Neural mechanisms underlying advanced cognition in humans

Alexandra O. Constantinescu

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the University of Oxford

Magdalen College

Trinity Term 2017

Neural mechanisms underlying advanced cognition in humans

Alexandra O. Constantinescu

Magdalen College **Trinity Term 2017**

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the University of Oxford

Abstract

Humans have a remarkable capacity for generalizing experiences to novel situations. In a brain network that has expanded the most during mammalian evolution, known as the default mode network, grid cells and place cells organize spatial representations into mental maps. The aim of this thesis is to use non-invasive functional magnetic resonance imaging and computational analyses to test how well-understood codes at the single-cell level in animals also generalize to complex human behaviours.

We first asked how brain mechanisms used for spatial navigation help us understand non-spatial cognition. We designed an equivalent computer task to the one used for real spatial navigation, but in an abstract space. Next, we tested for hexagonal signals using the same analyses as those developed for navigation in physical space. We found hexagonal grid-like signals in the same brain network for navigating in conceptual space and in physical space. This suggests that the brain organizes concepts into a mental map, allowing conceptual relationships to be navigated in a manner similar to that of space.

Next, we aimed to develop a new analysis method that is closer to electrophysiology, and thus could detect gradients in place and grid-cell coding of physical space with neuroimaging. We show that our analysis can detect such gradients in simulated data. Moreover, using a big data approach, we found behavioural evidence that humans form precise maps of our environment. We also show preliminary evidence of neural patterns at submillimeter resolution, suggestive of place and grid-cell coding. Such analyses open the possibility to investigate functional gradients in place and grid-cell coding during novel behaviours that are more easily accessible in humans, such as non-spatial memory, abstract reasoning and imagination.

Together, these studies illustrate the potential power of functional magnetic resonance imaging techniques to investigate precise computations in humans.

This thesis contains approximately 47000 words.

Table of contents

Chapter 1 : General introduction 1

Chapter 2 : Anatomy of the default mode network 3
2.1 Localizing the default mode network 3
2.2 Techniques for investigating the DMN neuroanatomy 4
2.3 The prefrontal cortex 7
2.3.1 Ventromedial prefrontal cortex 8
2.3.2 The orbitofrontal cortex 9
2.3.3 The anterior cingulate and dorsomedial prefrontal cortices 11
2.3.4 Homologies between humans, monkeys and rats 11
2.4 The medial temporal lobe 13
2.4.1 The hippocampus 14
2.4.2 The entorhinal cortex 16
2.5 The retrosplenial cortex 17
2.6 Conclusion 18

Chapter 3 : Cognitive maps for spatial and non-spatial knowledge 19
3.1 Cognitive maps 19
3.2 Spatial maps 22
3.2.1 Place cells 23
3.2.2 Grid cells 25
3.2.3 Computational models of grid cells 29
3.2.4 Grid cell coding in virtual reality 32
3.2.5 Replay and preplay 35
3.2.6 Other space cells and parallel navigation strategies 38
3.3 Non-spatial maps 41
3.3.1 Time cells 42
3.3.2 Place cells and grid cells for sound frequency 45
3.3.3 Concept cells 47
3.3.4 Discrete spaces 47
3.4 Implications for human cognition 49

Chapter 4 : Investigating the human brain at the mesoscale with high-resolution MRI and computational analyses 51
4.1 Scales and levels of description 51
4.2 High-resolution MRI 53
4.2.1 Functional gradients 54
4.2.2 Cortical layers and columns 55
4.3 Computational analyses for grid-like hexagonal symmetry 58

Chapter 5 : Basic principles of MRI physics, physiology and analysis 63
5.1 MRI physics 63
5.1.1 Magnetism 63
5.1.2 Resonance 65
5.1.3 Imaging 66
5.1.3.1 Contrast 66
5.1.3.2 Gradient fields 68

5.1.3.3 K-space ... 69
5.2 fMRI physics and physiology ... 70
5.2.1 Physics of the BOLD signal ... 70
5.2.1.1 Echo-planar imaging ... 70
5.2.1.2 Optimizing fMRI acquisition ... 72
5.2.2 Physiology of the BOLD signal ... 75
5.3 fMRI analysis ... 79
5.3.1 Preprocessing ... 79
5.3.1.1 Correcting distortion and dropout ... 80
5.3.1.2 Correcting slice timing ... 80
5.3.1.3 Correcting head motion artefacts ... 81
5.3.1.4 Correcting physiological and scanner artefacts ... 82
5.3.1.5 Spatial smoothing ... 82
5.3.1.6 Minimizing temporal smoothing ... 83
5.3.2 Spatial registration ... 83
5.3.2.1 Volume-based registration ... 84
5.3.2.2 Surface-based registration ... 85
5.3.3 Univariate statistics ... 87
5.3.3.1 General linear model and hypothesis testing ... 87
5.3.3.2 Significance testing ... 90
5.3.3.3 Group analyses ... 91
5.3.4 Cross-validation ... 92
5.3.5 Multivariate statistics ... 94
5.3.6 Encoding and decoding models ... 97

Chapter 6 : A novel behavioural method for navigating conceptual knowledge ... 101
6.1 Introduction ... 101
6.2 Methods ... 102
6.2.1 Participants and experimental design ... 102
6.2.2 Navigation in abstract space is similar to navigation in physical space ... 103
6.2.3 Match task ... 104
6.2.4 Explore task ... 106
6.2.5 Collect task ... 107
6.2.6 Recall task ... 107
6.2.7 Post-scanning debriefings ... 110
6.3 Results ... 110
6.3.1 Explore task ... 110
6.3.2 Collect task ... 112
6.3.3 Recall task ... 113
6.3.4 Post-scanning debriefings ... 113
6.4 Discussion ... 114

Chapter 7 : Organizing conceptual knowledge in humans with a gridlike code ... 117
7.1 Introduction ... 117
7.2 Methods ... 119
7.2.1 Participants and experimental design ... 119
7.2.2 MRI data acquisition ... 121
7.2.3 Pre-processing of functional images ... 122

7.2.4 Motion estimation 123
7.2.5 fMRI whole brain analyses 123
7.2.6 fMRI region-of-interest analyses 126
7.2.7 Control analyses for fMRI 127
7.2.8 Controls for the properties of the visual scene 128
7.3 Results 128
7.3.1 Hexagonal modulation 128
7.3.2 Hexagonal consistency within-day 131
7.3.3 Hexagonal consistency between-days 132
7.3.4 Replication of the results after a re-analysis 136
7.4 Discussion 139

Chapter 8 : A new experimental model for detecting grid-like coding as a function of location in space 145
8.1 Introduction 145
8.2 Methods 147
8.2.1 Participants and experimental design 147
8.2.2 Virtual reality environment 148
8.2.3 Encode task 148
8.2.4 Test and feedback task 149
8.2.5 Teleport task 151
8.2.6 Draw map task 151
8.2.7 Simulations using ideal grid cells 152
8.2.8 Gridness score 156
8.2.9 Simulations using random noise 157
8.2.10 Simulations using ideal place cells 158
8.3 Results 159
8.3.1 Encode task 159
8.3.2 Test and feedback task 159
8.3.3 Teleport task 161
8.3.4 Grid scale, grid angle and place field size estimation 161
8.4 Discussion 162

Chapter 9 : Developing a representational similarity analysis method to detect gradients in place-cell signals and grid-cell hexagonal symmetry in humans, using high-resolution fMRI 167
9.1 Introduction 167
9.2 Methods 170
9.2.1 Participants and experimental design 170
9.2.2 MRI data acquisition 173
9.2.3 Pre-processing of functional images 176
9.2.4 Segmentation of the hippocampal formation 177
9.2.5 fMRI analysis of inter-voxel similarity along the hippocampal long axis. 177
9.2.6 fMRI analysis of head-direction and grid-cell signals modulated by spatial direction 178
9.2.7 fMRI novel analysis 181
9.2.8 Tackling challenges for big data experiments 184
9.3 Results 186
9.3.1 Co-registration of high-resolution images 186
9.3.2 Gradients in inter-voxel similarity during immobility 187

9.3.3 Head-direction and grid-cell signals during immobility 189
9.3.4 Novel analyses 191
9.4 Discussion 192

Chapter 10 : General discussion and future directions 195
10.1 Developing new techniques for measuring grid-like and place-like representations in humans 195
10.2 Open questions 197

List of figures

Fig. 2.1: Anatomy of the default mode network in the human brain 4
Fig. 2.2. Anatomical connectivity of the DMN in primates 8
Fig. 2.3: Comparable anatomy of the human and monkey vmPFC 9
Fig. 2.4: The human OFC 10
Fig. 2.5: Architectonic maps of the medial prefrontal cortex in human, macaque and rat 12
Fig. 2.6: Subfields of the human hippocampal formation 13
Fig. 2.7: The hippocampus in rodents, monkeys and humans. 15
Fig. 2.8: Hexagonal anatomical patches in the rat ERH 17
Fig. 2.9: The human retrosplenial cortex 18
Fig. 3.1: Behavioural tests for the cognitive map 21
Fig. 3.2: Nobel-winning neurons. 23
Fig. 3.3: Functional gradient in the scale of place cells 24
Fig. 3.4: Grid cells. 26
Fig. 3.5: Computational models of grid cells 31
Fig. 3.6: VR in humans and rodents 34
Fig. 3.7: Replay in place cells and grid cells. 37
Fig. 3.8: Head-direction cells, boundary/border cells and conjunctive cells 39
Fig. 3.9: Time coding in the hippocampal formation 43
Fig. 3.10: Mapping of a sound frequency dimension by grid cells 46
Fig. 3.11: Navigation of discrete spaces 49
Fig. 4.1: Mesoscale 52
Fig. 4.2: David Marr's three levels of description 53
Fig. 4.3: Spatial specificity of 7T fMRI in single human subjects 54
Fig. 4.4: Functional gradients at 7T in humans 55
Fig. 4.5: Investigating individual layers of the entorhinal cortex with 7T in humans 57
Fig. 4.6: Functional maps of ocular dominance and orientation columns in the visual cortex with 7T in humans 58
Fig. 4.7: Encoding models for detecting grid-like hexagonal symmetry with fMRI 60
Fig. 4.8: Multivariate models for detecting grid-like hexagonal symmetry with fMRI. . 61
Fig. 5.1: Magnetism 64
Fig. 5.2: Resonance 66
Fig. 5.3: MRI contrast mechanisms 67
Fig. 5.4: Two example MRI contrasts 68
Fig. 5.5: K space 69
Fig. 5.6: Parallel imaging technique. 72
Fig. 5.7: Physiology of fMRI 76
Fig. 5.8: Example linear and non-linear transformations 85
Fig. 5.9: Visual areas in the brain are evident with cortical surfaces, rather than with slices through a 3D volume. 86
Fig. 5.10: Visualizing cortical surfaces 87
Fig. 5.11: Example fMRI data and model for a simple linear regression 89
Fig. 5.12: Comparison between univariate and multivariate analyses 94
Fig. 5.13: Comparison of voxel selection methods for searchlight analysis. 97
Fig. 5.14: Different meanings of the word "powder" across the human brain 99
Fig. 6.1: Birds vary naturally along the lengths of their neck and legs. 102
Fig. 6.2: Bird space 104
Fig. 6.3: Example screenshot of the match task 105

Fig. 6.4: Example screenshot of the explore task 106
Fig. 6.5: Example trial of the recall task 108
Fig. 6.6: Distribution of trajectory angles in the recall task 109
Fig. 6.7: Analysis of the explore task at the group level and in one representative subject 111
Fig. 6.8: Measures of learning during the collect task. 112
Fig. 6.9: Learning in the recall task 113
Fig. 7.1: Experimental design for navigation in abstract space 120
Fig. 7.2. Estimation of the grid angle 125
Fig. 7.3: Controls for the properties of the visual scene 128
Fig. 7.4: Identifying hexagonally symmetric signals across the whole brain. 130
Fig. 7.5: Correlation between hexagonal modulation and behavioural performance 131
Fig. 7.6: Grid angle consistency between separate sessions acquired within the same day 133
Fig. 7.7: The vmPFC consistency map does not depend on the outcomes 134
Fig. 7.8: Grid angle consistency between separate sessions acquired more than a week apart 135
Fig. 7.9: Distribution of grid angles across subjects 139
Fig. 7.10. Understanding the fMRI hexagonal symmetry signal 141
Fig. 7.11: The brain network for navigation in physical space 142
Fig. 8.1: VR environment 148
Fig. 8.2: Encode task 149
Fig. 8.3: Test and feedback task. Note that the trajectory is straight. 150
Fig. 8.4: Teleport task 152
Fig. 8.5: Simulation of ideal grid cells 153
Fig. 8.6: Novel analysis applied on simulated grid cells 155
Fig. 8.7: Validation of our novel method on the raw grid cell activity 156
Fig. 8.8: Estimation of gridness 157
Fig. 8.9: Novel analysis applied on noisy data. 157
Fig. 8.10: Novel analysis applied on ideal place cells 158
Fig. 8.11: Trajectories during the encode task in one representative subject. 159
Fig. 8.12: Performance in the test and feedback tasks 160
Fig. 8.13: Performance in the teleport task 161
Fig. 8.14: Detection of grid cell and place cell gradients with our novel analysis 162
Fig. 9.1: Rationale of the experiment 169
Fig. 9.2: Teleport task in the scanner. 171
Fig. 9.3: Rich stimulus set 172
Fig. 9.4: Behavioural schedules optimized for RSA 173
Fig. 9.5: Submillimeter fMRI data 174
Fig. 9.6: Inter-voxel similarity analysis 178
Fig. 9.7: RSA models for head-direction and grid-cell signals, as a function of spatial direction 180
Fig. 9.8: Comparison between the three analyses used in this chapter. 182
Fig. 9.9: The output of our GLM statistical analysis was 600 times bigger than the output of a typical study 185
Fig. 9.10: Example registrations of the hippocampal subfields 187
Fig. 9.11: Example gradient in inter-voxel similarity in CA1. 189
Fig. 9.12: Head-direction signals and grid-like hexagonal symmetry in one example subject. 191
Fig. 9.13: Example results of our novel analysis on real fMRI data from the ERH. 192

Acknowledgements

These amazing people contributed one way or another to this thesis. I've written it as I remembered all the good times we shared together that led up to this point.

Tim Behrens has made my DPhil a formative experience. Tim, thank you for being so generous with sharing your optimism and awesome insight on science and life. You helped me achieve my original goal to learn how to code, and so much more. Analyzing data together, but also being able to do my own thing, was a huge learning experience! And teaching me how to ski, fall, and get back up was quite a challenge, too ☺

Jill O'Reilly helped me discover how to carry out fMRI experiments, and how to deal with gremlins. Jill, thank you for your warm bonfires, early Christmas parties and ski moves!

Thanks to the Compneuro members for brainstorming together during lab lunches, meetings and retreats. Saad Jbabdi has made maths easy and guitar playing fun, and he always comes up with the right solution! Joe Nour used his hacking skills to push Dinoworld further. Junjing Wang helped setting up Dinoworld and training participants. Helen Barron and Mona Garvert helped me with Psychtoolbox, 7T, SPM and writing reviews on reverse inference. Emma Lawrence turned FMRIB into a massage salon whenever we needed to get brighter. Tim Muller and Alon Baram made my thesis writing time sweet. James Whittington contributed to my energy refill in China, which lasted me for the final months of thesis writing.

I thank FMRIB (now WIN) for being such a wonderful community and Heidi Johansen-Berg for an early advice to try out something different and fun.

Members of the Analysis Group helped me shift from my medical background into the world of coding. Eelke Visser showed great initiative in the registration task force. Matt Webster adapted the FSL pipelines to work on big datasets (and provided chocolate chunk cookies). Anderson Winkler, Moises Hernandez and Jonathan Hadida shared their

enthusiasm of Unix and statistics. Ludo Griffanti always found a way to FIX things and cheer me up. Dave Flitney and Duncan Mortimer offered last minute IT support.

Sharing office space with the Physics Group had its perks. Wenchuan Wu had great ideas about getting the best signal in the brain. Rob Brand has been distracting me from work whenever I got stuck – as you say, I would have finished writing much sooner had it not been for you... Johanna Vannesjo has been a constant source of kindness, as she lent me her flat while I was finishing my thesis, or as she helped me learn K-space and lindy hop. Stuart Clare made everything possible to acquire state-of-the art high resolution data in the 7T. Jon Campbell and Mike Sanders provided excellent radiographer support throughout extended piloting sessions, while chatting (and occasionally singing) about birds.

Thanks also to Carl Hodgetts and Natalie Voets for useful recommendations on hippocampal anatomy and manual segmentation. Ivan Alvarez showed me some cool demonstrations on automatic segmentation and surface analyses in Freesurfer – thanks!

Neil Burgess, James Bisby, Aidan Horner and Dan Bush deserve special thanks for helping me bring virtual reality to FMRIB. And thanks to Hamed Nili for being the RSA expert.

David Zollman has been by my side since we both started in Magdalen, and now is a fan of Easter celebrations. Miriam de Boer has been brightening up my days. Ioana and Dan Cioclu have been incredibly close even if we live far away. And Alina Scarlatescu is the heart doctor with the biggest heart.

Finally, thanks to my family for giving me the will to push that bit further, and for being supportive in every possible way. Mum, you are a fantastic listener – I always see things more clearly after talking to you. Thanks for encouraging me to listen to my intuition more and have a youthful spirit, like yours. Dad, thanks for bringing balance to Mum's advice by encouraging me to have my feet on the ground. Cris, Sorin, Lia, Tomita, Ana, Cris and Ralu, I know that you are proud of me and I can always count on you.

Chapter 1: General introduction

The aim of this thesis is to investigate precise neuronal mechanisms in behaviours that are most easily accessible in humans, such as abstract reasoning, memory, and imagination. Using sophisticated and non-invasive functional magnetic resonance imaging (fMRI) markers of cellular population codes, such as the Nobel Prize winning grid-cell code, I tested how precise cellular computations that are well-understood at the single-cell level in animals also generalize to complex human behaviours. This work has implications for understanding the mechanisms of advanced cognition in humans.

To address this aim, this thesis bridges different levels of understanding by relating work from several species, multiple brain regions and different neuroscience techniques. Consequently, **Chapters 2-5** introduce the reader to key concepts that set the context for the experimental work in **Chapters 6-9**, while **Chapter 10** presents general conclusions and future directions. In detail:

Chapter 2 is an introduction to the anatomy of the default-mode network, a set of brain regions that use the grid-cell code during advanced cognition in humans, and have shown the largest expansion and differentiation during mammalian evolution. An emphasis is made on inter-species differences in anatomical gradients.

Chapter 3 gives an overview of the role of grid cells and place cells in forming cognitive maps of spatial and non-spatial knowledge. The focus is on functional gradients at the cellular level and on other properties which are shared at the population level, and therefore, can be detected with fMRI.

Chapter 4 presents some key examples of how such population codes can be non-invasively investigated at the mesoscale in humans, with creative analysis techniques and submillimeter fMRI.

Chapter 5 explains in detail the physics and analysis methods of fMRI, and how they make the studies presented in **Chapter 4** possible.

Chapter 6 is the behavioural section of the *Stretchy Birds* study. Here, we developed a battery of computer games to train human participants to navigate continuous conceptual knowledge in two dimensions.

Chapter 7 presents the fMRI results of *Stretchy Birds*. We show a hexagonally-symmetric signal during navigation of conceptual dimensions, strongly suggestive of grid cells, in the default-mode network, which was also found during spatial navigation.

Chapter 8 is the behaviour and neural simulations section of the *Dinoworld* study. Here, we're developing a novel representational similarity analysis method for detecting place-like and grid-like signals, that is more akin to the autocorrelation used in electrophysiology studies in rodents. We show that humans can form very precise cognitive maps that allow them to know where they are located immediately, as they're teleported to various spatial locations. We also illustrate that our method can detect gradients in the sizes of place cell fields and the scales of grid cells, using simulated data.

Chapter 9 is the fMRI section of the *Dinoworld* study. Here, we built pipelines for working with big data. Using a rich stimulus set in intensively trained participants, high-resolution MRI data, and representational similarity analyses, we found preliminary evidence that place-like and grid-like signals underlie our spatial teleportation task, in the absence of movement. However, our novel analysis method is still work in progress, and we propose several plans that we hope would facilitate signal detection.

Chapter 10 draws conclusions on the previous chapters and proposes suggestions for future research.

Chapter 2: Anatomy of the default mode network

To understand the results of experiments that use functional magnetic resonance imaging (fMRI) in humans and how they relate to findings from electrophysiology recordings in animal models, we need to know how information flows between brain regions and how it translates across species. The "default mode network" is a set of brain regions where fMRI markers of grid cells were found. Despite no report in rodents of grid cells outside the hippocampal formation, direct recordings in humans have confirmed their existence in some of these areas. This chapters aims to dissect the anatomy of this brain network in primates and rodents, and highlight gradients in structural and functional connectivity. Anatomical similarities and differences between the species might explain how neural mechanisms have evolved to control behaviour.

2.1 Localizing the default mode network

Just like the motor or the visual systems, the default mode network (DMN) is a distinct brain system widely investigated in humans with fMRI (Fig. 2.1), where a group of brain areas are anatomically and functionally connected to each other (Buckner et al., 2008) and are paralleled by correlated glucose metabolism (Vogt et al., 2006) and gene expression (Richiardi et al., 2015; Wang et al., 2015). The DMN also exists in monkeys (Vincent et al., 2007) and mice (Stafford et al., 2014), although with some notable differences in the later species compared to humans.

The role of the DMN is intriguing. DMN is deactivated during most goal-oriented tasks, however it is active during rest (Smith et al., 2012) and, in humans, during many high level aspects of cognition (Smith et al., 2015), such as episodic and semantic memory (Binder et al., 2009), imagination (Schacter et al., 2012), value-guided decision making (Clithero

and Rangel, 2014), spatial reasoning (Doeller et al., 2010), theory of mind (Saxe et al., 2004) and conceptual knowledge (Kumaran et al., 2009; Constantinescu et al., 2016a). Although these high-level functions are seemingly disparate, they all involve different forms of abstract reasoning. Thus, the DMN can be identified functionally and its architecture can be informed by anatomical studies.

While subsequent chapters of this thesis will discuss how these functions are organized within the DMN, in the current chapter I ask how these brain regions are anatomically organized, how similar they are across species, and what this means when interpreting results across different species and experimental modalities.

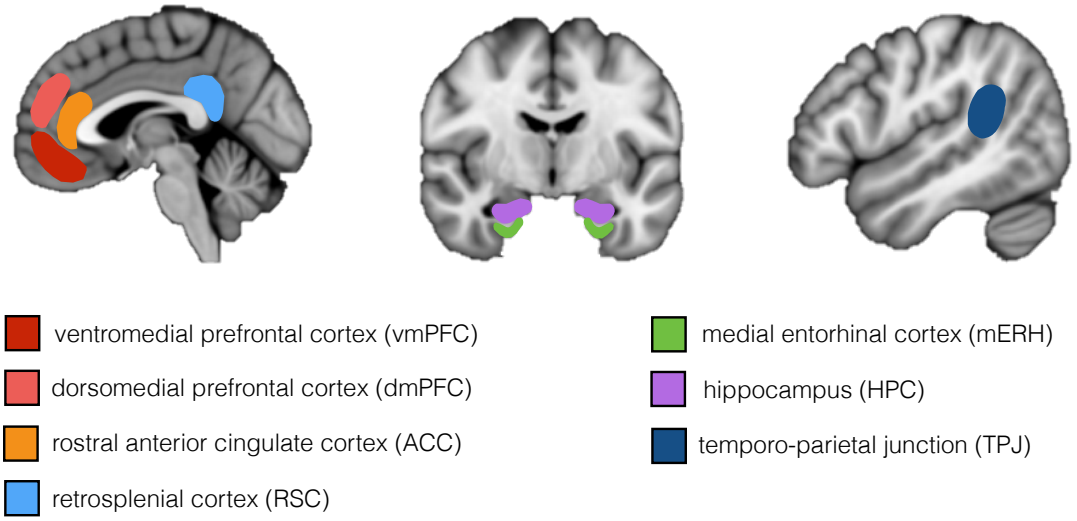


Fig. 2.1: Anatomy of the default mode network in the human brain. Adapted with permission from (Behrens et al., 2009).

2.2 Techniques for investigating the DMN neuroanatomy

The precise anatomical dissociations among the DMN regions is subject to debate in cognitive neuroscience and, in general, there are multiple ways to define a brain region, by combining a variety of research methods. For example, there are many continuous

anatomical gradients paralleled by functional gradients in the brain. I will briefly describe the strengths and limitations of most common neuroanatomical techniques.

Brodmann areas are regions in the cerebral cortex of the human, monkey and other animals, defined by their cytoarchitecture, myeloarchitecture and cell organization. The same area number in different species does not indicate homologous areas. Within a cortical area, we can describe its morphology according to the cortical layers that made up that area. For example, the prefrontal cortex has six layers because it is a neocortical region, whereas the hippocampus has three layers because it is an allocortex. In between these two types, transition regions have been identified where the number of layers increases from four to six, such as in the entorhinal and retrosplenial cortices.

The connectivity fingerprints can also define the anatomy and function of a brain area (Passingham et al., 2002). We can investigate brain connectivity using multiple complementary techniques. Two separate brain regions or two neurons can communicate directly via a monosynaptic connection, or indirectly via intermediate brain regions and polysynaptic connections, because the brain is highly wired. For example, the set of brain regions in the DMN are *functionally* (statistically) connected because their activities are correlated across time, however this can be due to either direct or indirect synaptic connections. Thus, to understand how information is passed within the DMN, we need to investigate the underlying *anatomical connectivity*. This can be done at different levels: macroscale (white matter tracts that connect brain regions formed by large number of neurons), mesoscale (fibers connecting neural columns, which are part of the same region and are formed by a small number of neurons with similar properties) and microscale (links that connect individual neurons within the same region or column). I will focus on the macroscale pathways because they form anatomical gradients that determine functional gradients within a brain region. This is interesting because we can use the bird's eye view

afforded by fMRI to measure such gradients at a global scale, and investigate how separate gradients in many different brain regions interact with each other.

Anatomical connectivity can be investigated at the macroscale with MRI diffusion tractography in humans, non-invasively (Smith et al., 2006). Tractography images white matter pathways in the whole brain simultaneously and quickly, and adds evidence to the existing atlases based largely on sparse, post-mortem observations (Behrens et al., 2003). Moreover, this technique can be used to monitor how these pathways change with experience (Scholz et al., 2009). However, tractography cannot differentiate between anterograde and retrograde connections, and thus cannot determine the direction of information flow from one brain region to the next.

Such directionality can be detected using histological studies of remote degeneration following focal excitotoxic lesions, or electrophysiological recordings at the location of the hypothesized axonal projections coming from the lesioned site. However, specific lesions are rare in the human brain, and therefore they can most often be used only in animals.

Another approach to detect directionality is to use anterograde and retrograde tract tracing, which uses dyes to label axons, and can also give information at the level of mesoscale (Sporns et al., 2005). However, tract tracing is invasive and it can only be used in animals. Moreover, it is limited to distances of tens of millimeters. Longer-distance tracts can therefore be investigated with MRI tractography or, in a limited number of cases, with brain dissections and histological studies of remote degeneration following a focal lesion (Behrens et al., 2003; Johansen-Berg and Behrens, 2006).

Finally, a modern alternative to tract tracing is to make use of genetics. For example, genetic tracing (Oh et al., 2014) uses adeno-associated viral tracers that express a green fluorescent protein either across all cells or target specific cell types, and can be combined

with 3D reconstructions. Specific circuits can be investigated using optogenetics, by injecting a virus in brain location A that tags only the direct connections that travel to location B, but not to other parts of the brain. We can then locate the tagged neurons in location B by stimulating that area with laser light with a certain frequency that is tuned selectively to those cells, but not other cells in the same location, and see if we find a neural response. Moreover, we can also use optogenetics to see if that circuit is causally involved in a specific function. Finally, a major advantage of using this techniques in humans is that they can be combined for a comprehensive dissection of the anatomy and mechanisms of different brain circuits. For example, such a multimodal study used a battery of genetic and tracer techniques to causally investigate the entorhinal network projections to the medial prefrontal and retrosplenial cortices, and their role in memory (Kitamura et al., 2017).

In the following sections, I will describe how various combinations of these techniques have been used to characterize regions that form the DMN, and I will aim to integrate among different findings to propose a common framework for the neuroanatomy of this brain network. I will focus on the parts of the DMN where neurons that contribute to the formation of cognitive maps have been recorded (see Chapter 3 on : Cognitive maps for spatial and non-spatial knowledge).

2.3 The prefrontal cortex

The part of the brain that has expanded and differentiated most during mammalian evolution is the prefrontal cortex, and is thought to contribute the most to high level cognition in humans (Passingham and Wise, 2012; Smaers et al., 2017). In this section, I will describe the anatomy of this region using the terminology of the Brodmann areas in humans and, unless otherwise specified, the connectivity profiles described in detail below

and summarized in Fig. 2.2 are based on research on non-human primates (Carmichael and Price, 1995; Ongür and Price, 2000; Mars et al., 2011; Haber and Behrens, 2014). I will explain how such comparisons between human and monkey prefrontal cortex are possible because the homologies between these species allow neuroscientists to draw analogies and make inferences for the human brain.

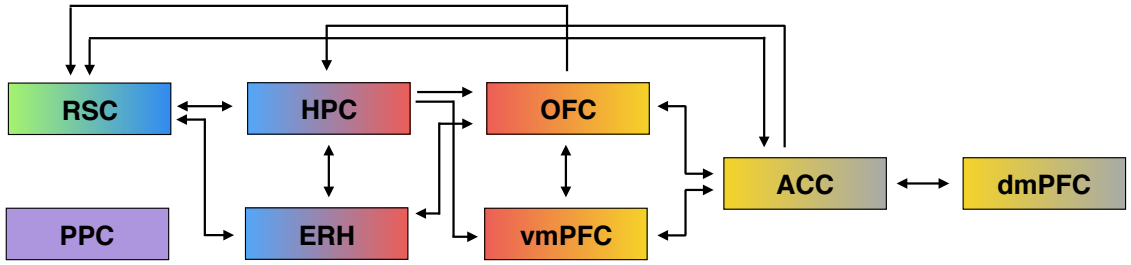


Fig. 2.2. Anatomical connectivity of the DMN in primates. Areas with similar colour gradients also have similar anatomical connectivity and functional gradients (for example, the HPC and ERH, or the OFC and vmPFC).

2.3.1 Ventromedial prefrontal cortex

The ventromedial prefrontal cortex (vmPFC) is a core region of the DMN and it corresponds to Brodmann area 14 (Rushworth et al., 2011) (Fig. 2.3). Tracing studies have shown that vmPFC receives inputs from the hippocampus (HPC), and is bilaterally connected to the orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC) (Haber and Behrens, 2014).

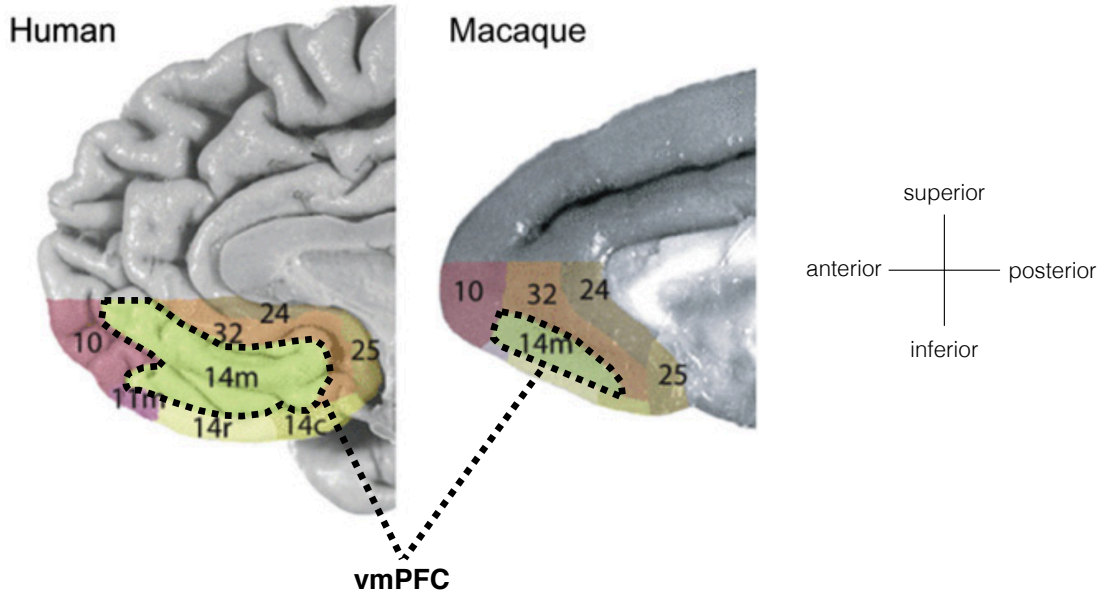


Fig. 2.3: Comparable anatomy of the human and monkey vmPFC. Adapted with permission from (Rushworth et al., 2009).

2.3.2 The orbitofrontal cortex

The OFC is a unique part of the prefrontal cortex because it receives inputs from all sensory modalities (Haber and Behrens, 2014) and has been proposed to have a flexible cognitive map of task space, representing the relevant variables in each task (Wilson et al., 2014). Although not formally part of the DMN and in general difficult to be detected with standard fMRI sequences because of signal loss, its medial subregions are in fact labeled as vmPFC. Its lateral subregions include Brodmann areas 11, 12, and 13 and, I will refer to these lateral areas as the OFC throughout this thesis (Rudebeck and Murray, 2014) (Fig. 2.4).

Lateral OFC was found to reflect precise, causal relationships between stimuli and outcomes, as measured in humans with fMRI. Lesion studies in macaques found that this behaviour was impaired when areas 11 and 13 were targeted, but not when the medial OFC/vmPFC was targeted (Noonan et al., 2010; Walton et al., 2010; Jocham et al., 2016; Noonan et al., 2017). This dissociation is important because a cognitive map of task space

requires knowledge of such precise, causal associations. In monkeys, OFC receives inputs from the HPC, sends outputs to the retrosplenial cortex (RSC), and has reciprocal connections with vmPFC, ACC and the entorhinal cortex (ERH). In rats, a recent study found that the OFC inputs to the ERH terminate in the upper layers (Kondo and Witter, 2014).

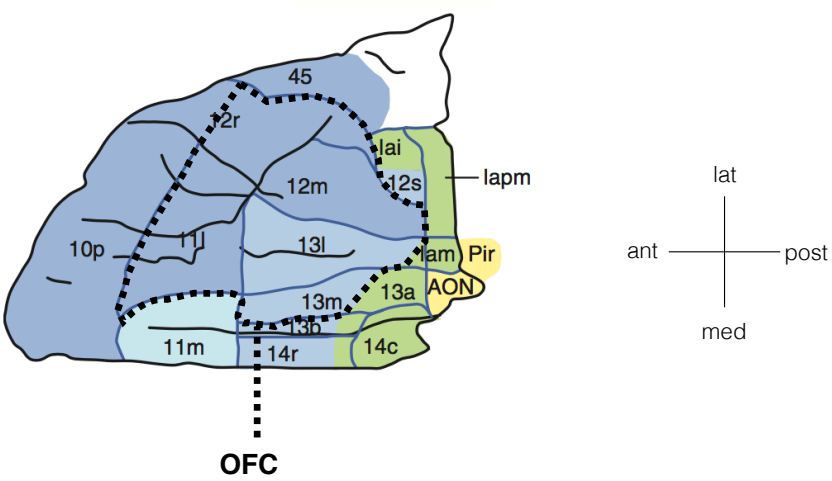


Fig. 2.4: The human OFC. Adapted with permission from (Wise, 2008).

The OFC and vmPFC have antero-posterior gradients in the anatomical connections and the underlying function (Sescousse et al., 2010; Clithero and Rangel, 2014). The connectivity of the anterior OFC and vmPFC with higher-cognitive areas, such as the prefrontal cortex and the hippocampal formation, makes sense for its role in processing higher-sensory stimuli and outcomes such as monetary rewards. This is in contrast with posterior OFC which connects to primary sensory cortex and, thus, processes primary rewards such as food.

2.3.3 The anterior cingulate and dorsomedial prefrontal cortices

Rostral ACC is another node of the DMN. It corresponds to areas 24 and 32, as well as area 25 which is labeled as a subgenual region of the ACC (Fig. 2.5, left). As mentioned above, the ACC is highly connected to the vmPFC and OFC, and it also has bilateral connections with the RSC. Even if the HPC sends many inputs to the mPFC, there are no major returning connections from the mPFC to the HPC (Vertes et al., 2007), although a recent study using optogenetics reported such a projection from the mouse ACC (Rajasethupathy et al., 2015).

The ACC has gradients along the antero-posterior axis, too (Beckmann et al., 2009). Just like vmPFC and OFC, rostral ACC is linked to areas in the PFC and is activated in reward tasks. However, in contrast to vmPFC and OFC, caudal ACC has a relative absence of primary sensory connections, and instead is tightly linked to motor control regions, thus activated during motor tasks. Thus, the ACC displays an antero-posterior gradient for reward and motor tasks. Activity in the neighbouring dorsomedial prefrontal cortex (dmPFC), which corresponds to area 9 and is linked to ACC, shows similar gradients for self-generated actions aimed to maximize reward (Walton et al., 2004).

2.3.4 Homologies between humans, monkeys and rats

How similar is the anatomy of the prefrontal cortex across different species and how does that impact its functional specialization? At the cytoarchitecture level, a clear distinction between humans, monkeys and rats has been made in the "granularity" of the prefrontal cortex, that is the number and density of cell bodies in the internal granular layer 4 (Fig. 2.5). For example, areas 10 and 14 are highly granular in humans and monkeys, but they do not exist in rodents. In fact, the rodent does not have any granular subregions of the prefrontal cortex. This distinction is thought to underlie differences in cognitive abilities

between rats and primates. Nevertheless, even if rodents lack certain areas that appeared during primate evolution, they still have homologies for many critical areas in the prefrontal cortex that evolved earlier (Wise, 2008).

