannot be explained by false positives due to random variation alone, as this would have resulted in a uniform distribution of significant voxels across the whole search space.

In addition to the voxel count difference of proportions test described above, I used a second analytic approach to test for a region x classification type interaction between the hippocampus and parahippocampal gyrus. Classifiers trained on A v B in the blue room (with average classification accuracy 66.1%, SD 6.4%, significantly above chance  $p=0.015$ ) performed at chance when discriminating the blue room and the green room contexts (average classification accuracy 51.9%, SD 3.0%, not significantly above chance  $p=0.27$ ). Similarly, classifiers trained on the blue room versus green room context (with average classification accuracy 54.9%, SD 0.8%, significantly above chance  $p=0.001$ ) performed at chance when discriminating A v B in the blue room (average classification accuracy 51.3%, SD 5.5%; not significantly above chance  $p=0.66$ ). This therefore confirms that there is a region x classification type interaction

between the hippocampus and parahippocampal gyrus and the position and context classifications.

### 6.3.7 Permutation testing and thresholds

The multivariate searchlight analysis approach provides prediction accuracy values for every voxel in the search space. In order to determine in an unbiased fashion whether these accuracies were significantly above chance (50% in the case of environment and pairwise position classification and 25% in the case of the 4-way position classification) 100 independent runs of the classification were performed with random permutations of the training labels. Voxel accuracies from these runs were concatenated into one distribution and the 95th percentile found. This accuracy value then represented an unbiased threshold above which voxel accuracy was considered to be significantly above chance. Table 14 summarises the threshold levels for each participant for each classification type.

**Table 14. 95<sup>th</sup> percentile accuracy value for prediction map thresholding\***

Classification	95 <sup>th</sup> percentile accuracy value (%) for each participant			
	s1	s2	s3	s4
Pairwise position <sup>†</sup>	66.07	66.07	66.07	66.07
4-way position (blue room)	33.04	33.04	33.04	33.04
4-way position (green room)	33.04	33.04	33.04	33.04
Environment	57.91	58.00	57.45	57.74

\*Chance is 50% in the case of pairwise position and environment classification, and 25% in the case of the 4-way position classifications. The values in the table represent accuracy thresholds above which a voxel was considered significant, equivalent to a confidence level of  $p < 0.05$  uncorrected in a standard t-test.

<sup>†</sup>For pairwise position classification target pairs were selected at random. Thus the 100 random runs were performed on: blue room AvB for s1; blue room AvB for s2; green room AvB for s3; blue room CvD for s4.

The overall consistency of the value of the 95th percentile accuracy for the permutation testing across participants shows that 100 independent random runs were sufficient to account for any random statistical variation in the voxel populations whilst remaining

computationally feasible on standard multi-processor computers (Nichols & Holmes 2002). Note that the environment thresholds show slight variability between participants compared to the position thresholds due to there being many more volumes to classify (~280 for environment compared to 56 for pairwise position and 112 for 4-way position) therefore allowing for a finer granularity of percentage accuracy values. The greater number of volumes to be classified is also the reason why the significance threshold for environment is lower than that for pairwise position.

Two final verification procedures were also performed. First, for every voxel in the search space I also averaged their individual accuracy values across all 100 random runs and found the 95th percentile of that distribution. Table 15 summarises these threshold values and, as can be observed, they are at chance with no voxel accuracies reaching significance thus indicating there was no systematic noise bias in the underlying imaging data.

**Table 15. 95<sup>th</sup> percentile averaged accuracy values for random runs\***

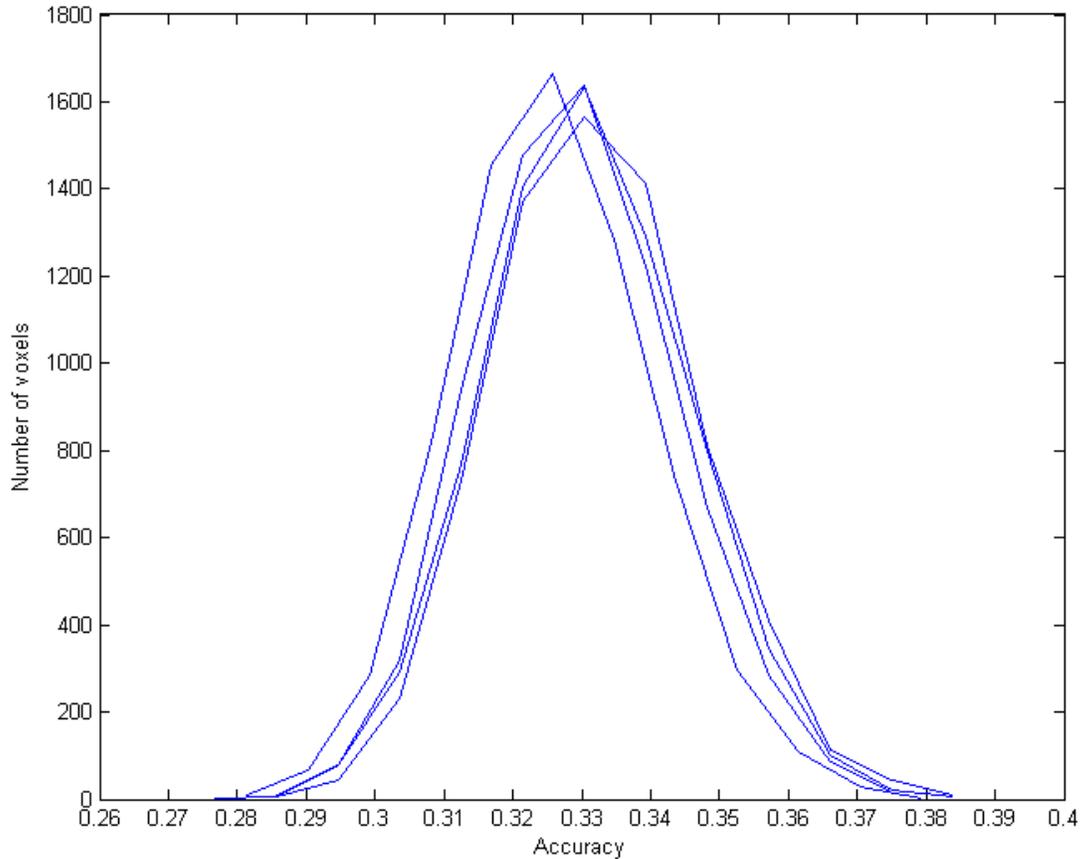
Classification	95 <sup>th</sup> percentile averaged accuracy value (%) for each participant			
	s1	s2	s3	s4
Pairwise position <sup>†</sup>	50.16	50.06	49.78	50.78
4-way position (blue room)	25.59	25.96	25.76	25.37
4-way position (green room)	25.32	25.58	25.66	25.53
Environment	49.45	50.03	49.51	49.92

\*Chance is 50% in the case of pairwise position and environment classification, and 25% in the case of the 4-way position classifications. Note that all the averaged accuracy values in the table are at chance with no voxel accuracies reaching significance thus indicating there was no systematic noise bias in the underlying imaging data.

<sup>†</sup>For pairwise position classification target pairs were selected at random. Thus the 100 random runs were performed on: blue room AvB for s1; blue room AvB for s2; green room AvB for s3; blue room CvD for s4.

Second, for each voxel individually the 95th percentile accuracy value across the 100 random runs for the 4-way classification in the blue room was taken and plotted (Figure 31). For all participants the values were normally distributed round a mean of approximately 33% with a small standard deviation of 1.46% thus demonstrating that

there is no systematic anatomical variation in the signal-to-noise ratio across the search space and justifying the use of a single global threshold value for all voxels in the search space.



**Figure 31. 95th percentile threshold for each voxel individually**

For each voxel in the search space (Figure 27b) the 95th percentile accuracy value across the 100 random runs for the 4-way classification in the blue room was taken and plotted as a population. Each blue line represents a single participant's population. The x-axis shows the 95th percentile accuracy value for a given voxel, and the y-axis displays how many voxels had 95th percentiles at that value. Populations were normally distributed for all participants (s1: mean 32.93%, SD 1.45%; s2: mean 33.20%, SD 1.48%; s3: mean 33.07%, SD 1.48%; s4: mean 32.59%, SD 1.46%) demonstrating there was very little variation in the signal-to-noise ratio across the voxels in the search space and therefore justifying our use of a single global threshold accuracy value for reprojection.

### **6.3.8 Additional multivariate control analyses**

An additional control binary classification for all participants was performed on volumes acquired during the navigation period in the blue room prior to the reaching position A versus those acquired prior to reaching position B. The thresholded reprojections did not reveal any consistent patterns of results anywhere in the search space. Therefore it is very unlikely that the activation patterns observed on the basis of the volumes acquired during the stationary phase at the destination locations (see Figure 28 and Figure 29) could be due to BOLD signal generated by any orienting objects systematically viewed en route.

A further set of additional control analyses were performed to confirm no other factors could have significantly contributed to the decoding performance (see Table 16 for individual subject data). Classifiers trained on destination labels (e.g. A v B) in the blue room performed at chance when discriminating destination labels (e.g. A v B) in the green room (average classification accuracy 51.3%, SD 5.9%, not significantly above chance  $p=0.68$ ) thus demonstrating that BOLD signals measured did not encode destination labels as a discriminable factor in this task. Classifiers trained on nearby landmarks (e.g. A v B) in the blue room also performed at chance when discriminating corresponding nearby landmarks (B v C) in the green room (average classification accuracy 52.5%, SD 5.6%, not significantly above chance  $p=0.45$ ) thus showing that the similarity of nearby orienting object exemplars was not a discriminable factor from BOLD signals recorded in this task. Finally, classifiers trained on relative position (e.g. C v D in the green room) performed at chance when discriminating A v D in the blue room (average classification accuracy 47.3%, SD 8.6%, not significantly above chance  $p=0.58$ ) thus demonstrating that the internal representation of spatial position was absolute and not relative to global room-centred coordinates anchored by the door.

**Table 16. Control analyses**

Participant	Classification accuracies			
	<i>Original (spatial)*</i>		<i>Destination labels</i>	
s1	bAvB	0.66	gAvB	0.52
s2	gAvB	0.61	bAvB	0.45
s3	bAvB	0.63	gAvB	0.58
s4	gAvB	0.75	bAvB	0.50
	<i>Original (spatial)*</i>		<i>Nearby objects</i>	
s1	bAvB	0.66	gBvC	0.50
s2	gAvB	0.61	bBvC	0.50
s3	bAvB	0.63	gBvC	0.48
s4	gAvB	0.75	bBvC	0.60
	<i>Original (spatial)*</i>		<i>Relative to the door</i>	
s1	gCvD	0.68	bAvD	0.43
s2	bCvD	0.70	gAvD	0.55
s3	gCvD	0.61	bAvD	0.53
s4	bCvD	0.63	gAvD	0.38

g=green room; b=blue room; \*all significantly above chance

## 6.4 Discussion

The present results demonstrate that fine-grained spatial information can be accurately decoded solely from the pattern of fMRI activity across spatially distributed voxels in the human hippocampus. This shows that the population of hippocampal neurons representing place must necessarily be large, robust and non-uniform. Thus these findings imply that, contrary to prevailing theories, there may be an underlying functional organisation to the hippocampal neural code. The present data also revealed a dissociation, permitting conclusions about anatomical specificity. While spatial positions were expressed in the hippocampus, by contrast voxels in the parahippocampal gyrus discriminated between the two environments.

Extending the pairwise position classification findings (Figure 28) to discriminate between four arbitrary environmental positions (Figure 29) revealed a region of the

hippocampus that is involved in the general storage and/or manipulation of position representations. The involvement of neuronal populations located specifically in the body-posterior of the hippocampus (Andersen et al. 2007) as indicated by the present data is highly consistent with findings from human and animal studies of spatial memory using other investigative techniques (Colombo et al. 1998; Maguire et al. 2006b; Moser & Moser 1998). It is therefore proposed that these individual abstracted position representations aggregated together form the basis of the allocentric cognitive map (O'Keefe & Nadel 1978), or the set of invariant spatial relationships (Cohen & Eichenbaum 1993), representing the layout of an environment.

Due to the constraint that pattern classifiers require a certain number of consistent examples for training purposes (Haynes & Rees 2006; Norman et al. 2006) discrete localised positions had to be used as target locations. However, there is nothing special about the target locations used in this study as any positions in the rooms could have been chosen. Indeed within each target location a participant's stationary position varied subtly trial-by-trial as the target area measured 1.5mx1.5m in size. Thus I suggest the spatial code for an environment is likely to be continuous with subtle differences in the neuronal code between adjacent positions.

The volumes acquired during an environment block whilst in the blue or green room (see Figure 26d) comprised fMRI activity from a large number of different 'snapshot' views of a room at numerous spatial positions within it (not only our four target positions). Hence the belief the classifier operating on hippocampal voxels did not discriminate between the two environments because this would have necessitated these voxels to have identifiably similar patterns of activity across environment block volumes (i.e. volumes acquired while in the blue or the green room). However, hippocampal voxels were instead acutely tuned to individual spatial positions within a block and therefore displayed differing patterns of activity during navigation in an environment block that encompassed numerous spatial positions.

By contrast, it is clear that the parahippocampal gyrus performed a distinct but complementary function. I speculate that this may have involved extracting the salient contextual features of each environment (Bar 2004; Epstein & Kanwisher 1998) such as object-in-place associations (Janzen & van Turennout 2004) and orienting wall object

configurations from multiple visual snapshots for input to the hippocampal place representations (Bird & Burgess 2008). Thus, the classifier operating on parahippocampal gyrus voxels was able to discriminate between the two environments, although the possibility that this region might have also been sensitive to the colour differences between the two environments cannot be excluded. Further studies will be needed to ascertain the exact nature and function of the representations in the parahippocampal gyrus during navigation and indeed other neocortical areas such as the prefrontal and parietal cortices also known to be involved in navigation (Spiers & Maguire 2006) but which were outside of the scanning coverage of this study.

The rigorous design of the paradigm, in particular the careful matching of visual input at the destination locations, the counterbalancing of starting and destination location combinations, and the use of an incidental visual task to maintain attention during the stationary phase, allows conclusions to be made that any informative patterns of voxels found by the multivariate analyses must code for the internal representation of spatial location only and not any other aspects of the task. In addition to these design features, the analysis was robust to any residual cognitive differences that may conceivably have occurred.

Classifiers can be thought of as distinguishing between learned commonalities across multiple training examples of two experimental conditions. Therefore in order for the classifier to successfully decode brain activity, the difference between two conditions must be systematic and consistent across the majority of the training examples. The paradigm was carefully designed to ensure that the only possible systematic difference between stationary periods was the internal representation of the current position. This was further confirmed by the additional control analyses that were performed to ensure that other factors such as the identity of the destination labels themselves or nearby orienting objects could not have significantly contributed to the successful decoding.

Hence it is with some confidence that the claim can be made that the hippocampal voxels that survived the rigorously controlled thresholding employed here, were associated with internal representations of absolute position within the environment alone. A further point to note specifically in relation to the effect of previously seen landmarks on the BOLD signal during the stationary phase is that paths and approaches

taken to target positions were not identical across trials and the timings of any views of landmarks en route varied widely. The effect of such substantial variability in paths to the target position in effect introduced a self-paced random jitter with respect to the influence of any landmarks seen on the BOLD signal during the stationary periods. Therefore landmarks cannot be a contributing factor to the successful performance of the classifier on the position discrimination.

The present finding that it is possible to distinguish between well-matched spatial positions with human fMRI, has significant implications for understanding the neuronal population code in the hippocampus. It has been proposed that information is encoded in the brain as a sequence of cell assemblies with each activated clique encapsulating a fundamental unit of information (Buzsaki 2004; Harris 2005; Hebb 1949; Pouget et al. 2000). Cell assembly synchronisation is thought to take place over timescales of ~30 milliseconds (Harris 2005), in contrast to the time frame of human neuroimaging which measures activity averaged over ~6 seconds. Although the BOLD signal is only an indirect measure of neuronal activity, and there is ongoing debate about the relationship between the two (Logothetis 2008), there is a robust correlation between BOLD responses and local field potentials (Goense & Logothetis 2008; Logothetis 2008). Therefore patterns of voxel activations acquired during a single fMRI volume and capable of discriminating between well-matched positions are likely to reflect the average synaptic activity within many cell assemblies that taken together can represent high level information such as spatial location within an environment.

Although neural codes in the hippocampus and wider MTL are generally considered to be sparse (see Section 1.5.4.1) (Quiroga et al. 2008; Quiroga et al. 2005; Waydo et al. 2006), this term has been used to describe a wide range of different representational scales from single 'grandmother' cells (Gross 2002) to more than two million cells in other accounts (Quiroga et al. 2008; Waydo et al. 2006). The human hippocampus contains approximately 40 million principal neurons (Andersen et al. 2007) and even at the high spatial resolution of the scanning employed here this translates to  $\sim 10^4$  neurons per voxel. Given the relatively coarse and noisy nature of human neuroimaging in both the temporal and spatial domains, it is striking that it was possible to robustly distinguish between positions of a participant in the environment that vary in only subtle ways.

To the extent that multivariate classification using fMRI reflects biased sampling of a distributed anisotropic neuronal representation (see Section 6.3.6) (Kamitani & Tong 2005), the results are consistent with the notion that hippocampal neuronal ensembles representing place memories are large and have an anisotropic predictable structure. Moreover, the prediction maps obtained indicated the presence of information sufficient to decode position from voxels distributed spatially throughout the hippocampus. The present data, therefore, are broadly supportive of two previous invasive studies that have suggested there may be some form of clustering (Eichenbaum et al. 1989) or topographical functional organisation (Hampson et al. 1999) in the hippocampus.

Although numerous invasive studies have reported the population code is random and uniformly distributed (Guzowski et al. 1999; Redish et al. 2001), a point often implicitly assumed by computational models (Hartley et al. 2000; Samsonovich & McNaughton 1997), this would result in uniform patterns of activity at the voxel level thus rendering classification impossible (Haynes & Rees 2006; Norman et al. 2006). However, there are ways in which these opposing views and our findings can be potentially reconciled. For instance the spacing of tetrodes randomly sampling single neurons (Redish et al. 2001) could be out of phase with the structure of the underlying functional organisation (Hampson et al. 1999), there may be differences in the clustering analyses used (see (Eichenbaum et al. 1989) compared with (Redish et al. 2001)), multiple factors other than space may be encoded by hippocampal cells including task (Pouget et al. 2000; Redish et al. 2001) and goal (Ekstrom et al. 2003), while the effect of cell assembly synchronisation on single cell spike output is unknown (Harris 2005).