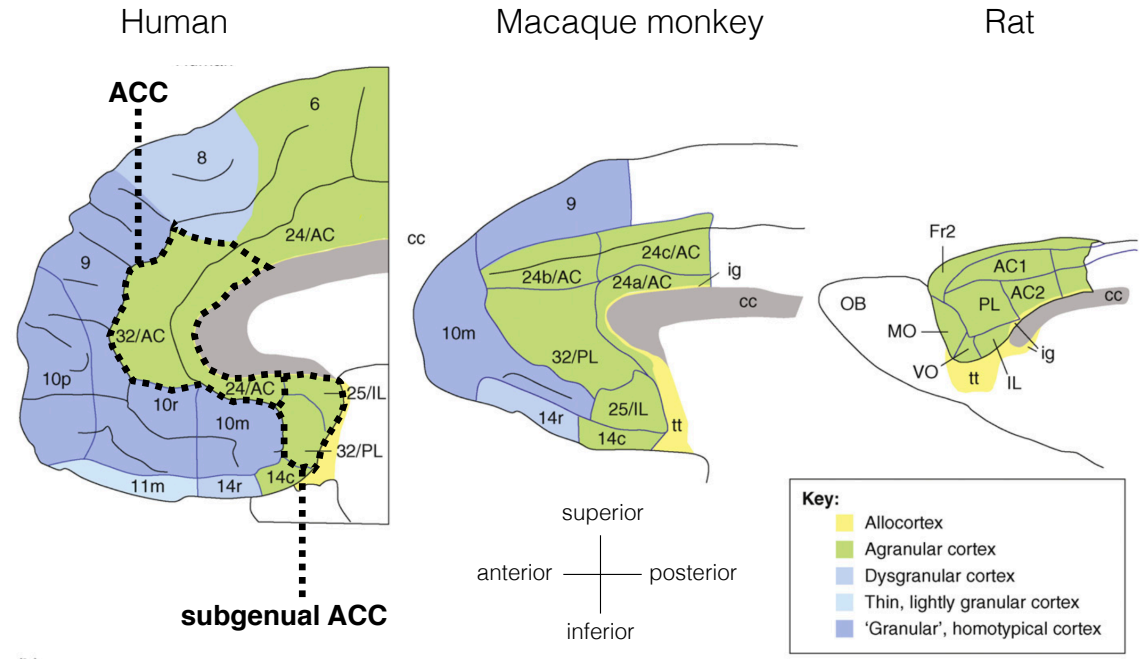


Fig. 2.5: Architectonic maps of the medial prefrontal cortex in human, macaque and rat. Note how a large part of the human and monkey prefrontal cortex is granular (shades of blue), whereas the rat prefrontal cortex is agranular (green). Adapted with permission from (Wise, 2008).

Other studies have investigated the anatomical and functional connectivity between human and monkey prefrontal cortex and found that it is very similar. A study that combined direct tracer measurements in monkeys with diffusion MRI tractography in both humans and monkeys found highly overlapping white matter tracts, illustrating homologies in anatomical connectivity between the two species (Jbabdi et al., 2013). Thus, the authors validated the use of these techniques across species as a valid strategy to determine the extent to which findings from macaques are informative of human brain anatomy. Another recent study used fMRI functional connectivity analysis and found that anatomical

subregions of the medial prefrontal cortex (including vmPFC, OFC and ACC) correspond to separate functional units, and these were also remarkably similar between humans and monkeys (Neubert et al., 2015). Taken together, these multimodal structural and functional studies show that we can make inferences about the anatomy of the human prefrontal cortex based on work done in monkeys.

2.4 The medial temporal lobe

For simplicity, I will focus on the anatomy of the hippocampus and the entorhinal cortex as the key regions in the medial temporal lobe because these are the locations where place cells and grid cells were originally found to organize spatial knowledge into cognitive maps (see Chapter 3 on : Cognitive maps for spatial and non-spatial knowledge). These brain regions are part of the hippocampal formation, which is also considered to include the dentate gyrus, the subiculum, presubiculum and parasubiculum (Amaral and Witter, 1989; Zeidman and Maguire, 2016) (Fig. 2.6). With the advance of high-field MRI, these subfields can be differentiated and visualized non-invasively in humans (see Chapter 4).

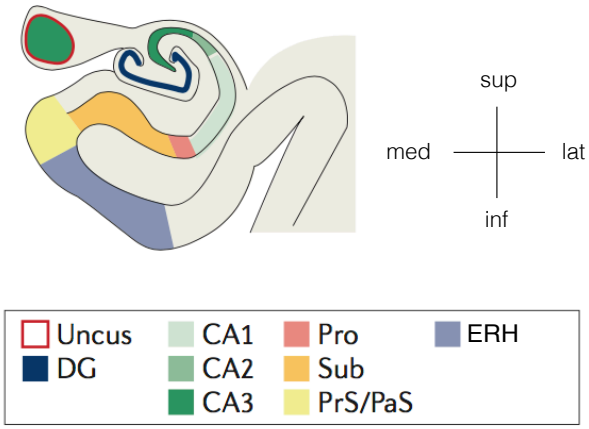


Fig. 2.6: Subfields of the human hippocampal formation. DG, dentate gyrus; Pro, prosubiculum; Sub, subiculum; PrS/PaS, presubiculum and parasubiculum. Adapted with permission from (Zeidman and Maguire, 2016).

2.4.1 The hippocampus

The hippocampus is divided into cytoarchitectonically defined subfields: the CA1, CA2, CA3 and dentate gyrus (DG) (Fig. 2.6, Fig. 2.7c). These are part of a microcircuit, where the ERH projects to the DG, which then projects to CA3, followed by CA2, CA1 and then the subiculum. The subiculum and CA1 then project back to the ERH (Strange et al., 2014). The HPC and ERH have mostly unidirectional connections originating in the ERH (Brun et al., 2002; 2008a; van Strien et al., 2009), although the HPC also sends projections selectively to the deep layers of the ERH (Derdikman and Moser, 2010a; Wheeler et al., 2015). Monosynaptic inputs from the ERH layer 3 to the CA1 have received particular interest because their disruption affects long-term spatial memory (Remondes and Schuman, 2004) and impairs the activity of place cells (Brun et al., 2008a). These results led to a set of theories and models about the formation of place cells in the HPC using inputs from grid cells in the ERH (see next chapter).

Critically, graded hippocampal connections with the ERH give rise to an anatomical gradient along the hippocampal long axis, which has a ventrodorsal orientation in rodents, and an anteroposterior orientation in primates (Fig. 2.7a, b and Fig. 2.2). Similarly, the rat ERH is also anatomically and functionally organized along a ventrodorsal axis, defined by its connections with the HPC but also other parts of the brain (Strange et al., 2014). For example, medial PFC inputs into the ventral ERH in rats will continue into the ventral hippocampus, and RSC inputs into the dorsal ERH will transition into the dorsal HPC (Witter, 1993). A homologous gradient exists in primates where, given the rotated orientation relative to the rats, it is the anterior portion of the ERH that connects to the medial PFC, while the posterior subregion is the one connected to the RSC (Aggleton, 2012; Aggleton et al., 2012; Navarro Schröder et al., 2015). The anatomical gradient in

the rodent HPC is mirrored by a functional gradient, such that place cells have firing fields that grow larger as one moves from posterior to anterior (Kjelstrup et al., 2008) (see next chapter). The experiment in Chapter 9 aims to find a similar gradient in humans.

Thus, the ERH serves as an intermediate region for pathways from the mPFC to the HPC. Even if the HPC sends many inputs to the mPFC, there are no major returning connections from the mPFC to the HPC (Vertes et al., 2007), even if a recent study reported such a projection from the mouse ACC (Rajasethupathy et al., 2015).

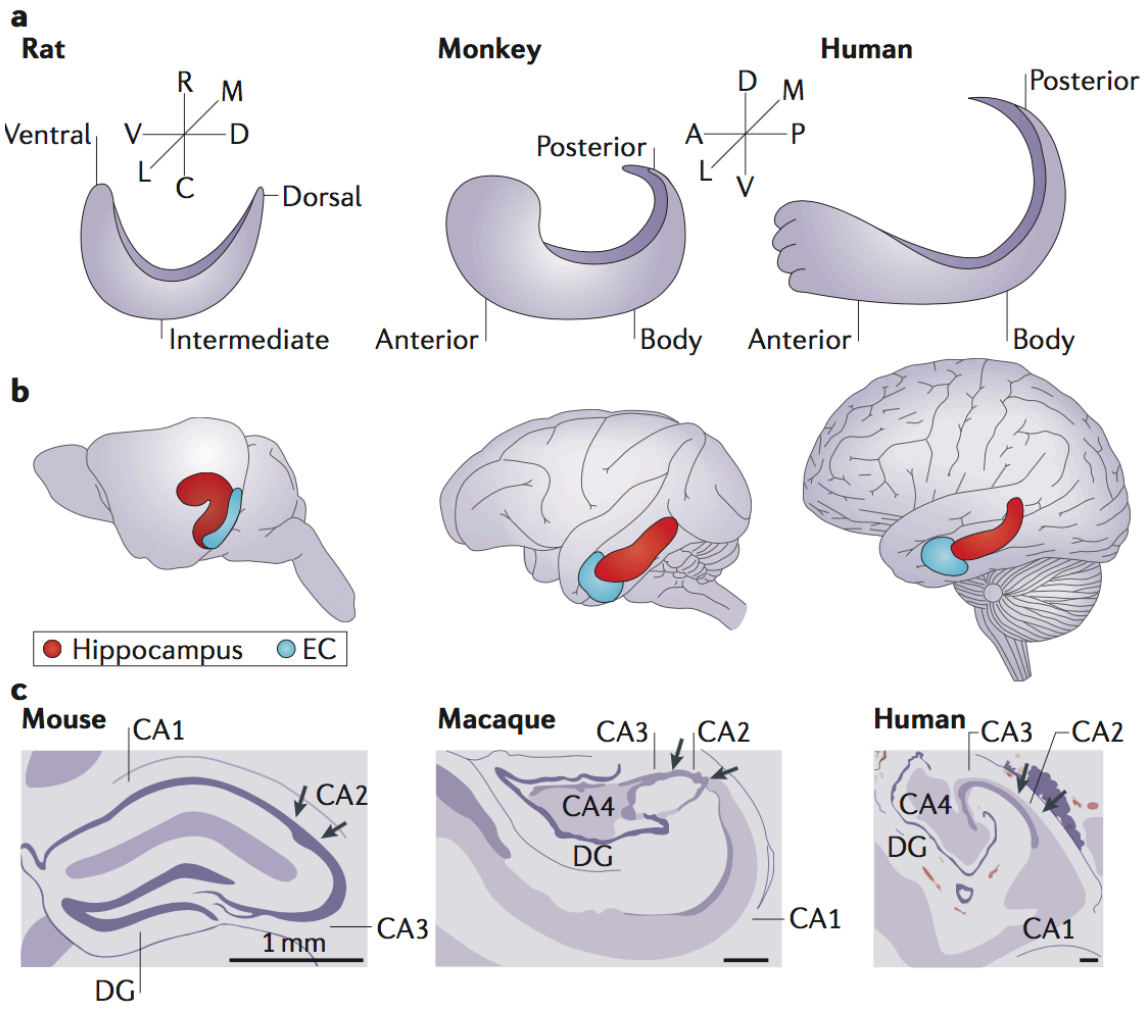


Fig. 2.7: The hippocampus in rodents, monkeys and humans. Note how the human and monkey HPC, together with the neighboring ERH, are rotated at 90° relative to their orientation in rodents. Taken with permission from (Strange et al., 2014).

2.4.2 The entorhinal cortex

Relative to rats, during evolution the human ERH has moved closer to the PFC and further away from the posterior cortical regions, and has increased in size and number of neurons (Insausti and Amaral, 2008; Strange et al., 2014). How this relates to inter-species differences in cognitive abilities, or in the granularity of the medial PFC (described earlier in this chapter) remains open to investigation.

Similarly to the HPC, the anatomical gradient in the ERH has a functional component. Here, the spacing between the grid cell firing fields, known as the grid scale, increases from the dorsal to the ventral subregions of the medial ERH (Brun et al., 2008b) (see next chapter). This means for example that, given the circuits between these neurons in the ERH and HPC (Zhang et al., 2013), grid cells with the smallest scale will connect to place cells with the smallest field. Again, we will aim to look for such gradients in humans, using the methods described in Chapter 9.

The ERH has been investigated at the level of individual layers in rodents and, recently, also in humans using high-field MRI (Fischl et al., 2009; Maass et al., 2014; Moser et al., 2014; Maass et al., 2015a) (see Chapter 4). Broadly, the ERH does not have a layer IV, and the superficial layers are the ones that give massive inputs to the HPC, whereas the deep layers receive input from the HPC (Canto et al., 2008). Layer II contains calbidin-positive pyramidal neurons that are arranged in anatomical patches with a hexagonal layout (Fig. 2.8) (Ray et al., 2014; Ray and Brecht, 2016). This is interesting because grid cells in the ERH also have a hexagonal grid pattern, at the functional level, raising the possibility that grid cells themselves are arranged in hexagons at the anatomical level (Heys et al., 2014a) (see Chapter 9). Overall, the existence of different connectivity patterns and mechanisms in different cortical layers might provide computational and developmental advantages (Raizada and Grossberg, 2003; Nikolaou and Meyer, 2015).

Moreover, laminar fMRI might give insight into brain function at a scale that is relevant for circuit dynamics.

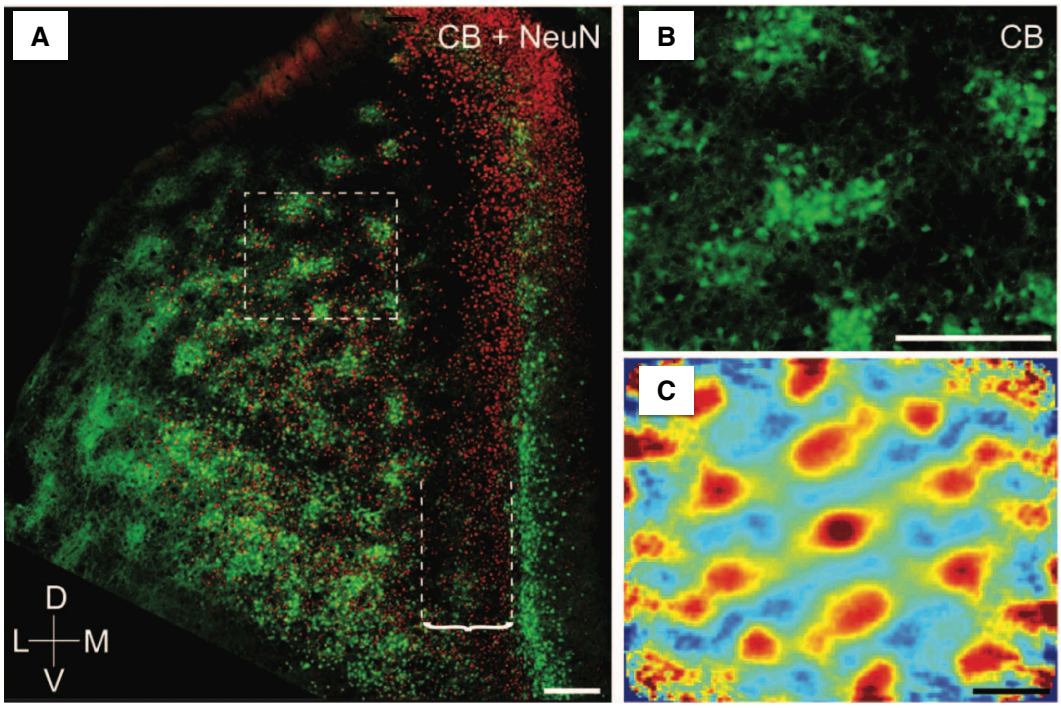


Fig. 2.8: Hexagonal anatomical patches in the rat ERH. (A) Patches of calbidin-positive pyramidal neurons in layer II have a hexagonal arrangement. **(B)** Inset of (A). **(C)** Two-dimensional spatial autocorrelation of (B) reveals the hexagonal pattern. Adapted with permission from (Ray et al., 2014).

2.5 The retrosplenial cortex

Another hub in the DMN is the RSC, corresponding to areas 29 and 30 (Fig. 2.9). The RSC is defined by its morphological and connectivity profiles that set it apart from the nearby regions. In monkeys, the number of layers increases from four to six, moving from the lateral to the medial sides of the brain, whereas the surrounding regions have six layers (Kobayashi and Amaral, 2003). Tract tracing in monkeys has illustrated bidirectional

connections with the ERH, HPC, OFC and ACC (Fig. 2.2) (Parvizi et al., 2006; Vann et al., 2009).

There are gradients in the RSC, as well. The anterior part projects preferentially to the ERH and the posterior RSC projects to the presubiculum (Sugar et al., 2011). These brain regions have head-direction cells (see next chapter on cognitive maps), raising the possibility that conjunctive grid by head-direction cell gradients might be driven by such anatomical gradients.

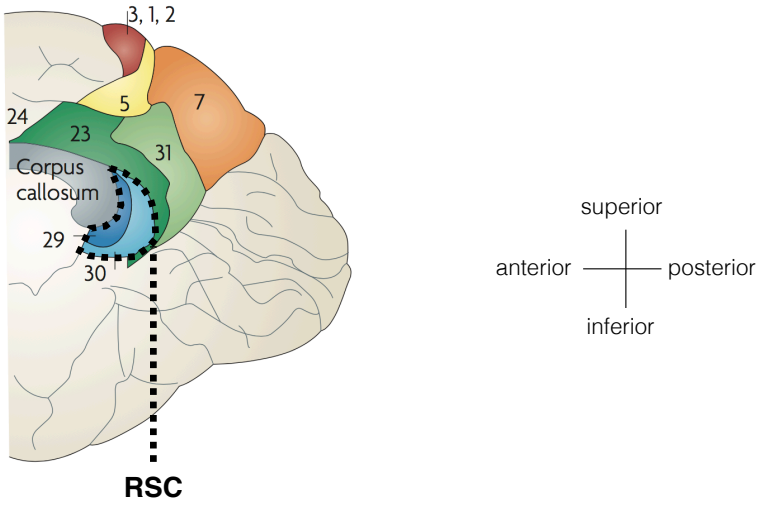


Fig. 2.9: The human retrosplenial cortex. Adapted with permission from (Vann et al., 2009).

2.6 Conclusion

This chapter presented the anatomy of the DMN, with an emphasis on the brain regions where spatially selective neurons, such as place cells, grid cells and head-direction cells, were found. The next chapter will describe the role of these neurons in forming cognitive maps.

Chapter 3: Cognitive maps for spatial and non-spatial knowledge

Spatial navigation is a great model to understand cognition at a more general level because a large body of literature has led the foundations for a bridge between single neurons, circuits and behaviours. This chapter describes well-understood cellular mechanisms for spatial and non-spatial cognition, such as the grid cell code, that served as inspiration for the human experimental work detailed in Chapters 6-9. An emphasis is made on functional gradients in the hippocampal formation, such as the grid cell scale, because such gradients can be mapped non-invasively across the whole brain in humans with careful fMRI analyses. Another focus is on the relationship between grid cell coding and physical immobility, presented from the points of view of computational models and studies of replay, preplay, and virtual reality. These findings have implications for understanding cognitive maps of non-spatial cognition and investigating their roles in novel situations that do not require movement.

3.1 Cognitive maps

How do we navigate our world and our memories? How do we take shortcuts over unseen routes and make novel inferences? A large body of literature shows that having an internal map of the world allows us to mentally navigate learned scenarios and imagine new ones. For example, if a roadblock in a well-known neighbourhood prevented us to return home, we should be able to seamlessly make a detour and reach our destination via a new path. This is known as the cognitive map theory.

In an engaging paper (Tolman, 1948a), Edward Tolman showed behavioural evidence for cognitive maps and "place learning", as opposed to "response learning". In a typical experiment, a hungry rat learns to navigate a maze to find food. One school of thought

argues that learning is driven by the relationship between stimuli (such as visual, auditory, olfactory) and the motor responses (turn left, turn right) they evoke (response learning) (Hull, 1950). Thus, specific routes that lead to a reward are frequently chosen and specific routes that lead to a punishment are often avoided, whereas the rest of the maze is considered uninformative and, hence, disregarded. However, the cognitive map school argues that the rat learns not only the trained stimulus-response pairs, but has knowledge of the entire maze which is built into a map-like representation in the brain. Tolman argued there are three main lines of evidence that support this theory. First, animals can learn the features of the environment even in the absence of rewards and, once rewards are introduced, they can find them much quicker than rats who had previously not been exposed to the same environment (latent learning). Second, animals can find a new, optimal path once a familiar one was blocked (place learning) (Fig. 3.1A, B). Third, in the natural setting, mammals can quickly find their way back to their nests after being released hundreds of miles away (O'Keefe and Nadel, 1978). These results are strongly in favour of the cognitive map theory.

However, some have proposed other possible mechanisms which might have been mislabeled as a cognitive map strategy in some contexts. For example, animals could use self-motion cues to keep track of their location relative to their home base, and not be familiar at all with the visual environment (path integration or vector navigation – red arrow in Fig. 3.1A, B) (McNaughton et al., 2006). Indeed, path integration can be done even during darkness. Path integration is egocentric, whereas map navigation is allocentric. However, these two strategies are not completely independent. Cognitive maps are formed through path integration. Moreover, these two strategies can be used simultaneously at different degrees, depending on task demands. For example, in cue-rich environments, cognitive maps would be an optimal strategy, whereas in environments with

few landmarks or in complete darkness, path integration would become the default (Fig. 3.1D) (Buzsáki and Moser, 2013).

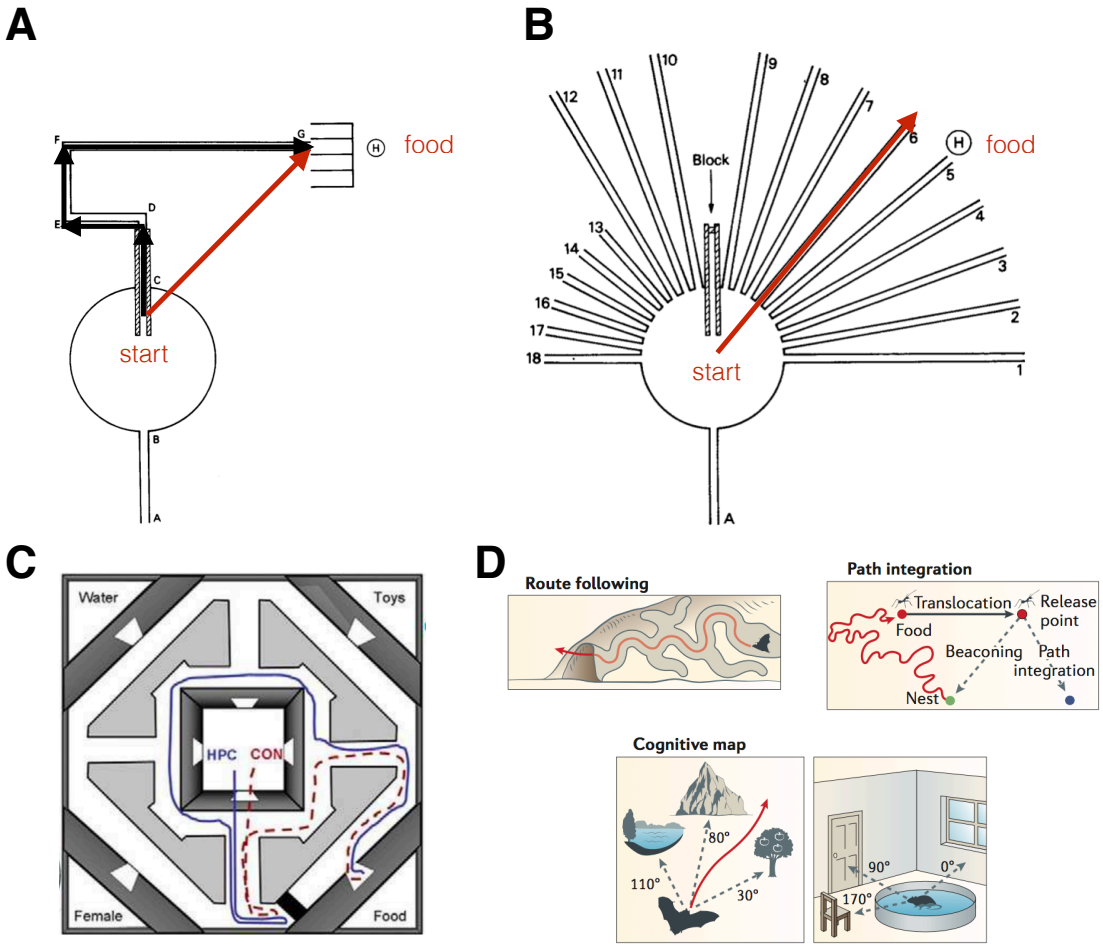


Fig. 3.1: Behavioural tests for the cognitive map. (A) Rats learn to follow a path to find food, as indicated by the black arrows. The shortest, Euclidean, distance to food is indicated by a red arrow. **(B)** When this path is blocked, rats learn to take a direct shortcut towards food, following a path they had not explored before: the sixth arm of the radial maze corresponding to the entrance to the food cup from (A). If rats followed only a stimulus-response strategy, then they should have chosen arms 9 or 10 because they were closest to the originally trained path. **(C)** After blocking the shortest path to food, control rats (CON) can find the optimal detour, whereas rats with hippocampal lesions (HPC) cannot. **(D)** Comparison between three types of navigation strategies in bats: route following, path integration and cognitive map. Route following is analogous to habitual response learning, where the bat needs to turn left or right at specific intersections. Path integration does not require any external landmarks and it only relies on self-motion cues. Thus, it can be done in darkness. If the bat uses path integration and is passively translocated by a distance *d* and an angle *a*, then when aiming to return to the nest, it would return to a location away from the nest by the same distance *d* and angle *a*. Cognitive maps combine path integration with information from landmarks. A bat with a cognitive map can locate itself in space by knowing the spatial relationships between different landmarks. Adapted with permission from (Tolman, 1948a; Winocur et al., 2010; Geva-Sagiv et al., 2015).

Some ways to investigate if animals are truly using a cognitive map strategy is to force them to make optimal detours, by blocking the shortest path to the goal (Fig. 3.1C). Other tests could be done in humans more easily. For example, we can ask human participants to recreate from memory a map of the environment or we could teleport them in random locations and test their knowledge of space (Chapter 8).

Cognitive maps are considered the most sophisticated navigational strategy that requires the highest cognitive abilities. For example, cognitive maps are thought to facilitate making novel inferences from memory (Eichenbaum and Cohen, 2014). However, cognitive maps are not precisely defined. The strongest evidence for the cognitive map comes from a neural perspective, such as place cells and grid cells, which are a focus of this thesis (see below).

3.2 Spatial maps

In 2014, The Nobel Prize in Physiology or Medicine was awarded to John O'Keefe, Edward Moser and May-Britt Moser for their discovery of place cells and grid cells "that constitute a positioning system in the brain" (Burgess, 2014) (Fig. 3.2). Together with head-direction cells and boundary cells, the hippocampal formation can flexibly represent the geometry of the environment and maintain it in long-term memory. In agreement with these, lesions of the hippocampal formation severely impair spatial navigation (Morris et al., 1982). Nevertheless, functional magnetic resonance imaging (fMRI) studies in humans have revealed that the hippocampal formation is part of a wider network of brain regions that are central to spatial navigation, including the prefrontal, retrosplenial and posterior parietal cortices (Spiers and Barry, 2015). This section discusses the neural bases of spatial and non-spatial maps.

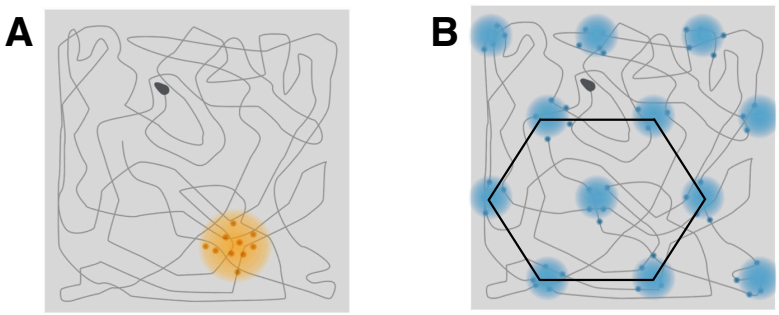


Fig. 3.2: Nobel-winning neurons. (A) As a rat runs around a box looking for chocolate treats, a place cell in the hippocampus fires whenever the rat crosses the area marked in yellow. **(B)** A grid cell in the entorhinal cortex fires when the rat crosses certain locations on the floor, marked in blue, which turn out to form a hexagonal grid. Adapted with permission from (Underwood, 2014).

3.2.1 Place cells

Initial interest in the hippocampal formation came after patient H.M. who suffered from seizures and, after surgical removal of the hippocampal formation bilaterally, he "could no longer recognize the hospital staff nor find his way to the bathroom, and he seemed to recall nothing of the day-to-day events of his hospital life" (Scoville and Milner, 1957). Motivated by the observed impact of hippocampal damage on memory, John O'Keefe recorded from single neurons in the rat hippocampus (HPC) during spatial navigation. They found that specific neurons fired only when the animal entered a specific place in the environment (O'Keefe and Dostrovsky, 1971), therefore they named these neurons "place cells" (Fig. 3.2A). Because different place cells represent different parts of space, together they can represent a continuous map of the environment, in agreement with the cognitive map theory (O'Keefe and Nadel, 1978).

Place cells are found in hippocampal areas CA1 (O'Keefe, 1976) and CA3 (Kjelstrup et al., 2008), and are likely to be pyramidal cells (Henze et al., 2000). Similar neurons were found in the HPC of other mammals, such as mice (Rotenberg et al., 1996), bats (Ulanovsky and Moss, 2007) and humans (Ekstrom et al., 2003), and they also exist when

bats are flying in 3 dimensions (Yartsev and Ulanovsky, 2013). Recently, 1-dimensional place cell signals have also been recorded in the retrosplenial cortex (Mao et al., 2017).

In rats, the topology of place cells in the brain has been thought to have no correspondence to the topology of the environment. Thus, adjacent place cells in the HPC were thought not to represent locations separated by large distances in the environment (O'Keefe, 1976; Wilson and McNaughton, 1994). However, a recent study found evidence for such a relationship as mice ran on a linear track (Dombeck et al, 2010). Another interesting feature of place cells is that their location along the hippocampal dorso-ventral axis has a correspondence to the size of place fields in the environment (Fig. 3.3). For example, the size of place fields is smallest in the dorsal HPC and can increase up to 10 folds in the ventral HPC (Jung et al., 1994; Kjelstrup et al., 2008). This functional gradient in the scale of place cells suggests that spatial environments might be represented at different resolutions across the hippocampal axis (van Strien et al., 2009).

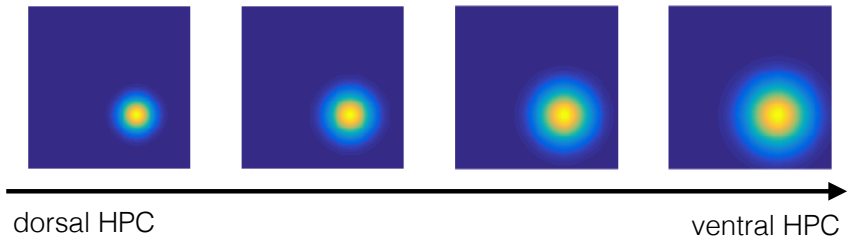


Fig. 3.3: Functional gradient in the scale of place cells. Simulated firing rate maps showing how the size of the place field increases from dorsal to ventral areas of the hippocampus. Color scales ranges from blue (no firing) to yellow (maximal firing rate).

It has been hypothesized that the HPC communicates with other brain regions to form maps of the world because the HPC alone cannot store a large number of unique spatial representations for each individual environment (Sharp, 1999). The same place cell will fire in completely different locations between different environments, a phenomenon known as remapping (Muller and Kubie, 1987). Thus, as a population, it has been shown

that place cells fire in orthogonal configurations between 11 different environments (Alme et al., 2014), implying that the HPC stores a unique map for each individual environment and this might lead to a storage problem. Based on the anatomical connectivity between the HPC and the entorhinal cortex (ERH) (see previous Chapter), it has been hypothesized that these brain regions communicate to form cognitive maps of the world.

3.2.2 Grid cells

One of the questions following the discovery of place cells in the hippocampus was how the place cell signal was generated. One synapse upstream from the HPC, in the rat dorso-medial ERH, Edvard and May-Britt Moser found neurons that also fire preferentially for certain locations in space (Fyhn et al., 2004; Hafting et al., 2005). However, unlike place cells, these entorhinal cells have multiple firing locations that take a remarkably periodic form of tessellating hexagonal grids, hence they were named grid cells (Fig. 3.2B). Moreover, whereas place cells remap between different environments, grid cells maintain the positional relationship between their firing fields (Fyhn et al., 2007). This rigid and periodic structure, together with persistent grid activity in darkness in the absence of visual cues, points to grid cells as a metric for cognitive maps not only in rats, but also in other mammals such as mice (Fyhn et al., 2008), bats (Yartsev et al., 2011) and humans (Jacobs et al., 2013). Such a metric allows the estimation of distances in the environment, for example by using the distance between the peaks of grid cells (Bush et al., 2015; Kraus et al., 2015), or the distance between the peaks of the eigenvectors obtained after the eigendecomposition of the place cell activity (Dordek et al., 2016; and section 3.2.3).

Grid cells are characterized by three main properties: scale (the distance between the grid fields), orientation (the rotation of the grid axes) and phase (the x-y coordinates of the grid fields).

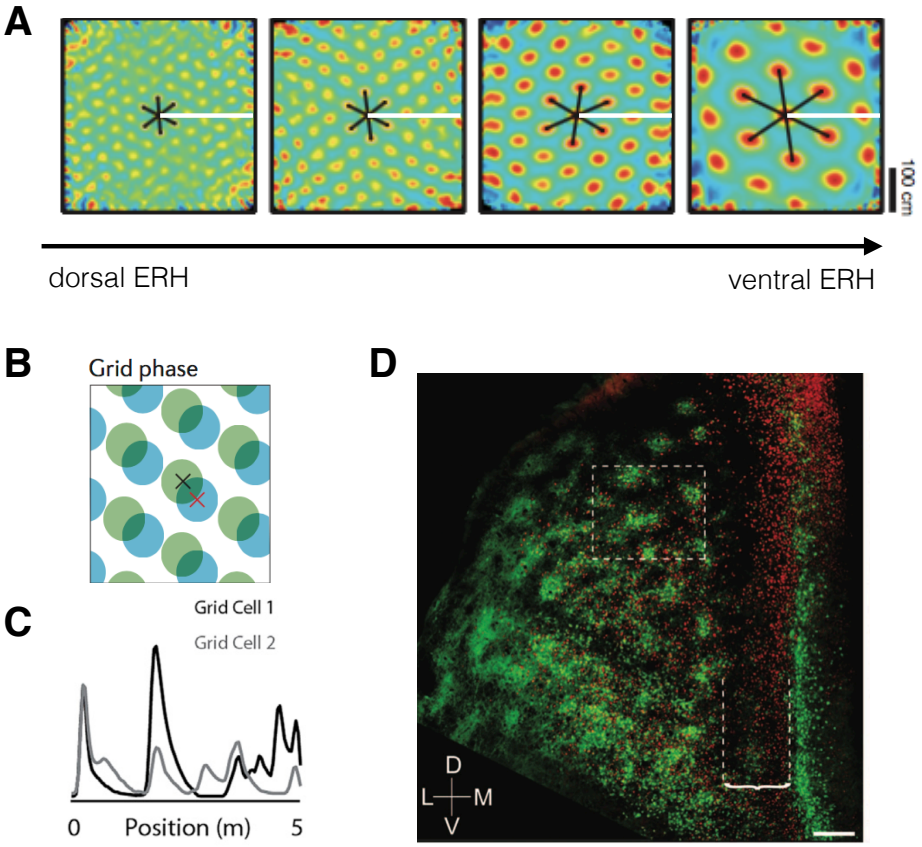


Fig. 3.4: Grid cells. (A) Autocorrelograms of grid cell activity in 2-dimensional physical space showing the discrete gradient in grid scale across different modules. In rats, small fields dominate in the dorsal ERH, and bigger fields in the ventral ERH. The black axes illustrate the orientation of the grid, with hexagonal symmetry, relative to the white line at 0°. Note how the grid orientation is similar between different modules. **(B)** Schematic illustrating the activity patterns of two different grid cells (green and blue) in 2-dimensional physical space, showing different phases. **(C)** Correlated activity of two different grid cells (black and grey) in a 1-dimensional linear track, indicative of having the same phase. These cells were recorded from nearby locations in the ERH. **(D)** Neurons in the ERH are anatomically arranged into clusters that form a hexagonal grid. It remains unknown whether the function of this neurons is grid-cell coding. Note that (A)-(C) show grid cell activity patterns in 2- and 1-dimensional physical space, whereas (D) illustrates anatomical space of the ERH. (A) Adapted with permission from (Stensola et al., 2012), (B) (Moser et al., 2014), (C) (Heys et al., 2014a) and (D) (Ray et al., 2014).

First, grid cells have a functional gradient in their grid scale that mirrors the increase in place field size along the dorso-ventral axis of the HPC. The grid scale increases in discrete steps ranging from 1.4 to 1.7 from the dorsal to the ventral subregions of the medial ERH, grouping grid cells into individual 4-5 modules of 300-500 μm in rats (Fig. 3.4A) (Barry et al., 2007; Brun et al., 2008c; Stensola et al., 2012). Given the anatomical connectivity between the HPC and ERH (see previous Chapter), place cells with the smallest fields will connect to grid scales with the smallest scales (Zhang et al., 2013). Similarly, place cells with the largest fields will connect to grid cells with the largest scales. The spacing between grid cell firing fields ranges between 25 cm to several meters in rats, and is thought to be used for decoding the location of the rat which is associated with a unique combination of active cells from different modules (Fiete et al., 2008; Stemmler et al., 2015a), and represent the environments at high spatial resolution (Mathis et al., 2012).