## **6.5 Conclusions**

In this study, I focused on the cross-species behaviour of navigation demonstrating that highly abstracted representations of space are expressed across tens of thousands of coordinated neurons in the human hippocampus in a structured manner. In so doing I have shown that, contrary to current consensus, neuronal ensembles representing place memories must be large, stable and have an anisotropic structure. Spatial representations of the type investigated here have been suggested to form the scaffold upon which episodic memories are built (see Chapter 8) (Bird & Burgess 2008; O'Keefe & Nadel 1978), but the precise mechanism by which the hippocampus achieves this is

still unknown. This crucial question is difficult to address in non-humans, where even the existence of episodic memory has been challenged (Tulving 2002b).

By showing it is possible to detect and discriminate between memories of adjacent spatial positions, the combination of non-invasive in vivo high resolution fMRI and multivariate analyses opens up a new avenue for exploring episodic memory at the population level. The question that naturally arises is whether it would be possible to decode individual episodic memory traces from the activity of neuronal ensembles in the human hippocampus. This was the subject of my next experiment.

# Chapter 7

*Experiment 4: Finding the Engram?*

## Precis

Experiment 3 demonstrated that high spatial resolution fMRI allied with sensitive multivariate pattern classification (see Section 3.6) were able to successfully decode fine-grained spatial representations in the human hippocampus. In my next exploratory study I extended these methods to the investigation of episodic memory. Specifically, I addressed the question of whether it would be possible to predict which of two episodic memories a subject was recalling solely on the basis of the activity patterns across voxels in the hippocampus. If this proved possible, it would be a step towards the tantalising prospect of ‘seeing’ a unique memory trace directly in the brain. As experimental tools, these techniques potentially pave the way for novel investigations into the properties of the hippocampal memory trace and the nature of episodic memory itself.

## 7.1 Introduction

The search for the elusive engram (Dudai 2004; Schacter 1996; Semon 1923) or memory trace in the brain has been an ongoing endeavour in neuroscience for nearly a century (Lashley 1950; Semon 1923). Although the biological existence of such engrams coding for memories is widely accepted, the precise mechanisms, locations and even the nature of the engram itself, in light of processes such as reconsolidation (Dudai 2004; Dudai 2006; Hupbach et al. 2007), is the subject of much debate. Clearly the components of a complex multi-modal memory such as a rich episodic memory (and thus one could argue parts of the engram) are likely to be widely distributed throughout the sensory cortices. For example, reminisce about the time you once gave a beautiful girl a red rose – a representation of her face would be stored in the fusiform face area (Kanwisher et al. 1997), the rose in the lateral occipital cortex and its scent in olfactory cortex (Gottfried et al. 2004). These components on their own are not enough, however, and something must bind the disparate elements of an episodic memory together to allow the relevant neural representations to co-activate thus facilitating the distinctive recollective experience Tulving described as ‘ecphory’ (Tulving 1983; Tulving 2002b).

Marr’s influential model of memory, discussed earlier in Section 2.2, proposed that the hippocampus provided this function by storing a memory ‘index’, in effect a simple or

distilled representation of the whole memory that contains the essence of the memory and which is synaptically linked to the full representation stored in the neocortex (Marr 1971). Indeed the hippocampus is ideally suited for this multi-modal binding role, given its location at the top of the sensory cortical hierarchy and widely acknowledged vital role in supporting episodic memory (Scoville & Milner 1957). Moreover, the high density of recurrent collaterals in its CA3 sub-region may support the process of pattern completion whereby a partial cue can reinstantiate the entire hippocampal index, which then in turn reactivates the complete neocortical representation of an episodic memory. This hippocampal index therefore can be thought of as the signature component that defines the engram or memory trace. The hippocampus and surrounding MTL would appear to be a prime candidate for investigating the nature of the engram, although clearly it does not support episodic memory recall on its own (Maguire 2001a; Spreng et al. 2008; Svoboda et al. 2006) but is part of a wider core network (see Section 1.4 and Chapter 5).

However, in order to reduce possible interference between similar memories Marr also suggested that such a memory index might be sparsely coded in the hippocampus involving only a small, distributed population of hippocampal neurons which is also computationally efficient in terms of storage capacity. Although sparsity is often assumed by theoretical and computational models (Hartley et al. 2000; Samsonovich & McNaughton 1997), empirical evidence to support this conjecture has been equivocal, with a recent single cell recording study in patients with medically intractable epilepsy showing a relatively large proportion of recorded cells in the hippocampus and entorhinal cortex consistently responded to free recall of a previously viewed episodic event (Gelbard-Sagiv et al. 2008). This suggests that it may be possible to extend the techniques pioneered for the spatial study described in Chapter 6 to episodic memory.

Specifically I wanted to test the hypothesis that individual episodic memories could be distinguished from one another on a single subject basis solely from patterns of activity across voxels in the human hippocampus and wider MTL, something not possible to do with conventional univariate techniques (see Section 3.6). Accordingly, subjects were asked to recall four distinct episodic memories a total of 12 times each in the scanner in a pseudorandom order. Of the four memories, two were real and two were imaginary, although all were recent, in order to mitigate against the possibility that the memory trace might have been consolidated out of the hippocampus to the neocortex (Squire et al.

2004). If this proved possible in this preliminary high spatial resolution fMRI study, it would potentially provide an exciting new tool with which to investigate many important properties of episodic memory and their traces non-invasively in healthy humans (see Section 7.5).

## **7.2 Methods**

### **7.2.1 Participants**

Three healthy, right-handed, native English speakers participated in the experiment (two males, mean age = 31.0 years (SD = 6.0), age range 25-37). All subjects gave informed written consent to participation in accordance with the local research ethics committee.

### **7.2.2 Pre-scan interview**

Interview sessions were conducted with subjects approximately one hour prior to scanning. Subjects sat facing the experimenter and were asked to verbally recount two recent episodic events that had occurred in the past six weeks. Memories had to be emotionally neutral, specific in time and place, and vividly recalled to be accepted as a stimulus. The two memories also had to be substantially different in terms of context and content. They were then asked to create two imagined events in response to short verbal cues describing scenarios (i.e. standing in a forest by a stream, standing in a bustling street market), following the protocol used in Chapters 4 and 5.

Subjects were asked to close their eyes and vividly imagine the scene and were then given 2-3 minutes to describe it in as much (multi-modal) detail as possible. They were instructed not to recount an actual memory or any part of one or something they planned to do but rather to create something completely new. After each scenario, subjects were asked to rate their imagined scenes for vividness (1 – not vivid...5 – very vivid) and for similarity to a real memory (1 – nothing like any memory...5 – exactly like a memory).

After a 30 minute break subjects were asked to recall and describe the four memories. They were told that they would be required to re-experience each memory a number of times in the scanner and that they therefore should settle on a prototypical 'template' version of each memory and try to recall each memory as consistently as possible each

time. Subjects were asked to think about each memory in turn and settle on template versions for each. They were then asked to describe their template versions twice each and if the second description was identical to the first, the template version of the memory was accepted.

### **7.2.3 Task and procedure**

Scanning consisted of a single main session lasting around 30 minutes during which 12 trials for each of the 4 memories were presented in a pseudorandom order with the constraint that the same memory was not presented twice in a row. The subject was informed as to which memory was to be recalled in that trial by a two-word text cue displayed in white text on a black background for 3.5s. This was then followed by a 'close your eyes' text instruction (see Figure 32). At this point subjects were instructed to close their eyes immediately and begin visualising the memory (the 'visualisation period') as vividly and consistently as possible for 16s. Subjects were told to maintain focus on the memory and keep re-experiencing it for the entire 16s duration.

Sixteen seconds was chosen because piloting for Experiment 2 (see Chapter 5) indicated that was sufficient time for a memory to be recalled comfortably in the scanning environment and the mental image of it to stabilise. A simple 1s audio tone signalled the end of the visualisation period (at which point subjects opened their eyes) and the start of the ratings phase. Using an MR-compatible 5-button key pad subjects scored their just visualised memory on 2 ratings: vividness (salience of the imagery: 1 – not vivid...5 – very vivid) and consistency (similarity to the template version of the memory: 1 – completely different...5 – exactly the same). For each rating subjects were given 4.5s to respond followed by 1s feedback confirming the button they had pressed. The rating period was followed by a 5s rest interlude before the next cue was presented.

### **7.2.4 Debriefing**

After scanning subjects were debriefed. They were asked about the methods they had used to recall the memories and the consistency with which they re-experienced each memory across different trials. The latter was verified by getting a participant to describe their template image for each memory and comparing it to what they had said in the pre-

scan interview. They were also questioned about their general concentration levels throughout the scanning session.

### **7.2.5 Image acquisition**

A 3T Magnetom Allegra head scanner (Siemens Medical Solutions, Erlangen, Germany) operated with the standard transmit-receive head coil was used to acquire functional data using a T2\*-weighted single-shot echo-planar imaging (EPI) sequence (in-plane resolution = 1.5x1.5 mm<sup>2</sup>; matrix = 128x128; field-of-view = 192x192 mm<sup>2</sup>; 35 slices acquired in an interleaved order; slice thickness = 1.5 mm with no gap between slices; echo time TE = 30 ms; asymmetric echo shifted forward by 26 phase-encoding (PE) lines; echo spacing = 560 μs; repetition time TR = 3.57 s; flip angle  $\alpha = 90^\circ$ ). All data were recorded in one single uninterrupted functional scanning session with 488 volumes acquired for each subject.

An isotropic voxel size of 1.5 x 1.5 x 1.5 mm was chosen for an optimal trade-off between BOLD sensitivity and spatial resolution. Further, the isotropic voxel dimension reduced resampling artifacts when applying motion correction. In order to minimise repetition time whilst also optimising coverage of the regions of interest in the medial temporal lobe, we captured partial functional volumes angled at 5 degrees in the anterior-posterior axis. Susceptibility-induced loss of BOLD sensitivity in the medial temporal lobe was intrinsically reduced by the high spatial resolution and adjusting the EPI parameters for the given slice tilt (z-shim gradient prepulse moment = 0 mT/m\*ms; positive PE polarity). A T1-weighted high-resolution whole-brain structural MRI scan was acquired for each subject after the main scanning session (1 mm isotropic resolution, 3D MDEFT).

### **7.2.6 Image data pre-processing**

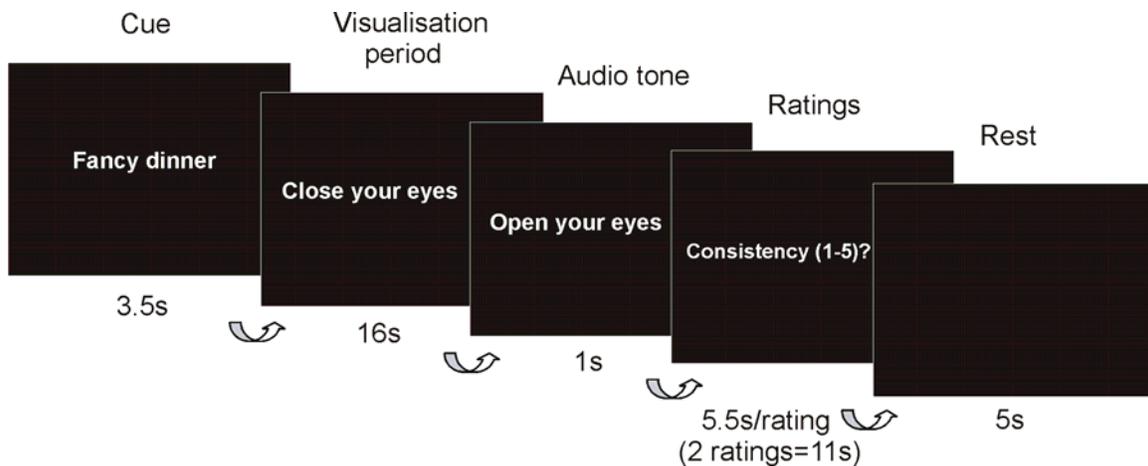
This consisted of realignment to correct for motion effects and minimal spatial smoothing with a small 3mm FWHM Gaussian kernel. The first six “dummy” volumes of the main session were discarded to allow for T1 equilibration effects (Frackowiak et al. 2004).

### **7.2.7 Multivariate pattern classification**

I performed a multivariate pattern analysis (see Section 3.6) (Haynes & Rees 2006; Norman et al. 2006) designed to identify brain regions where distributed fMRI activation patterns carried information about the specific episodic memory a participant was recalling. The procedures were similar to those outlined in Chapter 6.

For this preliminary analysis I focused on the two real memories. Using the standard leave one block out procedure this resulted in a 24-fold cross-validation step (12 trials of recalling real memory A and 12 trials of recalling real memory B). The cue was presented for 3.5s and the visualisation period lasted for 16s, with a TR of 3.57s this yielded approximately (depending on the rounding) one volume for the cue and five volumes for the visualisation period respectively. The classification, thresholding and reprojection procedures were performed in exactly the same way as in Chapter 6 (see Section 6.2).

I first performed the classification treating all six volumes of each trial, encompassing the cue onset to the end of the visualisation, as one homogeneous block. I next performed an additional series of classifications that made use of a two-volume 'sliding time window' to allow the temporal dynamics of the recall process to be appreciated. The re-experiencing of a past episodic memory is a complex process (Hassabis & Maguire 2007) and likely involves a series of processes evolving over time (Addis et al. 2007) underpinned by distinct sets of neural regions. Starting at the cue onset and ending on the last volume of the visualisation period a two-volume sliding time window was systematically applied resulting in 5 separate positions for the time window. Each two-volume time window was analysed separately yielding 5 separate classifications per subject. The 5 resultant prediction map outputs by the classifiers were each thresholded and reprojected separately (see Section 6.2.10).



**Figure 32. Example trial**

A two-word text cue was presented for 3.5s denoting the memory to be re-experienced. Subjects were then instructed to close their eyes and begin visualising the scene or object in as much detail as possible for 16s. A simple audio tone, played through headphones and lasting 1s, indicated the end of the visualisation period at which point the subject opened their eyes. Subjects then used a 5-button MR-compatible box to rate their just re-experienced memory on two five-point scales for vividness and consistency. Subjects were given 4.5s to respond per rating and feedback as to what button had been pressed was shown for 1s, resulting in an overall rating period of 11 seconds. This was followed by a 5s rest period where a blank screen was presented prior to the start of the next trial.

## 7.3 Results

Note that the following are initial results from preliminary analyses of the data performed on the two real memories A and B in each of three participants.

### 7.3.1 Behavioural data

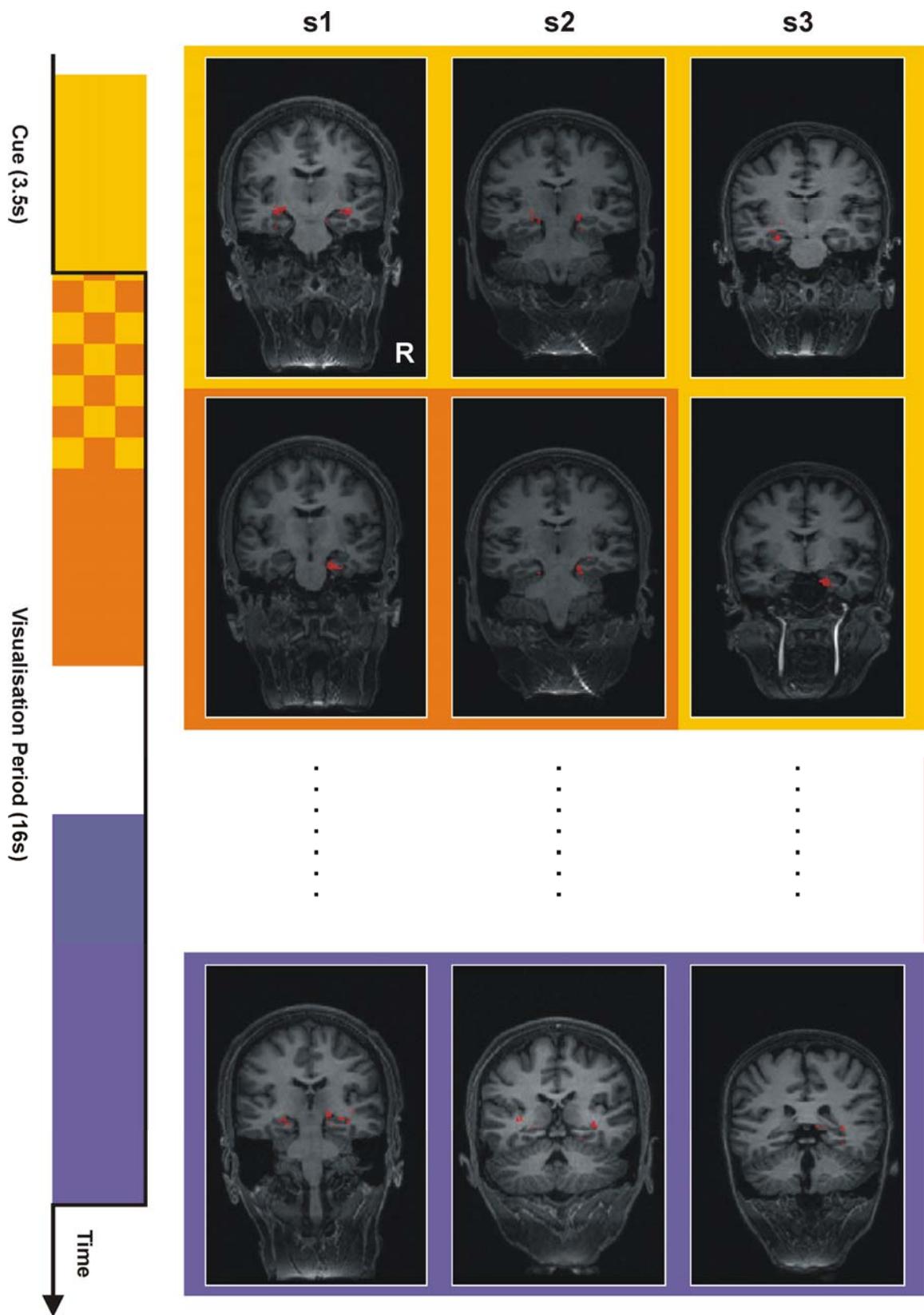
The two real memories (A and B) were well-matched behaviourally with no significant difference in vividness (s1 - A: mean 4.08, B: mean 3.75,  $p=0.17$ ; s2 - A: mean 3.67, B: mean 3.75,  $p=0.72$ ; s3 - A: mean 4.08, B: mean 3.92,  $p=0.17$ ) or consistency (s1 - A: mean 4.50, B: mean 4.33,  $p=0.50$ ; s2 - A: mean 4.50, B: mean 4.50,  $p=1$ ; s3 - A: mean 4.83, B: mean 4.67,  $p=0.49$ ) within a participant.

During the post-scan debriefing subjects all reported that they had consistently recreated each episodic memory with as similar a procedure as possible for each trial. This was verified by the post-scan descriptions of the memories which very closely matched the pre-scan descriptions. All subjects reported finishing the retrieval phase during the early part of the visualisation period and entering a 'maintenance' phase where they kept the image of the memory in mind and added small details to it. Good concentration was reported by subjects throughout the scanning session.

### **7.3.2 Multivariate fMRI data**

I investigated if it was possible to accurately predict which episodic memory a subject was recalling solely from the pattern of fMRI BOLD responses across multiple voxels in the hippocampus and MTL. We first performed a classification using all six volumes across the whole block and as expected this yielded no consistent results anywhere across the search space. This is because what subjects were doing across the whole block was not consistent across all volumes in the block across all trials. Richly recalling real episodic memories is a relatively unconstrained task from a temporal perspective. In order for a classifier to successfully decode brain activity, the difference between volumes acquired during two conditions must be systematic and consistent across the majority of the training examples.

Therefore I conducted a sliding time-window analysis systematically looking at each of the five two-volume mini-blocks covering the whole trial from cue onset to the end of the visualisation period (see Section 7.2.7). The reasoning behind this was that by using short time windows it might be possible to encapsulate the various discrete processes that comprise full episodic recollection (Addis et al. 2007) and that these discrete processes might be more consistent across trials than looking at the entire visualisation block as a whole. This analysis revealed an interesting temporal pattern in terms of the regions activated that appeared to be quite consistent across subjects (Figure 33).



### **Figure 33. Two-volume sliding time window classification**

Prediction maps showing the accuracies of the voxels at the centre of searchlight cliques that discriminated between the two real memories A and B significantly above chance (50%). The resultant prediction map thresholded at 67% (corresponding to  $p < 0.05$  in a standard t-test) for each subject and bounded by the search space (indicated by the red box in Figure 27b) is shown projected onto their respective structural brain images. All images are displayed on coronal sections, with a single subject shown separately in each vertical column. The bar on the left indicates the timeline of a single trial (black line). Classification was performed on two-volume time windows systematically applied across the whole trial (indicated by the solid coloured bars). The yellow bar indicates the first time window encompassing the cue onset and the first volume of the visualisation period. The second time window corresponding to the first two volumes of the visualisation period is indicated by the orange bar (note it overlaps by one volume with the first time window i.e. the chequered section). Dotted lines indicate there were no consistent activations anywhere in the search space for two middle time windows. Classification on the last time window is indicated by the purple bar and covered the final two volumes of the visualisation period.

Patterns of voxel activations in the body of the hippocampus bilaterally discriminated between the two memories during the first time window (yellow bar in Figure 33) encompassing the presentation of the cue and the start of the visualisation period. There was a left hippocampus bias in two of the three subjects. In the subsequent time window covering the first two volumes of the visualisation period (orange bar in Figure 33) the two memories were best discriminated by a highly focal pattern of voxel activations in the right entorhinal cortex. Note that for subject 3 this activation was actually more prominent during the first time window (hence the yellow background). In the next two time windows there were no consistent activations anywhere in the search space. Finally in the last time window encompassing the final two volumes at the end of the visualisation period (purple bar in Figure 33) the two memories were discriminated by voxel activation patterns in posterior hippocampus bilaterally.

## **7.4 Discussion**

By successfully discriminating between two well-matched episodic memories solely on the basis of voxel activation patterns in the hippocampus and MTL, in this study I have taken the first steps towards decoding the neural substrates of individual spatiotemporally specific episodic memories. Furthermore, an interesting set of temporal dynamics active during episodic memory recall have been revealed. This includes the

early engagement of the body of the hippocampus which may be supporting the reactivation of the memory index, and late activation of the posterior hippocampus. In between, the data highlighted the hitherto understudied involvement of the entorhinal cortex in episodic memory. This study was intended to be exploratory, however the significant nature of these results encourages further studies that utilise novel approaches to studying the properties and processes of episodic memory at a level of detail not possible with conventional fMRI techniques. This promises new insights into the nature of episodic memory and the memory traces that define them that may be difficult to discern in non-humans where the very existence of episodic memory has been questioned (Tulving 2002b).

At least two distinct phases of the episodic memory recall process have recently been suggested – construction and maintenance (also called elaboration) (Addis et al. 2007; Hassabis & Maguire 2007; Schacter et al. 2007). The construction phase involves the retrieval of the memory index and the reactivation and integration of the constituent components of the memory (Addis et al. 2007; Hassabis & Maguire 2007; Schacter & Addis 2007). The maintenance phase involves holding the, now coherent, whole memory online to allow for multi-modal re-experiencing (Hassabis & Maguire 2007) and for supplementary details to be optionally added (Addis et al. 2007). These two phases are dependent on a different network of brain regions, one for reconstruction and the other for elaboration of the memory, although both appear to rely on the hippocampus (Addis et al. 2007). Classification using all 6 volumes in a trial (as opposed to the two-volume time windows) as expected yielded no consistent patterns of results anywhere in the search space thus confirming a temporal evolution in brain region activities during the duration of the recall. This shows that previous traditional univariate imaging studies (Maguire 2001a; Svoboda et al. 2006) that have for reasons of design simplicity often implicitly assumed episodic memory recall to be unitary function are likely to be oversimplifications.

In a recent self-paced episodic memory recall fMRI study (Addis et al. 2007) the construction phase was found to last an average of 7.2s. In the present study this construction phase time period corresponds well with the first and second two-volume time windows (see the yellow and orange time periods in Figure 33) covering the cue onset and the start of the visualisation period. The time window analysis revealed an

interesting and consistent temporal evolution of brain region activities. In the first time window (yellow bar in Figure 33) encompassing the cue onset and the first volume of the visualisation period I found large activations in the body of the hippocampus bilaterally consistently across subjects, although with a clear left bias in two of the three subjects. This accords well with previous findings that have implicated the hippocampus (particularly on the left – (Maguire 2001a; Spiers et al. 2001)) during the early part of the construction phase (Addis et al. 2007). This seemingly transient early activation of the hippocampus has parallels in spatial memory recall where in a complex navigation task the hippocampus was found to be active transiently at the start of route-planning, presumably in support of the retrieval of the spatial memories needed for the subsequent navigation (Spiers & Maguire 2006). The body of hippocampus, perhaps predominately on the left for personal episodic memories, would therefore seem to be the most likely candidate for the location of the memory trace or index that defines the memory.

Classification on the second time window (orange bar in Figure 33), corresponding to the first two volumes of the visualisation period, revealed large focal activations in the right entorhinal cortex (note that in one subject this activation was more prominent in the first time window alongside the hippocampal activation). The role of the entorhinal cortex is essentially unknown in autobiographical memory (Maguire 2001a; Svoboda et al. 2006) and here its involvement in autobiographical memory recall in an imaging study is highlighted for the first time. However, a recent single cell recording study in humans (pre-surgical patients with drug resistant epilepsy) (Gelbard-Sagiv et al. 2008) investigating the free recall of recently viewed past events found that a high proportion of cells in the entorhinal cortex were reliably activated in response to the recall of specific episodes. Thus taken together it would seem that the entorhinal cortex is performing a, hitherto underappreciated, role in episodic memory recall. My results appear to demonstrate that the entorhinal cortex is temporally engaged after the initial hippocampal involvement although further analyses will be needed to confirm this. The important recent finding of grid cells in the rat entorhinal cortex (Hafting et al. 2005; Moser et al. 2008) perhaps points to a spatial role for this structure in episodic memory. It has been argued that spatial context provides the scaffold upon which episodic memories are built (Burgess 2002; Burgess et al. 2001; Hassabis & Maguire 2007) and it may be that the entorhinal cortex plays a critical part in setting up that spatial context (Moser et al. 2008). Alternatively involvement of the entorhinal cortex may reflect its

position as the main efferent pathway out of the hippocampus, thus conveying the products of hippocampal index processing downstream in order to drive cortical representations (Andersen et al. 2007).

Classification on the two time windows covering the middle portion of the visualisation period did not pick up any consistent patterns of activations anywhere in the search space. Of course this does not mean that the two memories are not differentiable elsewhere in the brain during this phase of recall. It could be that it is at this point that the other regions, such as ventromedial prefrontal, retrosplenial and parietal cortices, known to be implicated in episodic memory recall (Maguire 2001a; Svoboda et al. 2006) become involved in the refinement and guiding of the episodic memory. It was not able to address this question in the current study due to high resolution scanning constraints that necessarily limited coverage to the MTL and surrounding regions. However, we plan to empirically investigate this issue in future studies.

Classification on the last time window (purple bar in Figure 33), corresponding to the final two volumes of the visualisation period, revealed activations in the posterior hippocampus bilaterally and consistently across subjects. From their debriefings, I ascertained that subjects had by this stage fully recalled the memory and were now in a maintenance phase of holding attention on the visualised memory and thinking about further details. It is interesting here to connect this result with previous structural MRI findings that expert navigators have increased volume in the posterior hippocampus (Maguire 2001a). Navigation is known to involve the regular manipulation of complex imagery, and the posterior hippocampus is ideally situated anatomically, with its high degree of interconnectedness to the retrosplenial cortex, to be involved in the spatial visualisation pathway that runs through retrosplenial cortex and into parietal areas (Burgess 2002; Guariglia et al. 1993). Taken together this suggests a role for posterior hippocampus in complex visualisation of a scene or event, a critical feature of the recollective process.

In summary through the application of multivariate methods I have shed light on the complex temporal interplay between various subregions of the MTL during the recall of episodic memories. In successfully discriminating between two individual episodic memories in a single subject on the basis of patterns of voxels in the hippocampus I

have shown that the hippocampal memory index is activated early in the recall process, may reside in the body of the left hippocampus and must necessarily be large and robust, as otherwise our ability to decode would not have been successful. More work will be needed to ascertain the specific role of the entorhinal cortex highlighted here in episodic memory. Finally, the late activation of the posterior hippocampus once the memory recall process has stabilised and is simply being maintained suggests that there may be a dual role for the hippocampus in episodic memory recall. First, in the storage and reactivation of the memory index that instigates retrieval, and second, in supporting the visual re-experiencing of that memory within the framework of a coherent spatial context. This possibility is discussed further in Section 8.9.

## **7.5 Conclusions**

This exploratory study was primarily intended as proof of concept to show that multivariate searchlight techniques could be applied to interesting questions in the domain of episodic memory. By demonstrating that it is possible to successfully predict from brain activity alone the specific episodic memory a person is recalling, the way has been paved for new and potentially fruitful investigations into key issues and debates such as recent versus remote memories, real versus imaginary memories, content versus context differences, MTL region specialisation, temporal coordination between MTL regions, and free recall - avenues we plan to follow-up in future studies.

# Chapter 8

## *General Discussion*

## **8.1 Overview**

The aim of this thesis was to advance our understanding of the neural processes underpinning episodic recollection. A number of different approaches were employed ranging from neuropsychological testing of patients with hippocampal pathology to multivariate high spatial resolution functional MRI, covering a range of investigative levels from the population code to the brain networks, and with a specific focus on the hippocampus given its critical role in episodic memory. In so doing I succeeded in identifying and characterising a novel component process, namely scene construction (see Chapter 5) (Hassabis & Maguire 2007), that is involved in supporting rich episodic recollection and critically reliant on the hippocampus (Experiment 1).

I then went on to establish the fronto-temporo-parietal brain network underpinning scene construction and demonstrated that it could not only account for a large part of the classic episodic memory network but also constituted a major part of a core network that might lie at the heart of a set of disparate cognitive functions (Experiment 2). Finally, using sensitive multivariate techniques I sought to investigate the properties of the human hippocampal neural code for memory traces representing spatial and episodic memories (Experiments 3 and 4). In this final chapter, I will consider this work together with other recent advances in the episodic memory field, and then present a new perspective on the role of the hippocampus in episodic recollection. I discuss how this new (and preliminary) framework compares with current prevailing theories of hippocampal function, and suggest how it might account for some previously contradictory data.

## **8.2 Breaking down episodic memory using imagination**

Tulving (Tulving 1983; Tulving 2002b) defined three key properties of the recollective experience that characterise episodic memory recall: a subjective sense of time (mental time travel), connection to the self, and autonoetic consciousness – a special kind of consciousness that accompanies the act of remembering, enabling one to be aware of the self in subjective time (Tulving 2002b).

In this thesis I attempted to break down episodic memory into its constituent processes and in so doing make the first moves towards establishing the functions that they underpin, and the brain networks that support them. I did this by stepping outside mainstream memory research by focussing on imagination. Memory is widely accepted to be reconstructive (Bartlett 1932; Schacter et al. 1998) thus it might be expected that a constructive function such as imagination would utilise a similar set of processes, supported by the same brain regions. This was the hypothesis that led to the formulation of Experiment 1 and the testing of patients with hippocampal damage on their ability to imagine fictitious and future experiences.

In many ways imagining new experiences can be regarded as the purest expression of construction. All healthy volunteers can effortlessly use their imagination to a basic degree (indeed humans have told stories and delighted in fiction and narrative for thousands of years), and verbally-induced imagination of scenes has been shown to be possible and useful in the neuropsychological context (Bisiach & Luzzatti 1978). Experiences constructed by the imagination, while having much in common with episodic memories, have the advantage of being easier to systematise and experimentally manipulate (Hassabis & Maguire 2007). Crucially, tasks involving imagined scenarios can be designed to de-emphasise key features allowing insights to be gained into the neural substrates of these features when compared with episodic memories (see Chapter 4). For example, participants can be asked to construct fictitious experiences in their imagination that are atemporal (i.e. not set in the past or in the future) and with a low connection to the self. Being able to manipulate factors such as the level of self-relevance/involvement and the degree of overlap between memories and imagined experiences has the potential to progress our understanding of the core processes and brain areas involved (Hassabis & Maguire 2007). Thus imagination may prove to be a useful tool in the service of understanding the component processes that underpin episodic memory.

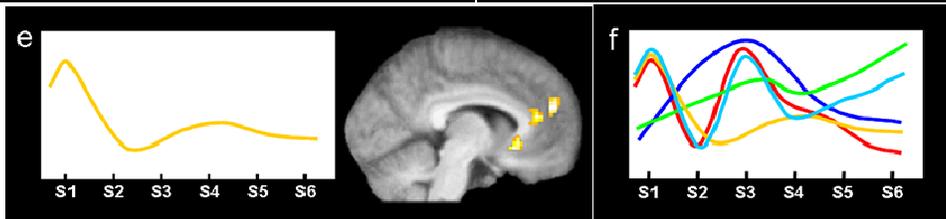
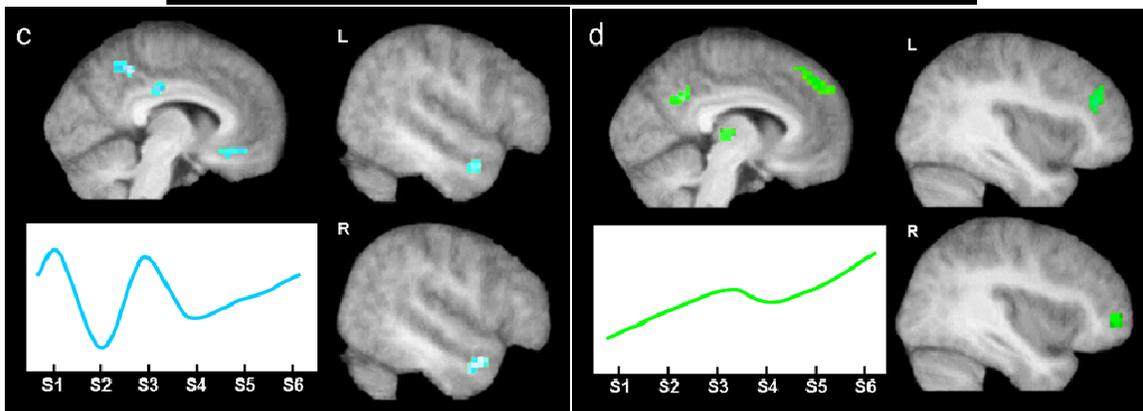
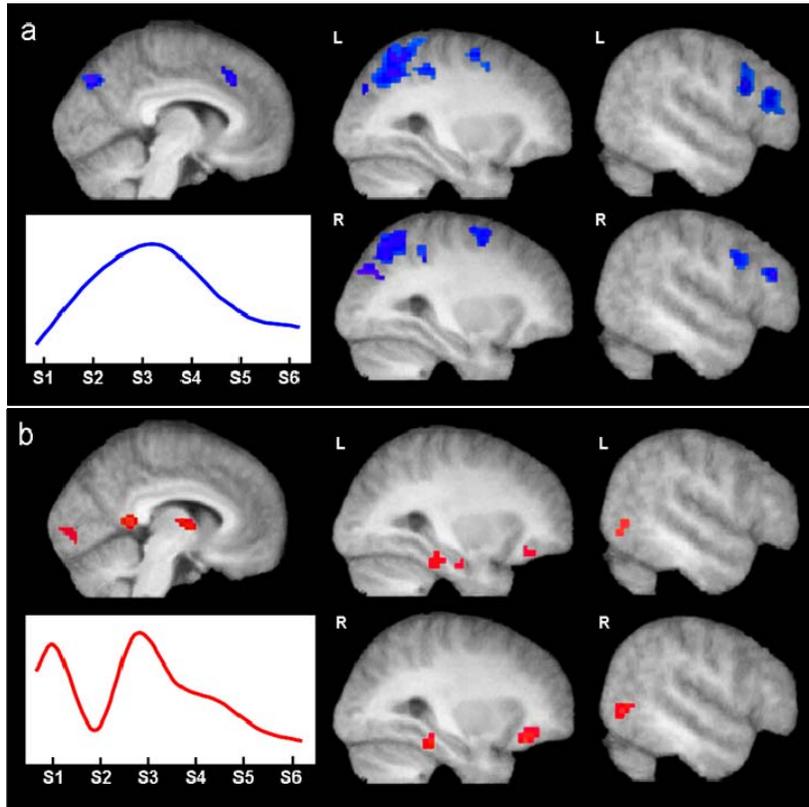
### **8.3 Scene construction**

The results from the patient imagination study (Chapter 4) demonstrate that hippocampal patients are impaired at imagining new experiences. Moreover, this deficit seems to be driven by an inability to form spatially coherent representations of the

imagined scenes. This deficit was encapsulated as the process of 'scene construction' (see Chapter 5), the mental generation and maintenance of a complex and coherent scene or event (Hassabis & Maguire 2007). This is achieved by the reactivation, retrieval and integration of relevant semantic, contextual and sensory components, stored in their modality specific cortical areas (Gottfried et al. 2004 (Deichmann et al. 2003; Wheeler et al. 2000), the product of which has a coherent spatial context, that can then later be manipulated and visualised (Hassabis & Maguire 2007). The hippocampus would appear to play a crucial role in scene construction by integrating the disparate elements of a scene into a spatially coherent whole either via its special spatial processing (O'Keefe & Nadel 1978) or associational (Cohen & Eichenbaum 1993) abilities.