Second, grid cells share a common grid orientation (Fig. 3.4A) (Barry et al., 2007; Stensola et al., 2012) and this has been shown to align with the walls of a square environment (Stensola et al., 2015) or with the peripheral cues of a polarized environment (Navarro Schröder et al., 2017). Because the population of grid cells has hexagonal symmetry, a common grid axis and fire faster when the rat moves aligned with the grid axes, then we can detect a hexagonally symmetric signal at the bulk level with fMRI in humans (Doeller et al., 2010; Kunz et al., 2015; Bellmund et al., 2016a; Constantinescu et al., 2016b; Horner et al., 2016) (see also next chapter). Critically, this has revealed that the hexagonal signal is not unique to the ERH, but it is also found in prescribed parts of the medial prefrontal, medial parietal and lateral temporal cortices, which form the default-mode network. Direct recordings in epilepsy patients have confirmed grid cell patterns in some of these areas (Jacobs et al., 2013), despite no report in rodents of grid cells outside the hippocampal

formation (Boccara et al., 2010). The previous chapter highlighted some interspecies differences in neuroanatomy that might explain these findings.

Third, grid cells have different phases, slightly shifted relative to each other (Fig. 3.4B), allowing the grid population to cover the entire space. When rats navigate in different environments, grid cells maintain the same phases relative to each other (Fyhn et al., 2007), in support of the idea of a stable metric for cognitive maps. As mentioned above, grid cells are anatomically clustered into modules, where they share the same scale and orientation. Interestingly, grid cells also form smaller, hexagonal anatomical clusters (~120 μm in mice) within a module, that also share the same phase (Heys et al., 2014a; Gu et al., 2017). These cells were found to be stellate neurons (Fig. 3.4C). Similar anatomical clusters with hexagonal symmetry have been reported before in the entorhinal cortex (Fig. 3.4D), however for pyramidal neurons (Ray et al., 2014). This property could be made use of for developing novel methods to detect grid scales in humans non-invasively, using high-resolution fMRI (Chapter 9).

Grid cells are identified at the functional level, based on their hexagonal symmetry. However, studies have started to also characterize them at a molecular level and use this information to refine computational models of grid cells, and understand the role of these neurons at the circuit level. Based on their spike waveforms, grid cells are excitatory neurons. The medial ERH has two main classes of excitatory neurons: stellate cells in layer II that mature before pyramidal cells (Donato et al., 2017) found in layers II, III, V and VI (Canto et al., 2008). Grid cells are abundant in superficial layers of the ERH (Sargolini et al., 2006), whereas in deep layers they are co-localized with conjunctive grid-by-direction cells (Boccara et al., 2010), which are further presented in the upcoming section on Other space cells and parallel navigation strategies. This would suggest grid cells are mostly stellate cells, however they can be both stellate and pyramidal neurons and might have

different roles (Domnisoru et al., 2013; Schmidt-Hieber and Häusser, 2013; Tang et al., 2014; Sun et al., 2015). To investigate the grid circuit in the ERH (Fig. 3.5A), simultaneous patch-clamp recordings of stellate cells with grid properties revealed that these cells were connected via inhibitory neurons (Couey et al., 2013). Such a recurrent inhibitory circuit stands in contrast with the more typical recurrent excitatory circuit found throughout the cortex (Lovett-Barron and Losonczy, 2013). However, additional evidence shows that such a recurrent local inhibition is indeed critical for grid cell activity because optogenetic silencing of inhibitory neurons in ERH layer II disrupts the grid pattern (Cao et al., 2016). Given that these stellate cells inhibit each other, what kind of excitatory inputs to the ERH drive the formation of the grid pattern? Place cells in the HPC are a strong candidate because inactivating the HPC has been shown to abolish the grid pattern in the ERH (Bonnevie et al., 2013). These results are interesting because they are evidence in support of continuous attractor networks (Burak and Fiete, 2009), one of the computational models of grid cells that can generate and inform new research hypotheses.

3.2.3 Computational models of grid cells

How is the grid pattern formed? And why is this peculiar, hexagonal code useful?

In agreement with the recurrent inhibitory connectivity of the stellate cells in the ERH (Fig. 3.5A), continuous attractor networks with inhibitory synaptic weights can give rise to hexagonal patterns characteristic of grid cells (Burak and Fiete, 2009). In this network model, stellate cells are arranged in a 2-dimensional neural sheet according to their phase differences, such that cells with more similar phases are closer together. This is equivalent to a mutual inhibition of the stellate cells that increases with distance in neural space, thus, cells with different phases will inhibit each other more than cells with similar phases (Fig. 3.5B left for 1-dimensional and middle for 2-dimensional neural sheet for one neuron).

The inhibition is achieved via a synaptic weight matrix in the shape of a Mexican hat. The model simulates movement using inputs from speed and head direction of the rat. Together with excitatory input from the hippocampus, an attractor state emerges, defined by a periodic hexagonal pattern in the neural sheet, where certain neurons (bumps in the attractor) will be active and will inhibit their neighbours (Fig. 3.5B right). This acts like an equilibrium state, and can be transformed into a nearby attractor state given new, strong inputs that generate a new, shifted hexagonal pattern. Wider widths of the Mexican hat will lead to wider spacings between the bumps, similar to the scale of the grid cell activity in physical space. Attractor models are a powerful working hypothesis for grid cells because they might explain the self-sustaining mechanisms and stable grid patterns between different environments (Moser et al., 2014).

While attractor models illustrate the need of movement inputs for the formation of the grid pattern (Barry and Burgess, 2014), another recent study shows that hexagonal patterns can emerge even in the absence of movement (Dordek et al., 2016). The authors created a biologically plausible single-layer neural network where a single grid cell received inputs from many place cells, following a standard Hebbian rule (Kropff and Treves, 2008; Stachenfeld et al., 2014), with the constraint that their inputs were only excitatory. Unlike the recurrent continuous attractor model, this network used feedforward connectivity (Fig. 3.5C) and was also based on experimental evidence that grid cells need inputs from place cells. Indeed, inactivating place cells disrupts grid cells (Bonnevie et al., 2013) and place cells develop before grid cells (Langston et al., 2010; Wills et al., 2010). This network implemented a non-negative principal component analysis (PCA) of the place cell activity. Projecting the main principal component to the place cell activity resulted in a remarkably periodic hexagonal pattern. In other words, the hexagonal activity of one grid cell was a

weighted sum of the activity of many place cells, where these weights were precisely the principal component of this PCA (Constantinescu and Behrens, 2016).

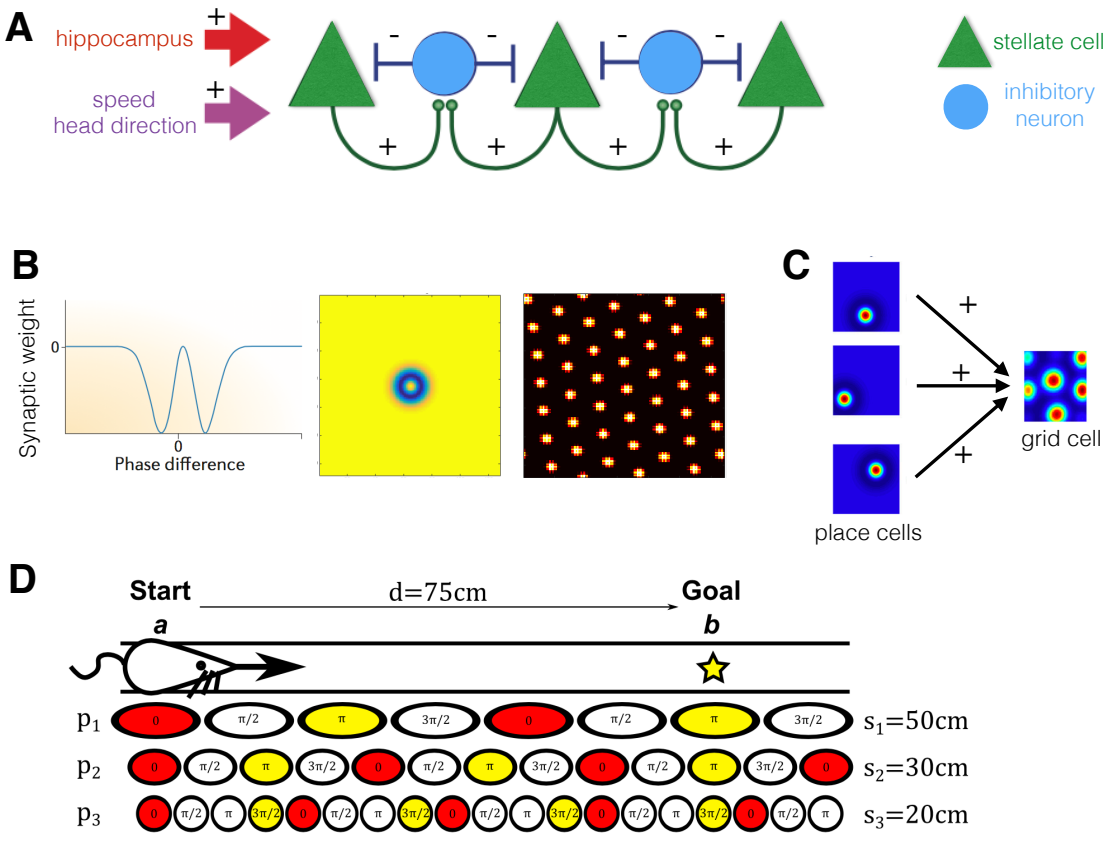


Fig. 3.5: Computational models of grid cells. (A) Neural circuits supporting the grid. Tonic excitation from the hippocampus, together with movement speed and head direction inputs, generate grid cell activity in layer II of the ERH, in a network of excitatory stellate cells connected with local inhibitory neurons, which in turn inhibit neighboring stellate cells. **(B)** Continuous attractor network using only inhibitory connectivity. Left, 1-dimensional synaptic weight matrix for one neuron, in the shape of a Mexican hat. Middle, 2-dimensional weight matrix. Note that neurons are arranged according to their phase difference in a neural sheet. Colour indicates the weights ranging from yellow (0) to blue (negative values). Right, Hexagonal symmetry in the activity of the neural sheet. Here, individual pixels correspond to individual neurons which are active (yellow) and inhibit neighbouring neurons (black). Note that these are neurons in a neural sheet space, as opposed to locations in physical space. **(C)** Neural network architecture with feedforward connectivity. The input layer corresponds to multiple place cells and the output layer corresponds to a single grid cell, given the constraint that the input weights are only positive. **(D)** Powerful, combinatorial code. Theoretical studies suggest that a spatial location can be determined by combining grids with different scales and finding the intersection (s1, s2, s3). Moreover, the distance between the animal's current location and the goal location can be modelled using both the grid scales and phases (red and yellow). (A) Adapted with permission from (Lovett-Barron and Losonczy, 2013), (B) from (Burak and Fiete, 2009; Moser et al., 2014), (C) from (Dordek et al., 2016) and (D) from (Bush et al., 2015).

Re-representing space by finding the dominant modes of variance in the data makes grid cells possible candidates for representing other types of dimensions that are not necessarily spatial, such as concepts (Constantinescu and Behrens, 2016; Constantinescu et al., 2016b). These results suggest that grid cells could support more general computations and have a broader role in memory and cognition, beyond physical space, because they can easily find the dominant variance in the data of other types of dimensions. This will be further discussed in the following section on Non-spatial maps.

Finally, why is this hexagonal code powerful (Fig. 3.5D)? Theoretical studies suggest that grid cells use a very efficient modulo code to decode unique locations (Fiete et al., 2008; Bush et al., 2015; Stemmler et al., 2015b). This potential power comes from the combination of activity from grid cells with different scales, such that each module provides a modulo code for location with modulus equal to the grid scale (Burgess, 2014). A small number of modules can have an exponential coding capacity, much higher than that of place cells. For example, N place cells could represent maximum N unique locations, whereas N grid cells could represent e^N locations. Moreover, by combining different scales with different phases, grid cells could also estimate distances between the current and goal locations (Bush et al., 2015).

3.2.4 Grid cell coding in virtual reality

How do computational models that predict grid cell activity in the absence of locomotion (Kropff and Treves, 2008; Dordek et al., 2016) agree with experimental data? How does an animal know where it is when it stops moving? Virtual reality (VR) experiments have offered some insight into these questions.

Spatial navigation and activity of the hippocampal formation are influenced by three broad set of cues (Ravassard et al., 2013) and various combinations of them have been used in

VR studies. First, distal visual cues are stationary landmarks that define the appearance of an environment. Second, self-motion cues are derived from physical movement, and include vestibular (head and/or body rotation), optic flow (the apparent motion of objects during the subject's movement) and proprioceptive cues (body awareness in space based on information coming from muscles and joints). Third, other sensory cues can be smell, sound and touch. Cognitive maps are thought to be driven by a combination of these cues which are available in the real world (O'Keefe and Nadel, 1978) implying that, together, they can trigger stronger neural responses. However, similar neural patterns continue to exist in VR, even though at a lesser extent (Ravassard et al., 2013).

It is well-known that grid cells use self-motion cues for navigation, making them good candidates for the path integration component of cognitive maps (Rowland et al., 2016). Any path integration mechanism requires running speed as a major input and, indeed, grid cells fire more when rats are running faster (Sargolini et al., 2006; Wills et al., 2012). However, it is also true that grid cells rely on landmarks and other sensory cues to a great extent (Moser et al., 2014; Pérez-Escobar et al., 2016). For example, the grid axis rotates with peripheral landmarks (Hafting et al., 2005), it aligns to the walls of square environments (Barry et al., 2007; Stensola et al., 2015) and to the cues of polarized environments (Navarro Schröder et al., 2017), while the grid fields are deformed by the shape of the environment (Krupic et al., 2015). This points towards a role of grid cells in the absence of physical movement, as it is often the case in VR.

VR enables well-controlled manipulations of space that are difficult or impossible to achieve in the real world (Doeller et al., 2012), such as 'teleportation' of a subject to a different place in the virtual environment (Jezek et al., 2011), or allowing human experiments with fMRI that require the head to be still. The first evidence of grid-like coding in the human brain was achieved using VR combined with fMRI (Doeller et al.,

2010). Participants played a computer game where they navigated a virtual arena and performed a spatial memory task (Fig. 3.6A). A follow-up study used the same analysis paradigm and found impaired grid-like coding in patients at genetic risk for Alzheimer's (Kunz et al., 2015). Neither of these experiments involved physical movement, therefore participants relied only on landmarks and optic flow to infer movement and spatial location, instead of proprioceptive and vestibular cues. Interestingly, grid-cell coding was also found to underlie imagined movement (Horner et al., 2016) and, furthermore, even mental simulation in the absence of imagined movement (Bellmund et al., 2016a) (the next chapter will present these techniques in detail).

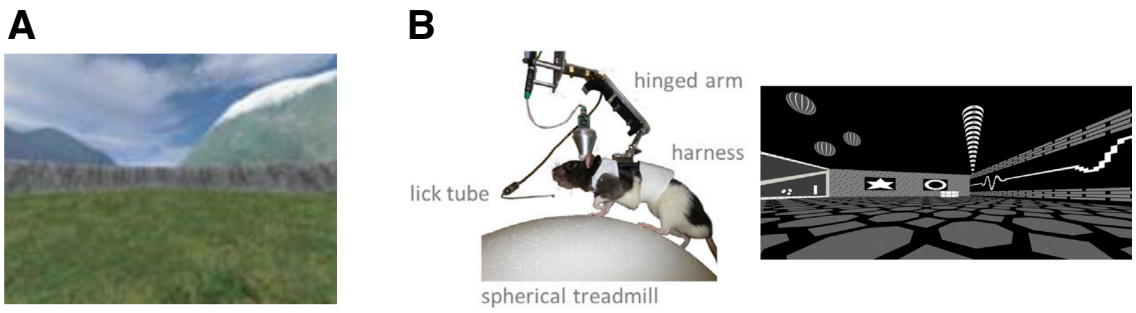


Fig. 3.6: VR in humans and rodents. (A) Human participants are playing video games while their neural activity is recorded with non-invasive techniques, such as fMRI. Note the peripheral landmarks (grassy hills and mountains) in the example screenshot. Participants press buttons to navigate and receive monetary rewards at the end of the scanning session. **(B)** Rodents are playing video games while their neural activity is recorded with electrodes or two-photon microscopy. The animals approximate natural motion by running on a large ball (spherical treadmill), while being secured with a harness. The speed and direction of locomotion are immediately translated into a visual projection on a wraparound screen that covers its visual field and displays the game. Rewards are delivered via a lick tube. Here, the rat can rotate its whole body during electrophysiology recordings. For intracellular recordings or two-photon microscopy, the animal's head need to be fixed. (A) adapted with permission from (Doeller et al., 2010) and (B) from (Aronov and Tank, 2014).

VR is also used in animals (Fig. 3.6B) to investigate the contribution of different types of cues to spatial navigation, or to add other techniques that require the head to be stationary, such as intracellular recordings and two-photon microscopy, to the existing repertoire of

experimental methods for grid cells. Advances in two-photon fluorescence microscopy and calcium imaging have enabled the ability to observe the activity of many genetically identified neurons at once (Chen et al., 2013). This is a shift from measuring a smaller number of neurons with electrophysiology, to investigating neural networks (Moser et al., 2014). In contrast, intracellular recordings allow individual synaptic measurements.

Such techniques have been used successfully to detect grid cells in VR. In 1-dimensional VR, calcium imaging (Low et al., 2014; Heys et al., 2014a) and intracellular recordings (Domnisoru et al., 2013) have been used to image grid cells while mice had their bodies and heads fixed. This suggests that, in a single dimension, rodent grid cells are active without using vestibular inputs coming from body rotations. Grid cells can also be recorded in 2-dimensional VR (Aronov and Tank, 2014). However, here they seem to require vestibular inputs, in agreement with an impaired place cell activity when rats navigate similar environments but with their bodies fixed (Aghajan et al., 2015). There is no such requirement for 2-dimensional VR in place and grid cells when humans navigate without vestibular inputs, using only landmarks and optic flow (Ekstrom et al., 2003; Jacobs et al., 2013). Possible explanations might be that humans can solve conflicts between cues and body movement more easily or they can pay attention to the task better than rodents who, for example, have disrupted grids when they are passively displaced in the environment without receiving rewards (Winter et al., 2015), hence without being motivated to pay attention to where they are in space.

3.2.5 Replay and preplay

While animals are immobile during sleep or wakefulness, place cells in the HPC activate briefly in the same sequence as they did during prior behavior, but much quicker. This is known as *replay* (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996;

Johnson and Redish, 2007), and occurs together with bursts of synchronous activity in the HPC, known as sharp-wave ripples (Buzsáki et al., 1983) (Fig. 3.7A). The order of the sequences is only forward during sleep (Skaggs and McNaughton, 1996), and both forward and backward during wakefulness (Foster and Wilson, 2006; Diba and Buzsáki, 2007; Davidson et al., 2009). Interestingly, if a specific place cell is replayed during sleep, then pairing that event with rewarding stimulation will cause mice to show a strong preference to the location in the environment corresponding to that specific place cell (de Lavilléon et al., 2015). This is direct evidence for a causal role of place cells and replay in navigation. Moreover, replay is believed to have a role in memory consolidation (Nakashiba et al., 2009) and the formation of cognitive maps (Derdikman and Moser, 2010b; Ólafsdóttir et al., 2017).

Replay sequences are analyzed using Bayes' theorem (Derdikman and Moser, 2010b). Briefly, using electrophysiology recordings, it is possible to compute the firing probability of each place cell when the rat is in a given position in the environment. With Bayes, we can ask the inverse question and find the probability of the rat being in a certain location given that place cells fired in a certain sequence (Fig. 3.7B). Thus, we can use this approach even when the rat is immobile while thinking about a trajectory, by decoding the trajectory from the neural activity.

Grid cells replay, too (Fig. 3.7D-F). In the deep layers of the medial ERH, they are coordinated with place cell replay (Ólafsdóttir et al., 2016a) whereas, in the superficial layers, they are replaying independently from place cells (O'Neill et al., 2017). Nevertheless, replay mechanisms have also been found in other parts of the brain outside of the hippocampal formation, such as the medial prefrontal cortex (Euston et al., 2007; Peyrache et al., 2009) and visual cortex (Ji and Wilson, 2007).

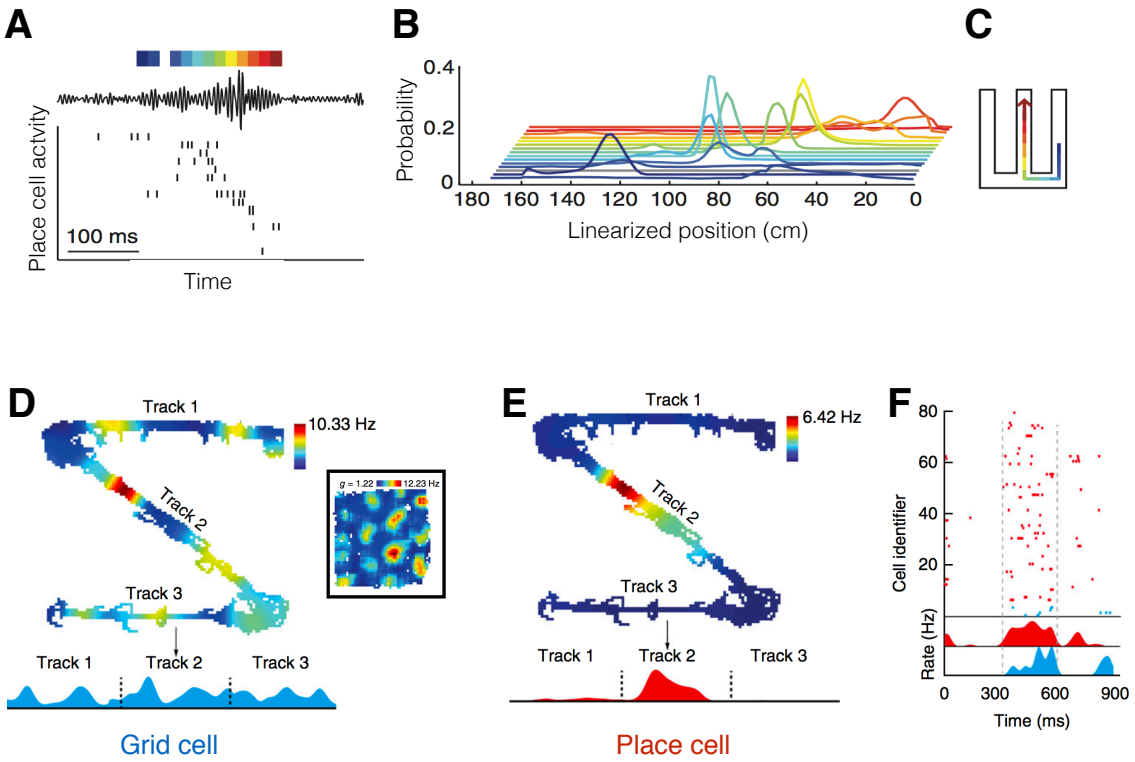


Fig. 3.7: Replay in place cells and grid cells. (A) Sequential firing of place cells during sharp wave ripples, shown in a black local field potential on top. Bottom, spike raster plots show sequential activity in a compressed timescale. The colorbar on top shows the colours associated with each time bin used for decoding. **(B)** Probability distribution of decoded locations, obtained with Bayes methods. **(C)** Schematic of the replayed trajectory, based on the decoded trajectory. **(D)** Example grid cell activity in a linear maze with three tracks. Note the periodic firing pattern. The inset shows the characteristic hexagonal pattern, as a validation in a two-dimensional environment. **(E)** Example place cell activity in the same maze. **(F)** Example replays of a grid cell from deep ERH layers (blue) and a place cell (red) happening at the same time. (A-C) Adapted with permission from (Carr et al., 2011) and (D-F) from (Ólafsdóttir et al., 2016b).

Preplay is a different process that refers to temporal sequences of firing of place cells also during periods of rest or sleeping, that precede a trajectory in a novel environment and correlate with the sequences during physical exploration of the novel environment (Dragoi and Tonegawa, 2011; Ólafsdóttir et al., 2015). This contrasts with replay, which recapitulates place cell sequences that have already occurred during previous exploration. Preplay is thought to underlie planning, imagination, and future thinking (Hassabis et al., 2007; Schacter et al., 2012).

Finally, a set of neurons in area CA2 of the hippocampus resemble place cells from areas CA1 and CA3, but are more active when the animals are immobile during wakefulness or sleeping, rather than during movement (Kay et al., 2016).

The persistence of firing of these neurons in the hippocampal formation, during VR, sleep, and awake rest, is exciting because it raises the possibility that they have evolved for a more general role in cognition that does not require physical movement (see section of Non-spatial maps).

3.2.6 Other space cells and parallel navigation strategies

Grid cells are co-localized in the brain together with other spatially selective neurons, which also exist in other parts of the brain (Fig. 3.8). In 2-dimensional environments, head-direction cells are neurons that act like a compass (Fig. 3.8A), being active whenever the subject's head faces a particular direction in the horizontal plane (Taube et al., 1990a). Different cells have different preferred directions such that, at the population level, all directions are equally represented (Taube et al., 1990b). In 3-dimensional environments, these neurons are also tuned to pitch (moving the head upwards or downwards) and roll (rotating the head along an axis from nose to tail) (Finkelstein et al., 2015). Originally found in the dorsal presubiculum (postsubiculum), head-direction cells are also found in in parts of the default mode network, such as the deep layers of the ERH, retrosplenial cortex and posterior parietal cortex (Grieves and Jeffery, 2017), where grid-like hexagonally symmetric signals have also been measured (Doeller et al., 2010; Constantinescu et al., 2016b). Moreover, a separate class of head-direction cells have bidirectional firing patterns (Jacob et al., 2017).

Boundary/border cells in the entorhinal cortex (Fig. 3.8B) are neurons that respond when the subject is close to environmental boundaries (Barry et al., 2006; Solstad et al., 2008;

Boccara et al., 2010), and have also been found intermingled with grid and head-direction cells in the pre- and parasubiculum (Boccara et al., 2010).

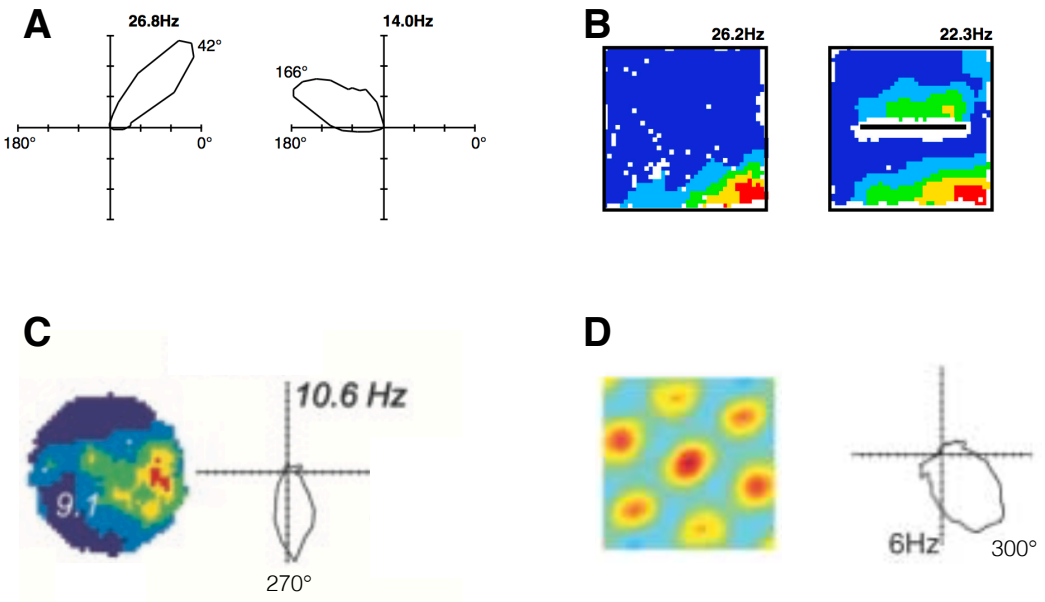


Fig. 3.8: Head-direction cells, boundary/border cells and conjunctive cells. (A) Two head-direction cells. The polar plots show firing rate as a function of head direction. For example, the cell of the left fires most when the animal's head is oriented at 42°. **(B)** A boundary vector cell that fires whenever there is an environmental boundary a short distance to the south. Introducing a wall in the environment causes a new border field to appear. **(C)** A conjunctive place by direction cell that prefers a location in the eastern side of the environment, whenever the rat's head is oriented at 270°. **(D)** A conjunctive grid by direction cell that shows hexagonal symmetry and prefers heading directions of 300°. (A-B) Adapted with permission from (Barry and Burgess, 2014), (C) from (Cacucci et al., 2004), and (D) from (Sargolini et al., 2006).

Some of these spatial cells have mixed selectivity (Hardcastle et al., 2017), therefore they fire in response to a conjunction of stimuli, hence the name conjunctive cells (Fig. 3.8C, D). For example, conjunctive place by direction were found in the dorsal presubiculum (Cacucci et al., 2004) and grid cells modulated by direction and speed in the deep layers of the ERH (Sargolini et al., 2006). This has implications for fMRI because it shows that a complex interplay between different neurons gives rise to a bulk, population response that controls behavior. Moreover, mixed selectivity is a characteristic of the association

cortex and it has the advantage of combining multiple types of inputs at the same time, to give rise to advanced cognition (Rowland et al., 2016). Thus, such a neural diversity in representing space in the brain can give rise to different neural mechanisms that operate in parallel to guide spatial and non-spatial cognition.

In agreement with the anatomical connectivity of the default-mode network (see previous chapter), other parts of this network apart from the hippocampal formation are often activated during spatial navigation (Vann et al., 2009; Doeller et al., 2012; Spiers and Barry, 2015) and have been shown to use a grid-like hexagonally symmetric code to organize spatial knowledge (Doeller et al., 2010). First, the ventromedial prefrontal cortex can generalize experiences to novel situations and, for example, construct novel, never experienced rewards (such as tea-jelly) by combining the values of their individual components (tea and jelly, separately) (Barron et al., 2013a).

Second, the nearby orbitofrontal cortex has been hypothesized to represent the structure of the task environment into an abstract cognitive map (Wilson et al., 2014; Schuck et al., 2016; Wikenheiser and Schoenbaum, 2016). In agreement with this theory, this brain region is essential for making precise associations between stimuli and outcomes (analogous to representing locations in physical space), and inferring information about the structure of the task that has never been experienced before (analogous to making shortcuts through unexplored parts of space) (Haber and Behrens, 2014). Indeed, these brain regions are often activated together with the medial temporal lobe during tasks that require making inferences, and imagining novel situations (Hassabis and Maguire, 2007). Moreover, the medial prefrontal cortex mediates between two parallel learning systems for spatial navigation: a hippocampal system that learns relative to a fixed environmental boundary, and a striatal, habitual system that learns relative to a landmark that changes position (Doeller and Burgess, 2008; Doeller et al., 2008a).

Third, the retrosplenial and posterior parietal cortices send information about navigation cues to the ERH (Epstein, 2008; Vann et al., 2009). Moreover, it has been proposed that the retrosplenial cortex can switch between allocentric representations (environment-bound; North, South), mediated by the medial temporal lobe, and egocentric representations (to my left, to my right), mediated by the parietal cortex (Vann et al., 2009). These results suggest that different parts of the brain communicate and bring their contribution to form cognitive maps of space. Interestingly, this same network is also regularly activated during non-spatial tasks, when the same brain regions communicate to underlie other types of cognitive functions, such as semantic memory (Binder et al., 2009), imagination (Schacter et al., 2012), valuation (Clithero and Rangel, 2014), and theory of mind (Saxe et al., 2004). All these seemingly disparate functions require the use of conceptual knowledge which, together with the finding that the same brain network uses a grid cell code to organize spatial knowledge (Doeller et al., 2010), inspired our hypothesis and findings of grid cell coding for organizing conceptual knowledge (Constantinescu et al., 2016b). This, together with other findings showing a role for cognitive maps in organizing non-spatial dimensions, is discussed next.

3.3 Non-spatial maps

It has been hypothesized that cognitive maps allow conceptual relationships to be navigated in a manner similar to that of space (Tolman, 1948b) and, thus, place cells and grid cells have a more general role for non-spatial cognition (O'Keefe and Nadel, 1978; Buzsáki and Moser, 2013; Eichenbaum and Cohen, 2014). Thus, just like cognitive maps allow taking shortcuts through unexplored spaces, they might also explain why humans have a remarkable ability to generalize their experiences to novel situations. For example,

if someone is trying a sport for the first time, and they like being active and spending time with a small group of friends, then they might infer that cycling or tennis is more fun than football. However, if one wants to step out of their comfort zone, then they might try playing football from time to time and discover that could be fun, too. This thesis has two experimental chapters dedicated to gridlike signals for navigation of 2-dimensional conceptual maps (Chapters 6 and 7). Therefore, in this section I will focus on single-cell results for 1-dimensional navigation of time and sound-frequency, as well as on concept cells, and I will discuss the role of cognitive maps for relationships that are discrete, rather than continuous.

3.3.1 Time cells

Time and space often go together: events happen at a particular time and in a particular space. Therefore, to investigate a role of the hippocampal formation in representing time, Howard Eichenbaum's lab used different behavioural tasks that aimed to dissociate time and space, by minimizing movement and distance traveled in rodents.

One of the earlier behavioural tasks involved rats learning that specific objects were paired with particular odours delivered 10 seconds later in a different location (Fig. 3.9A). Because the rats learnt to anticipate the correct odour when they were presented with an object, the delay period could be used to look for a sequence coding mechanism for time, similar to the replay mechanism seen for space (see the previous section on Replay and preplay). Such temporal signals were indeed observed (MacDonald et al., 2011). However, they could have been confounded by a representation of spatial location because the rat needed to move from the location of the object to the location of the odour. To account for this confound, a different behavioural task has been used instead, where rats were running in place, on a treadmill, during the delay (Fig. 3.9B).

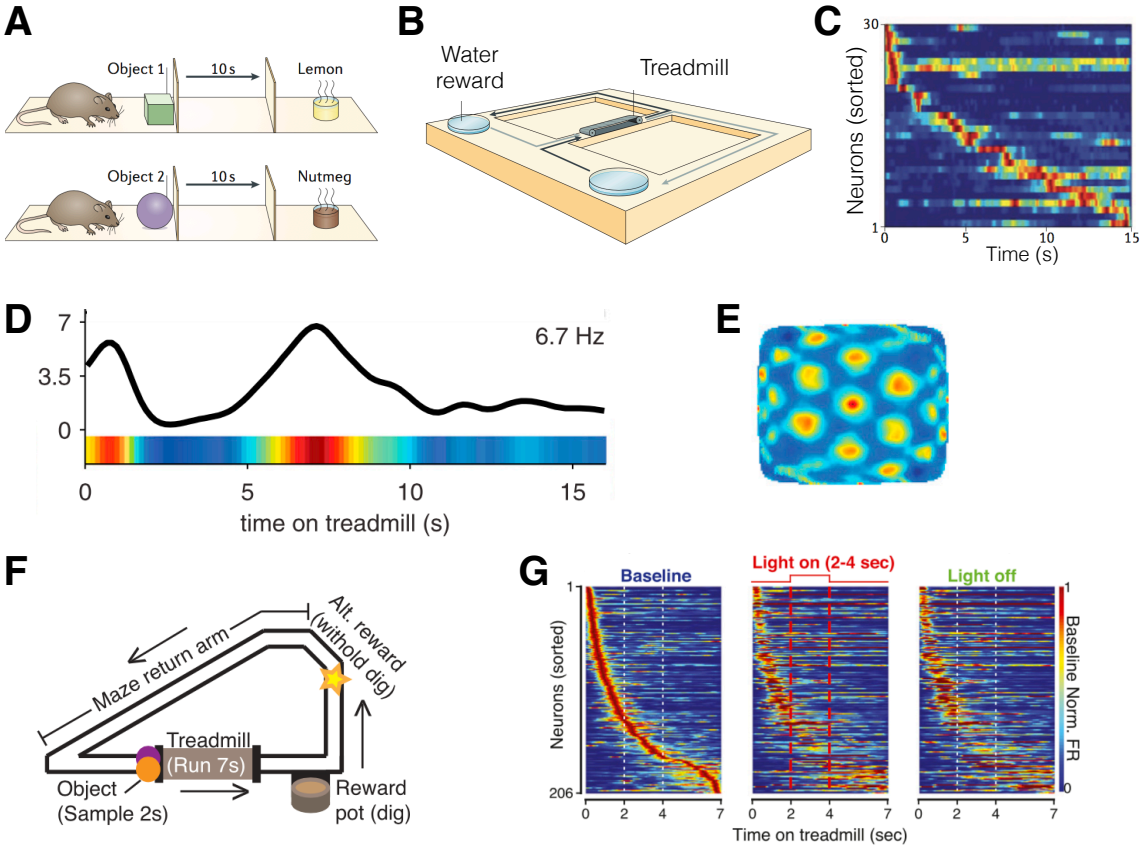


Fig. 3.9: Time coding in the hippocampal formation. (A) Delayed stimulus-outcome association task. Here, rats are rewarded if a lemon odour (as opposed to a nutmeg one) is delievered 10 seconds after the presentation of object 1. Rats are also rewarded if a nutmeg odour (as opposed to lemon) is delivered after the presentation of object 2. **(B)** Treadmill running task. Rats run on a treadmill, at different speeds, and then alternate between the two arms of the maze, as indicated by the black and grey arrows. **(C)** Hippocampal time cells during the treadmill running period. Each row represents once cell that fires at a different time and, together, they cover the entire delay period. **(D)** One time cell in the medial ERH during the treadmill delay. Note the multiple peaks, similar to the activity of one grid cell in 1-dimensional spatial environments (Yoon et al., 2016). **(E)** Same cell during navigation in a 2-dimensional environment showing a hexagonal pattern characteristic of grid cells. **(F)** Stimulus-reward association task with three stages: object sampling, treadmill running in place and spatial navigation in a 1-dimensional maze return arm. **(G)** Silencing the medial ERH during the treadmill stage impairs time cells in the HPC. (A-C) Adapted with permission from (Eichenbaum, 2014), (D-E) from (Kraus et al., 2015) and (F-G) from (Robinson et al., 2017).

To decouple elapsed time from the distance run, the speed of the treadmill was externally controlled (Kraus et al., 2013). The hippocampal activity during these task delays showed time cells (Fig. 3.9C), representing the amount of time the rat spent on the treadmill. If

these neurons were pure place cells, they should have been active during the entire delay period while the rat did not change location (Kraus et al., 2013). Critically, similar time cells were also found in the medial ERH (Fig. 3.9D) and the same cells showed hexagonal symmetry while rats navigated freely in 2-dimensional environments (Fig. 3.9E), suggesting that grid cells also code for time (Kraus et al., 2015).

These findings led to the hypothesis that medial ERH might have a non-spatial, rather than spatial, role in HPC coding, by temporally organizing HPC representations (Robinson et al., 2017). This might be possible because manipulations that impair grid cells do not also impair place cells (Koenig et al., 2011; Brandon et al., 2014), and near-complete lesions of the medial ERH result in broadening but not absence of place cell fields (Hales et al., 2014), however they do impair theta phase precession in the HPC (Schlesiger et al., 2015). To dissociate between time, space and object coding the medial ERH and its effect on the HPC, the medial ERH was silenced with optogenetics bilaterally during different stages of a task, while recordings were made from CA1 during the entire duration of the task (Robinson et al., 2017). Rats did a stimulus-reward association task that included three key stages: object sampling, treadmill running in place and spatial navigation on a 1-dimensional linear track (Fig. 3.9F). The results suggest that the medial ERH only during the treadmill running period is critical for making stimulus-reward associations. Moreover, silencing the medial ERH only during this period disrupts time cells in the HPC (Fig. 3.9G). However, silencing the medial ERH at any stage during the task does not impair object coding during the object sampling task, nor place coding during the spatial navigation task. The authors conclude that the role of the medial ERH in HPC coding is in temporal coding, rather than space or object coding (Robinson et al., 2017).

The positive results of this study are very interesting, showing that the medial ERH, probably through its own time cells, is critical for memory and time cells in the HPC. Also,

it makes sense not to observe an effect for object selectivity because that would rely on the lateral, but not the medial ERH. However, the study is not well-controlled for the spatial component of the task because, first, this stage in the task is not relevant for making the stimulus-reward associations, and second, the maze route was overlearnt and rats could navigate without needing to continuously pay attention and integrate information from landmarks and movement to locate themselves in space. Indeed, rodent place cells remain active during passive displacement that is task irrelevant (Terrazas et al., 2005), whereas grid cells are disrupted (Winter et al., 2015). Thus, whereas the non-spatial results are compelling, the findings do not rule out a spatial role of the medial ERH in HPC coding.

The finding of time cells poses the interesting question of whether these time fields expand across the dorso-ventral axes of the HPC and ERH, in a similar manner to the functional gradients in place fields and grid cell scales described in the previous section on Spatial maps.

3.3.2 Place cells and grid cells for sound frequency

Another experiment investigated if place cells and grid cells represent a continuous, 1-dimensional sound frequency dimension (Aronov et al., 2017). The authors trained rats to use a joystick to alter the frequency of a sound played, such that the sound got higher the longer the joystick was pressed. The rats were rewarded if they released the joystick when they reached a certain sound frequency. Thus, the animals were navigating in a 1-dimensional sound space, with minimal confounds caused by movement. To control for elapsed time, the frequency changed at different speeds. The animals also navigated freely a 2-dimensional space. The authors found place cells for sound, that is, neurons that responded to particular sound frequencies and also to particular locations in physical

space. The authors also found grid cells for sound, which often had multiple peaks in the 1-dimensional sound space and showed hexagonal symmetry in a 2-dimensional physical space (Fig. 3.10). Critically, these neurons responded to sound when the rats were engaged in the task, thus only when the sounds led to a reward, but not when the animals were listening passively to the same sounds. Thus, these neurons in the HPC and ERH are task-relevant, and are different from those in the primary auditory cortex that respond passively to sound.

These results are exciting because they strongly suggest place cells and grid cells also code for non-spatial dimension. To be sure this is the case, the same non-spatial navigation task could be done in two dimensions, such as sound frequency and sound amplitude. However, the overlap between these cells and spatial grid cells is strong evidence of place and grid-cell coding for non-spatial dimensions. Moreover, cells with larger sound frequency fields also corresponded to spatial grid cells with bigger scales, suggesting that functional gradients for non-spatial grid cells across the dorso-ventral axis of the medial ERH might parallel those of spatial grid cells.

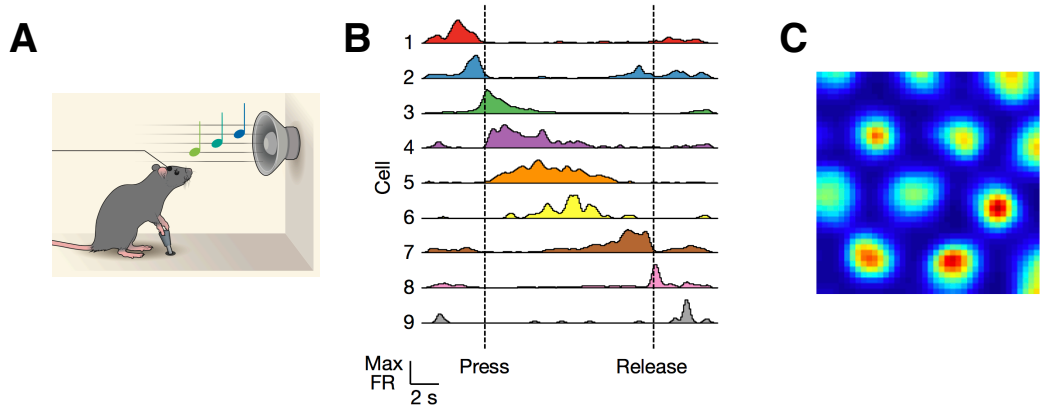


Fig. 3.10: Mapping of a sound frequency dimension by grid cells. (A) Navigation in 1-dimensional sound frequency space. **(B)** Medial ERH neurons responding to specific sounds. **(C)** Example neuron from (B) that shows hexagonal symmetry in 2-dimensional physical space, characteristic of grid cells. (A) Adapted with permission from (Rueckemann and Buffalo, 2017) and (B-C) from (Aronov et al., 2017).

3.3.3 Concept cells

Direct recordings during brain surgery in humans have shown that single neurons in the medial temporal lobe respond to specific familiar concepts, such as Jennifer Aniston (Quiroga et al., 2005), in a similar manner to place cells firing in response to specific locations. These neurons were named "Jennifer Aniston cells" or "concept cells" because they responded specifically to different pictures of Jennifer Aniston, but not to other people, animals, landmarks or objects. Other cells fired specifically for other celebrities or well-known architectural landmarks, depicted as photographs, drawn sketches, or letter strings (for example, "Halle Berry" or "Sydney Opera"), suggesting that such invariant responses cannot be attributed to common visual features.

Finding a Jennifer Aniston cell does not mean that this is the only neuron that responds to this concept (Quiroga, 2012). Such an extreme sparse coding mechanism is known as Grandmother cell coding, which seems impossible because there are not enough neurons in the brain to uniquely encode all possible concepts. Thus, there should be more than one neuron per concept. Indeed, the performance of decoding algorithms that predict presented stimuli from the firing pattern of a population of neurons, increases with the number of neurons included in the analysis (Quian Quiroga and Panzeri, 2009; Quiroga, 2012).

3.3.4 Discrete spaces

How does the brain navigate spaces that are discrete, rather than continuous? Take the following events: getting a pumpkin, carving it for Halloween, and making a pumpkin soup. They could be represented in a graph (or a "pumpkin space") where one could mentally navigate along certain paths, depending on the goals: eating the soup quickly, or also getting creative and carving the pumpkin beforehand (Fig. 3.11A). Indeed, research suggests that the hippocampal formation underlies navigation through such discrete spaces

(Schapiro et al., 2012; Wimmer and Shohamy, 2012; Schapiro et al., 2013; Horner et al., 2015; Garvert et al., 2017).

In a 1-dimensional discrete navigation experiment (Fig. 3.11B), participants learnt associations between stimuli by experiencing them together many times, probabilistically, such that some associations were stronger than others (Schapiro et al., 2012). This type of learning is implicit because the participants were not explicitly instructed to look for an order in the stimuli. Stimuli that occurred more frequently together had more similar neural representations in the HPC, ERH and parahippocampal cortex, as measured with fMRI multivariate analyses, suggesting that the neural distances between them were also shorter. To investigate if similar discrete stimuli are arranged in two dimensions, humans could navigate through a different space where the same stimulus was paired with more than one other stimulus, leading to a map (Garvert et al., 2017). Again, stimuli that were closer on the stimulus map also had more similar neural representations in the HPC and ERH, assessed by an independent technique, fMRI repetition suppression. This suggests that stimuli which were closer on the map were also closer in neural space. Notably, the map that explained these results was not Euclidean, but a predictive one, in agreement with the successor representation theory which has been proposed to account for a number of properties of place cell and grid cell activity (Stachenfeld et al., 2016). The advantage of such a map is that a place cell can predict future locations rapidly, by assessing the value of the current state (location) on the map, as a function of how likely it is to move to a goal state containing a reward, and how valuable the reward is (Garvert et al., 2017).

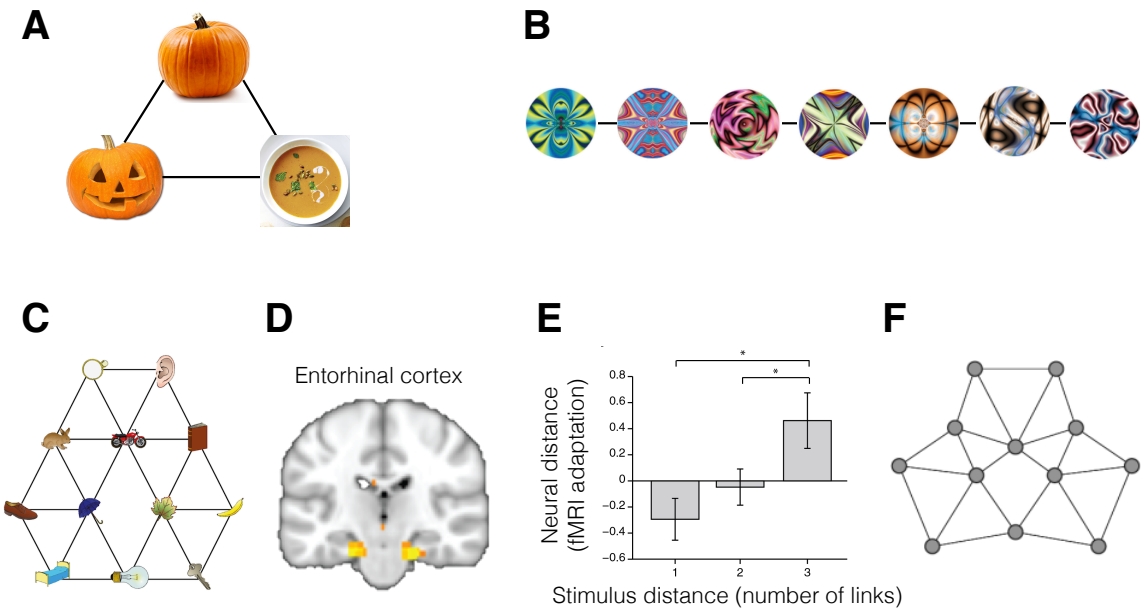


Fig. 3.11: Navigation of discrete spaces. (A) "Pumpkin space". **(B)** One-dimensional sequence of discrete stimuli. Some pairs have strong associations, whereas others have weak associations. **(C)** Two-dimensional discrete map of stimuli. **(D)** Distance effect in the hippocampus and entorhinal cortex. **(E)** The neural distance increases progressively with the number of links between the stimuli in stimulus space. **(F)** A predictive map explains the behavioural and neural data best. (B) Adapted with permission from (Schapiro et al., 2012) and (C-F) from (Garvert et al., 2017).

3.4 Implications for human cognition

Understanding such precise cellular computations allows the development of imaging markers of cellular population codes, that could be used to test how these mechanisms generalize to complex human behaviours, such as conceptual reasoning (Constantinescu et al., 2016c) and decision-making (Kaplan et al., 2017), and how functional gradients map across the entire brain. This has implications for the development of non-invasive markers that could be used to understand disrupted cognition in psychiatric and neurological disorders, such as Alzheimer's (Kunz et al., 2015).

Chapter 4 will describe the current state-of-the-art methods to investigate population codes with fMRI in humans using sub-millimeter voxels and sophisticated analyses, whereas

Chapter 5 will explain how fMRI works and how we can use it to address questions that are more easily tested in humans than animals. Chapters 6-9 will present experimental evidence for grid-cell coding in humans during navigation of conceptual dimensions and propose a new method to investigate grid-cell coding in a manner that is more akin to that used for analyzing electrophysiological data.

Chapter 4: Investigating the human brain at the mesoscale with high-resolution MRI and computational analyses

With continuous advances in MRI technology and computational analyses, it is now possible to investigate at the mesoscale, submillimeter level, precise mechanisms of the human brain that generate complex behaviours in humans. This chapter highlights state-of-the-art experiments that investigate functional gradients, laminar and cortical column processes, as well as grid-like coding in humans. Some of these techniques have been used in Chapters 7 and 9, and others could be used on our data in the future, as described in Chapter 10.

4.1 Scales and levels of description

How does the brain generate behaviour? In animal models, mechanisms underlying behaviour are investigated at the microscopic and mesoscopic scales. The microscopic scale involves individual neurons, whereas the mesoscopic scale concerns populations of neurons involved in a particular computation. In humans, most neuroimaging experiments investigate at the macroscopic scale, thus across distinct brain areas. However, with advances of magnetic resonance technology and computational analyses, analogous mechanisms from animals could be directly tested in the human brain.

In humans, there are ~10^{10} neurons in the cortex alone, and ~10^4 neurons per mm^3 (Logothetis, 2008). Therefore, investigating every single neuron would be difficult while, with functional magnetic resonance imaging (fMRI), even at a small voxel size we would record many neurons. However, there is overwhelming evidence that human cognitive functions depend on the activity of large populations of neurons in distributed networks, such as across cortical layers and columns at the mesoscale (Fig. 4.1) (Sporns et al., 2005).

Laminar organization shows clear differences in cytoarchitecture, connectivity and function of different layers, which might bring computational advantages (Dickson et al., 1997; Hamam et al., 2000; Raizada and Grossberg, 2003; Buffalo et al., 2011). Cortical columns also have a stereotypic internal processing with common anatomical and functional characteristics shared by their component neurons which tend to cluster (Rakic, 2008).

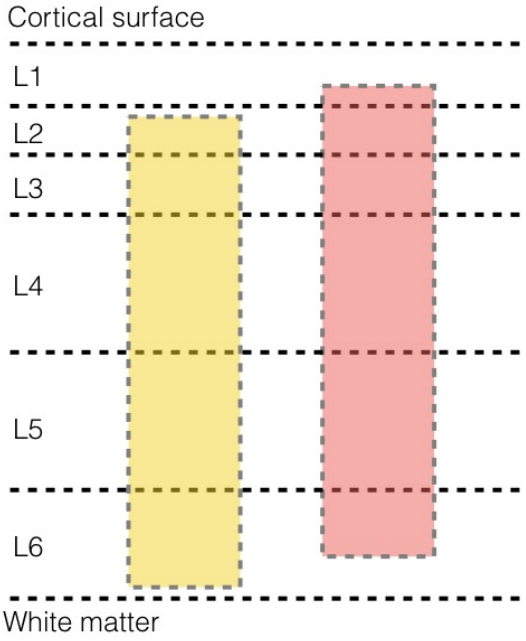


Fig. 4.1: Mesoscale. Cortical layers/laminae (L1-L6) and cortical columns that span the depth of the cortex (yellow and red).

Mesoscale is a relevant level for understanding behaviour because it can probe cognition and behaviour at the level of fundamental units of computation (Fig. 4.2). Studying the brain at the mesoscale can be achieved by various combinations of clever experimental designs, sub-milimetre fMRI and sophisticated analysis tools. These approaches are exemplified in this chapter by a selection of key neuroimaging experiments. First, I show that submillimeter neuroimaging can map functional gradients, cortical layers and columns. Then, I present how careful experimental designs combined with computational

analyses can detect hexagonal grid-like coding signals with fMRI. This chapter illustrates that translating mechanistic insights from animal studies into human neuroscience, and understanding how different levels of investigation relate to each other, is important both for human and animal research. Details of the fMRI physics, physiology and analysis that are the basis of these example experiments are presented in the next chapter.

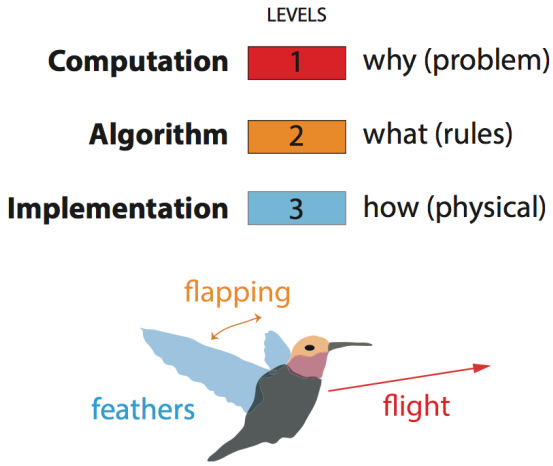


Fig. 4.2: David Marr's three levels of description. A bird flies (computation) by flapping its wings (algorithm) which are made of feathers (implementation). Understanding how the brain generates behaviour requires an understanding of these three levels together. For example, bats can fly even if they do not have feathers, or eagles fly by gliding, whereas finches fly by flapping their wings. Adapted with permission from (Krakauer et al., 2017).

4.2 High-resolution MRI

A 7 Tesla (7T) MRI scanner is strong enough to pick up a double-decker bus. It is currently used only in the experimental setting to safely and non-invasively image body tissues in humans. Moving to higher magnetic fields provides an increase in signal-to-noise ratio, which can be used to acquire fMRI images with greater spatial resolution and specificity of the underlying neural signals, with an unprecedented ability (Ugurbil, 2016; De Martino et al., 2017). This is a major technical advance because such anatomical resolution used to be possible only with invasive, electrophysiological techniques. Although an ambitious challenge, 7T MRI can give clear images at submillimeter scale (Fig. 4.3).

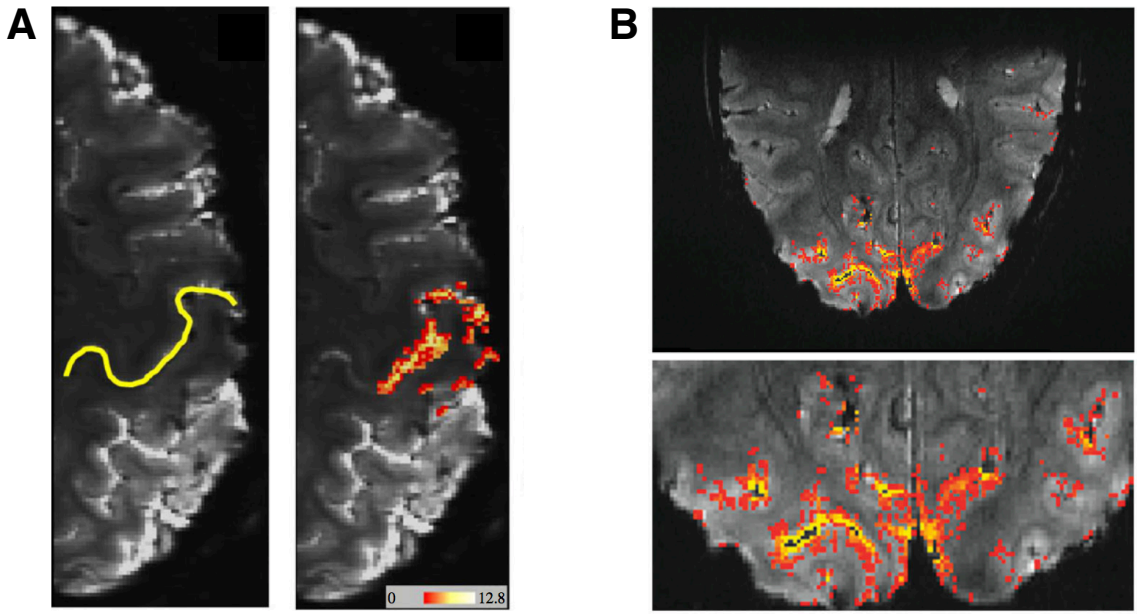


Fig. 4.3: Spatial specificity of 7T fMRI in single human subjects. (A) Raw fMRI data with an overlaid yellow contour highlighting the anatomical location of the motor hand knob and an activation map of a motor task in the same anatomical location. The spatial resolution is 0.75 mm isotropic. **(B)** Activation maps in the visual cortex of a single subject. The spatial resolution is 0.65 mm isotropic. Note how the functional activations follow precisely the shape of the sulci in the motor and visual cortices. Adapted with permission from (A) (Turner, 2016) and (B) (Heidemann et al., 2012).

4.2.1 Functional gradients

Increased resolution and specificity, together with the large coverage achievable in fMRI, facilitate the detection of functional gradients across the human brain. This is possible by moving mapping to individual subjects to resolve fine details of functional organization and test for inter-individual differences. For example, we can visualize at 7T cortical topographic maps of somatosensory, auditory and visual processes (Fig. 4.4).

The hand area of the somatosensory cortex contains highly detailed maps, were fingers are represented separately and next to each other (Kolasinski et al., 2016). With 7T, we can map such gradients not only in healthy individuals, but also in hand amputees who imagine moving their missing fingers, several decades after their hand loss (Kikkert et al., 2016) (Fig. 4.4A). It is also possible to visualize tonotopic maps (De Martino et al., 2015b) (Fig. 4.4B) in the auditory cortex and retinotopic maps in the visual cortex (Hoffmann et al.,

2009) (Fig. 4.4C). These gradients can be visualized even at voxel sizes bigger than 1 mm when employing advanced computational analysis techniques, such as representational similarity analysis of multivoxel patterns, population receptive field mapping, encoding and decoding models (see next chapter).

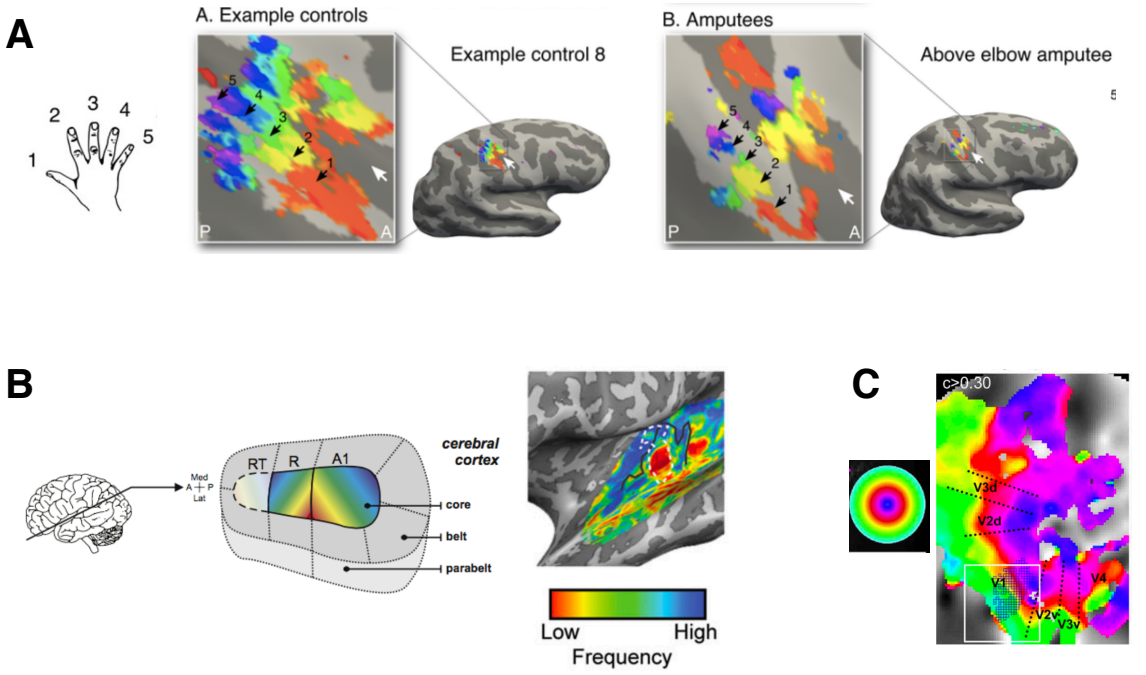


Fig. 4.4: Functional gradients at 7T in humans. (A) Individual finger representations in the sensorimotor cortex of a healthy individual and a hand amputee participant (voxel size 1.2 mm isotropic). **(B)** Schematic of an idealized sound frequency (tonotopic) map in the auditory cortex and real sound frequency representations in the auditory cortex which are visible on a single subject basis (voxel size 1.5 mm isotropic). **(C)** Colours are represented as retinotopic maps in the visual cortex (voxel size 2.5 mm isotropic). Note that all these maps are topographic because the order and organization of the stimuli (finger sensations, sounds and colours) is the same in the brain. That is, fingers/sounds/colours that are next to each other are also represented next to each other in the brain. (A) adapted from (Kikkert et al., 2016), (B) from (Saenz and Langers, 2014; De Martino et al., 2015b) and (C) from (Hoffmann et al., 2009).

4.2.2 Cortical layers and columns

With submillimeter fMRI, neural responses can be further characterized across the cortical depth, in different layers (Lawrence et al., 2017). This is possible because the blood supply for different cortical layers comes from capillaries, the smallest blood vessels which

regulate neurovascular coupling and, hence control the fMRI response (Goense and Logothetis, 2008; Logothetis, 2008; Goense et al., 2016). Hence, different layers have different metabolism, vascularization and oxygenation, but also different types of neurons (De Martino et al., 2017). For example, the entorhinal cortex has stellate cells in layer II and pyramidal cells in the other layers (Canto et al., 2008; Donato et al., 2017), and grid cells are abundant in superficial layers and co-localized with conjunctive grid-by-direction cells in deeper layers (Sargolini et al., 2006; Boccara et al., 2010). Moreover, grid cells in deep entorhinal layers replay in a coordinated fashion with place cells in the hippocampus, whereas those in superficial entorhinal layers replay independently (Ólafsdóttir et al., 2016b; O'Neill et al., 2017). Therefore, the ability to measure laminar signals opens the possibility to explore unique cognitive function in humans that cannot be easily tested in animals.

So far, the clear majority of laminar experiments are done with invasive techniques. However, pushing the limits of spatial resolution in MRI can get close to what is possible with invasive techniques, especially at the structural level (Fig. 4.5). This facilitates precise anatomical localization of neural activity obtained with fMRI.

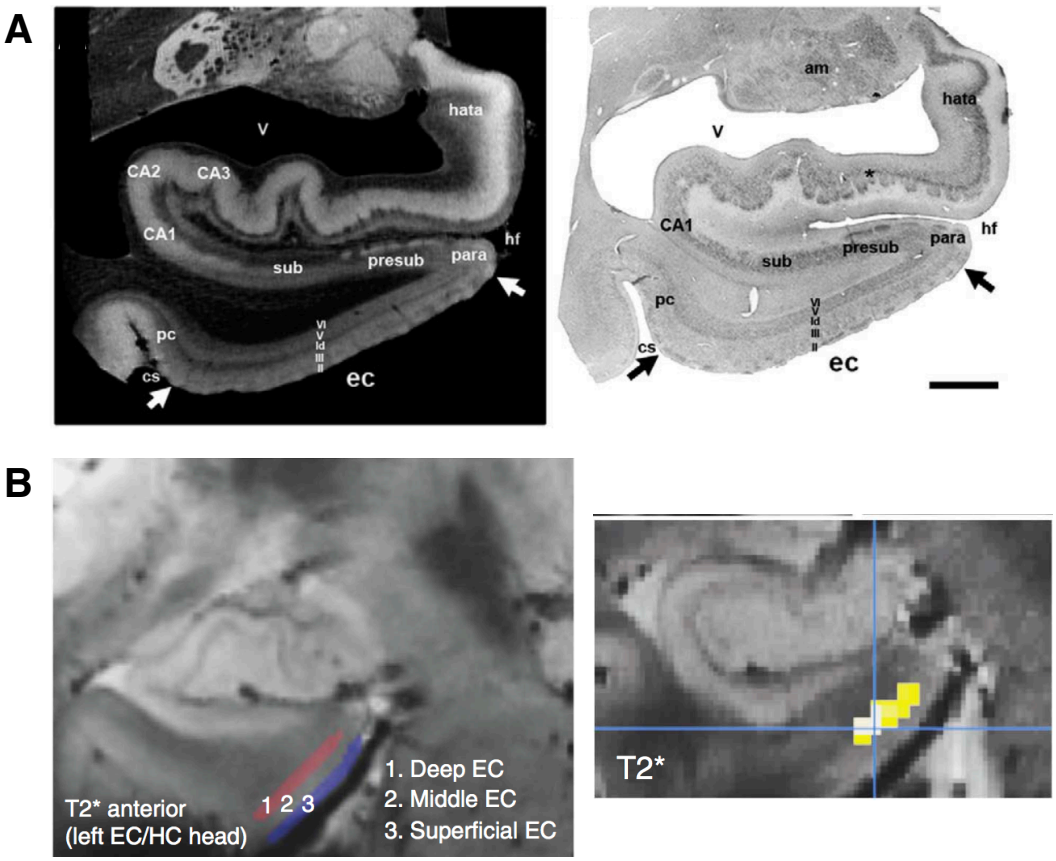


Fig. 4.5: Investigating individual layers of the entorhinal cortex with 7T in humans. **(A)** Comparison between high-resolution MRI (left; voxel size 0.1 mm isotropic) and a matched Nissl stained section (right). Note how individual layers of the entorhinal cortex (ec) can be seen. **(B)** The entorhinal cortex (EC) was divided equally into a deep, a middle and a superficial region (voxel size 0.33 mm isotropic in-plane). Single-subject fMRI activation related to memory is found only in the deep entorhinal region. (A) adapted with permission from (Fischl et al., 2009) and (B) from (Maass et al., 2014).

Remarkably, with 7T fMRI it is also possible to measure human markers of small cortical columns that function as fundamental computational units in the brain, such as for ocular dominance and orientation (Fig. 4.6), axis of motion, frequency and colour (Yacoub et al., 2008; Zimmermann et al., 2011; De Martino et al., 2015a; Nasr et al., 2016; Ugurbil, 2016).

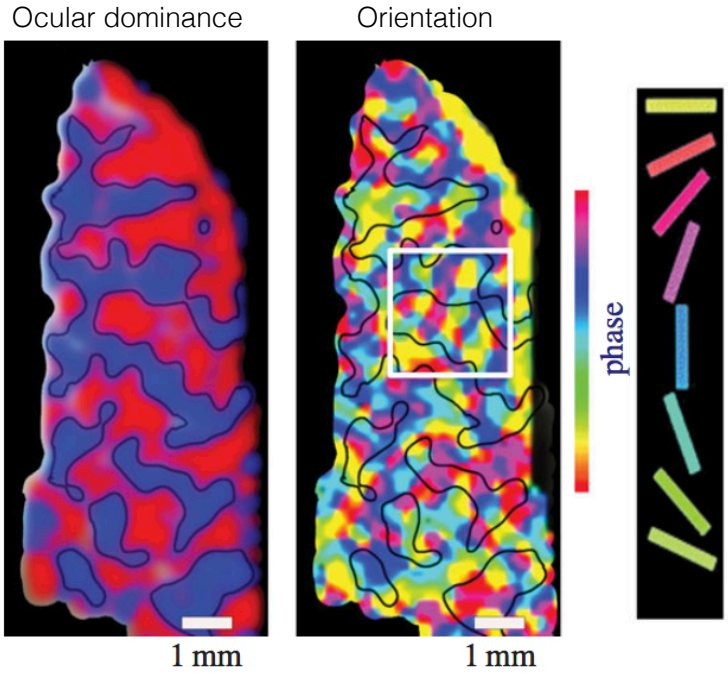


Fig. 4.6: Functional maps of ocular dominance and orientation columns in the visual cortex with 7T in humans. The red and blue colours in ocular dominance image represent preference to the right of left eye stimulation, whereas the rainbow colours in the middle image are indicative of the preferred stimulus orientation showed in the right image (voxel size 0.5 mm isotropic in plane, reconstructed to 0.25 mm isotropic). Adapted with permission from (Ugurbil, 2016).

4.3 Computational analyses for grid-like hexagonal symmetry

Another approach to investigate the human brain at the mesoscale with fMRI has been to combine clever experimental designs and computational analyses that take advantage of several characteristics that are shared between neurons at the population level. An exciting example, which is the basis of this thesis, was finding markers of grid cells with fMRI in humans using encoding models (Doeller et al., 2010) and multivariate models (Bellmund et al., 2016a). These computational analyses look for a precise hexagonally symmetric signal as a function of the running direction or imagined direction in the absence of movement, that cannot simply be explained by characteristics of the visual scene. These signals were found in a network of brain regions, including the medial prefrontal and

entorhinal cortices, which were validated by invasive recordings in humans (Jacobs et al., 2013).

Briefly, in the encoding model, human subjects navigate in a spatial environment and their directions are categorized as aligned or misaligned with the main axes of the population of grid cells (Fig. 4.7A-C). Because grid cells have hexagonal symmetry and consistent orientation across the population, finding a fMRI population response with hexagonal symmetry as a function of moving direction is strongly suggestive of grid cells (Fig. 4.7D). This is likely because running aligned with the grid activates the same cells more frequently than moving misaligned with the grid (Fig. 4.7E). The analysis used to detect this signal involves looking for a linear combination of two regressors, $\cos(6\theta)$ and $\sin(6\theta)$, where θ is the moving direction, and doing a cross-validation to test for consistent hexagonally symmetric signals between different datasets, using the regressor $\cos(6[\theta(t) - \varphi])$, where $\theta(t)$ is the moving direction in trial t and φ is the grid orientation estimated from a different dataset (Fig. 4.7E). In the multivariate model, human participants imagine directions in a well-learnt virtual environment and the hexagonal symmetry is probed at the level of similarity in voxel patterns (Fig. 4.8).

Such computational analyses, and a novel one we're developing in Chapters 8-9, have the advantage to allow us to infer mesoscale computations at a finer scale than the imaging resolution (3 mm isotropic in Chapter 7 and 0.9 mm in Chapter 9).

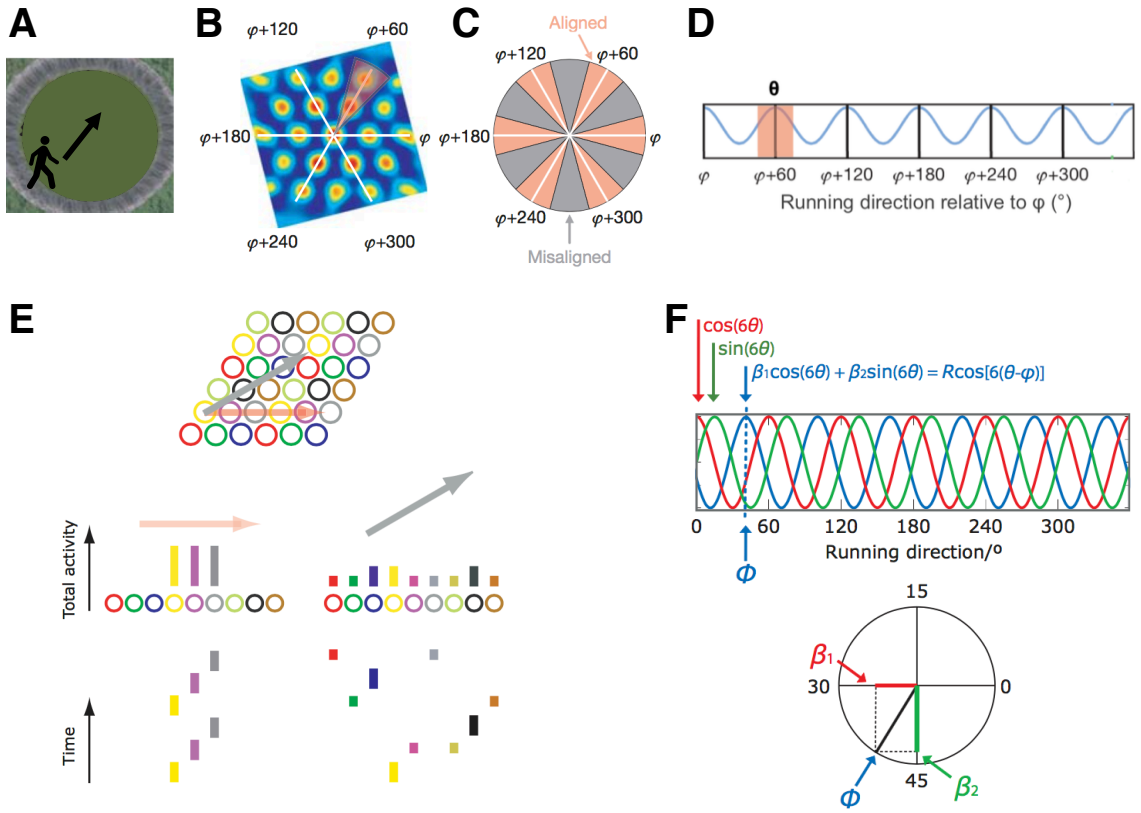


Fig. 4.7: Encoding models for detecting grid-like hexagonal symmetry with fMRI. **(A)** Spatial navigation in a virtual reality task. **(B)** Autocorrelation of a grid cell, showing the grid axis (white lines) and a sector aligned with the grid (red). **(C)** Running directions can be categorized as aligned (red) or misaligned (grey) with the grid. **(D)** Hypothesized hexagonal symmetry: a periodic signal with 6 peaks. **(E)** Proposed mechanism underlying the observed hexagonal symmetry. Running aligned with the grid (red arrow) means running through fewer firing fields (open circles) but repeatedly, leading to more frequent reactivation of the same cells, hence a bigger mean activity (the height of the bars). In contrast, running misaligned with the grid (grey arrow) means activating more grid cells, but less frequently, leading to a smaller mean activity. **(F)** Encoding model for finding hexagonal symmetry. β1 and β2 are the coefficients for the cos(6θ) and sin(6θ) regressors. Adapted with permission from (Doeller et al., 2010).