However, the hippocampus, although apparently vital, does not support scene construction on its own. The data from the conjunction analysis of the fMRI imagination study (Experiment 2 – see Chapter 5) revealed a network of regions including not only the hippocampus but also parahippocampal gyrus, retrosplenial cortex, and posterior parietal cortex, as well as the ventromedial prefrontal cortex (vmPFC, BA 12) and middle temporal cortices (see Figure 24) (see Section 5.5). Experiment 2 also demonstrated that there was no overlap between this scene construction network (see Figure 24) and the network supporting the visualisation of 'simple' objects (see Figure 23) (Sugiura et al. 2005), suggesting they represent dissociable cognitive processes with distinct neural bases. Nevertheless, complex scenes and experiences are clearly constructed out of simpler elements.

We recently investigated this dissociation further in a new fMRI study designed to break down the scene construction process itself into constituent sub-processes (Summerfield et al. in preparation). This time, instead of getting participants to imagine a full scene straight off the bat, they were required to build up a scene from individual elements provided by the experimenters. Short audio descriptions of six objects and background items were given (S1-S6), one at a time, and participants had to integrate them into a fully-fledged scene. We found that the usual brain areas in the core network were activated during this task, but critically using this design we were able to break the overall network into a set of smaller sub-networks, and identify different response profiles for these sub-sets of brain areas. We identified five main response profiles (Figure 34).



**Figure 34. Constructing a scene from its elements**

Five different response profiles for different regions of the scene construction network (see Figure 24) as a scene is built up from sequential statements (S1-S6) describing individual elements. (b), (c) and (e) show brain regions that are active initially at S1. (a), (b) and (c) show brain regions that are active at S3 when subjects subjectively report that they have constructed a fully formed scene. (c) and (d) show brain regions that tonically increased with sentence load. (f) shows these 5 different response profiles overlaid.

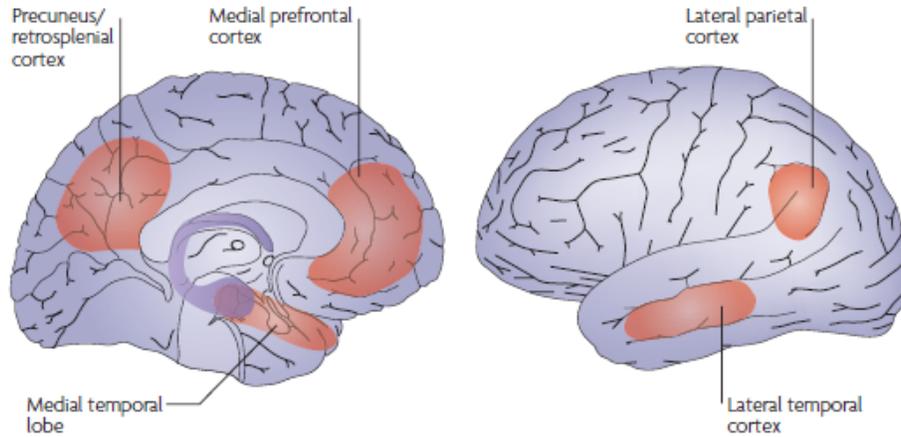
Parietal and dorsal prefrontal areas were most active when the scenes had got to three elements. For the hippocampus, parahippocampal gyrus, retrosplenial cortex, they were very active when participants heard the description of the first element, they were not so interested with the addition of a second element, then again became active at element three and then activity tailed off. Interestingly it was only once three elements had been provided that participants reported the presence of a scene. A different set of regions (including precuneus and vmPFC) showed a similar response but then continued to be active after element three, while activity in PCC and dorso-medial PFC increased the more elements that were added. Finally medial prefrontal areas, including anterior cingulate cortex, were active only for element one. It seems, therefore, that there are broadly three notable features: (i) activations at element one, such as the hippocampus and medial prefrontal cortex, are perhaps involved in task orientation and retrieval of a generic template spatial context; (ii) activations at element three, such as hippocampus and retrosplenial cortex, reflect the process of scene construction and are involved in maintaining and visualising the now reported scenes; (iii) activations increasing tonically with number of sentences, such as dorso-medial PFC, reflect increases in difficulty and working memory load (see Figure 34).

Previously, I have proposed that construction processes can integrate abstracted components no longer directly linked to a specific episodic event (see Chapter 4). It has been suggested by others, however, that past and future experiences draw primarily on information stored in episodic memory (Schacter & Addis 2007; Schacter et al. 2008). However, the component elements of constructions do not have to simply be fragments of past events, but can comprise elements that are more abstracted and semantic such as the sound of ocean waves crashing on the shore or the face of your best friend, and potentially learned over and shared across multiple episodic memories. More work will

be required to ascertain what constitutes an abstracted element that can be used for construction or reconstruction and when enough elements have been combined for the whole to be considered episodic in nature. It should also be noted that scene construction may in fact turn out to be a specific example of a more general 'associative construction' process involving the binding of disparate multimodal elements to (re)create an event as a whole. This includes contextual details such as sounds and smells in addition to visual inputs, people, objects, entities and their actions. This mirrors the debate between the cognitive map theory (O'Keefe & Nadel 1978) and the relational theory (Cohen & Eichenbaum 1993) on the elemental nature of the representations stored in the hippocampus and further work will be needed to establish whether a non-visuo-spatial construction has any practical validity.

## **8.4 The core network**

Interestingly, similar brain networks to the classic episodic memory network (see Figure 3 and Figure 22a) have also been activated in a variety of other circumstances including navigation (Burgess et al. 2002; Hartley et al. 2003; Spiers & Maguire 2006), spatial (Kumaran & Maguire 2005; Maguire et al. 2003) and place tasks (Epstein & Kanwisher 1998; Sugiura et al. 2005), episodic future thinking (Addis et al. 2007; Atance & O'Neill 2001; Schacter & Addis 2007; Szpunar et al. 2007), the default mode (Raichle et al. 2001) and daydreaming (Mason et al. 2007), leading to suggestions that it performs a general function (Buckner & Carroll 2007; Hassabis & Maguire 2007). This presents the intriguing possibility that these disparate cognitive functions, hitherto treated as distinct, might share common underlying processes (Buckner & Carroll 2007; Hassabis & Maguire 2007; Spreng et al. 2008). In fact the notion of a 'core network' of brain regions (Figure 35) underpinning all of these aforementioned cognitive functions has recently been advocated (Buckner & Carroll 2007; Hassabis & Maguire 2007; Schacter et al. 2007).



**Figure 35. The core network**

The core brain system that is consistently activated while remembering the past, envisioning the future and during related forms of mental simulation is illustrated schematically. Prominent components of this network include medial prefrontal regions, posterior regions in the medial and lateral parietal cortex (extending into the precuneus and the retrosplenial cortex), the lateral temporal cortex and the medial temporal lobe including the hippocampus. Moreover, regions within this core brain system are functionally correlated with each other and, prominently, with the hippocampal formation (Vincent et al. 2006). Note how closely this core network maps on to the scene construction network (see Figure 24). From (Schacter et al. 2007).

In view of the fMRI data from Experiment 2 (see Chapter 5) I advance the case for scene construction as a candidate for the common core process. As already discussed, the full recollective experience of a richly recalled episodic memory (Tulving 2002b), for example, remembering what you did last Saturday evening, is nearly always accompanied by complex mental imagery (Rubin et al. 2003) of that event played out within a spatial context (Burgess et al. 2001; O'Keefe & Nadel 1978) – likewise if you cast your mind forwards to what you might be doing next Saturday evening (Atance & O'Neill 2001). However, scene construction is not just limited to remembering the past (Tulving 2002b) or thinking about the future (Atance & O'Neill 2001). For example, nowhere is it more necessary than for navigation (Byrne et al. 2007). Think about how you would get to your favourite restaurant and chances are you will construct a mental representation of the street, the surrounding vicinity, and perhaps where you might park. Alternatively, imagine taking a pleasant walk through a forest that you have never been to, seen, or recognise but is instead purely a product of your imagination. Again, construction of a coherent scene would appear to be integral to the process of vividly

imagining such a fictitious experience (see Chapter 4). Thus, I argue that the cognitive functions outlined in Figure 36 are related because they all crucially rely on scene construction as a common core process (Buckner & Carroll 2007; Hassabis & Maguire 2007), and that this can account for a large proportion of the overlapping network found in neuroimaging studies of these functions (Buckner & Carroll 2007). If scene construction is a key underlying component process of episodic memory, this would accord well with theories of memory that propose the recollection of complex episodic memories is actually a (re)constructive process (Bartlett 1932; Dudai & Carruthers 2005; Schacter & Addis 2007; Schacter et al. 1998; Tulving 1983).

	Scene construction	Subjective time	Self	Autonoetic consciousness	Narrative	Familiarity	Visual imagery	Prospective planning	Task monitoring
Episodic memory recall	Y	Y	Y	Y	Y	Y	Y	N	Y
Episodic future thinking	Y	Y	Y	Y	Y	Y	Y	Y	Y
Navigation	Y	N	D	N	N	D	Y	Y	Y
Imagination	Y	N	N	N	D	D	Y	N	Y
Default network	Y	N	Y	D	D	D	D	N	U
Viewer replay	Y	N	N	N	Y	Y	Y	N	Y
Vivid dreaming	Y	N	D	N	Y	D	Y	N	U
Theory of mind	D	N	Y	N	N	D	D	D	D

**Figure 36. Mapping of component processes to cognitive functions**

Summary of the suggested mapping of component processes onto various related cognitive functions. Processes are labelled along the top and cognitive functions down the left hand side. **Abbreviations:** Y, yes - process is involved in that cognitive function; N, no - process is not involved in that cognitive function; D, depends - process involvement depends on the precise nature of the task and the content operated on; U, unknown - unclear if process is involved in that cognitive function. **Definitions of cognitive functions:** episodic memory recall - vivid recollection of a personal past event (Tulving 2002b); episodic future thinking - envisaging a plausible personal future event (Atance & O'Neill 2001). navigation - route planning and wayfinding (Burgess et al. 2002); imagination - richly imagining and visualising a new fictitious atemporal experience (see Chapter 5); default network - so-called default network activated when no overt task is assigned (Raichle et al. 2001), also suggested to be related to mind wandering or daydreaming (Mason et al. 2007); viewer replay - the vivid replay of an episodic-like memory even though the specific temporal time tag cannot be remembered, e.g. the vivid recollection of one's favourite episode from an old TV series, even though one can no longer remember when or where it was seen. **Definitions of processes:** scene construction - includes the retrieval of relevant semantic and sensory information (Greenberg & Rubin 2003; Wheeler et al. 2000), its integration into a coherent spatial context (see Chapter 4) and online maintenance for later manipulation and visualisation (Rubin et al. 2003) including possible viewpoint transformation

(Burgess 2006); subjective time - sense of the self in the context of the timeline against which the events in our lives play out (Tulving 2002b) or chronesthesia (Tulving 2002b); self - processes reliant on or affecting the concept of oneself (Conway & Pleydell-Pearce 2000) and thereby having a direct connection to or influence over our self-perception (Tulving 2002b); auto-noetic consciousness - a special kind of consciousness that accompanies the act of remembering, allowing one to be aware of the self in subjective time (Tulving 2002b); narrative – a story structure formed by the unfolding of a sequence of events (Rubin et al. 2003); familiarity – a subjective feeling or judgment of oldness (Wagner et al. 2005); visual imagery - complex visual imagery, involving the composition of simple imagery such as single objects (Rubin et al. 2003); prospective planning – making plans about how to achieve future goals (Simons & Spiers 2003); task monitoring – executive processes overseeing successful constructive task performance (Simons & Spiers 2003). From (Hassabis & Maguire 2007).

The interrelatedness of these cognitive functions and their reliance on the hippocampus is also evidenced by neuropsychological studies of patients with hippocampal damage. They confirm that the hippocampus is crucial not only for episodic memory (Scoville & Milner 1957; Spiers et al. 2001) and navigation (Maguire et al. 2006a) but also thinking about one's personal future (Atance & O'Neill 2001; Klein et al. 2002; Rosenbaum et al. 2005; Tulving 1985) and imagination as shown in Chapter 4. I propose that their common reliance on the hippocampus is due to the critical integrational role the hippocampus plays in scene construction (although it may also perform other additional mnemonic roles in episodic memory – see Section 8.9) and that hippocampal patients would likely be impaired on any task that required a rich 'recollection-like' experience.

Alternatives to scene construction have been proposed (Buckner & Carroll 2007; Schacter & Addis 2007; Schacter et al. 2008). The process of 'self-projection' (Buckner & Carroll 2007) defined as 'the shifting of the self to alternative perspectives in time or space' has been posited, like scene construction (Hassabis & Maguire 2007), as an underlying process common to a set of disparate cognitive functions including episodic memory recall, EFT and navigation. However, the findings from Experiment 2 (see Chapter 5) suggest that self-projection is in fact a multi-faceted concept that can be broken down into at least two distinct components with dissociable neural bases: a network centred on the hippocampus responsible for scene construction, with the amPFC/PCC and precuneus mediating self projection in time, sense of familiarity and self-schema. The disparate cognitive functions highlighted in Figure 36 may be differentially reliant on these two components (e.g. navigation likely predominantly relies

on the scene construction process, while episodic memory requires both). Therefore whilst self-projection is clearly an important concept, it conflates several distinct component processes including scene construction (Hassabis & Maguire 2007) and theory of mind (Amodio & Frith 2006). For the purposes of teasing apart the various component processes underpinning episodic memory it is clearly advantageous to consider constituent processes in as reduced a form as possible.

Taken together then, I believe the core network is most accurately characterised as the construction network. It is invoked whenever attention is directed away from the current external situation and instead focused inwards towards a rich internal (re)constructed representation of an event or scene, real or imagined. Processes such as theory of mind are only engaged if required, i.e. in the case of EFT or episodic memory recall but not necessarily in imagination or navigation. This may explain why the construction network has a similar pattern of activity to that associated with the default network (Buckner et al. 2008; Raichle et al. 2001) and mind-wandering (Mason et al. 2007), cognitive functions that involve minimal external stimuli combined with introspection and rich internal imagery. These constructed scenes or events, created and maintained by the construction network, can then be manipulated further by other processes, such as theory of mind, to allow shifting of the self to alternative perspectives in space or subjective time (Arzy et al. 2008; Buckner & Carroll 2007).

## **8.5 Other component processes of episodic memory**

Scene construction can, therefore, account for a good deal of the brain network consistently activated by episodic memory. Experiment 2 (see Chapter 5) not only allowed the commonalities across different types of tasks to be examined using a conjunction analysis, but also permitted direct comparison of the recall of recent real episodic memories with the recall of previously constructed fictitious experiences that were well-matched for difficulty, age, detail and vividness. This meant that the effects of scene construction could be factored out, thus revealing which brain regions might be responsible for other key aspects of the recollective experience, namely subjective sense of time, connection to self, and auto-noetic awareness (Tulving 2002b; Wheeler et al. 1997). The anterior medial prefrontal cortex (amPFC, BA 10) and regions of the posterior medial parietal cortex, including posterior cingulate cortex and precuneus, were

the only regions found to be preferentially engaged during the recall of real episodic memories as opposed to imaginary memories (see Figure 25).

This pattern of activation bears a striking resemblance to networks found to support self-reflection (Johnson et al. 2002), theory of mind (Amodio & Frith 2006; Kumaran & Maguire 2005) and episodic future thinking (Addis et al. 2007). The amPFC in particular has been heavily implicated in mentalising (Frith & Frith 2003) and self-perception processes (Conway & Pleydell-Pearce 2000; Gallagher 2000; Johnson et al. 2002), and has been predicted by Tulving (Tulving 2002b; Wheeler et al. 1997) as an area important for autothetic consciousness and mental time travel on the basis of findings from patients with frontal lesions (Wheeler et al. 1997). By contrast, activation of posterior parietal regions has previously been observed in studies of recognition memory, responding preferentially to old or familiar items (Vincent et al. 2006; Wagner et al. 2005), suggesting that real memories are in some sense more familiar than imagined fictitious memories, which is perhaps not surprising given their often highly familiar content. Therefore I suggested that during recall of real episodic memories the interaction or cooperation between these self-relevant and familiarity functions might be sufficient to give rise to the phenomenological feeling of whether a re-experienced event actually happened to one in the past or not (Tulving 2002b), or indeed whether it could plausibly happen to one in the future (Addis et al. 2007; Atance & O'Neill 2001). It is interesting to speculate whether this process might have malfunctioned in patients with schizophrenia (Simons et al. 2006) and in those who confabulate (Burgess & Shallice 1996).

These findings connect with those from a recent fMRI study (Addis et al. 2007) that explored episodic future thinking (Atance & O'Neill 2001). Addis et al. (2007) found comprehensive overlap between the brain networks supporting recall of past memories and thinking about plausible personal future events. Indeed, if considered from a process point of view, then all the processes highlighted so far (see Figure 36), including scene construction and the self, are present in both episodic memory recall and episodic future thinking. Therefore, one would predict the more or less complete overlap in brain networks that was indeed found by Addis et al. (2007). We suspect that the reason a study by Szpunar et al. (2007), which also examined episodic future thinking, did not observe some of the brain regions involved in scene construction that I highlight here is

because their baseline task (imagining Bill Clinton in various situations) would also have activated regions involved in scene construction, making it difficult for this factor to be examined. This is also true of other studies where scene construction might have been involved, but the process was not formally considered, controlled or manipulated (e.g. Ref. (Saxe & Kanwisher 2003)). This serves to highlight why using episodic future thinking as a comparison task may not prove to be fruitful as it is composed of practically the same component processes as episodic memory recall (Hassabis & Maguire 2007).