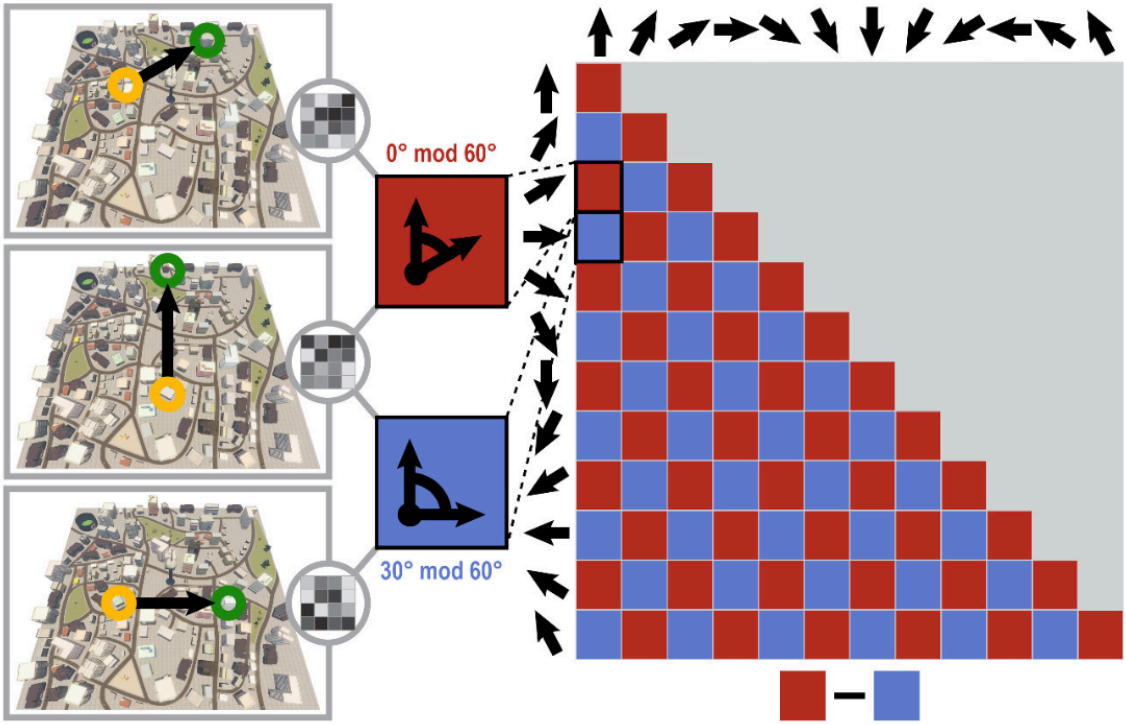


Fig. 4.8: Multivariate models for detecting grid-like hexagonal symmetry with fMRI. Participants are intensively trained to navigate a spatial environment. Then, in each trial, they are asked to imagine a spatial direction with 30° resolution. Pairs of trials are considered to have hexagonal symmetry if the remainder of their angle difference and mod 60° is 0°, and not 30°. To look for grid-like coding, we test if voxel patterns for hexagonally symmetric pairs of trials are more similar than those for the remaining pairs. Adapted with permission from (Bellmund et al., 2016a).

Chapter 5: Basic principles of MRI physics, physiology and analysis

Neuroscientists use functional magnetic resonance imaging (fMRI) together with behavioural tasks to build a picture of the areas of the brain involved with processing information in the task of interest. This chapter will describe the general principles of fMRI and will make references to various chapters in the thesis where specific imaging sequences and analyses were used. First, I will explain the physics of MRI and how we can use this knowledge to optimize the spatial and temporal resolution of fMRI. Second, I will discuss how fMRI relates to neural activity. And third, I will describe the main analysis steps used for task-based fMRI, including preprocessing and statistical models for univariate-multivariate and encoding-decoding approaches, with an emphasis on the importance of cross-validation.

5.1 MRI physics

MRI is an imaging technique used in research and clinical setting to form pictures of the anatomy and physiology of the body organs, such as the brain. MRI scanners use strong magnetic fields to generate these images at high spatial resolution, without using X-rays or radioactive substances. Therefore, MRI can be used safely and non-invasively to investigate the anatomy and function of the human brain. A set of physical principles underlies the generation of the MR signal.

5.1.1 Magnetism

The primary concept is that certain atomic nuclei act like small magnets, so we can use external magnetic fields to change their magnetic state and then measure this change. For a nucleus to be useful in MRI, it needs to have a magnetic field. The nucleus (Fig. 5.1A) is formed by protons and neutrons, and their sum gives the mass number. Protons and

neutrons spin around their own axis in the nucleus, with random directions. Therefore, if the nucleus has an odd mass number, the spins do not cancel each other out, so the nucleus spins. Because protons have a positive charge and neutrons do not, the nucleus will have a net positive charge. According to the laws of electromagnetic induction, a moving, unbalanced charge induces a magnetic field around itself, hence the nuclei are MR active (Fig. 5.1B).

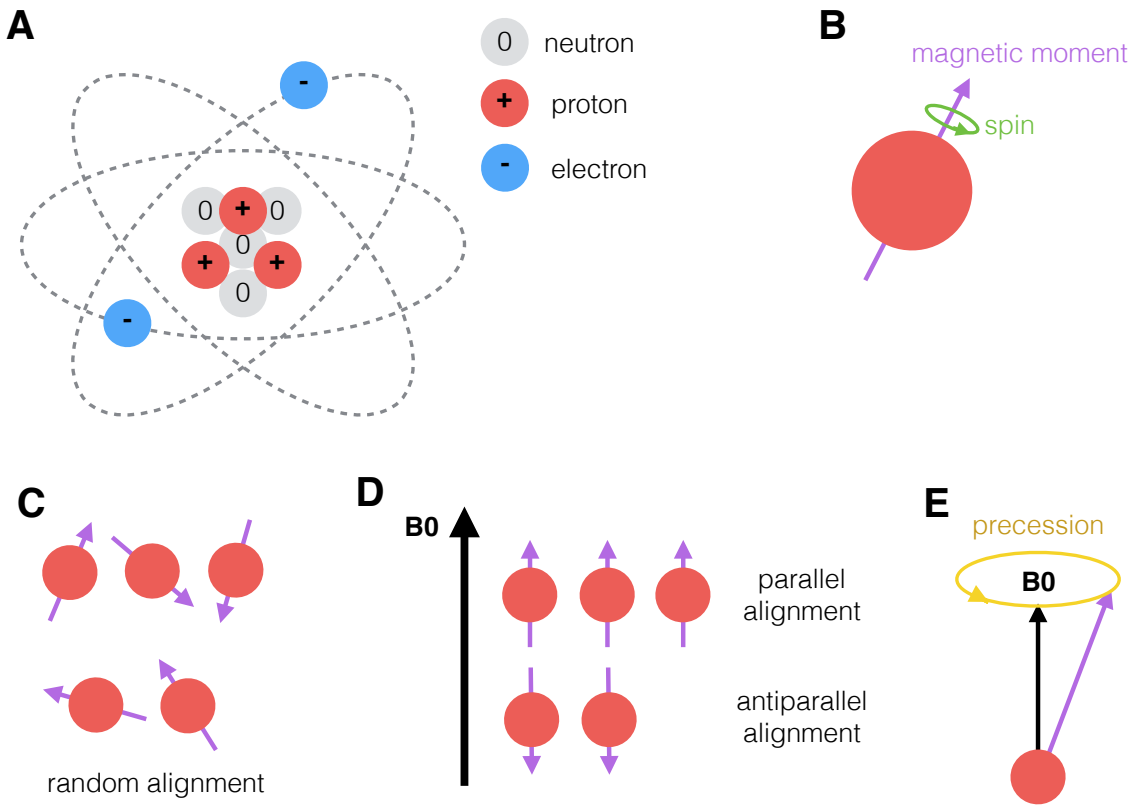


Fig. 5.1: Magnetism. (A) The atom. **(B)** The magnetic moment of the spinning proton. **(C)** In normal conditions, at room temperature, the protons are oriented randomly in 3 dimensions (for simplicity, the illustration is in 2 dimensions). **(D)** When a strong, external magnetic field is applied (B0), most of the protons reorient to a similar direction as B0 (parallel), whereas the remaining protons reorient in the opposite direction of B0 (antiparallel). **(E)** Precession. Adapted with permission from (Westbrook, 2002)

Neuroscientists are typically interested in hydrogen nuclei in water (protons) because they have a large magnetic moment and are numerous in the brain. Under normal conditions, the protons are oriented randomly and tend to cancel each other out (Fig. 5.1C). This means

they produce no overall magnetic effect and, thus, they cannot be detected with MRI. To interact with them, a very strong external magnetic field is required (B_0). The strength of B_0 defines the strength of the scanner and is measured in tesla (T). For example, we used a 3T scanner for the experiment in Chapter 7 and a 7T scanner in Chapter 9. More protons will be aligned parallel than aligned anti-parallel with the direction of B_0 (Fig. 5.1D). The balance between the parallel and anti-parallel magnetic moments will give the net magnetization vector (NMV).

Instead of aligning perfectly with B_0, the spinning protons will produce a secondary spin (precession) around the axis of B_0, like a spinning top (Fig. 5.1E). The frequency of the precession is determined by the strength of B_0 and a property that is fixed for each type of nucleus (gyromagnetic ratio), using the Larmor equation (Huettel et al., 2004):

precession_frequency = B0_strength * gyromagnetic_ratio

5.1.2 Resonance

Resonance occurs when a radiofrequency (RF) pulse is transmitted from the scanner coils at the same frequency as the one of the precessing protons. This means that the RF pulse will excite the protons selectively, allowing us to isolate protons from other magnetic nuclei in the brain that resonate at other frequencies.

The RF pulse will generate an additional external magnetic field (M1), which will flip the protons away from the direction of M_0 by an angle (flip angle) that is determined by the intensity and duration of the RF pulse (Fig. 5.2, left). Flipping the magnetic moment from the longitudinal direction to the transverse plane generates an MR signal that can be captured by the surrounding receiving coils of the scanner (Huettel et al., 2004).

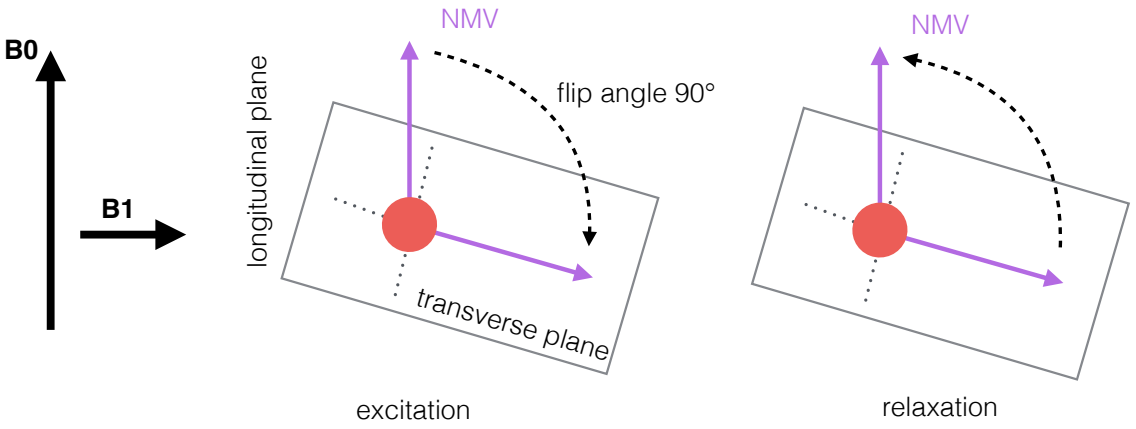


Fig. 5.2: Resonance. The radiofrequency (RF) pulse creates another external magnetic field (B1) and it flips the net magnetisation vector across all protons (NMV) from the longitudinal plane (B0) to the transverse plane. NMV in the transverse plane generates the MR signal. Here, the flip angle is at 90° relative to B0. When the RF pulse is switched off, the NMV in the transverse plane decreases and goes back to the longitudinal plane. Adapted with permission from (Westbrook, 2002).

5.1.3 Imaging

To generate detailed pictures of the brain from the measured MR signal, we need to generate a good contrast such that grey matter, white matter and cerebrospinal fluid show up with different colours in the final image. To know with millimeter precision which part of the brain the signal is coming from, we use external magnetic fields called gradient fields. Finally, we need to reconstruct the measured signal from an abstract space made up of frequencies, to a meaningful space of a 3D image of the brain.

5.1.3.1 Contrast

An image has good contrast if we can visualize different brain tissues easily. Thus, the image has areas of high intensity (white), as well as low intensity (black), and everything in between (shades of grey). To maximize the contrast between different tissues, we need to vary some parameters that facilitate a dissociation between signal intensities.

A key difference between different tissues is their rate of relaxation after the application of an RF pulse (Fig. 5.2, right). The two main types of relaxation are the recovery of NMV in the longitudinal plane, known as T1 recovery, (Fig. 5.3B) and the decay of NMV in the transverse plane, known as T2 decay (Fig. 5.3C). Another type of NMV decay is T2* which occurs when the magnetic field is not homogenous and is used in fMRI, so it will be detailed in the following section on fMRI physics. T1, T2 and T2* are fixed for each type of tissue and they vary differently across time. Therefore, adjusting the time between two successive RF pulses (repetition time - TR) and the time between the RF pulse and the signal measured with the coils (echo time - TE) will determine which tissues show up as bright or dark (Fig. 5.3C).

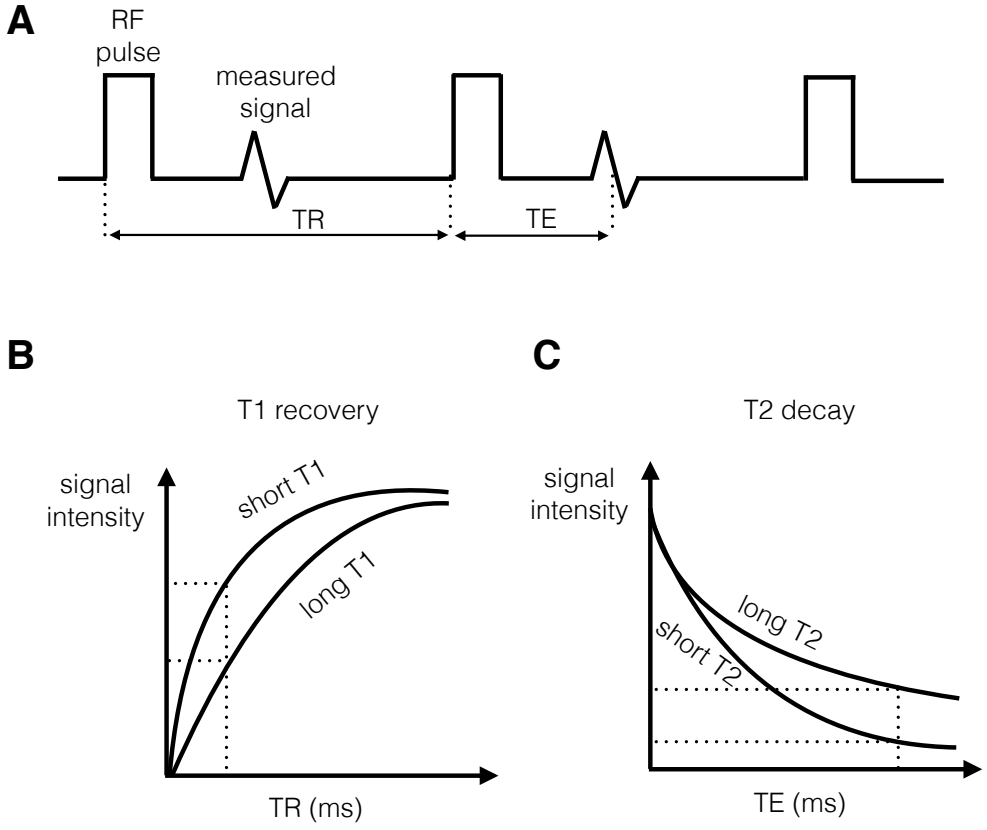


Fig. 5.3: MRI contrast mechanisms. (A) A basic pulse sequence. **(B)** T1 recovery and **(C)** T2 decay. Some example TR and TE values are highlighted for giving good contrast. Adapted with permission from (Westbrook, 2002).

To produce images where the contrast is predictable, we can vary the TR and TE and weight the image towards one type of contrast and away from the other two (Fig. 5.4). For example, in a T1-weighted image of the brain, we want an intermediate TR because it gives a large T1 contrast between grey matter, white matter and cerebro-spinal fluid, and a short TE because it gives very little contrast. T1-weighted images are often used to image the anatomy of the human brain. T2*-weighted images are commonly used for BOLD-contrast in fMRI.

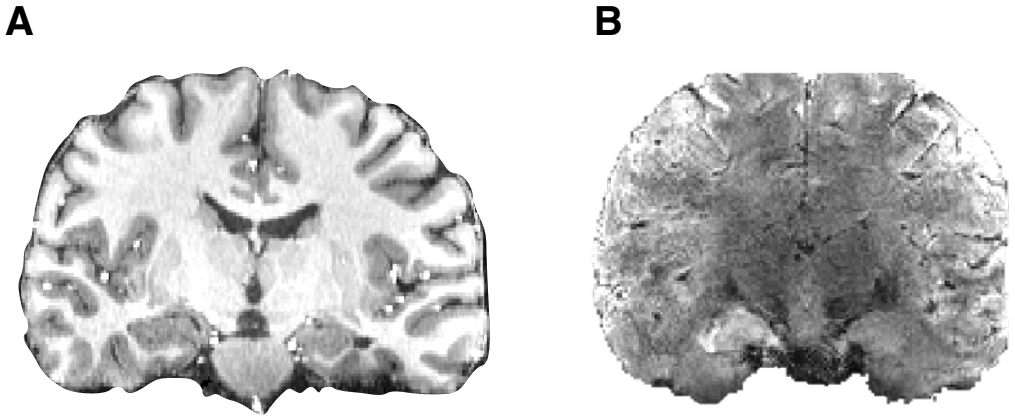


Fig. 5.4: Two example MRI contrasts. (A) T1-weighted anatomical image. **(B)** T2*-weighted fMRI image.

5.1.3.2 Gradient fields

To determine where in the brain the signal is coming from, we use three orthogonal gradient coils (Gx, Gy, Gz). Each coil generates an extra magnetic field on top of B0, that is carefully controlled to vary systematically across space. According to the Larmor equation, these extra magnetic fields will lead to different precession frequencies in different locations in the brain.

By convention, z refers to the length of the scanner, x refers to left-right axis and y to the anterior-posterior axis. Gz allows for slice selection along the axial plane, by applying an RF pulse that selectively matches the unique precessional frequency of a slice. For

example, if Gz creates a gradient field that is smaller at the bottom of the brain and larger at the top of the brain, and we wanted to excite just the bottom of the brain, then we would apply an RF pulse with a lower frequency. Gx (frequency-encoding gradient) and Gy (phase-encoding gradient) allow for spatial coding within a slice, where each x,y pixel has a unique frequency determined by Gx and a unique phase determined by Gy. This allows us to measure how much signal comes from particular locations in the brain.

5.1.3.3 K-space

The signal from each slice is stored as frequencies in k-space (Fig. 5.5A). The k-space does not correspond to the image, so there is not a one-to-one relationship between points in k-space and voxels (three-dimensional pixels) in image space. For example, the left side of the k-space is not the left side of the image. It is just a way of storing data and later reconstruct it into an image using Fourier analysis.

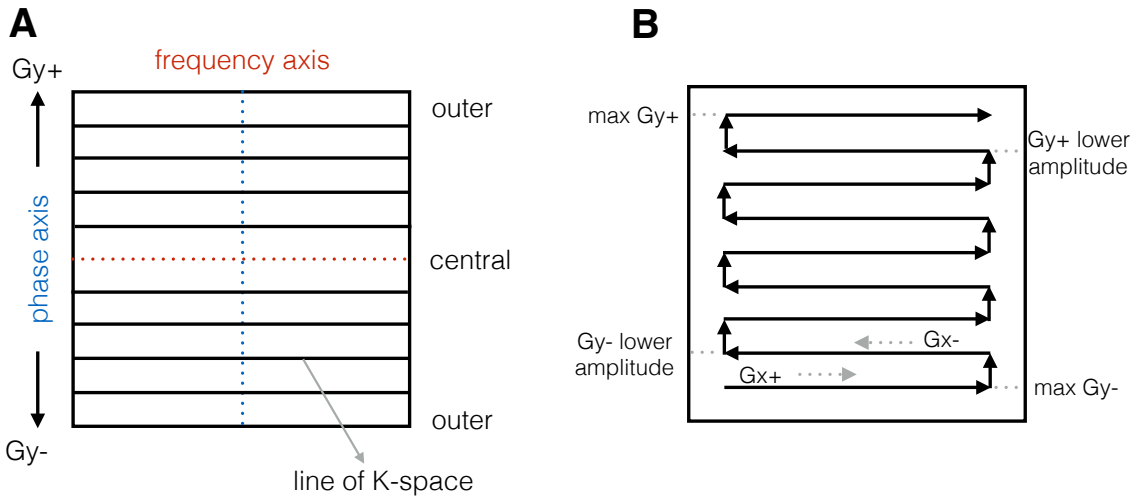


Fig. 5.5: K space. (A) MR data fills in specific locations in k-space according to signal frequency, amplitude and resolution. **(B)** Echo-planar imaging (EPI) is the top choice technique for filling in k-space during fMRI experiments. It fills in the k-space quickly, after a single RF pulse. Adapted with permission from (Westbrook, 2002).

K-space has two orthogonal axes, for representing frequency and phase. The polarity of Gy determines which part of space is filled. For example, a positive Gy will fill a location in the top of k-space. The central portion contains data with large signal amplitude and low resolution, whereas the outer portion contains data with small signal amplitude and high resolution.

5.2 fMRI physics and physiology

5.2.1 Physics of the BOLD signal

fMRI is a rapid MR imaging technique that acquires images of the brain during activity and at rest. When neurons "activate" or "fire" in a brain region, more oxygen is delivered by the blood to feed the increased demand and, consequently, the amount of deoxygenated blood decreases. The scanner can detect changes in the oxygen level in the blood of a person's brain using a sensitive contrast called blood oxygen level dependent contrast (BOLD). Thus, the magnetic state of the blood reflects its level of oxygenation.

The BOLD contrast can be measured because deoxyhaemoglobin creates an inhomogenous magnetic field nearby which affects the resonance frequency of the blood and, hence, shortens the T2* decay time (Ogawa et al., 1992). Therefore, changes in T2* reflect changes in the ratio between deoxyhaemoglobin to oxyhaemoglobin. Thus, we can use a T2* pulse sequence that is sensitive to such changes to measure local variations in brain activity. T2* is optimized using a TE equal to the T2* of grey matter (~35 ms at 3T and ~25 ms at 7T).

5.2.1.1 Echo-planar imaging

Different pulse sequences adopt different trajectories through k-space. For example, in a typical anatomical sequence, k-space is filled one line at a time. However, this is not

appropriate for fMRI experiments, where BOLD effects are very short lived and the goal is to acquire data from the whole brain almost simultaneously and quickly. An extremely rapid sequence most commonly employed in fMRI is echo-planar imaging (EPI, Fig. 5.5B). It was created by Peter Mansfield who was awarded the Nobel prize.

An EPI volume is a set of slices, each acquired after a single excitation. The technique uses very rapidly switching gradients to sample the whole k-space after a single RF pulse. Gx switches quickly between positive and negative to move backwards and forwards along rows of k-space, while Gy switches rapidly to move on to the next row in k-space. This speeds up the scanning session. EPI volumes are collected in sequence, representing time points in fMRI data.

We acquired fMRI data in chapters 7 and 9 using EPI sequences. In Chapter 9, we used a 3D EPI sequence (Poser et al., 2010) . This means we acquired data in a 3D k-space. Such sequences do not perform slice selection, and instead Gz excites the whole brain volume. This has the advantage of higher signal-to-noise ratio because more excited spins contribute to the MR signal in a large brain volume compared to a slice. However, to find out where the signal is coming from along the z direction, the sequence uses an additional phase-encoding gradient in the z dimension, on top of the phase-encoding gradient in the y direction and the frequency-encoding gradient in the x direction. During a TR, the same brain volume is excited multiple times, at different frequencies, and each time a 2D k-space is generated. For example, high spatial frequencies can be acquired at the start and end of the TR, while low spatial frequencies can be acquired in the middle of the TR.

However, there are some disadvantages of 3D EPI sequences, such as taking longer to fill in the entire k-space, hence increasing scan time. This is problematic because MR scanning is expensive and human participants become tired and less comfortable, which affects their performance during simultaneous behavioural tasks. To circumvent this, parallel imaging

techniques are often employed. Here, only a fraction of each k-space line is acquired, and the signal needs to be reconstructed for the datapoints that were not acquired, using sophisticated algorithms. For example, for a parallel factor of 4, only ¼ of the data is acquired. Because the acquired datapoints are superimposed, a multichannel coil is used to disambiguate spatial location, and have a better estimate of where the signal is coming from. Each channel will give a k-space, and which channel records a stronger signal is evidence for which part of the brain the signal was coming from (Fig. 5.6).

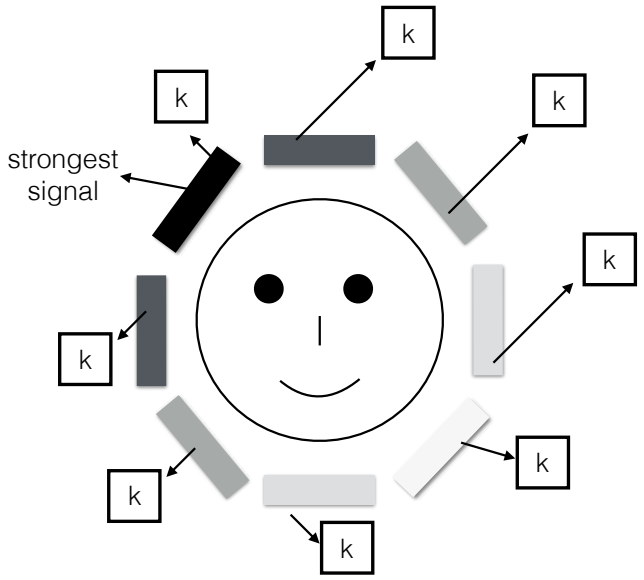


Fig. 5.6: Parallel imaging technique.

5.2.1.2 Optimizing fMRI acquisition

Knowledge of MR physics helps asking the right questions and finding the relevant parameters that need to be modulated to improve image quality. Signal-to-noise ratio (SNR), signal loss, spatial distortions, motion artefacts and image contrast determine image quality. The following equation defines the factors that influence SNR (Haacke et al., 1999; Craddock et al., 2013):

$$\text{SNR} \sim I_0 * \Delta_x * \Delta_y * \Delta_z * \sqrt{N_{acq} * N_x * N_y * N_z * \frac{1}{P_f}}$$

where, I_0 is the available signal; Δ_x, Δ_y, Δ_z are the voxel dimensions; N_{acq} is the number of image acquisitions; N_x, N_y, N_z are the number of samples in each dimension; P_f is the parallel imaging factor. In turn, the available signal I_0 is affected by other MRI parameters. I will next summarize here the main parameters that increase I_0, while also mentioning some of their side effects on other quality features of the image. Understanding the physics of MRI allows researchers to trade-off different parameters to improve I_0 (Westbrook, 2002).

At higher field strengths of the scanner, more spins will be parallel, rather than antiparallel, with M_0 therefore NMV becomes larger, thus the available signal I_0 increases. On the flip side, higher field strengths will also cause more image distortion and signal loss because of larger magnetic susceptibility effects.

Longer TRs allow for more longitudinal magnetization to recover and then be flipped in the transverse plane. Thus, more precession of the NMV in the transverse plane will give more MR signal. However, increasing the TR too much is suboptimal for T1-weighted images that require a short TR for good contrast, and fMRI experiments that need short TRs for good temporal resolution. Similarly, bigger flip angles allow for more longitudinal magnetization to be flipped in the transverse plane, and be read-out as MR signal. However, T2*-weighted images in fMRI need small flip angles and, thus, they often have lower SNR.

Adjusting these parameters to optimize the quality of an image requires careful thought. Fortunately, SNR for modern fMRI sequences is often greater than needed and it can be used as a trade-off to optimize other fMRI aspects, such as temporal resolution, spatial resolution, and correcting for various artefacts.

Temporal resolution depends on the TR, which in turn depends on the number of slices acquired. One way to increase the speed of data acquisition in k–space is to use parallel imaging.

Larger voxels improve the SNR because more spins can contribute to the signal. However, large voxels also mean decreased spatial resolution and less crisp images. Image resolution can be increased by reducing the field of view (Craddock et al., 2013).

The magnetic field becomes inhomogenous where air and tissue meet, such as the sinuses or ear canals, which lie close to the orbitofrontal cortex and medial temporal lobe, respectively. These brain regions thus become affected by spatial distortion and signal loss (dropout) (Theysohn et al., 2013). Spatial distortion means the position of voxels changes and this results in errors in the location of the structures in the resulting image. It can be corrected online using parallel imaging, or offline using maps of the spatial variation in magnetic field together with sophisticated mathematical algorithms (see the section in this chapter on fMRI analysis). Dropout occurs because various frequencies will cancel each other. It cannot be recovered offline, therefore the usual approach to tackle this issue is to optimize the MRI sequence to minimize the dropout, such as by using thin slices and tilting the slice orientation during scanning (Weiskopf et al., 2006a; 2007). Slice positioning will focus on the areas most vital to the experiment. For example, we had strong *a priori* predictions about the orbital and ventral parts of the medial prefrontal cortex in the experiment from Chapter 7, therefore we used a slice tilt of -30° to minimize dropout in this brain region (Weiskopf et al., 2006b). In contrast, in chapter 9 we aimed to minimize dropout in the medial temporal lobe, which suffers from the opposite polarity of susceptibility artefact, and thus we used a slice tilt of +30°. Another approach to minimize dropout is to use parallel imaging techniques, which decrease the time to fill in k-space, hence the dropout (Olman et al., 2009).

Finally, head motion induces artefacts in the fMRI signal. They can be minimized by acquiring brain slices in an interleaved order, because that allows the signal in a slice to fully relax before acquiring the neighbouring slice.

5.2.2 Physiology of the BOLD signal

The success of fMRI as an excellent tool to image the whole brain simultaneously and non-invasively with high spatial resolution, showing how different brain networks engage during a task, depends on understanding how the measured signal is related to the underlying neural activity that gives rise rapidly to the haemodynamic response. This is known as neurovascular coupling (Fig. 5.7). Electrophysiology is the only method to directly measure electrical activity of the neurons, by inserting electrodes in the brain, or picking up electrical signal from the outside of the skull. However, these approaches lack the depth and spatial resolution of fMRI. This section describes how changes in the oxygenation levels of the blood measured with fMRI are related to neural activity, and their implications for spatial and temporal resolution of fMRI.

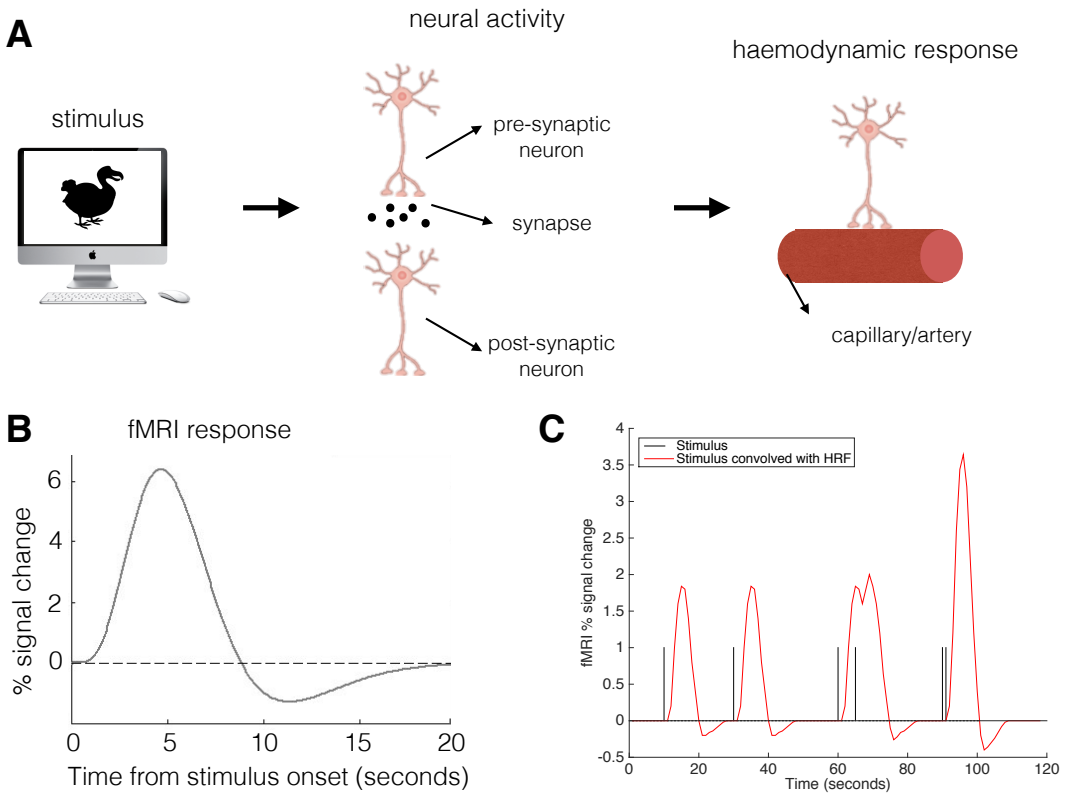


Fig. 5.7: Physiology of fMRI. (A) Neurovascular coupling. **(B)** BOLD response. **(C)** Convolving the stimulus with the HRF. Note how stimuli add linearly to get the BOLD activation.

The first step of neurovascular coupling is neural activity in response to a stimulus. Neural activity is reflected in three main processes: pre-synaptic action potentials that are passed down the axons of neurons, post-synaptic potentials transmitted via dendrites, and the molecular activity at the synapse that binds the pre- and post-synaptic neurons. Experiments that recorded simultaneously fMRI and electrophysiology data found that the BOLD response indeed reflects directly an increase in neural activity in the cortex of anesthetized (Logothetis et al., 2001) and awake monkeys (Goense and Logothetis, 2008). In particular, BOLD was correlated with local field potentials, which reflect post-synaptic potentials and synaptic activity, in contrast to single-cell recordings that measure pre-synaptic action potentials. Thus, the BOLD signal is more likely a measure of inputs into

an area and the processing of this information by the local circuits (Logothetis, 2003; 2008).

Whether these responses are excitatory or inhibitory remains debatable. Optogenetic stimulation of specific neurons during fMRI recordings suggests that the positive BOLD response is caused by activity of excitatory neurons, whereas the negative BOLD response reflects the activity of inhibitory neurons (Lee et al., 2010). However, this does not rule out the possibility that optogenetic stimulation can also activate nearby circuits that could lead to the BOLD signal via perisynaptic mechanisms (Logothetis, 2010). Moreover, excitation is immediately balanced by inhibition (Wehr and Zador, 2003; Haider et al., 2006). Thus, optogenetically activating the excitatory neurons will also stimulate the inhibitory neurons via the synapses from excitatory neurons. Moreover, we cannot completely rule out the contribution of inhibitory neurons to the positive BOLD response because, for example, the striatum contains a majority of inhibitory medium spiny neurons, estimated at ~95% (Kemp and Powell, 1971; Yager et al., 2015), yet it shows positive BOLD responses during the presentation of rewarding stimuli (Delgado et al., 2000). Thus, the BOLD response reflects neural dynamics through a complex interplay of excitatory and inhibitory neurons, in agreement with the global neural population dynamics giving rise to behaviour and complex brain processes.

The second step of neurovascular coupling is the haemodynamic response (Kisler et al., 2017). In response to neural activity and glutamate release, second messenger substance release will relax pericytes that surround capillaries, therefore capillaries will dilate (Hall et al., 2014). The increased oxygen demand will initially cause a small increase in deoxyhaemoglobin, often reflected as an initial dip in the BOLD response. Subsequently, the oversupply of oxygen will cause a decrease in deoxyhaemoglobin, which is reflected as in increase in the BOLD response, peaking at ~5 seconds (Fig. 5.7B and (Krüger et al.,

1996; Logothetis et al., 1999)). This peak is the most interesting feature because it is most directly related to neural activity (Logothetis et al., 2001). Approximately 20 seconds after the stimulus, the haemodynamic response returns to baseline, and if multiple instances of neural activity happen during this period, they are considered to add up linearly (Boynton et al., 1996).

The physiology of the BOLD response has several key implications for the spatial and temporal resolution of fMRI. First, fMRI has a spatial resolution far superior to the main techniques used to image the human brain non-invasively, such as positron emission tomography, electroencephalography and magnetoencephalography. For example, with the advance of high-resolution fMRI sequences and spatially unsmoothed data, we can use submillimeter voxel sizes to visualize different cortical layers. This is possible because most neurovascular coupling happens at the level of capillaries, the smallest blood vessels, and the ability to resolve cortical layers and columns depends on the scale of blood flow regulation (Goense et al., 2016). Capillaries are finely spread throughout the brain and interact with nearby neurons. Interestingly, neurovascular coupling varies across layers (Logothetis, 2008) because they differ in metabolism, vascularization and oxygenation (Hevner and Wong-Riley, 1992; Weber et al., 2008). For example, the BOLD response and the cerebral blood flow and volume are highest in the middle layers of the sensory cortices, compared to deep and superficial layers (Goense and Logothetis, 2006; Harel et al., 2006; Yu et al., 2014). Nevertheless, research suggests that differences in neural activity across layers also cause differences in neurovascular coupling (Zaldivar et al., 2015).

Second, even if the haemodynamic response is sluggish compared to neural activity which takes only a few milliseconds, as long as we can accurately estimate the shape of the haemodynamic response, then this lag does not need to dictate the useful temporal

resolution of fMRI. By convolving the expected time course with the haemodynamic response function (HRF), we can go back and forth between neural activity and the BOLD signal. This principle underlies statistical analysis of fMRI.

5.3 fMRI analysis

This section explains the steps required to go from raw data to the final results, which are usually statistical maps showing which brain regions respond to a behavioural task. fMRI data is noisy and different subjects have brains of different sizes and shapes. Thus, I will first describe how the data is prepared to correct for artefacts, reduce noise, and align it to a common space across multiple scanning sessions or subjects. Second, I will explain how the data is fitted with statistical models to find out which voxels are significantly activated, and I will describe how we can make inferences on the results, while accounting for the large number of statistical tests performed across the brain. This will include both statistics performed at each individual voxel level (univariate), as well as across multiple voxels (multivariate). Other concepts such as encoding-decoding models and cross-validation techniques are also discussed.

5.3.1 Preprocessing

Work in this thesis used FSL (Smith et al., 2004a) for preprocessing volumetric 3D MRI data and Freesurfer (http://surfer.nmr.mgh.harvard.edu) for preprocessing steps required for surface reconstruction.

5.3.1.1 Correcting distortion and dropout

As mentioned in the previous section on MRI physics, the orbitofrontal cortex and temporal cortex are affected by spatial distortion and signal dropout due to inhomogeneities in the main magnetic field B_0. We need to correct for dropout as early on as during data acquisition, using MR sequences that reduce dropout, because once signal is lost in cannot be recovered later offline analyses. We also need to correct for spatial distortion because this artefact mislocalizes the brain structures. By using field maps, we can correct imperfections in the magnetic field. Fieldmaps are images acquired at two different echo times, and they characterize the magnetic field. We then estimate the amount of B_0 inhomogeneity by computing the difference in phase between field maps, hence how much each voxel was shifted in space. This allows us to determine the original location of the voxels.

5.3.1.2 Correcting slice timing

In 2D EPI sequences, we systematically acquire different brain slices at different times throughout the TR (Chapter 7). They can be acquired in an ascending, descending or interleaved order, where the odd slices are acquired before the even slices This means we sample the BOLD signal at different layers of the brain at different time points. However, later fMRI analyses assume that all slices were acquired exactly at the same time.

Slice timing correction is an important preprocessing step that addresses this mismatch. It involves choosing a reference slice, usually one acquired half-way throughout the TR, and then interpolate the data in all other slices to match the timing of the reference slice. This results in a dataset where each slice represents activity at the same point in time. Short TRs and interleaved acquisitions are particularly robust to this method of correction (Poldrack et al., 2011).

5.3.1.3 Correcting head motion artefacts

During an fMRI experiment, there will be movement in the brain because subjects can move their head. This has two consequences: first, an individual voxel may not correspond to the same location in the brain throughout the experiment; and second, movement causes changes in voxel intensities and this can be a confounding factor if it correlates with the behavioural task. There are various methods to correct for these two effects of head motion.

In the experiments presented in this thesis we used linear registration to align all volumes from a dataset to a single reference volume (Jenkinson et al., 2002). Motion correction tools assume that head motion is a rigid body process, thus it can change its position and orientation, but not its shape (Fig. 5.8A, top row). After using a cost function to estimate how misaligned each volume is from the reference volume, this misalignment is minimized by applying a transformation model. Since the transformed voxels will have new coordinates in the new space, we also need to estimate what the intensities will be at those intermediate points using interpolation. For example, this can be achieved simply by using the intensities of the neighbouring voxels, or using more sophisticated weighted averages.

Moreover, to account for changes in voxel intensities caused by movement, we can include in the statistical model (see the section in this chapter on Univariate statistics) the estimates of motion as six regressors: translation and rotation across x, y and z dimensions. We can also address such potential concerns using well-controlled experimental designs.

5.3.1.4 Correcting physiological and scanner artefacts

The head can also move because of physiological pulsations caused by breathing and heart beating. One way to correct for these motion artefacts is to track them using a breathing belt and a pulse oximeter, respectively, and then include their motion estimates as confound regressors in the statistical model. Another way to clean the data from these movements, together with artefacts caused by the scanner, is by using independent component analysis (ICA). This is a very useful exploratory technique that detects systematic patterns in the data that are independent from each other. Automatic methods that classify these patterns as signal or noise, and then remove the noisy patterns from the data, have proven very efficient and accurate (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014).

5.3.1.5 Spatial smoothing

Spatial smoothing involves the application of a filter that blurs an image by removing high-frequency information and it is used in experiments where activations extend across many voxels (Chapter 7). First, this has the advantage of increasing SNR because noise that varies randomly across voxels will be canceled. In contrast, smoothing should not cancel out activation to the same extent as noise, if the activation is larger than the size of the smoothing kernel. Second, when data are combined across multiple subjects, the spatial location of functional regions can vary between subjects if registration is not perfect (see the next section on Spatial registration). Thus, smoothing can reduce this spatial mismatch between different brains and minimize potential biases due to individual variability. And third, some statistical analyses such as Gaussian Field Theory require data to be smoothed. In experiments that use multivariate pattern classification analyses, we use unsmoothed data because we want to keep data at high resolution and take advantage of the variations

in intensity across voxels (see the following section on Multivariate statistics). In Chapter 9, we used such an approach because we did not want to erase with smoothing fine details in the neural activity that may be important.

5.3.1.6 Minimizing temporal smoothing

Temporal smoothing means that the voxel intensities will be similar between moments that are close to each other in time because the BOLD response is slow. This means that nearby time points will be correlated to each other, known as temporal autocorrelation, and it needs to be corrected for optimal subsequent statistical analysis. During preprocessing stage, we remove slow drifts in the data by applying a high-pass filter. Then, during the statistics stage, we 'prewhiten' the data to remove the temporal autocorrelation (Woolrich et al., 2001). This is achieved in two steps. First, we ignore the temporal autocorrelation and we fit the general linear model (GLM) to obtain the error between the actual data and the estimated data, known as residuals. We then use the residuals to estimate the autocorrelation matrix, and then we run GLM again after prewhitening both the data and the design matrix (Poldrack et al., 2011). The reason we estimate temporal autocorrelation from the residuals is to remove it from the noise, but not from the task of interest. The section on General linear model and hypothesis testing includes a mathematical framework for prewhitening data.

5.3.2 Spatial registration

In fMRI we often want to transform the voxel coordinates and dimensions of an image to match the characteristics of a different image. Some of the common reasons are to match fMRI images from the same person scanned multiple times, when the head was inevitably positioned in a slightly different position in each scanning session, or to match images

from different people, who inevitably have brains of different sizes and shapes. This is often the case when we look for consistent brain activity patterns across different sessions and different subjects, therefore we need a common space to align different brain images. For example, the human brain is consistent in its overall structure across different healthy people, and everyone has major sulci such as the transverse gyrus in the temporal lobe, associated with the primary auditory cortex. However, some individuals have more transverse gyri (Penhune et al., 1996; Rademacher et al., 2001; Poldrack et al., 2011).

5.3.2.1 Volume-based registration

Volume-based registration is a method to transform a 3D volume of the brain. The two main types of volume-based registration are linear and non-linear (Jenkinson et al., 2002) (Fig. 5.8). The former involves linear transformations on the voxel coordinates, and are used for bringing images from the same modality or people into a common space (Fig. 5.8A). The latter involves complex transformations that can be more drastic on some parts of the image than others, and are used for bringing images from different modalities, spaces and people into a common space (Fig. 5.8B).

Some typical registration steps are linear registration of the fMRI data to the structural data from the same participant, followed by non-linear registration into the MNI152 brain template. This is an average image of T1-weighted structural scans across 152 individuals, co-registered together in the MNI space developed by the Montreal Neurological Institute.

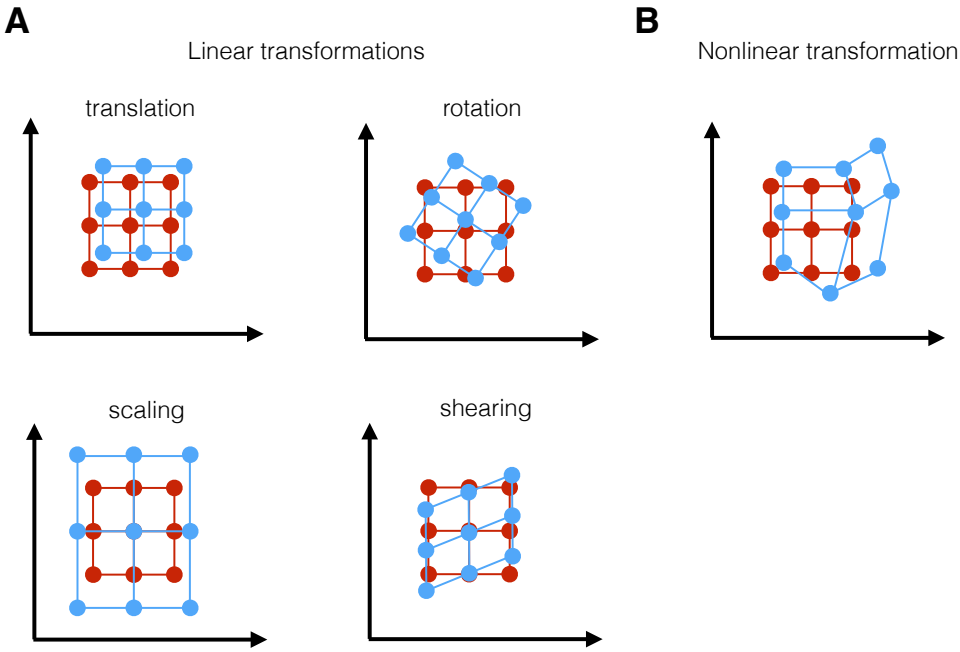


Fig. 5.8: Example linear and non-linear transformations. Red dots represent the original coordinate locations, and blue dots represent the new locations after applying the transformation. **(A)** Linear transformations involve a combination of translation, rotation, scaling, shearing. **(B)** Nonlinear transformations offer greater flexibility because they allow any kind of transformation and deform any part of the image. Adapted from (Poldrack et al., 2011)

5.3.2.2 Surface-based registration

Surface-based registration transforms brain surfaces, rather than volumes. The cerebral cortex is a sheet of tissue that is highly folded along gyri and sulci, making it difficult to understand how information is mapped across the cortical sheet. When the function follows the surface, this approach can reveal patterns of activity that are otherwise hidden when looking at the 3D volume data (Fig. 5.9). Recent work has developed methods for creating surfaces of deep, subcortical structures, which are typically more difficult to reconstruct (Postelnicu et al., 2009).

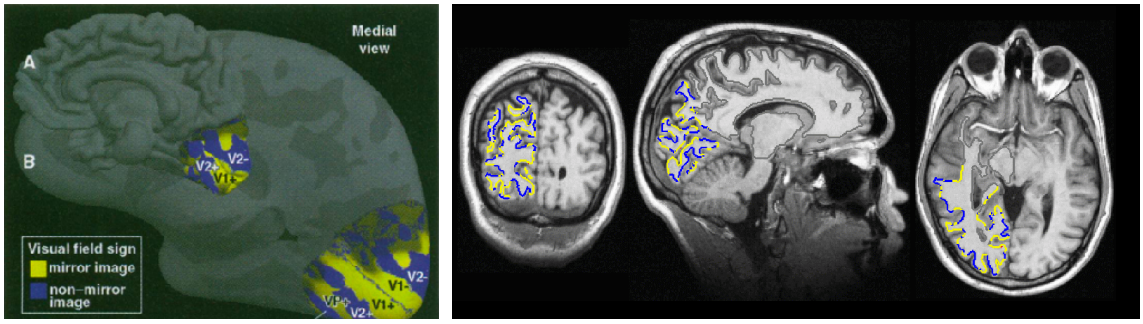


Fig. 5.9: Visual areas in the brain are evident with cortical surfaces, rather than with slices through a 3D volume. Adapted from (Sereno et al., 1995) and Freesurfer tutorials (https://surfer.nmr.mgh.harvard.edu/).

To reconstruct the cortical surface, grey matter is extracted as a separate volume, within the boundaries of the cerebrospinal fluid and the white matter (Fig. 5.10A). The resulting cortical volume is covered with a triangular tessellation according to the surface features, such as sulci and gyri, and deformed to produce an accurate and smooth representation of the grey/white interface as well as the pial surface (Fig. 5.10B) (Dale et al., 1999). Whereas volumes use voxel coordinates, surfaces are described by the vertices of the triangles, which are points in 3D space with x, y and z coordinates. There is not a one-on-one correspondence between the voxel and the vertex coordinates, and unlike voxels vertices are not uniformly distributed across the brain. However, the average distance between vertices is similar to the voxel size.

To visualize more easily patterns of activity that vary across the cortical surface, we can then computationally inflate the cortical volume (Fig. 5.10C) and flatten it out (Fischl et al., 1999). Next, we can project the fMRI data to this cortical surface and we can perform statistical analyses at the surface level. These also have the advantage of being less computationally expensive because we reduce the number of voxels inside the whole cortical thickness of a volume, to a much smaller number of vertices.

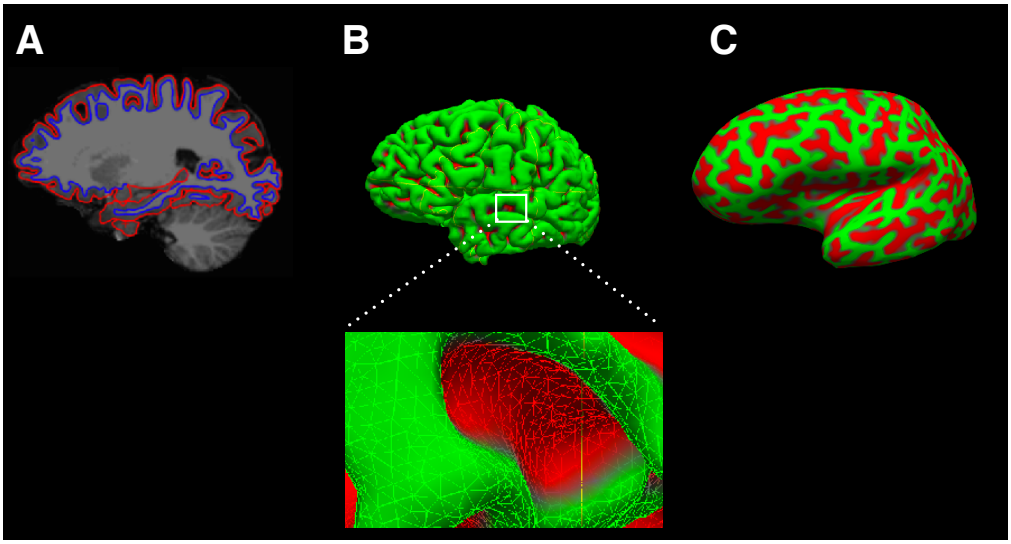


Fig. 5.10: Visualizing cortical surfaces. **(A)** To obtain the cortical surface, we first need to separate grey matter from white matter, shown here by the red and blue boundaries. **(B)** A cortical surface in one subject. The green regions are gyri and the red regions are sulci. Part of the surface is magnified to illustrate the triangular mesh. **(C)** The same surface, but inflated. The sulci now become visible.

5.3.3 Univariate statistics

Task-based fMRI analysis involves creating a model of some task-relevant parameters (stimuli) and then finding brain regions that respond most strongly to these parameters, hence where this model explains data well. The tool we used to fit the model to the data is the general linear model (GLM). When we do the GLM for each of the 100.000+ voxels separately, we perform a univariate analysis. After obtaining results for each individual subject, we can combine them across all subjects and do inferences at the group analysis. Work in this thesis used univariate statistics in Chapter 7.

5.3.3.1 General linear model and hypothesis testing

GLM is a generalization of linear regression, and the goal is to create a model (X) that fits well the actual fMRI data over time (Y), by computing some parameters that adjust the

intercept (β_0) and amplitude (β_1) of the model with respect to the actual data to minimize the error (ε):

$$Y = \beta_0 + \beta_1 * X + \varepsilon$$

Therefore, β_1 can estimate how well the model fits the actual data (Fig. 5.11). Because real data is noisy, the difference between the real data and the fitted data gives an error.

Complex signals can often be described by a combination of multiple independent variables, each making a different contribution to the signal. For example, if we have 4 variables, the model will have four components (X_1, X_2, X_3, X_4), each with its own parameter (β_1, β_2, β_3, β_4) that estimates how much that variable contributes to the signal. This can be written as:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * X_4 + \varepsilon$$

And it can be formalized concisely using linear algebra:

$$Y = X\beta + \varepsilon$$

To find the best estimates of β that minimize the error, we use the method of least squares:

$$\beta = (X^{'}X)^{-1}X^{'}Y$$

where $(X^{'}X)^{-1}X^{'}$ is the pseudoinverse of matrix X:

$$\beta = \text{pinv}(X)Y$$

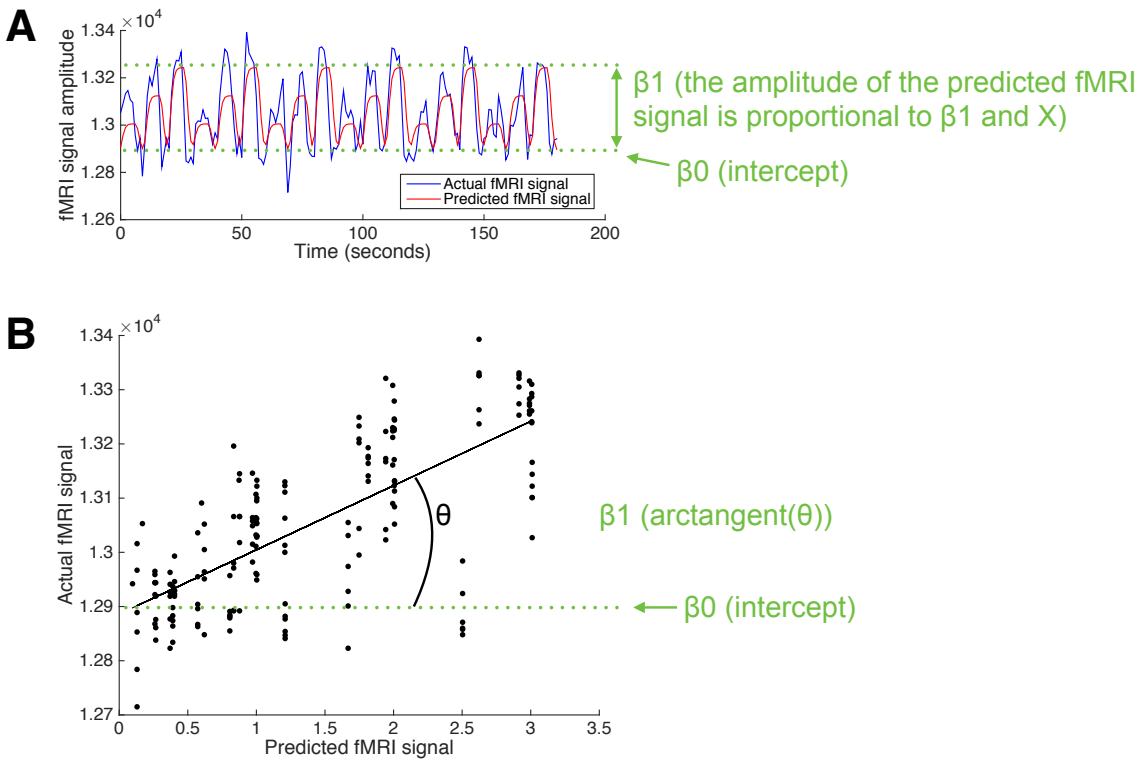


Fig. 5.11: Example fMRI data and model for a simple linear regression. (A) Here, the model fits the data well. Intuitively, to match the model with the actual data, β_0 shifts the model up and down and β_1 changes the amplitude of the model. **(B)** This means that when plotting the actual fMRI data ($y = \beta_0 + \beta_1*X + \text{error}$) against the predicted fMRI data ($y_{fit} = \beta_0 + \beta_1*X$), the slope ($\beta_1$) of the regression line will be steep and the variance will be small. Courtesy of Jill O'Reilly.

As mentioned in the section on Minimizing temporal smoothing, the noise in the fMRI data is temporally autocorrelated and we need to correct for it. Thus, the estimation of the β coefficients requires in fact a few intermediate steps (Poldrack et al., 2011). First, the covariance matrix of the noise cov(ε) can be expressed as a product of the symmetric correlation matrix of the noise V and its variance σ^2:

$$\text{cov}(\varepsilon) = \sigma^2 V$$

We can then prewhiten the data by finding a matrix W and multiplying both sides of the GLM by W, such that:

$$V^{-1} = W'W$$

$$WY = WX\beta + W\varepsilon$$

Which can be re-written as:

$$Y^*=X^*\beta + \varepsilon^*$$

We can solve that equation using the same GLM principles explained above because the errors are now independent:

$$cov(\varepsilon^*) = cov(W\varepsilon) = Wcov(\varepsilon)W' = W\sigma^2VW' = \sigma^2VV^{-1} = \sigma^2I$$

Finally, to test hypotheses and generate t-statistics for individual variables, we divide each β by its variance:

$$t = \beta/var(\beta)$$

This can also be generalized by defining a contrast vector c [1 0 0 0]:

$$t = c\beta/var(c\beta)$$

This formula allows us to perform t-tests across any two different variables. For example, if we want to test whether the activation for variable1 is greater than the activation of variable2 in a model with 4 variables, we will define a contrast vector c [1 -1 0 0]. Finally, to test if any linear combination across all variables explains the data, we will use an F statistic. The t and F statistical tests are associated with a probability (p-value).

5.3.3.2 Significance testing

After obtaining the GLM statistics for each individual voxel in the brain, we can display the voxel intensities as brain maps of β, t, F and p values. To test if these results are statistically significant, we can use the p-values for thresholding. Given that there are thousands of voxels in the brain, we cannot simply threshold at a conventional $p<0.05$ because of the problem of multiple comparisons. However, we also cannot simply correct for multiple comparisons with Bonferroni because this method is too stringent and does not account for the fact that nearby voxels have similar intensities. This happens because

of the nature of the MR signal and the choice of spatial smoothing during data preprocessing. Therefore, a sensitive and physiologically informed approach is to use Gaussian random field theory. The general convention is to threshold the image at $p<0.001$ and then test for significance of $p<0.05$ in the resulting clusters of voxels, depending on their size and the voxel intensities.

When we have clear *a priori* hypotheses about particular locations in the brain, then we can limit our search for activation and do statistics only in those regions: region-of-interest (ROI) analysis (Poldrack, 2007). This approach is advantageous because we can average the signal within the ROI across all voxels and, thus, treat it as an aggregate, rather than voxel-by-voxel. This also means that we can increase sensitivity by reducing the number of voxels tested to a few ROIs. Moreover, we can add strength to a study by examining the fMRI signal in-depth within an ROI defined independently, and illustrate the magnitude and time-courses of the effect. It is crucial that our ROIs are defined independently, to avoid circularity or "double dipping" (see next section on Cross-validation) which invalidates statistical inference (Kriegeskorte et al., 2009). Standard ways of defining unbiased ROIs are based the subject's anatomy, a standard atlas, an orthogonal functional localizer, or voxel coordinates from other related studies and meta-analyses.

5.3.3.3 Group analyses

To test hypotheses across a group of subjects, we can combine analyses from the single-subject level. The goal is to test whether an effect is consistent across the population. This is the usual approach for most fMRI studies, where we want to test if certain brain regions are reliably activated in response to the same condition, across multiple subjects. This involves spatially registering the brains along a common template, such as the standard

MNI152 system, to ensure that the same location in the brain is aligned between different participants.

We can use appropriate statistical models to do inferences not just across the particular subjects in the study, but rather the population from which they were sampled. When a subject is scanned multiple times, we take into account the within-subject variance by using a fixed effects analysis. When we test hypotheses at the group level and want to generalize the results beyond the tested individuals, we use both the within-subject (fixed effects) and the between-subject (random-effects) variances, and thus perform a mixed effects analysis.

Just like in single-subject statistics, the advantage of doing ROI statistics at the group level is that we can average the signal across voxels, hence avoid correcting for multiple comparisons. Moreover, we can define the ROI in each subject, hence we do not need to make sure all voxels across all subjects overlap in standard space.

5.3.4 Cross-validation

Doing unbiased tests for statistical inference is a critical feature of rigorous fMRI analysis that avoids "double dipping" (Kriegeskorte et al., 2009; Poldrack et al., 2011). "Double dipping" means using the same data for selection and selective analysis, or looking for a significant effect in a brain region using the same analysis and data as the ones used for defining that brain region in the first place. This means the analysis becomes circular and this is problematic because it greatly inflates the estimates of the effect sizes.

For example, if we choose an ROI based on statistically significant voxels that respond more strongly to stimulus A than B, and then we want to further characterize that signal in the ROI by averaging across all voxels and doing a t-test for the response to A-B, then we can be certain that the t-test will be positive and significant. There is no way this would

not be the case, even if the voxels had noise instead of real signal. This is simply because we had already selected the voxels based on their statistical significance to a test of A-B. A correct, independent analysis would be to define the ROI according to significance to A+B, thus an ROI that has a large average response to both A and B, but then test inside the ROI for a difference between conditions, hence A-B. This is equivalent to using two separate orthogonal contrasts, A+B and A-B, for data selection and analysis. However, even this approach can be problematic if the actual regressors used to define the contrast are not orthogonal themselves, which can happen if conditions A and B are correlated (Kriegeskorte et al., 2009).

Similarly, a biased approach in multivariate analyses (see the following section on Multivariate statistics) would be to train a classifier on one dataset and then test the classifier on examples from the same dataset. If we use the training data to assess its accuracy, we would overestimate the accuracy because we would overfit the model capturing noise and categorizing it as a neural pattern for a stimulus (Kriegeskorte et al., 2009).

The gold-standard tool for unbiased testing is cross-validation, that is, generating a specific hypothesis in one dataset and testing for it in a completely different dataset (Kriegeskorte, 2015). The hypothesis can be anything from finding significant effect in an ROI or training a classifier to discriminate between different conditions. If we can validate the hypothesis in the other independent dataset, then we can be sure that our analysis was correct and the statistics efficient. Common examples are twofold cross-validation, when we split the data in half, and leave-one-out cross-validation, when we split the data in multiple sessions and we use all sessions but one for training and we use the left-out session for testing.

5.3.5 Multivariate statistics

Multivariate techniques in fMRI are known as multivoxel pattern analysis (MVPA) and rather than asking which brain regions respond most strongly to a stimulus, they use both strong and weak responses to identify more subtle patterns of activity. MVPA takes advantage of small biases in the distribution of functionally specific neurons across neighbouring voxels (Haxby, 2012). They are often used to find the model parameters in an observed dataset and make predictions for a new dataset that has not been observed, yet. This contrasts with univariate statistics which most often find the model parameters that best characterize the observed data, although univariate models can also be used to make predictions about new data (Huth et al., 2016).

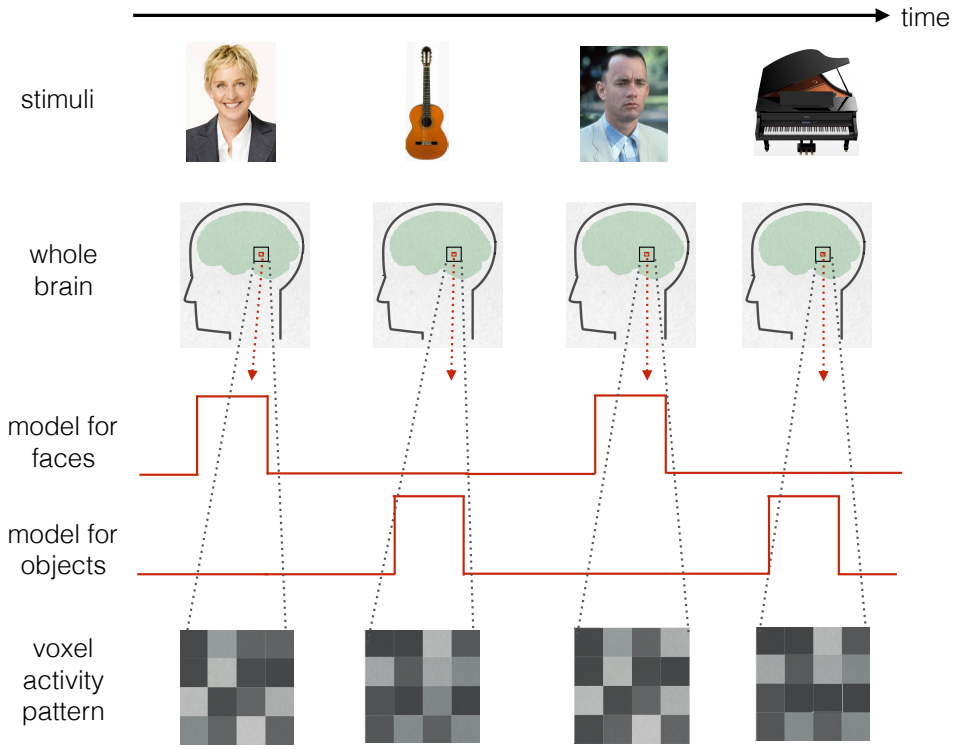


Fig. 5.12: Comparison between univariate and multivariate analyses. Schematic of brain activity inside a voxel (red circle) and inside a ROI (black square). Here, the red voxel responds to faces but not objects, and inside the whole ROI some voxels respond more to faces and others respond more to objects. In univariate analyses, we look for voxels where the model we created matches the data well. Here, the model for faces will predict data better than the model for objects. In multivariate analyses, we look for patterns of activity across voxels inside an ROI. Note that voxel patterns for faces are more similar to each other, and voxel patterns for objects are more similar to each other, whereas voxel patterns for faces and objects are slightly different from each other.

A key advantage of MVPA over traditional univariate analyses is that it facilitates our understanding of functional activations that overlap (Peelen and Downing, 2007). For example, when a group of voxels are activated during two different conditions, there are two main possibilities: either there is a single neural population that uses the same mechanism for both conditions, or there are two different neural populations with distinct mechanisms which co-exist in the same voxel. MVPA can dissociate between these two possibilities because different populations of neurons with different mechanisms will result in different neural patterns across voxels. Nevertheless, with careful univariate analyses using a technique called repetition suppression it is also possible to tackle the same issue (Barron et al., 2016). However, MVPA remains the top choice technique if neural activity patterns extend across large areas, when even single-cell recordings cannot capture some of this variance. For example, this method has shown that faces and other object categories are widely distributed and overlapping in the ventral visual cortex (Haxby et al., 2001; Haxby, 2012). This is interesting because it stands in contrast to an earlier idea that the brain has separate, specialized areas for different categories (Kanwisher et al., 1997).

MVPA is most sensitive for brain regions where neurons with similar function are grouped in clusters that are large enough relative to the voxel size, or are non-uniformly distributed inside the voxel, such that some voxels contain more neurons sensitive to one stimulus and other voxels contain more neurons sensitive to another stimulus (see Chapter 9 for a more in depth explanation). For example, studies have used MVPA in the visual cortex to distinguish neural responses to different stimulus orientations, even if the neural columns that prefer the same orientation are only 500 micrometers width and the fMRI voxel width was six times bigger (Haynes and Rees, 2005; Kamitani and Tong, 2005). However, this technique works even in the prefrontal cortex which does not have such columnar

organization and the individual neurons have mixed selectivity, responding to multiple stimulus features rather than preferring a single one (Zeithamova et al., 2012; Rigotti et al., 2013; Howard et al., 2015; Barron et al., 2016; Brown et al., 2016).

A popular variant of MVPA is representational similarity analysis (RSA), and this was the experimental approach we used in Chapter 9. Instead of evaluating the characteristics of the pattern itself, RSA looks for similarities between different patterns (Kriegeskorte et al., 2008a; Diedrichsen and Kriegeskorte, 2017). For example, animate and inanimate stimuli form separate patterns of activity in the human inferior temporal cortex, as measured with fMRI, and this categorization is remarkably similar in monkey data when using a completely different technique, that is, single-cell recordings (Kriegeskorte et al., 2008c). This study suggests that RSA can be used as a tool that connects different species and different experimental modalities, including fMRI, electrophysiology, behaviour and computational models (Kriegeskorte et al., 2008b).

RSA often uses searchlight procedures, doing analyzes inside a small sphere ROI centered on each voxel in the brain, and assigning the results to the centre voxel. This procedure outputs a brain map showing regions where the analysis is significant, an is useful as an overview or as a functional localizer. However, this analysis can be computationally expensive, with processing time increasing for bigger searchlight spheres and smaller voxels. Smaller spheres are also preferred to larger ones because they are more spatially selective and will not integrate information between different brain regions that might use different neural mechanisms. A common way to minimize the interference from different brain regions and tissues is to perform the searchlight at the cortical surface-level, rather than at the volume-level (Oosterhof et al., 2011). For example, gyri between different brain regions are folded and nearby in volumetric space, but when flattened out on a surface, they will be further apart and hence not included in the same searchlight sphere. Moreover,

the surface includes only grey matter voxels, hence cleaning the data from white matter noise (Fig. 5.13). Finally, surface-based searchlight analyses also reduce the computational load, as the tests will be performed across fewer locations, at the surface vertices instead of at the voxel level.

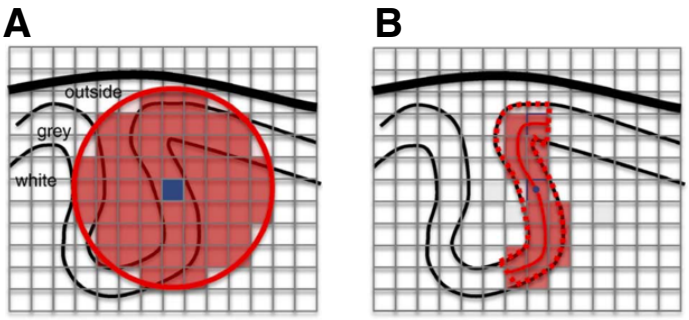


Fig. 5.13: Comparison of voxel selection methods for searchlight analysis. (A) The volumetric searchlight (red) includes voxels from grey matter, white matter and cerebrospinal fluid, from two neighbouring gyri. **(B)** Cortical surface-based searchlights will include only grey matter voxels and will follow the shape and distance of the gyri, minimizing confounding effects from neighbouring gyri that become further away in surface space. Adapted with permission from (Oosterhof et al., 2011).

5.3.6 Encoding and decoding models

Another way of thinking about fMRI analyses is in the ability to give insight into neural *codes*, and the directionality in the relationship between stimuli and brain activity, rather than the number of voxels we use to draw inferences on (Naselaris et al., 2011). This chapter has discussed how we can use stimuli to predict the brain activity using GLM. This is the principle of *encoding* analysis, with the notable characteristics of using low-level features of stimuli, rather than stimuli *per se*, and cross-validation, to infer neural codes. This is the approach used in Chapter 7, where we modelled the movement direction corresponding to a particular visual scene, rather than whether that visual scene *per se* was on or off, and then we cross-validated our results using independent datasets. Other examples of stimulus features used in encoding models are gabor wavelets for detecting scenes in early visual areas (Naselaris et al., 2015) or face shape and appearance (such as

roundness of face, distance between the eyes and skin colour) for facial recognition in the temporal lobe (Chang and Tsao, 2017). The opposite approach is to use the brain activity to mind-read or *decode* what stimulus the participant saw, by inverting the encoding models using a Bayesian framework.

Encoding-decoding models can be combined with univariate-multivariate approaches. For example, most encoding models are indeed univariate, however encoding models that predict patterns of activity across many voxels also exist (Naselaris et al., 2009). Each approach has its own advantages and can be used depending on the research questions, and some studies have combined both approaches (Kay et al., 2008; Naselaris et al., 2009). Encoding models can be related to other representational models, such as multivariate RSA and univariate pattern component modeling using common mathematical frameworks that allow going back and forth between them, when including all data across time and voxels (Diedrichsen and Kriegeskorte, 2017). Each have different constraints, but when combined with careful experimental design and analyses, they can give similar descriptions of the neural population response.

An exciting recent study that used encoding models mapped the meaning of words across the whole brain while human participants listened to hours of stories during fMRI acquisition (Huth et al., 2016). To estimate the semantic selectivity of each voxel, the authors used a univariate approach to model neural responses to semantic features of words, defined as words belonging to the Wikipedia's List of 1000 Basic Words, which includes "weather", but not "foggy" or "breezy". For example, if within a short sequence of words from the story, "foggy" or "breezy" occurred frequently together with the word "weather", then they were considered to belong to the semantic feature "weather", hence they were modeled as the "weather" feature. The features were modeled in a GLM which was run for each individual voxel in the brain. This resulted in brain maps of β values for

each feature. To validate the model, they did cross-validation using the resulting maps to predict the neural responses to a new story, and then correlated the predicted fMRI timeseries with the actual timeseries. The prediction performance was good and remarkably consistent across participants. The authors found that different parts of the brain responded to certain words or concepts, so they grouped them in categories shown in different colours. For example, red areas were associated with social concepts and green areas with visual words. Different participants had similar concepts in similar brain areas. This study used a combination of creative analyses to bring novel evidence on how the semantic system is organized in the brain.