More work examining these brain areas and their roles in supporting the 'selfness' and 'realness' of memories and constructions is clearly required and beginning to emerge (Abraham et al. 2008). We investigated these issues further in a recent fMRI study (Summerfield et al. 2009) that required participants to recall self (autobiographical) and non-self (movie or news) events that were either real or previously imagined. The results confirmed that anterior and ventral medial prefrontal regions were primarily modulated by connection to the self and that parts of the posterior cingulate and posterior parietal cortex were primarily modulated by realness (Summerfield et al. 2009).

It should be noted that while I have proposed that intact self-reflection and theory of mind processes are also recruited in addition to scene construction to support episodic memory, this does not imply that the episodic memory system as a whole is required for the operation of any individual component process. This erroneous logic was applied recently in a study showing, not surprisingly, that a patient with amnesia retained intact theory of mind abilities (Rosenbaum et al. 2007).

## **8.6 Downgrading the importance of time**

In all previous studies and discussions about prospective memory or thinking about the future (Buckner & Carroll 2007; Gilbert & Wilson 2007), it is assumed that imagination necessarily involves mental time travel involving the self into the future, and is a form of episodic future thinking (Atance & O'Neill 2001). By contrast, this assumption was not made in the patient imagination study in Chapter 4, and fMRI study reported in Chapter 5. In both cases, I maintain that it is possible to richly imagine and visualise a new fictitious experience that is not explicitly temporal in nature, and that is not necessarily self-relevant, plausible or even possible.

In the classic taxonomy of episodic memory (Tulving 1972; Tulving 2002b), what, where and when, have been traditionally regarded as equally important properties of an episodic memory (Tulving 1983; Tulving 2002b). Indeed, a memory of an event is only defined as truly episodic if one remembers when and where it happened in addition to what happened (Tulving 2002b). The rich recollective experience that accompanies the recall of an episodic memory (Tulving 1983; Tulving 2002b) has content and a spatial context, the what and the where – the central contribution of the scene construction process to the provision of this coherent spatial context has been discussed. But what role does time have, and is it really as important to the recollective experience as the other two aspects of an episodic memory?

Much has been written about time (Clayton et al. 2003; Coull et al. 2004; Friedman 1993; Tulving 2002b) and there are at least two types of time that have relevance to the recollective experience of episodic memory. There is the moment by moment order in which an event or sequence unfolds (Ferbinteanu et al. 2006) or 'micro-time'. Then there is subjective time (Tulving 2002a; Tulving 2002b), the timeline against which the events in our lives play out, the conscious awareness of which has been dubbed chronesthesia (Tulving 2002a), and which can be thought of as 'macro-time'. Micro-time, then, is an intrinsic property of an episodic memory that when recalled ensures it is played back in the same (or reverse (Foster & Wilson 2006)) temporal sequence in which it was recorded (Eichenbaum 2004). It is related to the physical laws of nature, for example, one must open a door before going through it, and it might be reliant on the special anatomical properties of the hippocampus (Eichenbaum 2004; McClelland et al. 1995).

By contrast, macro-time is a concept that, unlike micro-time (Ferbinteanu et al. 2006), perhaps only humans possess (Suddendorf & Busby 2003; Tulving 2002b). But should it be elevated to the status of a special sense or process as some have suggested (Buckner & Carroll 2007; Tulving 2002a), and is it necessary for the full recollective experience of an episodic memory? Recent empirical evidence does not support the idea that chronesthesia is an independent process with a distinct neural signature. Near total overlap has been found in brain regions supporting past episodic memory recall and episodic future thinking (Addis et al. 2007; Szpunar et al. 2007). Besides the subjective sense of past/future as described by Tulving (Tulving 1983), macro-time has

also been commonly interpreted to mean that there are specific time-tags associated with each episodic memory. Interestingly, studies testing memories from distinct time periods have failed to show consistent neural correlations in response to macro-time modulation (Addis et al. 2007; Gilboa et al. 2004; Maguire 2001a; Maguire et al. 2001a; Moscovitch et al. 2005; Nadel et al. 2000; Steinworth et al. 2006), and time is also a poor retrieval cue for episodic memory (Jones 1976). Finally, if the age of a memory has some intrinsic neural basis then it seems plausible it might be related to the age of its memory trace (Marr 1971; Moscovitch et al. 2005). However, in light of current ideas regarding reconsolidation (Dudai 2006; Nadel & Land 2000; Nader et al. 2000), it is not clear whether one can meaningfully talk about the age of a memory because the act of recalling it refreshes the trace anew (Dudai 2006; Hupbach et al. 2007).

Thus, although macro or subjective time is frequently part of a true episodic memory it is not clear whether it can be meaningfully separated from ideas of the self and auto-noetic consciousness (Tulving 2002a; Tulving 2002b). This is in contrast to micro-time, which appears to be an intrinsic property of an episodic memory (Eichenbaum 2004; Foster & Wilson 2006). By this view then, the timestamp of an event simply becomes another piece of semantic knowledge that might or might not be retrieved or logically deduced at the point the event is recalled, perhaps not too dissimilar to the name of the place it occurred in or the names of any people involved (Hassabis & Maguire 2007).

## **8.7 Evolutionary function of construction**

If constructive processes (i.e. the ability to put informational components together in novel ways in the service of a goal) underlie episodic memory and many other related high-level cognitive functions (see Figure 36), what, if any, evolutionary advantages does this confer (Tulving 2002b)? From a computational perspective, reconstructing a memory from its components is more efficient in terms of storage capacity than the alternative of storing each memory separately as an intact record (Schacter 2001; Schacter & Addis 2007). Furthermore, this kind of storage structure lends itself conveniently to making abstraction and generalisation inferences across distinct experiences (McClelland et al. 1995).

One can speculate as to whether episodic memory evolved as the ultimate expression of these underlying constructive processes (Tulving 2002b) or whether these processes developed for other purposes (e.g. spatial navigation) and were then later co-opted for by episodic memory (Tulving 2002b). Either way it seems clear that the resultant flexible system is useful for survival (Buckner & Carroll 2007; Schacter & Addis 2007). Consider an organism that, in their present situation, is confronted by several choices of what to do next. Being able to accurately and richly mentally simulate (Buckner & Carroll 2007) or construct what those possible future states might be like, before making the decision, would aid both the evaluation of the desirability of those outcomes and the planning processes needed to make them happen. The applications of construction go beyond just planning for and anticipating possible future events (Atance & O'Neill 2001; Buckner & Carroll 2007; Emery & Clayton 2004; Schacter & Addis 2007). It forms the basis of imagination and possibly creativity, where constructions are envisaged that are not directly related to the future or the past, or to prediction per se, but for general problem-solving and invention purposes such as tool manufacture or art.

Viewed in this way imagination simply becomes the flipside of memory where the goal now is to create something unfamiliar and new, rather than familiar, from the same set of building blocks. Information about the world we live in, accumulated through multiple experiences, is stored as components in the sensory cortices of the brain. As I have already discussed, when we richly recall an autobiographical memory what we are really doing is reconstructing this memory from its fragmentary components. The building blocks of a memory are reassembled with the goal of transiently recreating a whole that seems coherent, familiar and self-relevant.

By contrast, the creative process attempts to connect these components together in new unfamiliar ways, the more creative the person the more unusual and lateral the connections they can make. But this is not enough; to be a Mozart, a Picasso or a Da Vinci one must also have good aesthetic judgment. It is not sufficient to be able to just connect things in new ways. These connections must also make sense and be elegant in the context of the task domain e.g. aesthetic beauty in art, consumer preferences in business. One can speculate that this is potentially where the brain's emotional system is required (Gilbert & Wilson 2007; Sharot et al. 2007), acting as an evaluation function filtering out the bad or ugly new connections from the good. Great artists and creative

people are usually excellent lateral thinkers with allied innate taste. So the creative process can be summarised in three steps: (i) the acquisition of components through exposure to experiences - nothing is created in a vacuum, (ii) making novel connections between these components, and (iii) selecting the best fitting connections from amongst them. This allows humans to be limitlessly creative and inventive even though constrained by a basic set of raw component elements gleaned over a lifetime of experiences.

It has been suggested that the episodic memory system is unique to humans (Suddendorf & Busby 2003; Tulving 2002b). This might be true for properties such as autothetic consciousness and self-awareness that are difficult to categorically test for in animals given the absence of language (Emery & Clayton 2004). However, there is considerable evidence that certain animals do possess at least some of these constructive episodic capabilities (Buckner & Carroll 2007; Clayton et al. 2003; Emery & Clayton 2004). This has been demonstrated most persuasively in scrubjays, a member of the corvid family of birds (Emery & Clayton 2004). They flexibly cache food for future consumption (Clayton et al. 2003; Raby et al. 2007) and display impressive 'causal reasoning, imagination and prospection abilities' (Emery & Clayton 2004). It might be that some animals have scene construction capabilities to help them formulate plans for the future but they lack some of the introspective add-ons (see Section 8.5) that give true episodic memory their distinctive human quality (Tulving 2002b). The complexity of the constructive episodic memory system and its underlying processes might be the key to its flexibility (Tulving 2002b). It could also explain its vulnerability (see main text) and why it is relatively late developing, suggested by some to only become fully operational in humans at the age of about four years (Tulving 2002b).

## **8.8 Interim summary**

The process of 'scene construction' (see Chapter 5) involves the mental generation and maintenance of a complex and coherent scene or event (Hassabis and Maguire 2007). The scene construction network highlighted in Chapter 5 (see Figure 24) and Section 8.3 supports the construction system of the brain allowing for the internal rehearsal of events or scenes (Schacter et al. 2008). Scene construction underpins the process of creating a setting in which a simulated event can unfold whether past, present, future, atemporal or

hypothetical. As such it may be involved not only in episodic memory and imagination (see Chapters 4 and 5) but also in many other cognitive functions including future thinking, navigation and daydreaming (see Figure 36). Undoubtedly we still have a long way to go to understand exactly how scenes and events are constructed, and the precise role of each brain area in the system. Nevertheless it is clear that the ability to pre-experience hypothetical events confers an evolutionary advantage in planning for the future and imagining alternative scenarios (Hassabis & Maguire 2007; Schacter et al. 2008).

## **8.9 The dual function theory**

I believe that the results from this thesis and other recent work (Buckner & Carroll 2007; Hassabis & Maguire 2007; Schacter & Addis 2007) potentially prompt a rethink about the function of the hippocampus in episodic memory recall. In this concluding section I tentatively outline a preliminary framework for a new theory of hippocampal function in episodic memory, the 'Dual Function Theory' (DFT), centred on the premise that the hippocampus performs two distinct and potentially separable roles in support of the recollective process. Namely, its traditionally understood mnemonic role involving the reactivation of cortical representations by a memory indexing mechanism, and a second non-mnemonic role involving the dynamic integration of active cortical representations into a spatially coherent whole. This second non-mnemonic role for the hippocampus is critical in supporting the process of scene construction (see Chapters 4 and 5) and when I mention the term scene construction below in the context of hippocampal function I use it as a short-hand for this non-mnemonic integration role. In light of this theory, I also review some of the more controversial results in the episodic memory field and show how seemingly opposing conclusions can be potentially reconciled by this view.

### **8.9.1 Comparisons to the standard model and multiple trace theory**

As summarised in Section 2.3, the standard model of memory (Cohen & Squire 1980; Squire 1992; Squire et al. 2004) states that the MTL is primarily concerned with memory. During episodic memory recall a simple representation or memory index in the hippocampus (Marr 1971) reinstatiates the full representation of the memory in the neocortex. Later, through the process of consolidation, connections between these

neocortical representations are progressively strengthened until the memory can be supported and accessed independently of the hippocampus. This consolidation process is held to explain the phenomenon of temporally graded retrograde amnesia (McClelland et al. 1995; Squire & Alvarez 1995; Squire et al. 2004).

Experiment 1, the patient study on imagination (see Chapter 4), provided evidence that the hippocampus might be directly involved in a purely constructive process thus posing a problem for the standard model view that the MTL/hippocampus is solely involved in memory (Squire et al. 2004). This was confirmed by a follow-up imaging study in healthy subjects showing that the hippocampus was indeed recruited when imagining new experiences for the first time in the scanner (see Chapter 5), and a conjunction analysis from the same study demonstrated that the hippocampus formed a critical part of the brain network responsible for supporting the process of scene construction.

Controversial evidence has been found recently for perceptual deficits following lesions of various subregions of the MTL (Gaffan 2002; Graham & Gaffan 2005; Murray & Bussey 2001). Specifically it has been claimed that the hippocampus may have a crucial role in higher level perception (i.e. of higher order complexity than an object), for example for spatial scenes (Bird & Burgess 2008; Bussey & Saksida 2007; Lee et al. 2005a; Lee et al. 2005b). However, if viewed within the context of the DFT framework presented here, then this higher level perceptual deficit could instead be expressed as a more general deficit in the process of scene construction, the only extension being that the components that make up the spatial scene are externally generated via sensory perception rather than internally generated via imagination. In both cases, the dynamic integration of the disparate elements that make up a spatial scene into a coherent whole, a process critically reliant on the fast binding capabilities of the hippocampus, is fundamentally required for the later generation, manipulation and perhaps even 'perception' of a holistic scene. These downstream scene manipulation and perception processes are perhaps mediated by other regions in the scene construction/perception network such as prefrontal, retrosplenial and parietal cortices, supporting functions such as viewpoint transformation and the directing of attention (Burgess 2002; Guariglia et al. 1993).

The multiple trace theory (see Section 2.4) (Moscovitch et al. 2006; Moscovitch et al. 2005; Nadel & Moscovitch 1997) argues that retrograde amnesia following lesions of the MTL differentially affects episodic and semantic memory, with the deficit observed in episodic memory not temporally graded (Squire et al. 2004) but global in nature. There is compelling empirical evidence from the majority of neuroimaging studies of episodic memory recall e.g. (Steinvorth et al. 2006)(but see: (Maguire 2001a)) that it engages the hippocampus irrespective of the age of the memory, with the recent suggestion that the richness or detail level of the recollection may modulate hippocampal engagement (Addis et al. 2004; Gilboa et al. 2004; Moscovitch et al. 2005). Superficially, therefore, the present findings from the imagination studies (see Chapters 4 and 5) would appear to be supportive of these notions.

However, the central tenet of the multiple trace theory is that each time an old memory is retrieved, a new hippocampally-mediated trace is created so that old memories are represented by more or stronger HC/MTL-neocortical traces than new memories (Moscovitch et al. 2006; Moscovitch et al. 2005; Nadel & Moscovitch 1997). It is hard to conceive of a process based on the concept of multiple traces that could readily account for constructive or even reconstructive processes (i.e. imagination and episodic memory recall respectively). It is interesting to note that even the original proponents recently admitted that the theory would need modification to take recent findings, such as the importance of richness (Gilboa et al. 2004), into account (Moscovitch et al. 2005).

Moreover, there is little direct empirical evidence for the multiple trace theory from several neuroimaging studies that have systematically varied the age of the recalled episodic memories (Addis et al. 2004; Maguire & Frith 2003; Moscovitch et al. 2005; Steinvorth et al. 2006). These have failed to show any consistent modulation of hippocampal activity when other factors such as richness were controlled. The original formulation of the multiple trace theory predicted that a reverse gradient might have been observed in such studies (Nadel & Moscovitch 1997) with older memories perhaps eliciting greater hippocampal activity due to the multiple traces. The results obtained in my multivariate pattern classification investigation into individual episodic memories (described in Chapter 7) could also be interpreted as evidence against the idea of multiple traces, as the paradigm necessarily required the participants to repeatedly recall the same memory many times in order for there to be enough training examples for the

classification procedure. However, if the memory trace or set of traces had been continually changing or added to as implied by the multiple trace theory, then the classifier would not have been able to identify a consistent pattern of activity in hippocampal voxels on which to accurately predict the identity of the memory being recalled (see Section 3.6).

So whilst wholeheartedly agreeing with all of the criticisms of the standard model articulated by the multiple trace theory (Nadel & Moscovitch 1997) I would like to respectfully suggest that DFT potentially provides a more parsimonious explanation for these discrepancies than does the notion of multiple traces. The idea that the hippocampus is perhaps performing two distinct roles in support of episodic recollection, one mnemonic (i.e. its traditional memory index role) and one non-mnemonic (i.e. its dynamic integrational role that underpins the process of scene construction – see Section 8.3), is potentially a helpful one from an interpretative point of view and can account for some of the seemingly conflicting findings in the memory literature. Firstly, it accounts for the involvement of the hippocampus in complex constructive tasks such as imagination, that are not purely mnemonic (see (Squire et al. 2004)). Secondly, it readily explains why richness of recollection appears to strongly modulate hippocampal involvement (Gilboa et al. 2004) as this is likely to be highly correlated with the coherence and complexity of the constructed scene underpinning the recollected memory. Thirdly, it explains why full episodic memory recollection, as measured by a more in-depth scoring system (Bayley et al. 2003; Levine et al. 2002; Moscovitch et al. 2005), is always dependent on the hippocampus if it is to be truly relived even if the original memory index has long since been consolidated to neocortex (Squire et al. 2004). Fourthly it explains why the fate of semantic memory differs from episodic memory in hippocampal amnesia, as by definition retrieval of semantic memory does not require rich reliving of an episodic event and therefore does not need to engage scene construction processes. Thus once consolidated, semantic memory becomes hippocampal-independent (Squire et al. 2004). Finally, the idea that a constructive process such as scene construction might be an important component process underpinning episodic recollection relates to psychological notions of memory being a reconstructive process (Bartlett 1932; Schacter et al. 1998) to cognitive neuroscience theories of hippocampal function.