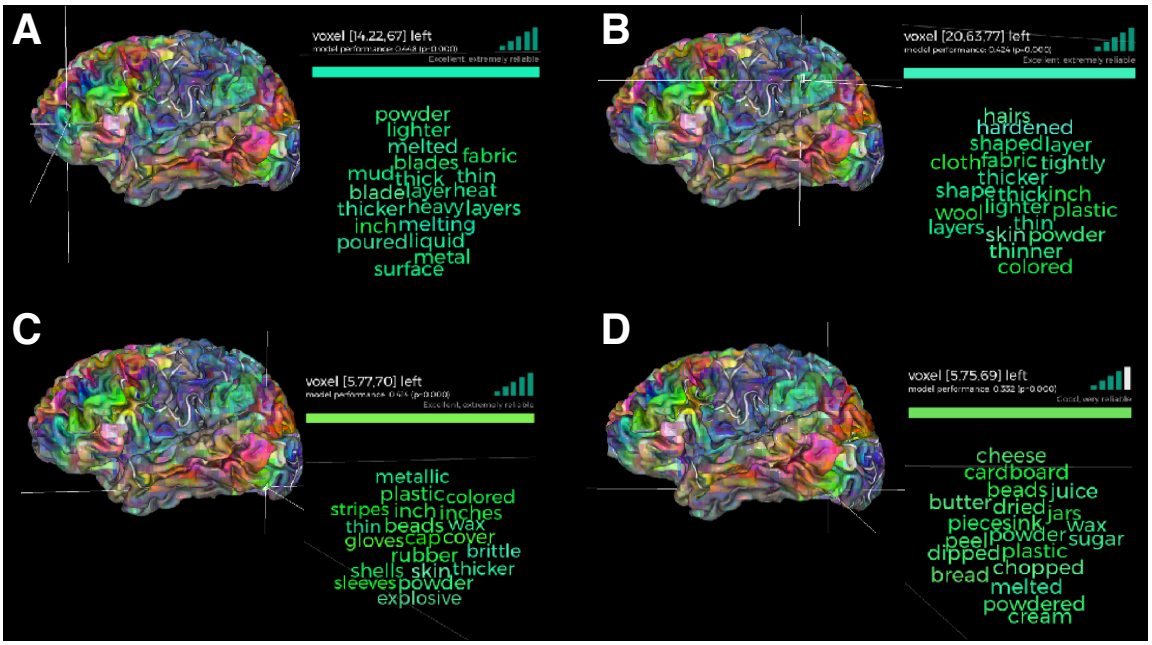


Fig. 5.14: Different meanings of the word "powder" across the human brain. (A) The word "powder" can be found together with words associated with skiing, such as "blades", "thicker" and "layers". **(B)** In a different voxel, "powder" is found together with words from the semantic domain of cosmetics, such as "skin" and "colored". **(C)** "Powder" and "explosive" are suggestive of dynamite. **(D)** "Powder", "sugar", "cream" and "jars" are indicative of "confectionery". The voxels were discovered with the brain viewer (http://gallantlab.org/huth2016/) based on (Huth et al., 2016).

Other studies have used encoding models in combination with decoding models to identify (Kay et al., 2008) and even reconstruct (Naselaris et al., 2009) natural images from the human brain activity. Similarly to (Huth et al., 2016), a model is estimated for each voxel,

in this case using Gabor wavelets as features of images from dataset A. Second, the subject views novel images from dataset B and the model estimated from dataset A is used to predict brain activity. However, instead of correlating the predicted fMRI activity across time with the actual activity for model validation as in (Huth et al., 2016), the authors correlated the predicted voxel activity patterns that would be evoked by each of the new images, with the actual voxel activity patterns (Kay et al., 2008). The image whose correlation was strongest was selected as the decoded image seen by the observer. Thus, identification is a type of decoding where we try to determine which specific stimulus a subject has seen, when the stimulus is drawn from a pool of known stimuli. A subsequent, ambitious study showed that it is possible to decode using reconstruction of a new image from the voxel patterns, without searching through a list of possible images (Naselaris et al., 2009). This latter approach has implications for "mind reading" in situations when we do not know what possible stimuli the participant saw or imagined, and some early attempts have been made to decode dreams (Horikawa et al., 2013).

Chapter 6: A novel behavioural method for navigating conceptual knowledge

We developed a battery of behavioural tasks ("Stretchy Birds") that allowed human participants to navigate a non-spatial, conceptual environment. These tasks were analogous to the ones used for navigation in physical space (Doeller et al., 2010), with the notable difference that our dimensions were organized in an abstract, rather than physical, space. The abstract space was continuous, two-dimensional, and consisted of bird-shaped stimuli, with the legs and neck lengths being the features of the two orthogonal dimensions. Subjects navigated this "bird space" by morphing these features to generate bird shapes previously associated with specific outcomes, by analogy to a standard stimulus-outcome task. We aimed to use these tasks to look for fMRI grid-like signals of conceptual knowledge in humans (see next chapter). The work in this chapter is published in (Constantinescu et al., 2016a).

6.1 Introduction

This chapter explains in detail how we developed a set of behavioural tasks, as a novel method of human navigation in a continuous, two-dimensional abstract space. The next chapter explains how we let participants navigate in this abstract environment while scanning their brains with fMRI, aiming to detect grid-cell signals which can be identified by a characteristic hexagonal symmetric pattern. In this chapter, I will refer briefly to this hexagonal symmetry when showing that we were careful not to make our tasks hexagonal, hence any hexagonal patterns in the fMRI signals were caused by the bulk population of grid cells instead of by something special inherent to the tasks.

We aimed to develop an abstract navigation task that is similar to that used for navigation in physical space (Doeller et al., 2010). We used birds as inspiration because they vary

naturally across two dimensions, the lengths of the neck and legs. We then aimed to develop a set of behavioural tasks that would allow moving through this abstract environment and learning the locations of landmarks using a non-spatial controller. I will present multiple measures of learning that show how participants, after intensive training, knew how to navigate precisely towards specific locations in this environment. This paradigm facilitated the detection of grid-like signals with fMRI (next chapter).



Fig. 6.1: Birds vary naturally along the lengths of their neck and legs.

6.2 Methods

6.2.1 Participants and experimental design

28 participants (16 men; mean age 26) reported to be in good health with no history of neurological or psychiatric disease and with normal or corrected-to-normal eye vision participated in this experiment at the Oxford Centre for Functional MRI of the Brain (FMRIB). One participant was excluded from all further analyses because he did not learn how to use the controller for navigation in abstract space. Four participants were trained less than the rest of the group. To avoid a confounding effect of the training regimes, we excluded them from the correlations with behavior. Subjects were trained outside the MRI

scanner to navigate in a continuous abstract "bird space". All participants gave written informed consent and were paid for participating, as approved by the ethics committee of the University of Oxford (ref. number 2013-171).

Subjects performed a stimulus-outcome (S-O) learning task in which they learned that bird stimuli (Davis and Poldrack, 2014) were associated with different Christmas symbols (Fig. 6.2A). The study therefore resembles other S-O learning tasks except that here the bird stimuli were not independent fractals or symbols but, instead, they varied according to two continuous dimensions: the lengths of the neck and legs. Each stimulus could therefore be described within a two-dimensional conceptual "bird space" (Fig. 6.2B). Even though the features of this bird space were lengths as in physical maps, here they had to be extracted from a one-dimensional (vertical) visual space where different stimuli did not form any angles (Fig. 6.2A) and transformed into a two-dimensional conceptual map (Fig. 6.2B).

6.2.2 Navigation in abstract space is similar to navigation in physical space

We developed a set of novel tasks for navigation in an abstract space that were designed to be analogous to those used for navigation in a physical space. We trained the participants extensively to be able to navigate precisely in this abstract space. Using keyboard button presses, subjects learned to morph the birds, to explore the abstract space and they learned the locations of the outcomes. Each location in this abstract space was represented by a bird stimulus. A trajectory through the bird space was equivalent to morphing one bird stimulus into another, that is, watching the neck and legs stretching vertically. The direction $\boldsymbol{\theta}$ of the trajectory was the ratio between how much the neck and legs were changing relative to each other. Subjects were instructed to learn which bird shape was associated with each outcome and we purposefully did not instruct them that these associations could actually be organized in a continuous 2D "bird space". The video in

(Constantinescu et al., 2016d) illustrates a recording of the task and the rationale of our experimental design.

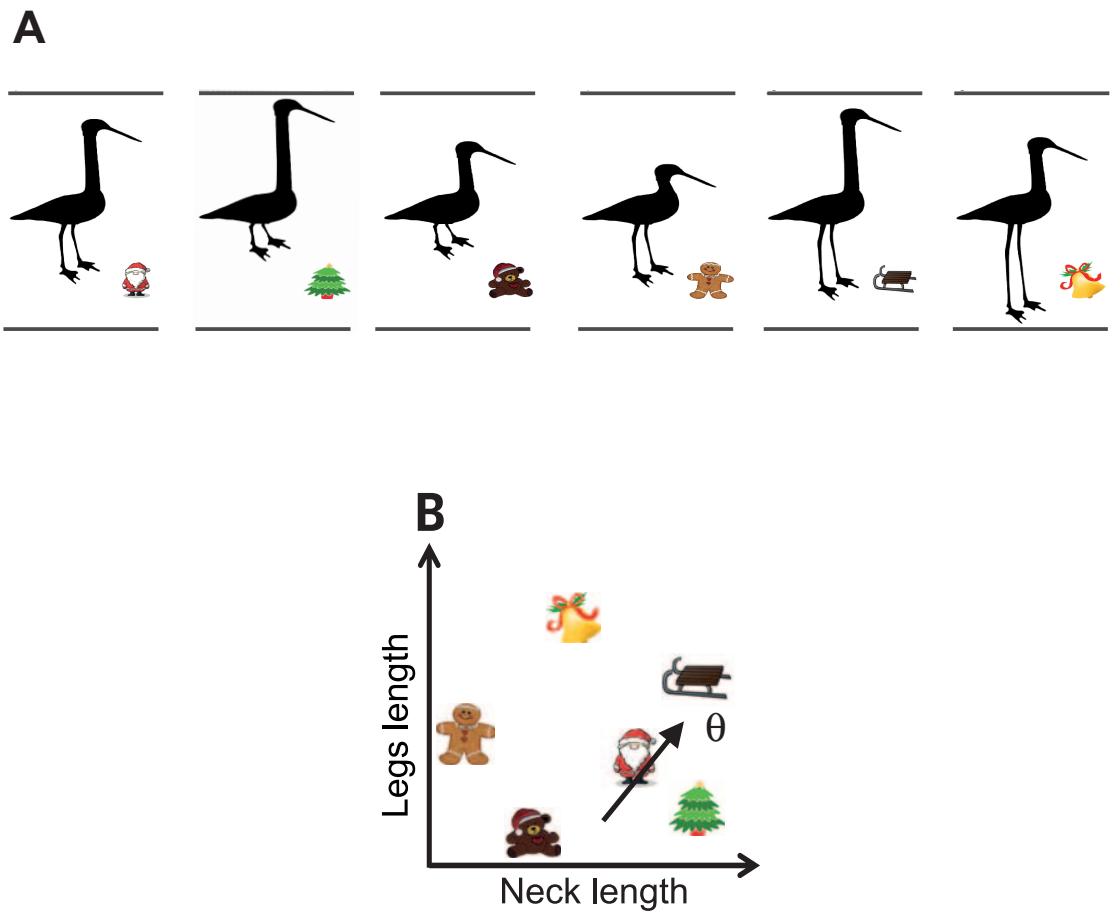


Fig. 6.2: Bird space. (A) Subjects were trained to associate stimuli (birds) with outcomes (Christmas symbols). **(B)** Example trajectory in abstract space. A location in this abstract space was represented by a bird stimulus. A trajectory was equivalent to visually morphing one bird into another. The direction θ of the trajectory depended on the ratio of the rates of change of the legs and the neck (Constantinescu et al., 2016d). Subjects were not consciously aware that these associations could be organized in a continuous "bird space".

6.2.3 Match task

First, subjects played the *match task* where they were trained to use a controller to morph the bird silhouette shown on the left side of the screen to match the bird on the right side of the screen (Fig. 6.3). The two birds varied only in two dimensions: the lengths of the neck and legs. Thus, the participants could morph the bird by changing the lengths of the neck and legs. To ensure that participants integrated between these two dimensions

simultaneously, we precisely instructed them to morph the bird by choosing the ratio between how much the neck and legs change relatively to each other only once and as correctly as possible (as opposed to changing the lengths of the neck and legs separately). To ensure that subjects did not do this task spatially, we created a non-spatial controller for choosing this ratio. The participants pressed buttons on the keyboard to choose the ratio and then to morph the bird with that ratio. The morphing was continuous and they saw a bird shrinking or stretching its neck and legs. Each of the two black bars of the controller signified how much that dimension would change. If the black bar was positioned on the midline, the corresponding dimension did not change. The closer to the top line the bar was, the more that dimension would increase. Conversely, the closer to the bottom line the bar was, the more that dimension would decrease.

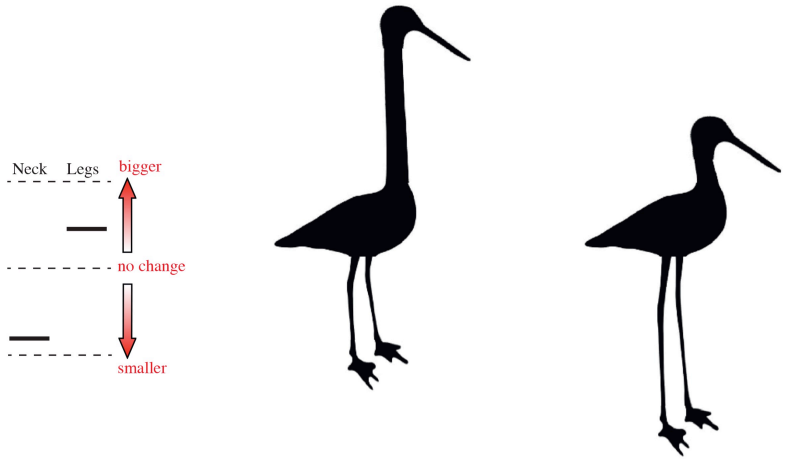


Fig. 6.3: Example screenshot of the match task. Here, to morph the bird on the left into the one on the right, the neck needed to decrease more than the legs needed to increase. The correct neck:legs ratio to achieve this is shown on the controller on the left side of the screen (notice the position of the two black bars).

6.2.4 Explore task

After the participants learned how to perform the *match task* correctly, they continued with the *explore task* (Fig. 6.4). In this game, subjects needed to find six outcomes (images of Santa Claus, a Christmas Tree, a gingerbread man, a sledge, a bell and a teddy-bear) by freely morphing the bird. Each outcome was associated with a different bird (Fig. 6.2). Participants knew they found an outcome because it would pop up next to its associated bird. Critically, this ensured that subjects became familiar with the entire bird space and not just with the stimuli paired with outcomes and that they had a clear representation of what one unit distance on each axis (neck, legs) looked like. Thus, we did not simply train the participants on the mere associations between the six birds and the six outcomes (i.e. only the locations paired with outcomes in the abstract space).

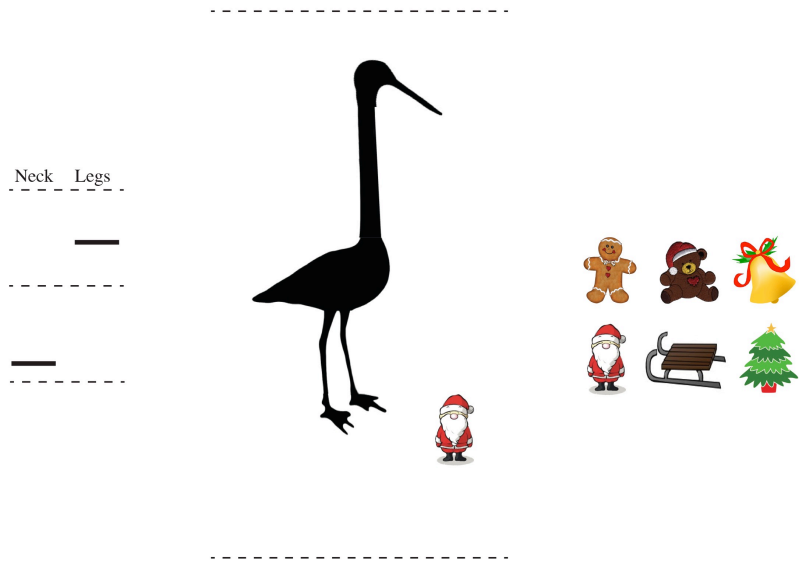


Fig. 6.4: Example screenshot of the explore task. Here, the subject found Santa Claus.

This approach is analogous to exploration tasks in open environments in rodents (Dupret et al., 2010). Indeed, participants did not know a priori which birds were paired with outcomes, hence, they had to morph through many different unpaired birds to find the paired ones (i.e. they explored many locations in the abstract space that were not paired

with outcomes). The locations of the outcomes were identical in all participants. This allowed us to identify learning patterns of the environment across participants, and rule out the possibility that inter-individual differences between participants were caused by the outcome locations. We carefully selected these locations to rule out any potential hexagonal arrangements.

6.2.5 Collect task

When subjects knew the stimulus-outcome associations well, they continued to train with the *collect task*. In this task, we periodically tested their knowledge of the bird space by asking them to find specific outcomes from random start positions, by correctly choosing the neck:legs ratio only once. Thus, they had to recall the correct neck/legs ratio associated with that outcome, and generate the abstract trajectory to that bird correctly in one go.

6.2.6 Recall task

Subjects were trained with the version of the task that would also be used subsequently during scanning (Constantinescu et al., 2016d) and the fMRI results are presented in detail in the next chapter. During fMRI scanning, we wanted to present participants with a controlled set of trajectories, thus, we devised a task in which participants watched videos of birds morphing into different birds according to pre-defined neck:legs ratios. Each trial corresponded to one trajectory and consisted of several stages: morphing, imagination and, in a subset of trials, choice (Fig. 6.5).

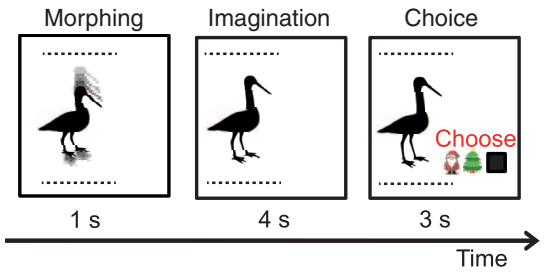


Fig. 6.5: Example trial of the recall task. This trial corresponds to the trajectory with direction θ shown in Fig. 6.2.

During the morphing stage, the participants saw only a single trajectory in the abstract space. Next, they were instructed to imagine the bird continued to morph with the same neck:legs ratio, for the same amount of time and the same speed, and recall if any of the imagined bird stimuli was associated with an outcome. This means they did not only need to estimate equivalents of angles of moving in space, but also equivalents of distances. In other words, we split the trajectories in half, and subjects saw the first half (visual morphing of the bird) and had to imagine the second half (imagined morphing of the bird). Finding grid-like representations during imagined navigation in physical space is indeed possible, as shown in a recent study (Horner et al., 2016).

Finally, subjects could choose one out of three possible outcomes (outcome 1, outcome 2 or no outcome) by pressing one out of three keys on a button box. To control for the effect of the outcomes, we designed two versions of the task. In one version, these outcomes were displayed during the entire duration of the trial and subjects made a choice immediately after the imagination period. In the other task version, participants only saw the outcomes at the end of the imagination period, with a 50% probability, and then were asked to make a choice.

We deliberately designed the trajectories to sample all angles uniformly, across all trials, as well as separately for trials paired or unpaired with outcomes (Fig. 6.6). Trajectory

angles were evenly sampled across the 360° space (varying between 0°-360°). Histograms are shown for trajectory angles passing through all locations in conceptual space, and categorized as a function of passing through an outcome or not (Fig. 6.6A; Rayleigh's test, p=1, p=0.626 and p=0.64, respectively). To investigate the distribution of angles relative to hexagonal symmetry, we transformed the angles into a 60° space (varying between 0°-60°), using the modulo operation, which computes the remainder after division of the angles by 60. For example, a 70° trajectory in 360° space corresponds to a 10° trajectory in 60° space. We found that the trajectory angles were evenly sampled across the 60° space for all locations in sum, and separately for the locations paired and unpaired with outcomes (Fig. 6.6B; Rayleigh's test, p=1, p=0.623 and p=0.637, respectively).

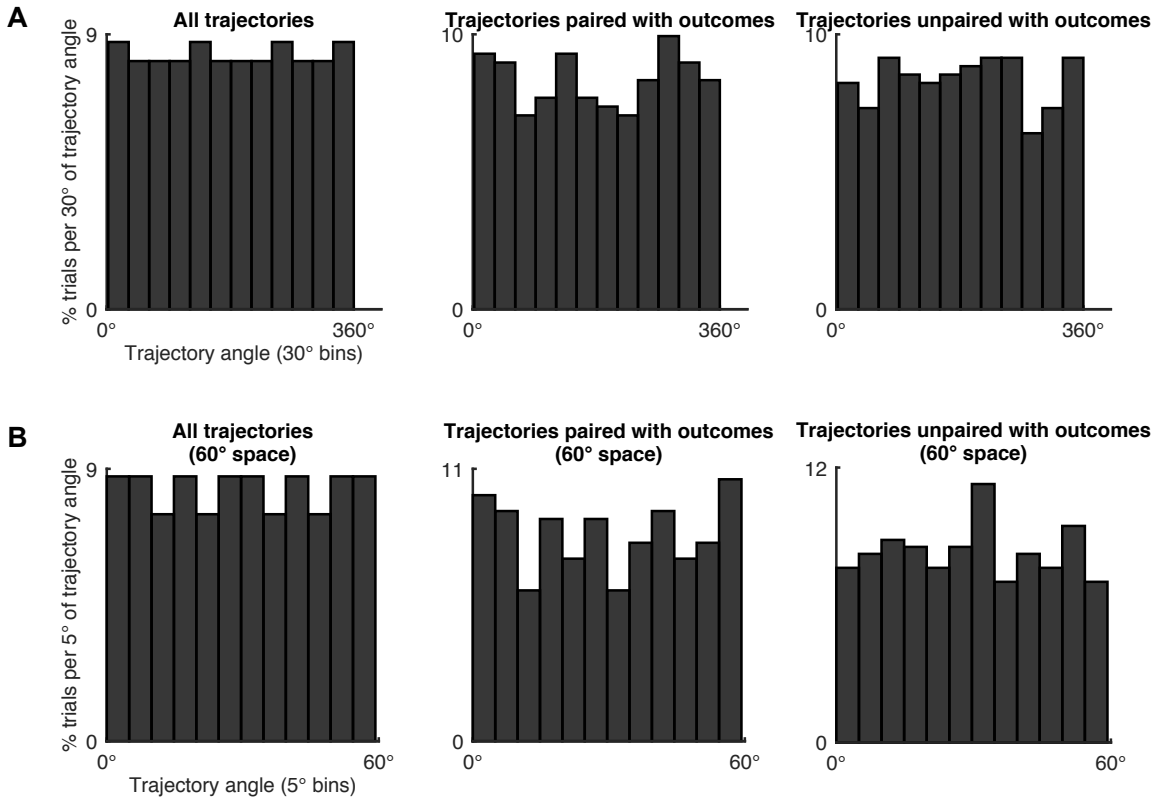


Fig. 6.6: Distribution of trajectory angles in the recall task. Angles in **(A)** 360° space and **(B)** 60° space are uniformly distributed across all trials, as well as separately across trials paired and unpaired with outcomes.

6.2.7 Post-scanning debriefings

Even though the stimuli were not 2D spatial, it would have been possible for the subjects to solve the problem using conscious and explicit analogy to 2D space. Such a spatial strategy would go beyond the scope of previous reports of the function of grid cells but would retain space as a 2-dimensional anchor for the grid. Alternatively, subjects may have had no conscious knowledge of the spatial arrangement of the birds. To gain some handle on this interesting question, we asked subjects after scanning to describe the strategy they used to do the task and to remember the associations between the birds and the outcomes.

6.3 Results

6.3.1 Explore task

We tested for measures of learning during the *explore task*, where subjects navigated freely to look for the outcomes. We concatenated the data from all training sessions in each subject, and then split it into five equal parts (quantiles). We then computed the amount of time spent in each part of the environment during each quantile and plotted it as color-coded trajectory maps (Fig. 6.7A, yellow is maximum, dark blue is 0). Note how subjects first navigated in most parts of the abstract space, choosing random trajectories to find the outcomes. As training progressed, subjects navigated more precisely towards the outcomes.

Next, we estimated the amount of time spent navigating at the edges ("time at edges") and at the locations paired with outcomes ("time at outcomes"), relative to the total time spent navigating. At the group level, we averaged these measures across participants. As training progressed, participants spent less time at the edges (times at edges, quantile1 vs quantile5, t_{22}=3.1776, ** $p<0.01$) (Fig. 6.7B) and more time navigating directly between locations

paired with outcomes (time at outcomes, quantile1 vs quantile5, t22=-3.1773, ** p<0.01) (Fig. 6.7C). These results demonstrate that the identities of the birds paired with outcomes were well learnt in the subject cohort, conceptually equivalent to learning object locations in a 2D spatial environment. This means that subjects did not only need to indicate if an outcome occurred somewhere on the imagined trajectory, but also at which distance.

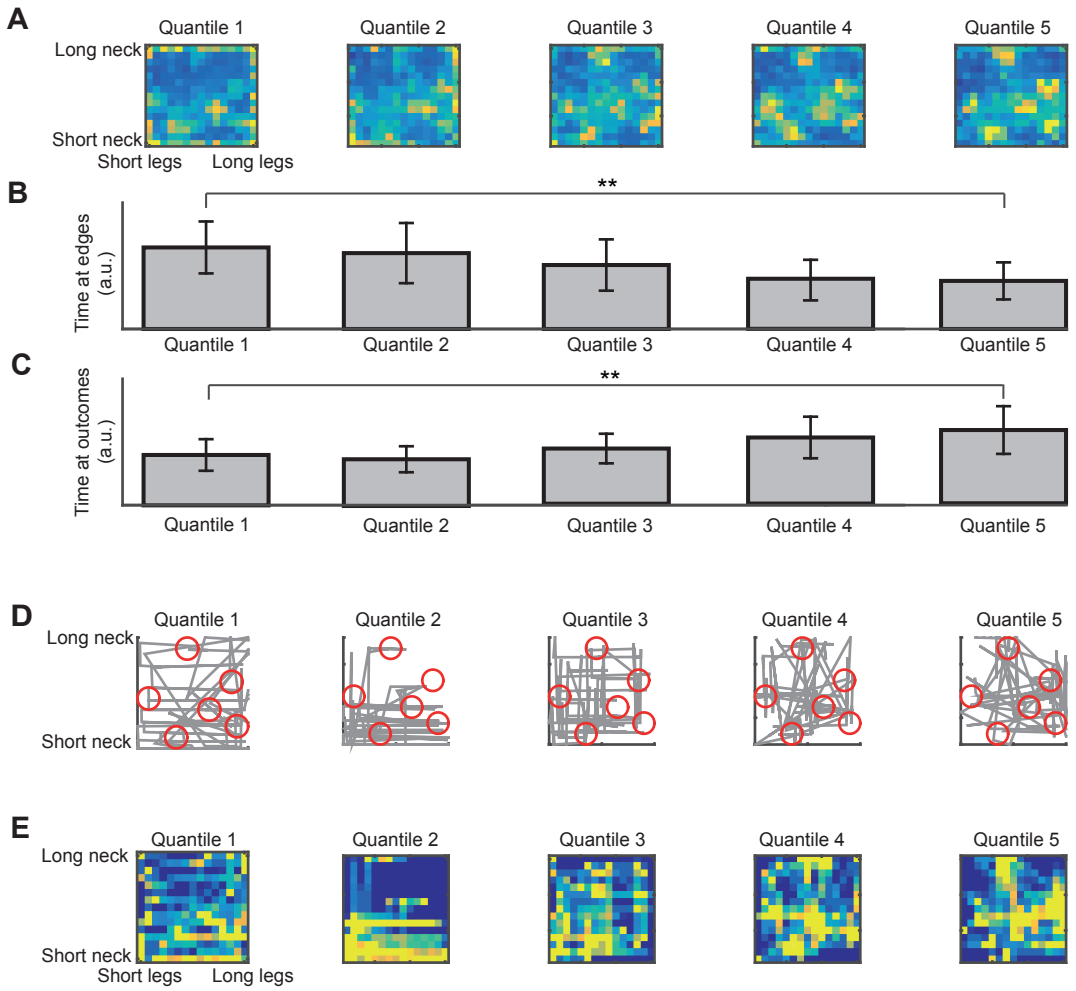


Fig. 6.7: Analysis of the explore task at the group level and in one representative subject. **(A)** As training progressed, participants navigated towards the outcomes more directly, **(B)** thus they spent less time navigating around edges and **(C)** more time at the outcome locations. **(D, E)** One example subject at the beginning of training explored more the edges in quantile 1 and the lower part of the environment in quantile 2.

To give an example in one representative subject, we plotted the trajectories (grey lines) and the locations paired with the outcomes (red circles) across time (Fig. 6.7D), as well as the corresponding color-coded trajectory maps (Fig. 6.7E). Note how this subject spent

more time at the edges in quantile1 and explored more the lower area of the environment in quantile 2.

6.3.2 Collect task

In the *collect task*, subjects were precisely instructed to find specific outcomes. We computed the mean number of transitions needed to find an outcome, in each trial. A perfect performance is equivalent to making only one transition to find the target outcome, thus, choosing the correct neck:legs ratio only once and generating the abstract trajectory to the target bird in one go. In contrast, a poor performance means making more than one transition. Subjects made fewer transitions (t33=-2.89, ** p<0.01) in day 2 than day1 (Fig. 6.8A).

Moreover, a precise trajectory would have a small angle error, defined as the angle between the ideal trajectory angle and the angle of the first transition. Thus, we next computed the percentage of trials where participants had an angle error < 15°. The percentage of angle errors < 15° increased in day 2 compared to day 1 (t_{33}=2.37, * p<0.05) (Fig. 6.8B).

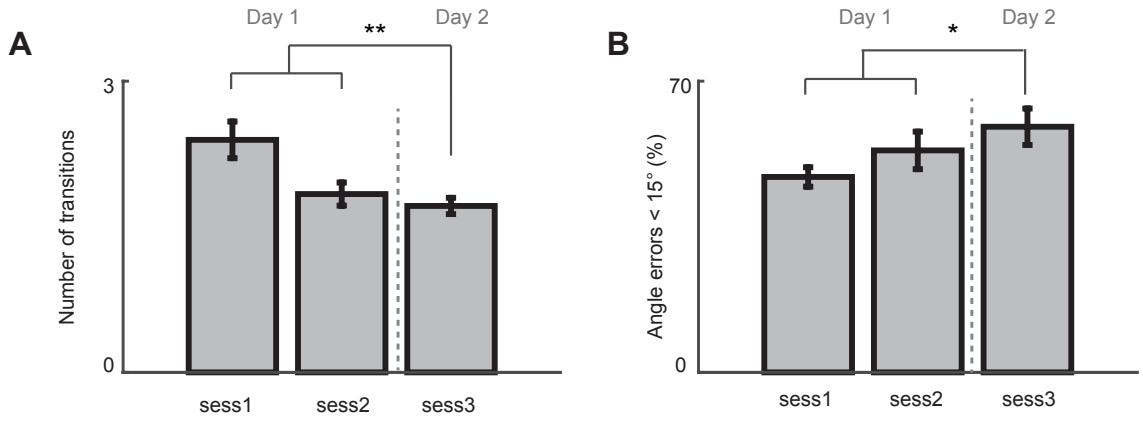


Fig. 6.8: Measures of learning during the collect task. Participants made significant improvements in training day 2 compared to training day 1. **(A)** They made fewer transitions, and **(B)** the percentage of angle errors < 15° increased.

6.3.3 Recall task

In the recall task in the scanner, we computed the performance as percentage of correct responses in each scanning session. Participants continued to improve performance each day, possibly because they received fresh training before each day's scanning (Fig. 6.9; day 2 vs day 1, t41=3.89, *** p<0.001). Across all sessions, participants reached a performance of 72.8 ± 1.0% accuracy in predicting the outcomes.

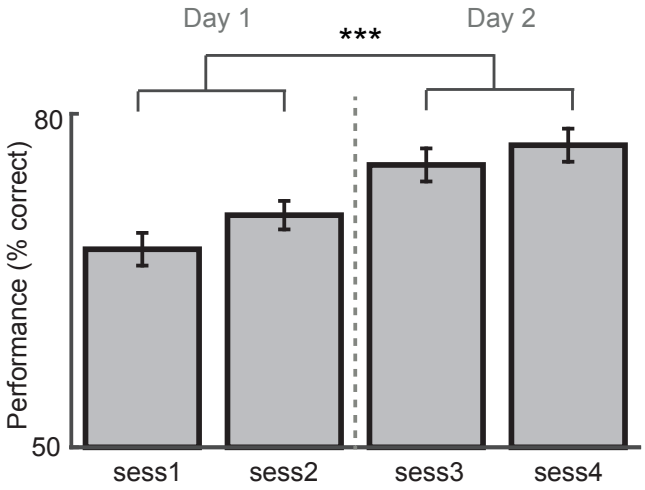


Fig. 6.9: Learning in the recall task. Effects of fresh training on subsequent performance in the scanner. Participants made significantly more correct responses in day 2 compared to day 1.

6.3.4 Post-scanning debriefings

When the subjects were asked what strategy they used to do the task and to remember the associations between the birds and the outcomes, none of them reported conceiving of the relationships between birds or outcomes as lying in a spatial map. For example, some strategies for remembering what bird gave the Christmas Tree were: visualizing the precise lengths of the neck and legs as a mental picture, using internal metrics such as the bird has 4 units of neck and 1 unit of legs, verbally repeating in their minds "long neck, short legs", finding similarities between the characteristics of the bird and the outcome such as the bird

had a long neck (trunk) and short legs (base) like a true Christmas Tree and, also, finding similarities between different birds associated with outcomes (e.g. the Christmas Tree bird had longer neck but similar legs with the Teddy Bear bird). Thus, no participant described being aware that birds could be associated spatially, on a two-dimensional map.

For certainty, we subsequently showed them the map of the abstract space (Fig. 6.2B) and asked them whether this spatial strategy was "easier", "harder" or "the same" as their own strategy. Two participants reported that the spatial strategy had the same level of difficulty as their own strategy, however they clarified that their approach was not spatial and were unaware that the birds could be arranged into a 2D map. Everyone else reported that the spatial strategy was easier or harder than their own. Overall, no subject reported doing the task spatially.

6.4 Discussion

This chapter has shown how a standard stimulus-outcome task, but in a structured stimulus space, can be used to train human participants to navigate in a two-dimensional, continuous abstract space, just like they would navigate in a two-dimensional, continuous physical space (Doeller et al., 2010). For the purposes of our experiment, the abstract space was formed by birds with different body shapes, where the x and y axes corresponded to the lengths of the neck and legs, respectively. A bird stimulus could be morphed into many different, discrete stimuli, creating a conceptual space for the brain to navigate. For example, a bird silhouette resembling a duck could be morphed into a swan by stretching its neck. We reasoned that if humans exploring this "bird space" showed grid cell signals while their brains were scanned with fMRI, then this suggests grid cells organize conceptual knowledge, just like they organize physical space (see next chapter).

To facilitate the detection of these signals with fMRI we first made sure our participants navigated this conceptual space with accuracy, as shown in our multiple measures of learning. Subjects had to know exactly where they were in conceptual space at each time point, by constantly integrating between the absolute lengths and legs of the birds. This allowed them to know precisely at what distance on each trajectory an outcome occurred. In fact, participants would have not been able to perform the task correctly had they not estimated the distances accurately. For example, if they miss-estimated the trajectory as being too short, then a trajectory paired with an outcome would have been represented as a non-outcome trajectory because they did not reach the location of the reward. Thus, they would have chosen the *no outcome* option (which was always one of the available options) instead of choosing the correct outcome option. Similarly, if they miss-estimated the trajectory as being too long, then the trajectory could have intersected more than one outcome. This would have led the participants to make again an incorrect response, by choosing an incorrect reward option.

Moreover, during training, participants first estimated the neck:legs ratio (the equivalent of moving direction) and then how much to change the size of the bird (the equivalent of the distance on the trajectory where the rewards were located). With further training, subjects refined their estimates of reward locations from looking for them incorrectly near the boundaries of the environment, to finding them closer to their actual locations, towards the centre of the environment. Critically, by the end of training participants knew which birds led to the outcomes, not only in terms of the ratio of neck:legs but also in terms of the actual appearance of the bird. This is equivalent to knowing a location in a two-dimensional space, which can be reached from any point by estimating both the angle of movement and the distance.

Second, we made sure that the stimulus space was well-controlled. Indeed, the angles of moving in conceptual space were sampled across all directions. This was also the case for trajectories that went through an outcome or not. This means that the angles were uniformly distributed across space and did not have any hexagonal arrangements. Therefore, any fMRI hexagonal signals would be indicative of grid cells, rather than mirror some particular features of the task that were independent from grid cells (see next chapter).