### **8.9.2 Evidence from patients**

The dual function theory is therefore able to reconcile some of the seemingly opposing views in the hippocampal amnesic literature. The finding of spared remote episodic memory in hippocampal patients (Bayley et al. 2005; Bayley et al. 2003; Squire et al. 2004) really comes down to what qualifies as 'spared function'. Tulving considered rich recollection to be a 'defining characteristic of episodic memory' and thus the act of reliving should feature highly in any definition of spared episodic memory function (Tulving). However, most of the simple probing protocols that have classically been used in the majority of patient studies to date (Kopelman et al. 1989) do not have the requisite granularity to satisfactorily quantify this critical aspect of a recalled memory. Instead they often consist of a binary decision based on the provision of just a single semantic fact relating to the episodic memory in question, which is more akin to a test of semantic memory (Bayley et al. 2003; Kopelman et al. 1989; Squire et al. 2004). Numerous studies from multiple groups (Moscovitch et al. 2005; Squire et al. 2004; Vargha-Khadem et al. 1997) have reached a broad consensus that remote semantic memory is spared in hippocampal amnesia so perhaps it is not surprising under the coarsely grained conditions of classic scoring systems (Kopelman et al. 1989) that remote episodic memory may erroneously be reported to be intact. Indeed when more sophisticated scoring systems and testing protocols (see Section 4.2) (Levine et al. 2002; Moscovitch et al. 2005), specifically designed to quantify the richness of the recollected memory, are employed it appears that hippocampal amnesics do in fact have impaired remote episodic memory (Gilboa et al. 2004; Moscovitch et al. 2005; Steinworth et al. 2005). DFT accounts for this because even if the original memory index for an episodic memory has been consolidated to neocortex (Squire et al. 2004), thus eschewing the need for the traditional mnemonic role of the hippocampus, the hippocampus will still be required if the memory is relived in a truly episodic manner due to its non-mnemonic integrational role that underpins the process of scene construction (see Section 8.3).

If the key distinction, as advanced here, between episodic memory and semantic memory in terms of the processes involved is considered to be the rich recollective experience as underpinned by scene construction, then this could potentially account for other dichotomies drawn between different types of episodic memory (Arzy et al. 2008; Klein et al. 2002; Rosenbaum et al. 2005). For example two hippocampal patients DB

(Klein et al. 2002) and KC (Rosenbaum et al. 2005) were separately tested on future thinking tasks, both personal and non-personal (public), and were reported to only be impaired on personal future tasks thus leading to the conclusion that the personal/non-personal distinction was somehow crucially dependent on the hippocampus (Arzy et al. 2008; Klein et al. 2002; Rosenbaum et al. 2005)(also see (Buckner & Carroll 2007)). However, closer inspection of the paradigms employed revealed that the probes used for non-personal future questions only required semantic-like answers for satisfactory performance without the need for active reliving, whereas the personal future probes were different in character and generally did require reliving. A typical example of a non-personal future probe used in the Klein et al (2002) study was “can you tell me what you think will be the most important issues facing the environment in the next 10 years?”. The answer given by DB which was deemed satisfactory was “the threat that weather and rainfall patterns are going to change because of industrial pollution”, a perfectly adequate answer but one that is clearly semantic and factual in nature and not one that requires the use of recollection supported by scene construction processes. By contrast in the systematic patient group study (Chapter 4), I found that hippocampal patients were impaired not only on imagined scenarios but on future events whether personal or not (see Chapter 4) as long as the probes elicited truly episodic-like answers.

A similar distinction can also be made in the spatial memory domain. In a single case study the navigational abilities of hippocampal patient EP were tested (Teng & Squire 1999) and findings appeared to demonstrate that remote spatial memory for his hometown was intact, the implication being that it had been fully consolidated and was thus hippocampal-independent. However, in another study, where the navigation ability of hippocampal patient TT, a London taxi driver and expert navigator, was tested revealed a more complex pattern of performance (Maguire et al. 2006a). He was tested during navigation around a virtual reality version of central London. Intriguingly this revealed that TT’s navigational performance was spared on routes that only required the use of major roads (called ‘A’ roads in the UK) but was impaired when routes required the use of minor roads. The DFT view would interpret this pattern of performance as due to the likely consolidation or semanticisation of the major road routes by virtue of their high frequency of usage (Maguire et al. 2006a) whereas routes that involved less well-known and used minor roads may not have been as heavily schematised thus requiring fully operational scene construction processes to allow for successful navigation.

Therefore when TT and EP were tested on routes extremely well-known to them, other, possibly frontal or striatal-mediated strategies might have been adopted that did not require the process of scene construction and visualisation and thus their consolidated remote spatial memories were sufficient for successful performance (Doeller et al. 2008).

Note that the multiple trace theory would also predict the pattern of spatial memory impairment seen in patient TT (Maguire et al 2006a) but for different reasons to DFT (Moscovitch et al. 2005). The multiple trace theory would ascribe this deficit in navigating less well-known routes to a failure to retrieve specific episodic memories of those routes (Moscovitch et al. 2005). However, it seems highly implausible that veteran taxi drivers, such as TT, with many decades of driving experience and thus many thousands of overlapping episodic route memories would use the explicit recall of a single specific episodic memory in order to navigate even a less well-known route. Furthermore, as part of their job requirement they would from time to time inevitably be asked to navigate to a novel destination they had never experienced before. Therefore the DFT view that less well-known (or new) routes simply need to be dynamically and coherently visualised, a function that requires the process of scene construction and hence the integration abilities of the hippocampus (see Chapter 4), would seem to be a more parsimonious explanation for the pattern of impairment seen in patient TT (Maguire et al. 2006a).

In the context of DFT it is also worth revisiting the performance of the outlier patient KN in the imagination task (see Chapter 4). KN was the most densely amnesic for recent memories of all the patients and yet performed better than his controls when asked to imagine new experiences. This was attributed to the sparing of some function in the residual tissue of right hippocampus (see Section 4.3.3 and Figure 20). The performance of KN, a patient with damage limited to the hippocampus (see Section 4.2.2), potentially provides a very important dissociation between episodic memory function and construction function thus supporting the notion advanced by DFT that the hippocampus might be performing two separable roles in support of episodic memory only one of which, the mnemonic role, has been damaged in KN. Of course caution must always be applied when making any inferences on the basis of a single case, but it would be potentially interesting to investigate this finding further perhaps in patients with unilateral hippocampal lesions to see if a more patients can be uncovered with this type of dissociation.

### **8.9.3 Evidence from neuroimaging**

As discussed earlier, recent neuroimaging studies of episodic memory recall in healthy subjects have shown correlations between hippocampal activation and the richness and detail level of the recollective experience (Addis et al. 2004; Gilboa et al. 2004). Hippocampal involvement in episodic memory recall has not been shown to be robustly or consistently modulated by any other factor including the age of the memory (Maguire & Frith 2003; Steinvorth et al. 2006), temporal orientation to the past or future (Addis et al. 2007), realness of the event or degree of connectedness to the self (Summerfield et al. 2009)(but see (Maguire & Mummery 1999)). Moreover in the studies described in Chapters 4 and 5 I have demonstrated the critical involvement of the hippocampus in imagining new experiences and in underpinning the process of scene construction. These findings dovetail nicely with the notion that this kind of scene construction process might critically support the richness of episodic recollection (Gilboa et al. 2004) and potentially provides a unifying account for the involvement of the hippocampus in the large set of seemingly disparate, and not explicitly mnemonic, tasks including navigation, episodic future thinking and the default network (Hassabis & Maguire 2007).

Furthermore it has been suggested that episodic recollection can be temporally split into at least two distinct phases, a reinstatement phase and a visualisation phase (also called construction and elaboration phases respectively (Addis et al. 2007)), lasting around 5-8s each (Addis et al. 2007). The hippocampus has been shown to be active during both phases for both past and future episodic thinking (Addis et al. 2007). This sort of timeline and hippocampal activation is mirrored in the study described in Chapter 7 with separate early and late onsets of the hippocampus. Within the DFT framework, hippocampal activity in the two phases of recollection reflects the performance of differing functions. Initially, during the reinstatement phase, the hippocampal index or memory trace reactivates the component representations in the cortical regions (Marr 1971; Squire et al. 2004). This is perhaps done iteratively and directed by the prefrontal cortex with gradually the correct set of active cortical representations being assembled (Rugg & Wilding 2000). Once the active set has stabilised, full reliving of the memory can begin usually involving rich visualisation of a replay of the past event. Hippocampal involvement at this stage is now performing the online integration function binding the

active elements of the memory together so it can be relived as a spatially coherent whole. It should be noted here that for the recall of remote memory if the standard model view of consolidation is correct (Squire et al. 2004), DFT would predict hippocampal activation only for this latter stage.

The reprojection data from Chapter 7 (see Figure 33) suggest an anatomical segregation within the hippocampus for these two distinct functions with the body of the hippocampus supporting the early onset indexing function and the posterior hippocampus supporting the late onset spatial integration function. Anatomically this makes sense as the posterior hippocampus has a high degree of connectivity with the retrosplenial cortex (Maguire 2001b), a region thought to be important in navigation and viewpoint transformation (Burgess 2006; Burgess et al. 2001; Maguire 2001a) and has also been implicated in animal studies of spatial memory (Colombo et al. 1998; Moser & Moser 1998). Structural neuroimaging studies in expert human navigators have shown increases in grey matter in the posterior hippocampus seemingly at the cost of a decrease in grey matter in the anterior hippocampus, both correlating with degree of navigational experience accrued (Maguire et al. 2006b). Further work will be needed ascertain exactly how these coarse anterior-posterior distinctions map to specific subregions of the hippocampus (Andersen et al. 2007) to further unravel the exact temporal dynamics of the recollective process.

#### **8.9.4 Compatibility with the cognitive map theory**

The ideas put forward here are most compatible with the views expressed in the cognitive map theory (see Section 2.5)(Burgess 2002; Burgess et al. 2001; O'Keefe & Nadel 1978). However, it may prove to be useful to draw an explicit distinction between spatial memory and cognitive maps on the one hand, and the dynamic ability to bind incoming cortical information into a spatially coherent whole for non-mnemonic purposes such as scene construction. This distinction is much harder to draw in animals, with the behavioural correlates of such constructive processes unclear and indeed the very existence of functions such as episodic recollection questioned (Tulving 2002b) (but see (Clayton et al. 2003; Suddendorf & Busby 2003)). In fact, it is not clear how dynamic constructive processes (Schacter et al. 1998) can be explained at all by the cognitive map theory in its current formulation. Given that the cognitive map theory was borne out

of observations from animal studies it is interesting to consider whether rats have this online integration ability and scene construction function. A recent study has provided evidence for planning and prospection abilities in rats, where they were seemingly capable of imagining future locations and decisions prior to reaching them (Johnson & Redish 2007).

It is possible that scene construction may prove to be a specific example of a more general associative construction process (Cohen & Eichenbaum 1993) and more experiments will be needed to clarify this. However, it is worth noting that the patients in the imagination study (Chapter 4) did have a specifically spatial bias to their deficit. Also in an imaging study involving two well-matched spatial and non-spatial complex associational tasks hippocampal activity was only found for the spatial associational task (Kumaran & Maguire 2005). So it does seem as if there may be something special about space over and above its associational properties (Cohen & Eichenbaum 1993; O'Keefe & Nadel 1978).

### **8.9.5 Conclusion**

In summary then, according to Tulving's definition (Tulving 1983), the key properties of episodic memory (Tulving 1983) include mental time travel, the self, auto-noetic consciousness, specificity in time and space, and an accompanying rich recollective experience (Tulving 2002b). In this thesis I argue that rich recollection should be considered as *the* defining characteristic of episodic memory that marks it out as unique and moreover is underpinned by the process of scene construction which is critically reliant on the fast binding capabilities of the hippocampus (see Chapter 4). Consequently, the hippocampus is required for true episodic recollection irrespective of the age of the memory and even if its memory index has been fully consolidated to neocortex (Squire et al. 2004).

This observation forms the crux of the proposed dual function theory of the hippocampal role in episodic recollection. The hippocampus has long-been known (Scoville & Milner 1957) to be fundamental in the declarative memory system (Squire et al. 2004). It is also widely accepted that hippocampal memory traces (or indexes or representations) can over time be consolidated to neocortex so that they are no longer hippocampal

dependent (see (Squire et al. 2004)). However, in addition to this well-established mnemonic role we also ascribe a second potentially dissociable dynamic integration function to the hippocampus. This facilitates the binding of online cortical information into a spatially coherent whole, a process that underpins scene construction and is key to the rich recollective experience that accompanies episodic memory recall.

These two hippocampal roles, mnemonic and online spatial integration, are separable and dissociable, and can be engaged individually or in conjunction by different tasks, with the mnemonic role for declarative memory possibly time-dependent following consolidation (Squire et al. 2004). For example, the following tasks all require the dynamic integration function: rich reliving of a remote episodic memory with a fully consolidated memory index, imagination of new experiences, and perhaps even scene perception. Whereas purely semantic tasks such as word-pair recall that do not require reliving only engage the mnemonic function. However, the re-experiencing of a recent memory or navigating in a recently learned location would engage both hippocampal functions.

Thus, humans with lesions to the hippocampus are likely to be impaired on any task that requires either the mnemonic *or* integration hippocampal functions *or* both. It is the loss of these two functions, not other task properties advanced by some such as past/future, personal/non-personal, episodic/semantic, recent/remote distinctions, that best characterises patient deficits across a diverse range of tasks. Perhaps the possibility that there may be two separate functions performed by the hippocampus in support of episodic recollection, with the integration function hitherto largely overlooked by current hippocampal theories, could explain the confusion surrounding the status of episodic memory across hippocampal patients.

I believe that the dual function theory best accounts for the available data on episodic memory, and offers a promising avenue for further experimental exploration. Clearly much more work will be needed to address outstanding questions concerning issues such as hippocampal sub-region specificity and lateralisation, temporal dynamics of the involvement of the hippocampus in the retrieval process (see Chapter 7), spatial versus non-spatial integration, and under what circumstances does integration lead to encoding. By identifying hippocampal processes such as dynamic integration, that underlies scene

construction, and separating them out from traditional mnemonic processes, we may come closer to the goal of accurately characterising what makes episodic memory the special kind of memory that Tulving so presciently argued for more than three decades ago (Tulving 1972; Tulving 1983; Tulving 2002b).

## References

- Abraham, A., von Cramon, D. Y. & Schubotz, R. I. 2008 Meeting George Bush versus meeting Cinderella: the neural response when telling apart what is real from what is fictional in the context of our reality. *J Cogn Neurosci* **20**, 965-76.
- Addis, D. R., Moscovitch, M., Crawley, A. P. & McAndrews, M. P. 2004 Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus* **14**, 752-62.
- 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**, 1363-77.
- Aggleton, J. P. & Brown, M. W. 1999 Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav Brain Sci* **22**, 425-44; discussion 444-89.
- Aggleton, J. P., 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* **43**, 1810-23.
- Aho, A. V., Ullman, J. D. & Hopcroft, J. E. 1983 *Data Structures and Algorithms*: Addison-Wesley.
- Allwein, E., Schapire, R. & Singer, Y. 2001 Reducing multiclass to binary: a unifying approach for margin classifiers. *Journal of Machine Learning Research* **1**, 113-141.
- Alvarez, P. & Squire, L. R. 1994 Memory consolidation and the medial temporal lobe: a simple network model. *Proc Natl Acad Sci U S A* **91**, 7041-5.
- Amaral, D. G. 1993 Emerging principles of intrinsic hippocampal organization. *Curr Opin Neurobiol* **3**, 225-9.
- Amaral, D. G. 1999 Introduction: what is where in the medial temporal lobe? *Hippocampus* **9**, 1-6.
- Amodio, D. M. & Frith, C. D. 2006 Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* **7**, 268-77.
- Andersen, P., Bliss, T. & Skrede, K. 1971 Lamellar organisation of hippocampal excitatory pathways. *Experimental Brain Research* **13**, 222-238.
- Andersen, P., Morris, R., Amaral, D. G., Bliss, T. & O'Keefe, J. 2007 *The Hippocampus Book*. New York: Oxford University Press.
- Arzy, S., Molnar-Szakacs, I. & Blanke, O. 2008 Self in time: imagined self-location influences neural activity related to mental time travel. *J Neurosci* **28**, 6502-7.
- Atance, C. M. & O'Neill, D. K. 2001 Episodic future thinking. *Trends Cogn Sci* **5**, 533-539.
- Bar, M. 2004 Visual objects in context. *Nat Rev Neurosci* **5**, 617-29.
- Barnes, C. A., Rao, G. & McNaughton, B. L. 1987 Increased electrotonic coupling in aged rat hippocampus: a possible mechanism for cellular excitability changes. *J Comp Neurol* **259**, 549-58.
- Bartlett, F. C. 1932 *Remembering*: Cambridge University Press.
- Bayley, P. J., Gold, J. J., Hopkins, R. O. & Squire, L. R. 2005 The neuroanatomy of remote memory. *Neuron* **46**, 799-810.
- Bayley, P. J., Hopkins, R. O. & Squire, L. R. 2003 Successful recollection of remote autobiographical memories by amnesic patients with medial temporal lobe lesions. *Neuron* **38**, 135-44.

- Biederman, I. 1987 Recognition-by-components: a theory of human image understanding. *Psychol Rev* **94**, 115-47.
- Bird, C. M. & Burgess, N. 2008 The hippocampus and memory: insights from spatial processing. *Nat Rev Neurosci* **9**, 182-94.
- Bisiach, E. & Luzzatti, C. 1978 Unilateral neglect of representational space. *Cortex* **14**, 129-33.
- Bliss, T. 2006 Synaptic Plasticity in the Hippocampus. In *The Hippocampus Book* (ed. P. Andersen, R. Morris, D. G. Amaral, T. Bliss & J. O'Keefe), pp. 343-475. New York: Oxford University Press.
- Bloch, F., Hansen, W. W. & Packard, M. 1946 Nuclear Induction. *Physical Review* **70**.
- Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. 2008 The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci* **1124**, 1-38.
- Buckner, R. L. & Carroll, D. C. 2007 Self-projection and the brain. *Trends Cogn Sci* **11**, 49-57.
- Burgess, N. 2002 The hippocampus, space, and viewpoints in episodic memory. *Q J Exp Psychol A* **55**, 1057-80.
- Burgess, N. 2006 Spatial memory: how egocentric and allocentric combine. *Trends Cogn Sci* **10**, 551-7.
- Burgess, N., Becker, S., King, J. A. & O'Keefe, J. 2001 Memory for events and their spatial context: models and experiments. *Philos Trans R Soc Lond B Biol Sci* **356**, 1493-503.
- Burgess, N., Maguire, E. A. & O'Keefe, J. 2002 The human hippocampus and spatial and episodic memory. *Neuron* **35**, 625-41.
- Burgess, P. W. & Shallice, T. 1996 Confabulation and the control of recollection. *Memory* **4**, 359-411.
- Bussey, T. J. & Saksida, L. M. 2007 Memory, perception, and the ventral visual-perirhinal-hippocampal stream: thinking outside of the boxes. *Hippocampus* **17**, 898-908.
- Buzsaki, G. 2004 Large-scale recording of neuronal ensembles. *Nat Neurosci* **7**, 446-51.
- Byrne, P., Becker, S. & Burgess, N. 2007 Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol Rev* **114**, 340-75.
- Cabeza, R. & St Jacques, P. 2007 Functional neuroimaging of autobiographical memory. *Trends Cogn Sci* **11**, 219-27.
- Cermak, L. S. & O'Connor, M. 1983 The anterograde and retrograde retrieval ability of a patient with amnesia due to encephalitis. *Neuropsychologia* **21**, 213-34.
- Cipolotti, L., Shallice, T., Chan, D., Fox, N., Scahill, R., Harrison, G., Stevens, J. & Rudge, P. 2001 Long-term retrograde amnesia: the crucial role of the hippocampus. *Neuropsychologia* **39**, 151-72.
- Clark, R. E., Broadbent, N. J. & Squire, L. R. 2005 Impaired remote spatial memory after hippocampal lesions despite extensive training beginning early in life. *Hippocampus* **15**, 340-6.
- Clayton, N. S., Bussey, T. J. & Dickinson, A. 2003 Can animals recall the past and plan for the future? *Nat Rev Neurosci* **4**, 685-91.
- Cohen, N. J. & Eichenbaum, H. 1993 *Memory, Amnesia and the Hippocampal System*. Cambridge, MA: MIT Press.
- Cohen, N. J. & Squire, L. R. 1980 Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* **210**, 207-10.

- Colombo, M., Fernandez, T., Nakamura, K. & Gross, C. G. 1998 Functional differentiation along the anterior-posterior axis of the hippocampus in monkeys. *J Neurophysiol* **80**, 1002-5.
- Conway, M. A. & Pleydell-Pearce, C. W. 2000 The construction of autobiographical memories in the self-memory system. *Psychol Rev* **107**, 261-88.
- Corkin, S. 2002 What's new with the amnesic patient H.M.? *Nat Rev Neurosci* **3**, 153-60.
- Cortes, C. & Vapnik, V. 1995 Support-Vector Networks. *Machine Learning* **20**, 273-297.
- Coull, J. T., Vidal, F., Nazarian, B. & Macar, F. 2004 Functional anatomy of the attentional modulation of time estimation. *Science* **303**, 1506-8.
- Cox, D. D. & Savoy, R. L. 2003 Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **19**, 261-70.
- Crovitz, H. H. & Schiffman, H. 1974 Frequency of episodic memories as a function of their age. *Bull. Psychonom. Soc.* **4**, 517-518.
- Daniel, W. W. & Terrell, J. C. 1995 *Business Statistics for Management and Economics*. Boston: Houghton Mifflin.
- deCharms, R. C. & Zador, A. 2000 Neural representation and the cortical code. *Annu Rev Neurosci* **23**, 613-47.
- Deichmann, R., Gottfried, J. A., Hutton, C. & Turner, R. 2003 Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* **19**, 430-41.
- Dietterich, T. D. & Bakiri, G. 1995 Solving multiclass learning problems via error correcting output codes. *Journal of Artificial Intelligence Research* **2**, 263-286.
- Doeller, C. F., King, J. A. & Burgess, N. 2008 Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc Natl Acad Sci U S A* **105**, 5915-20.
- Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. 2007 Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43-57.
- Duda, O. R., Hart, P. E. & Stork, D. G. 2001 *Pattern Classification*. New York: Wiley.
- Dudai, Y. 2004 The neurobiology of consolidations, or, how stable is the engram? *Annu Rev Psychol* **55**, 51-86.
- Dudai, Y. 2006 Reconsolidation: the advantage of being refocused. *Curr Opin Neurobiol* **16**, 174-8.
- Dudai, Y. & Carruthers, M. 2005 The Janus face of Mnemosyne. *Nature* **434**, 567.
- Eichenbaum, H. 2004 Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* **44**, 109-20.
- 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* **23**, 209-26.
- Eichenbaum, H., Stewart, C. & Morris, R. G. 1990 Hippocampal representation in place learning. *J Neurosci* **10**, 3531-42.
- Eichenbaum, H., Wiener, S. I., Shapiro, M. L. & Cohen, N. J. 1989 The organization of spatial coding in the hippocampus: a study of neural ensemble activity. *J Neurosci* **9**, 2764-75.
- Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L. & Fried, I. 2003 Cellular networks underlying human spatial navigation. *Nature* **425**, 184-8.
- Emery, N. J. & Clayton, N. S. 2004 The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* **306**, 1903-7.
- Epstein, R. & Kanwisher, N. 1998 A cortical representation of the local visual environment. *Nature* **392**, 598-601.

- Ferbinteanu, J., Kennedy, P. J. & Shapiro, M. L. 2006 Episodic memory--from brain to mind. *Hippocampus* **16**, 691-703.
- Fisher, R. A. 1935 *The design of experiments*. Edinburgh: Oliver Boyd.
- Foster, D. J. & Wilson, M. A. 2006 Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680-3.
- Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., Price, C. J., Zeki, S., Ashburner, J. T. & Penny, W. D. 2004 *Human Brain Function*. New York: Elsevier Academic Press.
- Friedman, W. F. 1993 Memory for the time of past events. *Psychological Bulletin* **113**, 44-66.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D. & Turner, R. 1998 Event-related fMRI: characterizing differential responses. *Neuroimage* **7**, 30-40.
- Friston, K. J., Frith, C. D., Turner, R. & Frackowiak, R. S. 1995 Characterizing evoked hemodynamics with fMRI. *Neuroimage* **2**, 157-65.
- Friston, K. J., Holmes, A. P. & Worsley, K. J. 1999 How many subjects constitute a study? *Neuroimage* **10**, 1-5.
- Friston, K. J., Penny, W. D. & Glaser, D. E. 2005 Conjunction revisited. *Neuroimage* **25**, 661-7.
- Frith, U. & Frith, C. D. 2003 Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci* **358**, 459-73.
- Fuster, J. M. 1997 *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. New York: Lippincott-Raven.
- Gaffan, D. 2002 Against memory systems. *Philos Trans R Soc Lond B Biol Sci* **357**, 1111-21.
- Gallagher, S. 2000 Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn Sci* **4**, 14-21.
- Gegenfurtner, K. R. & Kiper, D. C. 2003 Color vision. *Annu Rev Neurosci* **26**, 181-206.
- Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R. & Fried, I. 2008 Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* **322**, 96-101.
- Gilbert, D. T. & Wilson, T. D. 2007 Propection: experiencing the future. *Science* **317**, 1351-4.
- Gilboa, A. 2004 Autobiographical and episodic memory--one and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia* **42**, 1336-49.
- Gilboa, A., Winocur, G., Grady, C. L., Hevenor, S. J. & Moscovitch, M. 2004 Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. *Cereb Cortex* **14**, 1214-25.
- Goense, J. B. & Logothetis, N. K. 2008 Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr Biol* **18**, 631-40.
- Gottfried, J. A., Smith, A. P., Rugg, M. D. & Dolan, R. J. 2004 Remembrance of odors past: human olfactory cortex in cross-modal recognition memory. *Neuron* **42**, 687-95.
- Graham, K. S. & Gaffan, D. 2005 The role of the medial temporal lobe in memory and perception: evidence from rats, nonhuman primates and humans. *Q J Exp Psychol B* **58**, 193-201.
- Greenberg, D. L. & Rubin, D. C. 2003 The neuropsychology of autobiographical memory. *Cortex* **39**, 687-728.
- Gross, C. G. 2002 Genealogy of the "Grandmother cell". *Neuroscientist* **8**, 512-518.
- Guariglia, C., Padovani, A., Pantano, P. & Pizzamiglio, L. 1993 Unilateral neglect restricted to visual imagery. *Nature* **364**, 235-7.

- Guzowski, J. F., McNaughton, B. L., Barnes, C. A. & Worley, P. F. 1999 Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. *Nat Neurosci* **2**, 1120-4.
- Hafting, T., Fyhn, M., Molden, S., Moser, M. B. & Moser, E. I. 2005 Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801-6.
- Hamming, R. W. 1950 Error detecting and error correcting codes. *The Bell System Technical Journal* **29**, 147-160.
- Hampson, R. E., Simeral, J. D. & Deadwyler, S. A. 1999 Distribution of spatial and nonspatial information in dorsal hippocampus. *Nature* **402**, 610-4.
- Hannula, D. E., Tranel, D. & Cohen, N. J. 2006 The long and the short of it: relational memory impairments in amnesia, even at short lags. *J Neurosci* **26**, 8352-9.
- Harris, K. D. 2005 Neural signatures of cell assembly organization. *Nat Rev Neurosci* **6**, 399-407.
- Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsaki, G. 2003 Organization of cell assemblies in the hippocampus. *Nature* **424**, 552-6.
- Harris, K. D., Henze, D. A., Hirase, H., Leinekugel, X., Dragoi, G., Czurko, A. & Buzsaki, G. 2002 Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* **417**, 738-41.
- Hartley, T., Bird, C. M., Chan, D., Cipelotti, L., Husain, M., Vargha-Khadem, F. & Burgess, N. 2007 The hippocampus is required for short-term topographical memory in humans. *Hippocampus* **17**, 34-48.
- 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**, 369-79.
- Hartley, T., Maguire, E. A., Spiers, H. J. & Burgess, N. 2003 The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron* **37**, 877-88.
- Hassabis, D. & Maguire, E. A. 2007 Deconstructing episodic memory with construction. *Trends Cogn Sci* **11**, 299-306.
- Hasselmo, M. E. & Schnell, E. 1994 Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: computational modeling and brain slice physiology. *J Neurosci* **14**, 3898-914.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. & Pietrini, P. 2001 Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425-30.
- Haynes, J. D. & Rees, G. 2005a Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* **8**, 686-91.
- Haynes, J. D. & Rees, G. 2005b Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol* **15**, 1301-7.
- Haynes, J. D. & Rees, G. 2006 Decoding mental states from brain activity in humans. *Nat Rev Neurosci* **7**, 523-34.
- Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C. & Passingham, R. E. 2007 Reading hidden intentions in the human brain. *Curr Biol* **17**, 323-8.
- Hebb, D. O. 1949 *The Organization of Behavior*. New York: Wiley.
- Heeger, D. J. & Ress, D. 2002 What does fMRI tell us about neuronal activity? *Nat Rev Neurosci* **3**, 142-51.
- Henze, D. A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K. D. & Buzsaki, G. 2000 Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol* **84**, 390-400.
- Hirsh, R. 1974 The hippocampus and contextual retrieval of information from memory: a theory. *Behav Biol* **12**, 421-44.

- Hodges, J. R. 1995 Retrograde amnesia. In *Handbook of memory disorders* (ed. A. D. Baddeley, B. A. Wilson & F. N. Watts), pp. 81-107. Chichester: Wiley.
- Hornberger, M., Rugg, M. D. & Henson, R. N. 2006 fMRI correlates of retrieval orientation. *Neuropsychologia* **44**, 1425-36.
- Hupbach, A., Gomez, R., Hardt, O. & Nadel, L. 2007 Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learn Mem* **14**, 47-53.
- Huxter, J., Burgess, N. & O'Keefe, J. 2003 Independent rate and temporal coding in hippocampal pyramidal cells. *Nature* **425**, 828-32.
- Jacobs, W. J. & Nadel, L. 1998 Neurobiology of reconstructed memory. *Psychology, Public Policy and Law* **4**, 1110.
- Janzen, G. & van Turenout, M. 2004 Selective neural representation of objects relevant for navigation. *Nat Neurosci* **7**, 673-7.
- Jeffery, K. J. 2007 Integration of the sensory inputs to place cells: what, where, why, and how? *Hippocampus* **17**, 775-85.
- Johnson, A. & Redish, A. D. 2007 Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J Neurosci* **27**, 12176-89.
- Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E. & Prigatano, G. P. 2002 Neural correlates of self-reflection. *Brain* **125**, 1808-14.
- Jones, G. V. 1976 A fragmentation hypothesis of memory: cued recall of pictures and sequential position. *Journal of Experimental Psychology: General* **105**, 277-293.
- Jung, M. W., Wiener, S. I. & McNaughton, B. L. 1994 Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J Neurosci* **14**, 7347-56.
- Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z. & Buckner, R. L. 2008 Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. *J Neurophysiol* **100**, 129-39.
- Kamitani, Y. & Tong, F. 2005 Decoding the visual and subjective contents of the human brain. *Nat Neurosci* **8**, 679-85.
- Kanwisher, N., McDermott, J. & Chun, M. M. 1997 The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* **17**, 4302-11.
- Kapur, N. 1999 Syndromes of retrograde amnesia: a conceptual and empirical synthesis. *Psychol Bull* **125**, 800-25.
- Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. 2003 Spontaneously emerging cortical representations of visual attributes. *Nature* **425**, 954-6.
- Klein, S. B., Loftus, J. & Kihlstrom, J. F. 2002 Memory and temporal experience: the effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition* **20**, 353-379.
- Kopelman, M. D., Wilson, B. A. & Baddeley, A. D. 1989 The autobiographical memory interview: a new assessment of autobiographical and personal semantic memory in amnesic patients. *J Clin Exp Neuropsychol* **11**, 724-44.
- Kosslyn, S. M., Ganis, G. & Thompson, W. L. 2001 Neural foundations of imagery. *Nat Rev Neurosci* **2**, 635-42.
- Kosslyn, S. M., Thompson, W. L., Kim, I. J. & Alpert, N. M. 1995 Topographical representations of mental images in primary visual cortex. *Nature* **378**, 496-8.
- Kreiman, G., Koch, C. & Fried, I. 2000 Imagery neurons in the human brain. *Nature* **408**, 357-61.
- Kriegeskorte, N., Goebel, R. & Bandettini, P. 2006 Information-based functional brain mapping. *Proc Natl Acad Sci U S A* **103**, 3863-8.

- Kumaran, D. & Maguire, E. A. 2005 The human hippocampus: cognitive maps or relational memory? *J Neurosci* **25**, 7254-9.
- Kumaran, D. & Maguire, E. A. 2006 The dynamics of hippocampal activation during encoding of overlapping sequences. *Neuron* **49**, 617-29.
- Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R. & et al. 1992 Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci U S A* **89**, 5675-9.
- LaConte, S., Strother, S., Cherkassky, V., Anderson, J. & Hu, X. 2005 Support vector machines for temporal classification of block design fMRI data. *Neuroimage* **26**, 317-29.
- Lashley, K. S. 1950 In Search of the Engram. *Society of Experimental Biology* **4**, 454-482.
- Lee, A. C., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., Bussey, T. J., Davies, R. R., Kapur, N., Hodges, J. R. & Graham, K. S. 2005a Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus* **15**, 782-97.
- Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., Hodges, J. R. & Graham, K. S. 2005b Perceptual deficits in amnesia: challenging the medial temporal lobe 'mnemonic' view. *Neuropsychologia* **43**, 1-11.
- Levine, B., Svoboda, E., Hay, J. F., Winocur, G. & Moscovitch, M. 2002 Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychol Aging* **17**, 677-89.
- Lin, L., Osan, R., Shoham, S., Jin, W., Zuo, W. & Tsien, J. Z. 2005 Identification of network-level coding units for real-time representation of episodic experiences in the hippocampus. *Proc Natl Acad Sci U S A* **102**, 6125-30.
- Logothetis, N. K. 2003 The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci* **23**, 3963-71.
- Logothetis, N. K. 2008 What we can do and what we cannot do with fMRI. *Nature* **453**, 869-78.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. 2001 Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150-7.
- Magee, J. C. 2000 Dendritic integration of excitatory synaptic input. *Nat Rev Neurosci* **1**, 181-90.
- Maguire, E. A. 2001a Neuroimaging studies of autobiographical event memory. *Philos Trans R Soc Lond B Biol Sci* **356**, 1441-51.
- Maguire, E. A. 2001b The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand J Psychol* **42**, 225-38.
- Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D. & O'Keefe, J. 1998 Knowing where and getting there: a human navigation network. *Science* **280**, 921-4.
- Maguire, E. A. & Frith, C. D. 2003 Aging affects the engagement of the hippocampus during autobiographical memory retrieval. *Brain* **126**, 1511-23.
- Maguire, E. A. & Frith, C. D. 2004 The brain network associated with acquiring semantic knowledge. *Neuroimage* **22**, 171-8.
- Maguire, E. A., Frith, C. D., Rudge, P. & Cipolotti, L. 2005 The effect of adult-acquired hippocampal damage on memory retrieval: an fMRI study. *Neuroimage* **27**, 146-52.

- Maguire, E. A., Henson, R. N., Mummery, C. J. & Frith, C. D. 2001a Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *Neuroreport* **12**, 441-4.
- Maguire, E. A. & Mummery, C. J. 1999 Differential modulation of a common memory retrieval network revealed by positron emission tomography. *Hippocampus* **9**, 54-61.
- Maguire, E. A., Nannery, R. & Spiers, H. J. 2006a Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain* **129**, 2894-907.
- Maguire, E. A., Valentine, E. R., Wilding, J. M. & Kapur, N. 2003 Routes to remembering: the brains behind superior memory. *Nat Neurosci* **6**, 90-5.
- Maguire, E. A., Vargha-Khadem, F. & Mishkin, M. 2001b The effects of bilateral hippocampal damage on fMRI regional activations and interactions during memory retrieval. *Brain* **124**, 1156-70.
- Maguire, E. A., Woollett, K. & Spiers, H. J. 2006b London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. *Hippocampus* **16**, 1091-101.
- Manns, J. R., Hopkins, R. O. & Squire, L. R. 2003 Semantic memory and the human hippocampus. *Neuron* **38**, 127-33.
- Mansfield, P. 1977 Multi-planar imaging using NMR spin echoes. *Journal of Physics C* **10**.
- Marr, D. 1971 Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci* **262**, 23-81.
- Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T. & Macrae, C. N. 2007 Wandering minds: the default network and stimulus-independent thought. *Science* **315**, 393-5.
- Mayes, A., Montaldi, D. & Migo, E. 2007 Associative memory and the medial temporal lobes. *Trends Cogn Sci* **11**, 126-35.
- McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. 1995 Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* **102**, 419-57.
- McKenna, P. & Gerhand, S. 2002 Preserved semantic learning in an amnesic patient. *Cortex* **38**, 37-58.
- 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'. *Nat Rev Neurosci* **7**, 663-78.
- Mehta, M. R., Lee, A. K. & Wilson, M. A. 2002 Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741-6.
- Morris, R. G. 2006 Theories of Hippocampal Function. In *The Hippocampus Book* (ed. P. Andersen, R. Morris, D. G. Amaral, T. Bliss & J. O'Keefe), pp. 581-715. New York: Oxford University Press.
- Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. 1982 Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681-3.
- Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A. & Rosenbaum, R. S. 2006 The cognitive neuroscience of remote episodic, semantic and spatial memory. *Curr Opin Neurobiol* **16**, 179-90.
- Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G. & Nadel, L. 2005 Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *J Anat* **207**, 35-66.
- Moser, E. I., Kropff, E. & Moser, M. B. 2008 Place cells, grid cells, and the brain's spatial representation system. *Annu Rev Neurosci* **31**, 69-89.

- Moser, M. B. & Moser, E. I. 1998 Functional differentiation in the hippocampus. *Hippocampus* **8**, 608-19.
- Moses, S. N. & Ryan, J. D. 2006 A comparison and evaluation of the predictions of relational and conjunctive accounts of hippocampal function. *Hippocampus* **16**, 43-65.
- Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G. & Brammer, M. 2006 The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. *Neuroimage* **33**, 1055-65.
- Muller, R. U., Bostock, E., Taube, J. S. & Kubie, J. L. 1994 On the directional firing properties of hippocampal place cells. *J Neurosci* **14**, 7235-51.
- Mumby, D. G., Astur, R. S., Weisend, M. P. & Sutherland, R. J. 1999 Retrograde amnesia and selective damage to the hippocampal formation: memory for places and object discriminations. *Behav Brain Res* **106**, 97-107.
- Murray, E. A. & Bussey, T. J. 2001 Consolidation and the medial temporal lobe revisited: methodological considerations. *Hippocampus* **11**, 1-7.
- Nadel, L. & Land, C. 2000 Memory traces revisited. *Nat Rev Neurosci* **1**, 209-12.
- Nadel, L. & Moscovitch, M. 1997 Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* **7**, 217-27.
- Nadel, L. & O'Keefe, J. 1974 The hippocampus in pieces and patches: an essay on modes of explanation in physiological psychology. In *Essays on the nervous system* (ed. R. Bellairs & E. G. Gray). Oxford: The Clarendon Press.
- Nadel, L., Samsonovich, A., Ryan, L. & Moscovitch, M. 2000 Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus* **10**, 352-68.
- Nader, K., Schafe, G. E. & LeDoux, J. E. 2000 The labile nature of consolidation theory. *Nat Rev Neurosci* **1**, 216-9.
- Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., Kato, A., Carr, C. A., Johnston, D., Wilson, M. A. & Tonegawa, S. 2002 Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* **297**, 211-8.
- Nichols, T. E. & Holmes, A. P. 2002 Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* **15**, 1-25.
- Norman, K. A. & O'Reilly, R. C. 2003 Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol Rev* **110**, 611-46.
- Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. 2006 Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* **10**, 424-30.
- O'Keefe, J. 2006 Hippocampal Neurophysiology in the Behaving Animal. In *The Hippocampus Book* (ed. P. Andersen, R. Morris, D. G. Amaral, T. Bliss & J. O'Keefe), pp. 475-549. New York: Oxford University Press.
- O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J. & Maguire, E. A. 1998 Place cells, navigational accuracy, and the human hippocampus. *Philos Trans R Soc Lond B Biol Sci* **353**, 1333-40.
- O'Keefe, J. & Dostrovsky, J. 1971 The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* **34**, 171-5.
- O'Keefe, J. & Nadel, L. 1978 *The hippocampus as a cognitive map*. Oxford, England: Oxford University Press.
- O'Keefe, J. & Recce, M. L. 1993 Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317-30.
- Obermayer, K. & Blasdel, G. G. 1993 Geometry of orientation and ocular dominance columns in monkey striate cortex. *J Neurosci* **13**, 4114-29.

- Ogawa, S. & Lee, T. M. 1990 Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. *Magn Reson Med* **16**, 9-18.
- Ogawa, S., Lee, T. M., Kay, A. R. & Tank, D. W. 1990a Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A* **87**, 9868-72.
- Ogawa, S., Lee, T. M., Nayak, A. S. & Glynn, P. 1990b Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med* **14**, 68-78.
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H. & Ugurbil, K. 1992 Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A* **89**, 5951-5.
- Otto, T. & Eichenbaum, H. 1992 Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: evidence for hippocampal processing in recognition memory. *Hippocampus* **2**, 323-34.
- Pauling, L. & Coryell, C. D. 1936 The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proc Natl Acad Sci U S A* **22**, 210-6.
- Payne, J. D., Nadel, L., Allen, J. J., Thomas, K. G. & Jacobs, W. J. 2002 The effects of experimentally induced stress on false recognition. *Memory* **10**, 1-6.
- Penfield, W. & Mathieson, G. 1974 Memory. Autopsy findings and comments on the role of hippocampus in experiential recall. *Arch Neurol* **31**, 145-54.
- Penfield, W. & Milner, B. 1958 Memory deficit produced by bilateral lesions in the hippocampal zone. *AMA Arch Neurol Psychiatry* **79**, 475-97.
- Polyn, S. M., Natu, V. S., Cohen, J. D. & Norman, K. A. 2005 Category-specific cortical activity precedes retrieval during memory search. *Science* **310**, 1963-6.
- Pouget, A., Dayan, P. & Zemel, R. 2000 Information processing with population codes. *Nat Rev Neurosci* **1**, 125-32.
- Prabhakaran, V., Narayanan, K., Zhao, Z. & Gabrieli, J. D. 2000 Integration of diverse information in working memory within the frontal lobe. *Nat Neurosci* **3**, 85-90.
- Proust, M. 1927 *Remembrance of Things Past*.
- Purcell, E. M., Torrey, H. C. & Pound, R. V. 1946 Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review* **69**.
- Quiroga, R. Q., Kreiman, G., Koch, C. & Fried, I. 2008 Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* **12**, 87-91.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. 2005 Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102-7.
- Raby, C. R., Alexis, D. M., Dickinson, A. & Clayton, N. S. 2007 Planning for the future by western scrub-jays. *Nature* **445**, 919-21.
- 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. *Proc Natl Acad Sci U S A* **98**, 676-82.
- Redish, A. D., Battaglia, F. P., Chawla, M. K., Ekstrom, A. D., Gerrard, J. L., Lipa, P., Rosenzweig, E. S., Worley, P. F., Guzowski, J. F., McNaughton, B. L. & Barnes, C. A. 2001 Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *J Neurosci* **21**, RC134.
- Rempel-Clower, N. L., Zola, S. M., Squire, L. R. & Amaral, D. G. 1996 Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. *J Neurosci* **16**, 5233-55.

- Ress, D., Backus, B. T. & Heeger, D. J. 2000 Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci* **3**, 940-5.
- Rolls, E. T., Treves, A., Robertson, R. G., Georges-Francois, P. & Panzeri, S. 1998 Information about spatial view in an ensemble of primate hippocampal cells. *J Neurophysiol* **79**, 1797-813.
- Rosenbaum, R. S., Kohler, S., Schacter, D. L., Moscovitch, M., Westmacott, R., Black, S. E., Gao, F. & Tulving, E. 2005 The case of K.C.: contributions of a memory-impaired person to memory theory. *Neuropsychologia* **43**, 989-1021.
- 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**, 1619-35.
- Rosenbaum, R. S., Stuss, D. T., Levine, B. & Tulving, E. 2007 Theory of mind is independent of episodic memory. *Science* **318**, 1257.
- Rubin, D. C., Schrauf, R. W. & Greenberg, D. L. 2003 Belief and recollection of autobiographical memories. *Mem Cognit* **31**, 887-901.
- Rugg, M. D., Otten, L. J. & Henson, R. N. 2002 The neural basis of episodic memory: evidence from functional neuroimaging. *Philos Trans R Soc Lond B Biol Sci* **357**, 1097-110.
- Rugg, M. D. & Wilding, E. L. 2000 Retrieval processing and episodic memory. *Trends Cogn Sci* **4**, 108-115.
- Ryan, L., Nadel, L., Keil, K., Putnam, K., Schnyer, D., Trouard, T. & Moscovitch, M. 2001 Hippocampal complex and retrieval of recent and very remote autobiographical memories: evidence from functional magnetic resonance imaging in neurologically intact people. *Hippocampus* **11**, 707-14.
- Samarasekera, S. R., Vincent, A., Welch, J. L., Jackson, M., Nichols, P. & Griffiths, T. D. 2007 Course and outcome of acute limbic encephalitis with negative voltage-gated potassium channel antibodies. *J Neurol Neurosurg Psychiatry* **78**, 391-4.
- Samsonovich, A. & McNaughton, B. L. 1997 Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* **17**, 5900-20.
- Saxe, R. & Kanwisher, N. 2003 People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* **19**, 1835-42.
- Schacter, D. L. 1996 *Searching for memory: The brain, the mind and the past*. New York: Basic Books.
- Schacter, D. L. 2001 *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton Mifflin.
- Schacter, D. L. & Addis, D. R. 2007 The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos Trans R Soc Lond B Biol Sci* **362**, 773-86.
- Schacter, D. L., Addis, D. R. & Buckner, R. L. 2007 Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci* **8**, 657-61.
- Schacter, D. L., Addis, D. R. & Buckner, R. L. 2008 Episodic simulation of future events: concepts, data, and applications. *Ann N Y Acad Sci* **1124**, 39-60.
- Schacter, D. L., Norman, K. A. & Koutstaal, W. 1998 The cognitive neuroscience of constructive memory. *Annu Rev Psychol* **49**, 289-318.
- Schacter, D. L. & Slotnick, S. D. 2004 The cognitive neuroscience of memory distortion. *Neuron* **44**, 149-60.
- Scoville, W. B. & Milner, B. 1957 Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry* **20**, 11-21.
- Semon, R. 1923 *Mnemic Psychology*. London: George Allen & Unwin.

- Sharot, T., Riccardi, A. M., Raio, C. M. & Phelps, E. A. 2007 Neural mechanisms mediating optimism bias. *Nature* **450**, 102-5.
- Sidtis, J. J., Strother, S. C. & Rottenberg, D. A. 2003 Predicting performance from functional imaging data: methods matter. *Neuroimage* **20**, 615-24.
- Simons, J. S., Davis, S. W., Gilbert, S. J., Frith, C. D. & Burgess, P. W. 2006 Discriminating imagined from perceived information engages brain areas implicated in schizophrenia. *Neuroimage* **32**, 696-703.
- Simons, J. S. & Spiers, H. J. 2003 Prefrontal and medial temporal lobe interactions in long-term memory. *Nat Rev Neurosci* **4**, 637-48.
- Spiers, H. J. & Maguire, E. A. 2006 Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage* **31**, 1826-40.
- Spiers, H. J., Maguire, E. A. & Burgess, N. 2001 Hippocampal amnesia. *Neurocase* **7**, 357-82.
- Spreng, R. N., Mar, R. A. & Kim, A. S. 2008 The Common Neural Basis of Autobiographical Memory, Propection, Navigation, Theory of Mind and the Default Mode: A Quantitative Meta-analysis. *J Cogn Neurosci*.
- Squire, L. R. 1992 Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol Rev* **99**, 195-231.
- Squire, L. R. & Alvarez, P. 1995 Retrograde amnesia and memory consolidation: a neurobiological perspective. *Curr Opin Neurobiol* **5**, 169-77.
- Squire, L. R., Cohen, N. J. & Zola-Morgan, L. 1984 The medial temporal region and memory consolidation: a new hypothesis. In *Memory Consolidation: Psychobiology of Cognition* (ed. H. Weingartner & E. S. Parker), pp. 185-210. Hillsdale: Lawrence Erlbaum Associates.
- Squire, L. R., Stark, C. E. & Clark, R. E. 2004 The medial temporal lobe. *Annu Rev Neurosci* **27**, 279-306.
- Squire, L. R. & Zola, S. M. 1996 Ischemic brain damage and memory impairment: a commentary. *Hippocampus* **6**, 546-52.
- Steinvorth, S., Corkin, S. & Halgren, E. 2006 Ecphory of autobiographical memories: an fMRI study of recent and remote memory retrieval. *Neuroimage* **30**, 285-98.
- Steinvorth, S., Levine, B. & Corkin, S. 2005 Medial temporal lobe structures are needed to re-experience remote autobiographical memories: evidence from H.M. and W.R. *Neuropsychologia* **43**, 479-96.
- Suddendorf, T. & Busby, J. 2003 Mental time travel in animals? *Trends Cogn Sci* **7**, 391-396.
- Suddendorf, T. & Corballis, M. C. 1997 Mental time travel and the evolution of the human mind. *Genet Soc Gen Psychol Monogr* **123**, 133-67.
- Sugiura, M., Shah, N. J., Zilles, K. & Fink, G. R. 2005 Cortical representations of personally familiar objects and places: functional organization of the human posterior cingulate cortex. *J Cogn Neurosci* **17**, 183-98.
- Summerfield, J. J., Hassabis, D. & Maguire, E. A. 2009 Cortical midline involvement in autobiographical memory. *Neuroimage* **44**, 1188-200.
- Summerfield, J. J., Hassabis, D. & Maguire, E. A. in preparation How does the brain construct scenes?
- Sutherland, R. J., Weisend, M. P., Mumby, D., Astur, R. S., Hanlon, F. M., Koerner, A., Thomas, M. J., Wu, Y., Moses, S. N., Cole, C., Hamilton, D. A. & Hoising, J. M. 2001 Retrograde amnesia after hippocampal damage: recent vs. remote memories in two tasks. *Hippocampus* **11**, 27-42.
- Suzuki, W. A. & Amaral, D. G. 1994 Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J Neurosci* **14**, 1856-77.

- Svoboda, E., McKinnon, M. C. & Levine, B. 2006 The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia* **44**, 2189-208.
- Szpunar, K. K., Watson, J. M. & McDermott, K. B. 2007 Neural substrates of envisioning the future. *Proc Natl Acad Sci U S A* **104**, 642-7.
- Takehara, K., Kawahara, S. & Kirino, Y. 2003 Time-dependent reorganization of the brain components underlying memory retention in trace eyeblink conditioning. *J Neurosci* **23**, 9897-905.
- Teng, E. & Squire, L. R. 1999 Memory for places learned long ago is intact after hippocampal damage. *Nature* **400**, 675-7.
- Teyler, T. J. & DiScenna, P. 1986 The hippocampal memory indexing theory. *Behav Neurosci* **100**, 147-54.
- Treves, A. & Rolls, E. T. 1991 What determines the capacity of autoassociative memories in the brain? *Network* **2**, 371-397.
- Treves, A. & Rolls, E. T. 1992 Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* **2**, 189-99.
- Tsodyks, M., Kenet, T., Grinvald, A. & Arieli, A. 1999 Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **286**, 1943-6.
- Tulving, E. 1972 Episodic and semantic memory. In *Organisation of Memory* (ed. E. Tulving & W. Donaldson), pp. 381-403. New York: Academic.
- Tulving, E. 1983 *Elements of Episodic Memory*. Oxford: Clarendon.
- Tulving, E. 1985 Memory and consciousness. *Canadian Psychologist* **26**, 1-12.
- Tulving, E. 2002a Chronesthesia: conscious awareness of subjective time. In *Principles of Frontal Lobe Function* (ed. D. T. Stuss & R. C. Knight), pp. 311-25: Oxford University Press.
- Tulving, E. 2002b Episodic memory: from mind to brain. *Annu Rev Psychol* **53**, 1-25.
- Turner, R., Le Bihan, D., Moonen, C. T., Despres, D. & Frank, J. 1991 Echo-planar time course MRI of cat brain oxygenation changes. *Magn Reson Med* **22**, 159-66.
- 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**, 376-80.
- 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. *J Neurosci* **24**, 6489-96.
- Vincent, A., Buckley, C., Schott, J. M., Baker, I., Dewar, B. K., Detert, N., Clover, L., Parkinson, A., Bien, C. G., Omer, S., Lang, B., Rossor, M. N. & Palace, J. 2004 Potassium channel antibody-associated encephalopathy: a potentially immunotherapy-responsive form of limbic encephalitis. *Brain* **127**, 701-12.
- Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E. & Buckner, R. L. 2006 Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J Neurophysiol* **96**, 3517-31.
- Viskontas, I. V., McAndrews, M. P. & Moscovitch, M. 2000 Remote episodic memory deficits in patients with unilateral temporal lobe epilepsy and excisions. *J Neurosci* **20**, 5853-7.
- Wagner, A. D., Shannon, B. J., Kahn, I. & Buckner, R. L. 2005 Parietal lobe contributions to episodic memory retrieval. *Trends Cogn Sci* **9**, 445-53.
- Wang, G., Tanaka, K. & Tanifuji, M. 1996 Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* **272**, 1665-8.
- Warrington, E. K. 1996 Studies of retrograde memory: a long-term view. *Proc Natl Acad Sci U S A* **93**, 13523-6.

- Waydo, S., Kraskov, A., Quiñero, R., Fried, I. & Koch, C. 2006 Sparse representation in the human medial temporal lobe. *J Neurosci* **26**, 10232-4.
- Wechsler, D. A. 2001 *Wechsler Test of Adult Reading*. New York: Harcourt.
- Weingardt, K. R., Loftus, E. F. & Lindsay, D. S. 1995 Misinformation revisited: new evidence on the suggestibility of memory. *Mem Cognit* **23**, 72-82.
- Wheeler, M. A., Stuss, D. T. & Tulving, E. 1997 Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychol Bull* **121**, 331-54.
- Wheeler, M. E., Petersen, S. E. & Buckner, R. L. 2000 Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci U S A* **97**, 11125-9.
- Wills, T. J., Lever, C., Cacucci, F., Burgess, N. & O'Keefe, J. 2005 Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873-6.
- Wilson, M. A. & McNaughton, B. L. 1993 Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055-8.
- Winocur, G., Moscovitch, M., Fogel, S., Rosenbaum, R. S. & Sekeres, M. 2005 Preserved spatial memory after hippocampal lesions: effects of extensive experience in a complex environment. *Nat Neurosci* **8**, 273-5.
- Witter, M. P. & Amaral, D. G. 1991 Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. *J Comp Neurol* **307**, 437-59.
- Witter, M. P., Van Hoesen, G. W. & Amaral, D. G. 1989 Topographical organization of the entorhinal projection to the dentate gyrus of the monkey. *J Neurosci* **9**, 216-28.
- Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. 1999 The global record of memory in hippocampal neuronal activity. *Nature* **397**, 613-6.
- Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauve, M. J., Widaman, K. F. & Knight, R. T. 2002 Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nat Neurosci* **5**, 1236-41.
- Zola-Morgan, S., Squire, L. R. & Amaral, D. G. 1986 Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *J Neurosci* **6**, 2950-67.