Finally, the participants were not aware that these stimuli could be organized in two dimensions. What participants saw was only the visual space of the complex one-dimensional bird stimuli. These visual stimuli were morphed only along one dimension, vertically. The reason we designed the task this way was to make it as non-spatial as possible. Indeed, in the visual space, the bird is not moving in two dimensions at all and there are no angles in one dimension. This means there are no angles in the bird stimuli themselves. However, angles exist in a two-dimensional conceptual space and our fMRI experiment relied on finding those angles. Indeed, we needed a two-dimensional space to find grid-cell patterns. To make a two-dimensional space out of our bird stimuli, participants had to realize that there are two independent components of the bird, the neck and the legs, and map them onto two orthogonal dimensions of a conceptual space.

Thus, it is possible that the birds are subconsciously mapped onto existing spatial grids, though this task would require very complex processing and it would be extremely interesting if such a mapping could be achieved. Indeed, this would be just as interesting as the ability to form new grids for conceptual space. This point is completely independent of the particular dimensions we chose to use because this would be equally true for dimensions that are measured in lengths, as it would be for dimensions that are measured in colour intensity, or sound frequency (Aronov et al., 2017).

Chapter 7: Organizing conceptual knowledge in humans with a gridlike code

It has been hypothesized that the brain organizes concepts into a mental map, allowing conceptual relationships to be navigated in a manner similar to that of space. Grid cells use a hexagonally symmetric code to organize spatial representations and are the likely source of a precise hexagonal symmetry in the functional magnetic resonance imaging signal. Humans navigating conceptual two-dimensional knowledge showed the same hexagonal signal in a set of brain regions markedly similar to those activated during spatial navigation. This gridlike signal is consistent across sessions acquired within an hour and more than a week apart. Our findings suggest that global relational codes may be used to organize nonspatial conceptual representations and that these codes may have a hexagonal gridlike pattern when conceptual knowledge is laid out in two continuous dimensions. The work in this chapter is published in (Constantinescu et al., 2016a).

7.1 Introduction

Humans have a remarkable capacity for generalizing experiences to novel situations (Kumaran et al., 2009; Barron et al., 2013b). It has been hypothesized that this capacity relies on a "cognitive map," allowing conceptual relationships to be navigated in a manner similar to that of space (Tolman, 1948a; O'Keefe and Nadel, 1978; Buzsáki and Moser, 2013; Eichenbaum and Cohen, 2014). Grid cells use a hexagonally symmetric code to organize spatial representations (Hafting et al., 2005). Here we ask whether conceptual knowledge may also be organized by gridlike codes.

Human grid cells have been identified during intraoperative recordings (Jacobs et al., 2013) and are the likely source of a precise sixfold (hexagonal) symmetry in the functional

magnetic resonance imaging (fMRI) signal, as a function of movement direction during virtual navigation (Doeller et al., 2010; Kunz et al., 2015; Horner et al., 2016). This hexagonal signal varies depending on whether the direction of moving in space is aligned or misaligned with the orientation of the grid (Fig. 7.1 D and E). Such a signal is a precise and unusual prediction for fMRI: It is predicted in the bulk activity because grid cells share a common grid axis (Barry et al., 2007; Stensola et al., 2012; 2015), and conjunctive grid cells fire faster on average when movement is aligned to this axis (Doeller et al., 2010). It pertains to the moving direction and cannot be explained by any characteristics of the currently experienced visual scene. It is hexagonal and therefore does not align on average to the cardinal directions. Its temporal waveform is different in every scan and every subject and thus cannot be easily predicted by imaging artefacts (see discussion).

This gridlike signal is not unique to the entorhinal cortex, but can be measured during spatial navigation in prescribed parts of the medial frontal, medial parietal, and lateral temporal cortices (Doeller et al., 2010). Despite no report in rodents of grid cells outside the hippocampal formation, direct recordings during brain surgery in humans have confirmed gridlike firing patterns in some of these areas (Jacobs et al., 2013). This same network of brain regions, often referred to as the "default mode network" (Fox and Raichle, 2007), is also regularly activated in nonspatial tasks that involve the manipulation of conceptual knowledge, such as memory (Binder et al., 2009), imagination (Schacter et al., 2012), scene construction (Hassabis and Maguire, 2007), valuation (Clithero and Rangel, 2014), and theory of mind (Saxe et al., 2004), and in situations when subjects must generalize learned concepts to novel situations (Kumaran et al., 2009; Barron et al., 2013b; Benoit et al., 2014).

The ability to interact with knowledge in this flexible and generalizable way is the central advantage of maintaining an explicit cognitive map (Tolman, 1948a). Together with the

regions' established role in nonspatial conceptual generalization, the finding of gridlike activity in these brain regions during spatial navigation therefore raises the possibility of common neural coding mechanisms for storing spatial and conceptual representations. Indeed, this hypothesis is strengthened by the findings that hippocampal cells (analogous to rodent place cells) encode individual concepts in humans (Quiroga et al., 2005) and sound frequency in rodents (Aronov et al., 2017), and that rodent grid cell coding may not be restricted to spatial dimensions, but also represent time (Kraus et al., 2015).

We used fMRI to test if humans use a hexagonally symmetric code when navigating through abstract conceptual representations. We designed a task analogous to the one used for navigation in physical space (Doeller et al., 2010), with the notable difference that our dimensions were organized in an abstract, rather than physical, space.

7.2 Methods

7.2.1 Participants and experimental design

Participants had extensive experience of navigation in a continuous "bird space" the day before scanning (Fig. 7.1 A and B). The previous chapter has an in depth description of the rationale of the "bird space" and the battery of tasks we used to train our subjects. In brief, participants learned which bird stimuli were paired with outcomes by freely morphing the neck and legs dimensions. This ensured that subjects became familiar with the entire bird space and not just with the stimulus-outcome pairings. Indeed, participants progressively refined the locations of the outcomes through training (Fig. 7.1F). We periodically tested their knowledge of the bird space by asking them to find specific outcomes from arbitrary start positions, by correctly choosing the appropriate neck:legs ratio. When asked to make such ballistic movements, subjects significantly increased their precision through training, and therefore learned to visualize the target bird (Fig. 7.1G). In

the scanner, participants continued to improve performance each day, possibly because they received fresh training before each day's scanning (Fig. 7.1G).

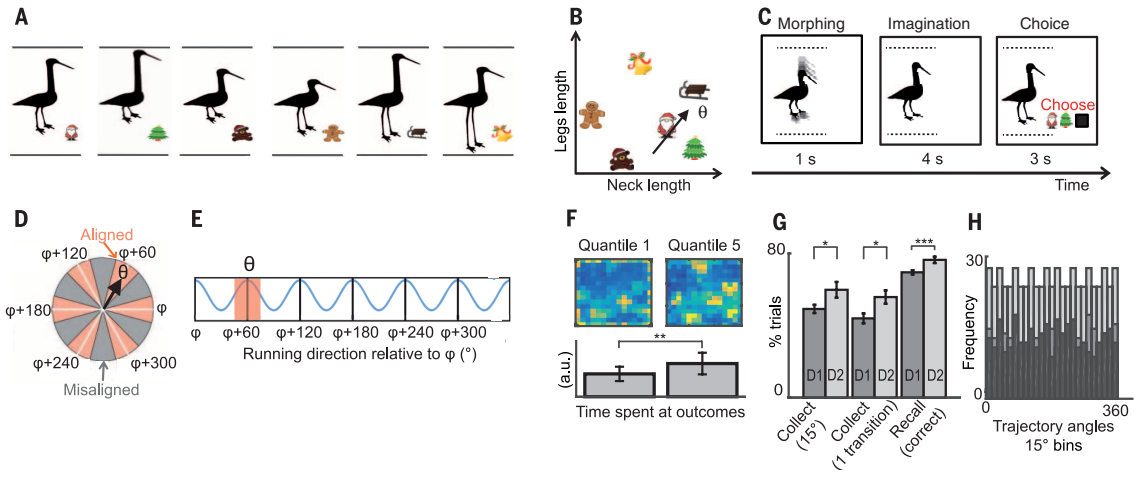


Fig. 7.1: Experimental design for navigation in abstract space. (A) Subjects were trained to associate stimuli (birds) with outcomes (Christmas symbols). **(B)** Example trajectory in abstract space. A location in this abstract space was represented by a bird stimulus. A trajectory was equivalent to visually morphing one bird into another. The direction θ of the trajectory depended on the ratio of the rates of change of the legs and the neck (supplementary video in (Constantinescu et al., 2016d)). Subjects were not consciously aware that these associations could be organized in a continuous "bird space." **(C)** Example trial corresponding to the trajectory with direction θ. **(D)** Trajectories can be categorized as aligned (red sectors) or misaligned (gray sectors) with the mean orientation φ of the hexagonal grid. Note that φ is different for each participant (see the methods for details on how φ was calculated). Here, the direction θ is aligned with the grid. **(E)** fMRI markers of grid cells showing hexagonal symmetry: The signal is bigger for trajectories aligned versus those misaligned with the grid. **(F)** (Top) Color-coded trajectory maps illustrating time spent in each part of the environment during the "explore" task in the first (quantile1) and last parts of training (quantile5). Yellow is maximum and dark blue is 0. (Bottom) Barplots showing the amount of time spent at the locations/stimuli paired with outcomes in each epoch relative to the total time spent navigating ("time at outcomes" quantile1 versus quantile5, $t_{22} = -3.17$, **P < 0.01). **(G)** In the "collect" task, participants made significant improvements in training day 2 compared to training day 1: The percentage of trials with an angle error < 15° ($t_{33} = 2.37$, *P < 0.05) and with only one transition increased ($t_{33} = 2.55$, *P < 0.05). In the "recall" task, participants made significantly more correct responses in day 2 compared to day 1 ($t_{41} = 3.89$, ***P < 0.001). **(H)** Example data from the most commonly used schedule: even distribution of trajectory angles across all trials (light gray), outcome trials (medium gray) and nonoutcome trials (dark gray).

Subjects then participated in 1-4 separate fMRI sessions, spanning two separate days, at least 1 week apart. In each trial during scanning, subjects watched a video of a bird

morphing according to a predefined neck:legs ratio (Fig. 7.1C and supplementary video found in (Constantinescu et al., 2016d)). They were then instructed to imagine the outcome if the bird continued to morph with the same neck:legs ratio. In some trials, they had to choose one of three offered outcomes: two outcomes they were trained with and a "no outcome" option (black square). Participants reached a performance of 72.8 ± 1.0% accuracy in predicting outcomes, and 0 out of 28 reported conceiving of the relationships between birds or outcomes as lying in a spatial map (as detailed in the previous chapter). We created two versions of the task in the scanner and half of the participants performed one version and the other half performed the other version.

Twenty-two subjects participated in both days, whereas the remaining 6 subjects participated only in the first day and did not return for the second testing day. We split in half the data of the subjects who were scanned only once for subsequent analyses that required cross-validation between separate session. In total, we acquired 94 separate datasets. One participant was excluded from the consistency analysis in the prefrontal cortex as a result of parameter estimates being more than 4 s.d. away from the group mean. All participants gave written informed consent and were paid for participating, as approved by the ethics committee of the University of Oxford (ref. number 2013-171).

7.2.2 MRI data acquisition

We acquired T2-weighted functional images on a 3 Tesla scanner. We used a gradient-echo echo-planar imaging (EPI) pulse sequence that sets the slice angle of 30° relative to the anterior-posterior commissure line, minimizing the signal loss in the orbitofrontal cortex region (Weiskopf et al., 2006b). We acquired 45 slices, 3mm thick, in an interleaved order. We used the following parameters: repetition time (TR) = 3000 ms, echo time (TE)

= 30 ms, flip angle = 87°, field of view (FoV) = 192mm, voxel size = 3 x 3 x 3 mm^3. To correct for deformations in the inferior prefrontal cortex, we also acquired a field map with dual echo-time images covering the whole brain, with the following parameters: TR = 500 ms, TE1 = 5.19 ms, TE2 = 7.65 ms, flip angle = 60°, FoV = 224mm, voxel size = 3.5 x 3.5 x 3 mm^3. To facilitate an accurate registration of the EPIs to the standard space, we also acquired a T1-weighted structural image using a magnetization-prepared rapid gradient echo sequence (MPRAGE) with the following parameters: TR = 2040 ms, TE = 4.7 ms, flip angle = 8°, FoV = 192mm, voxel size = 1 x 1 x 1 mm^3. Stimulus presentation and subject button presses were registered and time-locked to the fMRI data.

7.2.3 Pre-processing of functional images

We analyzed whole-brain MRI data with FMRIB's Software library (FSL) (Smith et al., 2004a). We segmented brain matter from nonbrain (Smith, 2002). We corrected functional data for motion artefacts and we removed low frequency signals using a high-pass filter at 1/100 Hz. We smoothed the data using a gaussian filter of 7 mm full width at half maximum and we corrected for slice time acquisition differences. We also corrected for geometric distortions in the EPI images using the acquired fieldmaps (Jenkinson et al., 2002). We registered the EPI images to the high-resolution anatomical image using boundary-based reconstruction and then we normalized them into standard space (Montreal Neurological Institute – MNI152) using non-linear registration. Because of the notable breathing- and susceptibility-related artefacts in the enthorinal cortex, we ran a separate analysis where we pre-cleaned the data with FMRIB's ICA tool, FIX (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014). This approach is more effective than separately tracking breathing and pulse, and using them as confound regressors in the GLM (Klein-Flugge et al, 2011).

7.2.4 Motion estimation

We computed four parameters to assess motion in each participant and identify movements bigger than the voxel size (3mm). First, we computed the mean displacement, defined as the amount of head motion (sum of the absolute translational movement in the x, y and z directions) relative to the middle time point across the entire scanning session, averaged across all time points.

Second, we examined the total displacement (TD) from the beginning to the end of the scanning session.

Third, we examined the data for rapid movements because these are the ones that could have disrupted signal intensities. We computed the mean framewise displacement, as the amount of head motion relative to the previous time point, averaged across all time points. Finally, we computed the maximum framewise displacement (MaxFD) to address the concern of sudden movement more conservatively. We looked for subjects who moved more than 3mm at any time point.

7.2.5 fMRI whole brain analyses

After pre-processing, we modeled the fMRI time series using two general linear models (GLMs) in FSL. Both models included regressors for the main effects of the morphing stage and the response stage of the trial and six nuisance regressors to account for motion-related artifacts. Each individual model had various parametric modulations of the regressors for the morphing stage (see below). All these regressors were convolved with the FSL default hemodynamic response function and filtered by the same high pass filter as the fMRI data before entering the GLM.

GLM 1: hexagonal modulation analysis (Fig. 7.4):

We created two parametrically modulated regressors (Doeller et al., 2010): the sine and cosine of the direction of each trajectory in each trial, θ(t), with a periodicity of 60°, that is, sin(6θ(t)) and cos(6θ(t)). The factor 6 means that these regressors will produce coefficients with high amplitude $\left(\sqrt{\beta_{\sin}{}^2 + \beta_{\cos}{}^2}\right)$ for brain regions that are sensitive to hexagonal symmetry. To calculate the magnitude of the hexagonal modulation, we used an F test to investigate which brain regions were significantly modulated by a linear combination of these two regressors: β_{sin}*sin(6θ) + β_{cos}*cos(6θ). We then transformed this statistic to a Z-statistic in each subject, using an asymptotic approximation (http://www.fmrib.ox.ac.uk/analysis/techrep/tr00mj1/tr00mj1/) and, together with a behavioral covariate, we performed a 1-sample t-test across the group. For visualization purposes, the statistic images were thresholded using clusters determined by Z-statistic = 3.1 and the supra-threshold clusters were corrected for family-wise error using a cluster significance threshold of p = 0.05. For the entorhinal cortex, we used a more lenient cluster forming threshold at Z = 2.3 and p = 0.05 because this brain region is prone to signal loss and, thus, it is difficult to image with fMRI.

This test has the major benefit of identifying periodic signals across the whole brain without knowing the phase of these signals. However, although the reported clusters survive multiple comparison correction across the whole brain, this test is unusual and should not be used for statistical inference because it can overestimate the Z-scores. Thus, we used it instead as a method to identify informative brain regions that we could then use to create ROIs for subsequent unbiased tests.

In detail, the first level statistic that is being brought to the group level relies on the mean and variance of the first-level estimates in such a way that misestimating the variance can

cause a bias in the mean (for example if the variance is under-estimated, the F-statistic will be overestimated, and the expectation of the transformed Z-statistic will be greater than zero even in null data). Estimating unbiased variances from autocorrelated fMRI data is notoriously difficult. Therefore, the reader should be cautious in interpreting the results of this test as statistically significant. In the current manuscript, we use this test as a method for generating candidate regions of interest for tests that are certainly unbiased (consistency tests). This allows us to extract the grid angle from these ROIs in one dataset and test it in another. These subsequent tests do not use first level variances and are therefore unbiased tests performed with the standard fMRI GLM machinery.

GLM 2: hexagonal consistency analysis (Fig. 7.6A):

We created a parametric regressor cos(6[**θ**(t) – **φ**]), where **θ**(t) is the trajectory angle for each trial t within one session and **φ** is the grid angle from another session acquired on the same day (Doeller et al., 2010; Kunz et al., 2015) (Fig. 7.2). This regressor had higher values for grid-like activity with the same grid angle as in the other session. Thus, if the trajectories from one session aligned with the predicted grid angle taken from a different session, then this regressor would have high values. This resulted in subject-specific parameter estimates for each experimental session.

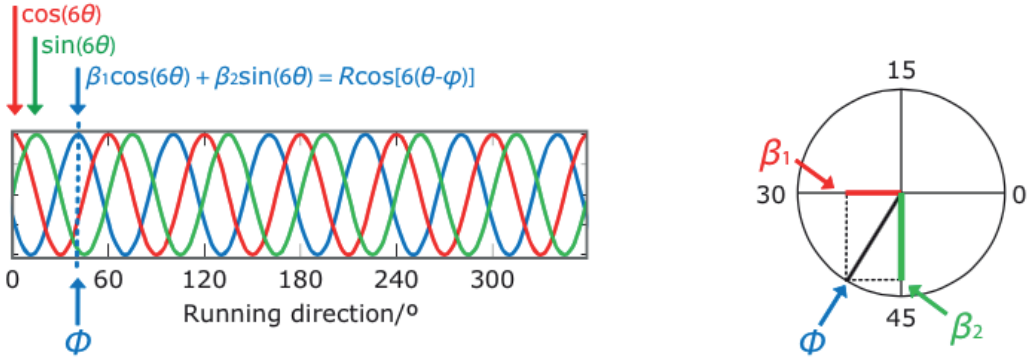


Fig. 7.2. Estimation of the grid angle. Reproduced with permission from Nature, Doeller et al, 2010.

In the next stage, participant-specific linear contrasts of these parameter estimates were averaged across the experimental sessions and, together with a behavioral covariate represented by the behavioral accuracy, they were entered into a series of one-sample t-tests, each constituting a group-level statistical parametric map. The statistic images were thresholded using clusters determined by Z-statistic = 2.3 and the supra-threshold clusters were corrected for family-wise error using a cluster significance threshold of $p = 0.05$.

To estimate the grid angle $\boldsymbol{\varphi}$, we used a similar approach to the one for navigating in physical space (Doeller et al., 2010). We first averaged the coefficients for the sine (β_{sin}) and cosine (β_{cos}) regressors across all voxels in the ROI. Then, we calculated the grid angle $\boldsymbol{\varphi}$ (varying between 0 and 59°) as arctan $[(\beta_{sin}/\beta_{cos})]/6$ within regions of interest (ROIs), where arctan was mapped into the 360° space, varying between -30° to 30°, according to the signs of β_{sin} and β_{cos}.

7.2.6 fMRI region-of-interest analyses

We defined all ROIs from the whole-brain quadrature filter analysis for hexagonal modulation (GLM 1 and Fig. 7.4), which was orthogonal to the contrasts of interest. This allowed statistical tests to be performed in an unbiased fashion. To make it clear why this is the case: in the null case, two scans that both had high hexagonal magnitude would have randomly oriented hexagonal phase (grid angle) and so all analyses that relied on aligned grid angle would result in mean zero signal. The ROIs were 5-mm spheres centered on the voxels with the highest hexagonal magnitude activity (GLM 1) in the ventromedial prefrontal (vmPFC) and the entorhinal cortex (ERH) (Fig. 7.4A). We used these ROIs as

seeds for estimating the values of the grid angle and to test for grid angle consistency between separate experimental sessions.

To examine the pattern underlying the hexagonal magnitude effect, we ran separate regression analyses for trajectories that were aligned and misaligned with the grid, using fMRI time series from the vmPFC and ERH ROIs. We aligned the trajectory angles to the grid angle φ and we split these aligned trajectories into 12 equal bins of 30°. We thus created 12 separate regressors for trajectories that belonged to one of these 12 bins, resampled them at a resolution of 100ms and convolved them with the hemodynamic response function. To match the resolution of the fMRI timeseries, we then resampled the regressors to the duration of the fMRI repetition time. Next, we ran 12 separate regression analyses in each individual experimental session. The 12 resulting regression coefficients were then averaged separately across various sessions, and expressed as mean ± sem (Fig. 7.6B-C and Fig. 7.8). Given our strong *a priori* hypotheses of hexagonal symmetry, we performed a t-test on the differences between the resulting betas aligned and misaligned to the grid angle φ, across subjects, to investigate if these differences were significantly different from zero.

We tested for consistency within-day (Fig. 7.6), across-day (Fig. 7.8A), and all data combined within- and across-day (Fig. 7.8B-C). Again, to allow statistical unbiased testing, we aligned data from one session to the grid angle from the other session.

7.2.7 Control analyses for fMRI

To test for the specificity of the six-fold symmetric sinusoidal modulation, we used the same approaches as above, but using control models with directional periodicities of 90°, 72°, 51.4° and 45° (that is, four-fold, five-fold, seven-fold and eight-fold rotational symmetry).

7.2.8 Controls for the properties of the visual scene

To test whether the fMRI signal had hexagonal symmetry, as a proxy for grid cells, we ensured that the orientation of the trajectories for movement in bird space was dissociated from the properties of the visual scene. That is, trajectories with the same orientation were formed by different stimuli, and trajectories with different orientations could pass through the same stimulus. Indeed, the critical hexagonal symmetry regressors described above, sin(6*θ) and cos(6*θ), did not share more than 5% variance with any tested basic visual property of the stimuli, outcomes, or behavioral accuracy in any subject (Fig. 7.3). The trajectories were sampled evenly across directions both in sum and when separated according to outcome (Fig. 7.1H).

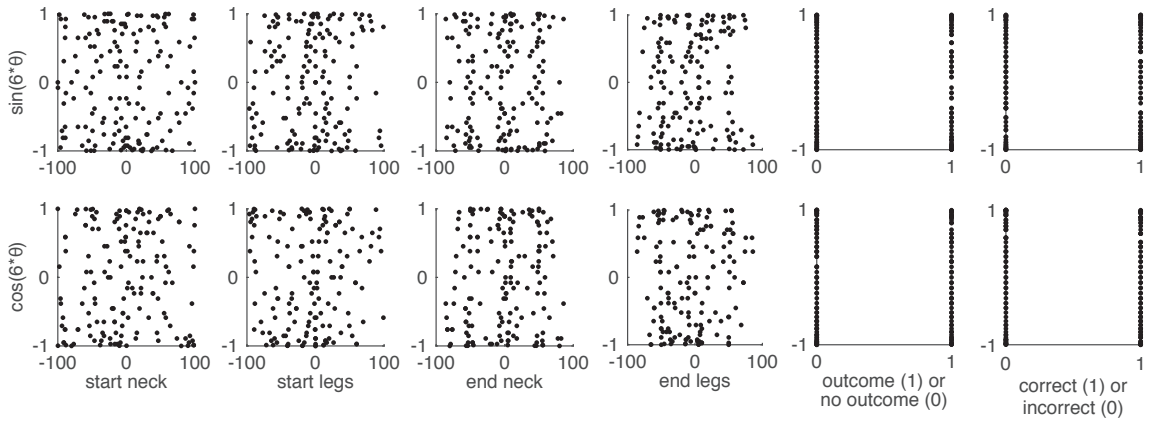


Fig. 7.3: Controls for the properties of the visual scene. Example data from the most commonly used schedule. We tested if the sin(6θ) and cos(6θ) regressors correlated with multiple confounding factors. These regressors did not correlate with the start neck, start legs, end neck, and end legs lengths, whether the subject responded accurately or whether the morph passed through an outcome (all coefficients of determination R^2 averaged across all subjects < 0.02).

7.3 Results

7.3.1 Hexagonal modulation

Despite the absence of any hexagonal modulation of these confounding factors, we found a hexagonal modulation effect in the neural activity. We first identified hexagonally

symmetric signals across the whole brain and then we focused on those regions where the effect was strongest. This approach allowed us to test in an unbiased way if this hexagonal symmetry had a consistent grid angle across two sessions acquired on the same day and more than a week apart.

To identify brain regions sensitive to hexagonal symmetry, we used a Z-transformed F-statistic to test for a significant modulation of the fMRI signal by any linear combination of $\sin(6\theta)$ and $\cos(6\theta)$, where θ is the trajectory angle in bird space (Fig. 7.2). We found hexagonal symmetry in a network of brain regions that overlapped anatomically with the network found during navigation in physical space (Doeller et al., 2010; Jacobs et al., 2013) and with the default-mode network (Fox and Raichle, 2007) (Fig. 7.4). However, although these brain regions all survived whole-brain cluster correction, this quadrature test could overestimate the Z scores. This is because the Z scores are computed by an F-to-Z transform which will be biased by any mis-estimation of the data autocorrelation (see section 7.2.5), and precise estimation of data autocorrelation is notoriously difficult in fMRI studies which have limited temporal samples (Woolrich et al, 2001). Thus, we used this test not for statistical inference *per se*, but rather to create orthogonal regions of interest that allowed us to test in an unbiased way if the grid angle was consistent across separate experimental sessions. This was possible because the quadrature test was independent from the phase of the periodic signal, that is, the grid angle. We focused on brain regions where grid cells have been recorded in humans during spatial tasks (Jacobs et al., 2013): that is, the anterior cingulate/medial pre-frontal cortex (mPFC) and entorhinal cortex (ERH).

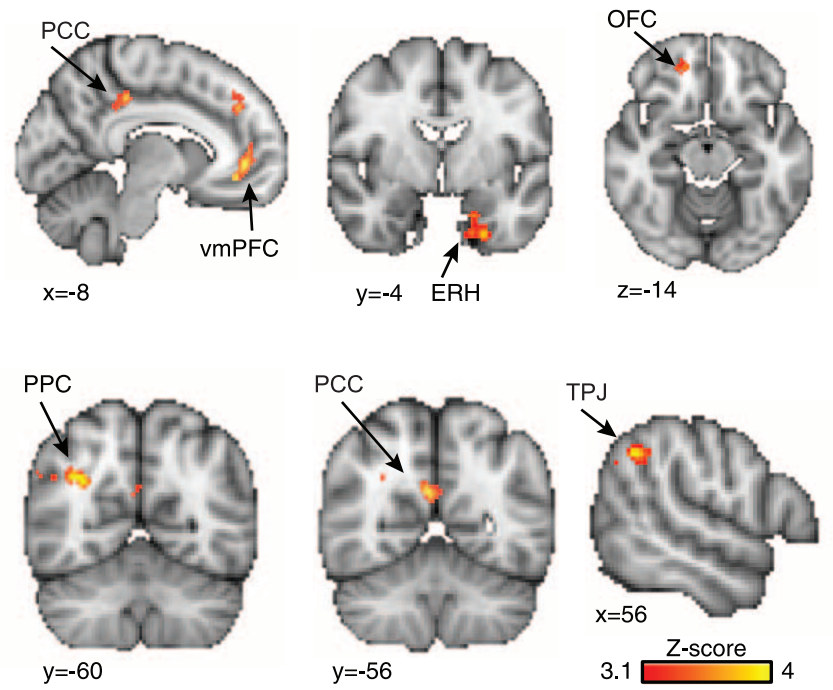


Fig. 7.4: Identifying hexagonally symmetric signals across the whole brain. Hexagonal modulation in a network of brain regions including the medial prefrontal cortex, with a peak in its ventral region (vmPFC; peak Montreal Neurological Institute coordinates –8/42/0, peak Z score = 4.09), the medial entorhinal (ERH; –18/0/ –38; Z = 4.41), the orbitofrontal (OFC; 6/44/–10; Z = 4.27), the posterior cingulate (PCC; 0/–32/28; Z = 4.3), retrosplenial (RSC; 6/–52/24; Z = 4.73), and lateral parietal cortices (LPC; 30/ –62/28; Z = 4.96) and the temporoparietal junction (TPJ; 52/–42/40; Z = 4.13). For visualization purposes, the maps are cluster corrected at a cluster threshold Z = 3.1 and P < 0.05 for all brain regions apart from the ERH, where we used a more lenient threshold of Z = 2.3 and P < 0.05.

Using the peak coordinate of the hexagonal modulation signal in the mPFC located ventrally (vmPFC), we found that subjects with greater hexagonal modulation had a more accurate performance at the task (Fig. 7.5). This region has also been shown to correlate with the performance in memory and conceptual knowledge tests (Kumaran et al., 2009; Benoit et al., 2014).

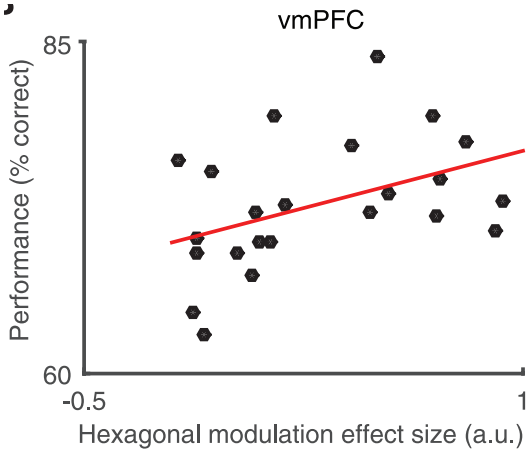


Fig. 7.5: Correlation between hexagonal modulation and behavioural performance. Subjects who performed better at the task had significantly more hexagonal signal modulation in the vmPFC (correlation coefficient r = 0.432, P = 0.039).

7.3.2 Hexagonal consistency within-day

As previously described for spatial hexagonal symmetries (Doeller et al., 2010; Kunz et al., 2015), we next asked whether the grid angle to which this hexagonal modulation was aligned was consistent between separate experimental sessions. We thus used the data from one session to estimate the grid angle for a given participant, using the beta coefficients for the sin(6θ) and cos(6θ) regressors (Fig. 7.2). We then took the data from a separate session and looked for differences in activation between trials in which the trajectories were aligned versus misaligned to this hexagonal grid. This was achieved with the regressor cos(6[θ(t) – φ]), where θ(t) is the trajectory orientation in trial t and φ is the mean grid orientation across the region of interest (Doeller et al., 2010; Kunz et al., 2015). This "cross-validation" procedure was counterbalanced across sessions. We then performed a one-sample t test across the group on the resulting regression coefficients.

Using this approach, we tested for hexagonal consistency between separate sessions acquired half an hour apart. We found such an effect at the whole-brain level in the vmPFC (Fig. 7.6A, left), which was not a function of the presence of the outcomes. Again, the consistency effect correlated with behavior in subjects who performed better at the task

(Fig. 7.6A, right). Thus, we replicated the finding that participants with more gridlike representations performed better at the task, using two independent analyses. The same neural signal was not predicted by the speed of learning of the task during training (Fig. 7.1, F and G). Gridlike activity was therefore related to the current performance rather than the trajectory of learning.

Next, to examine the pattern underlying this hexagonal effect, we separated all aligned (red) and misaligned (gray) trajectories (Fig. 7.1D). The signal in the vmPFC was significantly higher for aligned than misaligned trajectories (Fig. 7.6B). The same pattern appeared in ERH (Fig. 7.6C). This effect was significant only for sixfold but not control four-, five-, seven- and eight-fold symmetries (Fig. 7.6, B and C, right panels).

7.3.3 Hexagonal consistency between-days

We also tested for hexagonal consistency between separate sessions acquired more than a week apart. We found a significant effect in the vmPFC (Fig. 7.8A). When grouping together within- and between-day data, the consistency effect was strongest in the vmPFC (Fig. 7.8B). Moreover, we found hexagonal consistency also between the ERH and vmPFC, suggesting that different brain regions may contain gridlike activity that is aligned to the same angle (Fig. 7.8C). Again, all these effects were significant only in hexagonal but not control symmetries (Fig. 7.8, right panels).

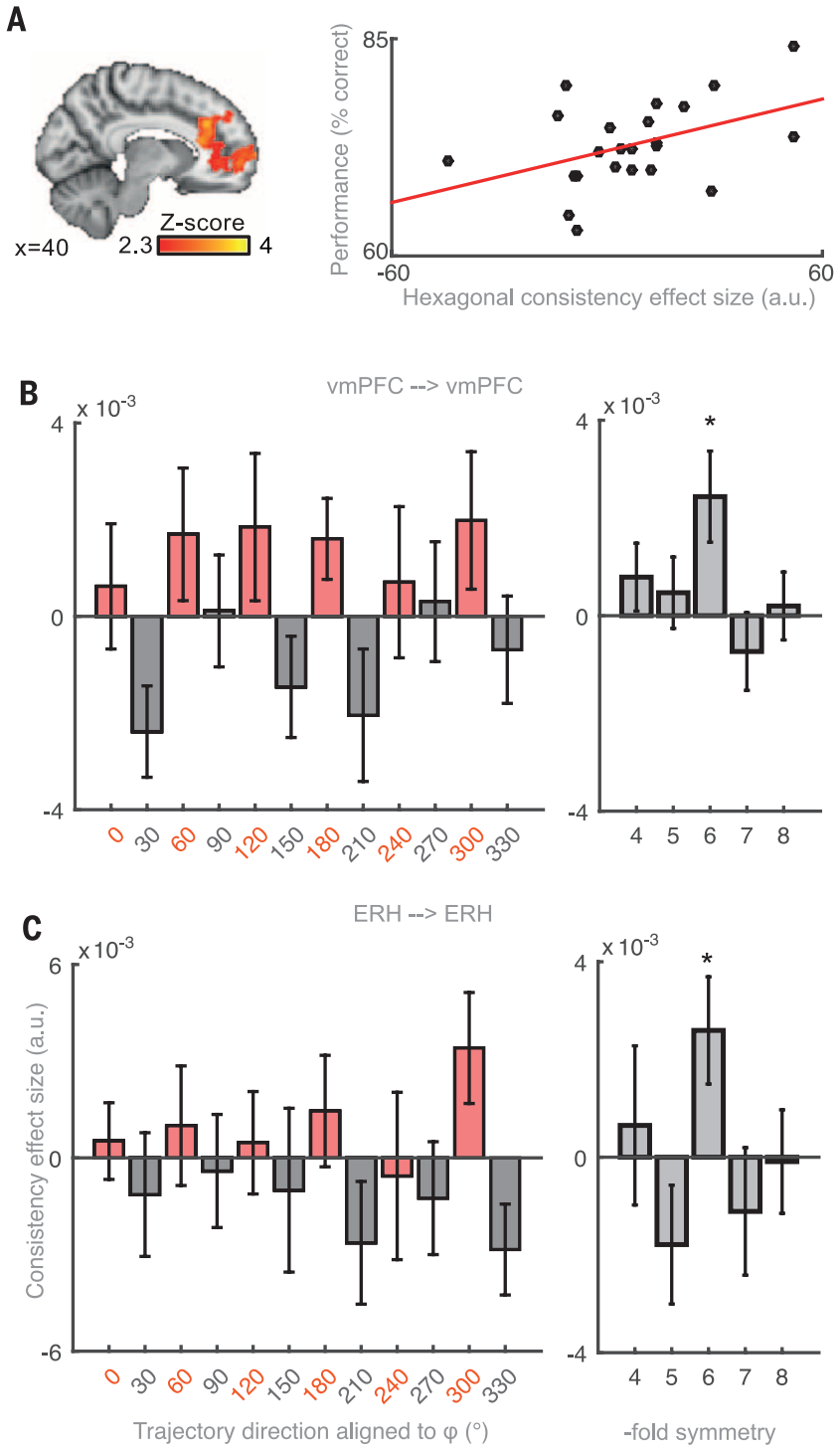


Fig. 7.6: Grid angle consistency between separate sessions acquired within the same day. (A) Left: Whole-brain level grid angle consistency in vmPFC (cluster corrected Z = 2.3 and P < 0.05; 16/54/–2; Z = 3.76, P < 0.0001). Right: participants with higher hexagonal consistency per- formed more accurately on the task (r = 0.431, P = 0.039). **(B and C)** Left panels: Sixfold modulation signals aligned to the same grid angle in the vmPFC (t_{26} = 2.61, *P < 0.05) and ERH (t_{27} = 2.36, *P < 0.05). The effect is plotted separately for all aligned (red) and misaligned (gray) trajectories. Right panels: This effect was specific for sixfold, but not any other control periodicities between four- and eightfold (all P > 0.15).

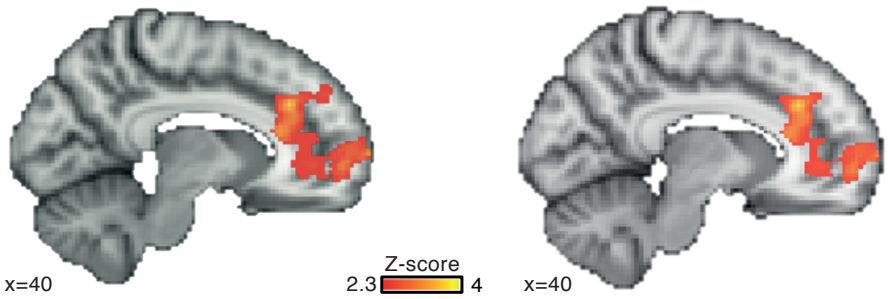


Fig. 7.7: The vmPFC consistency map does not depend on the outcomes. The vmPFC consistency map (left and also shown originally in Fig. 3A) did not change after including a confound regressor that modeled out the effect of outcomes (right).

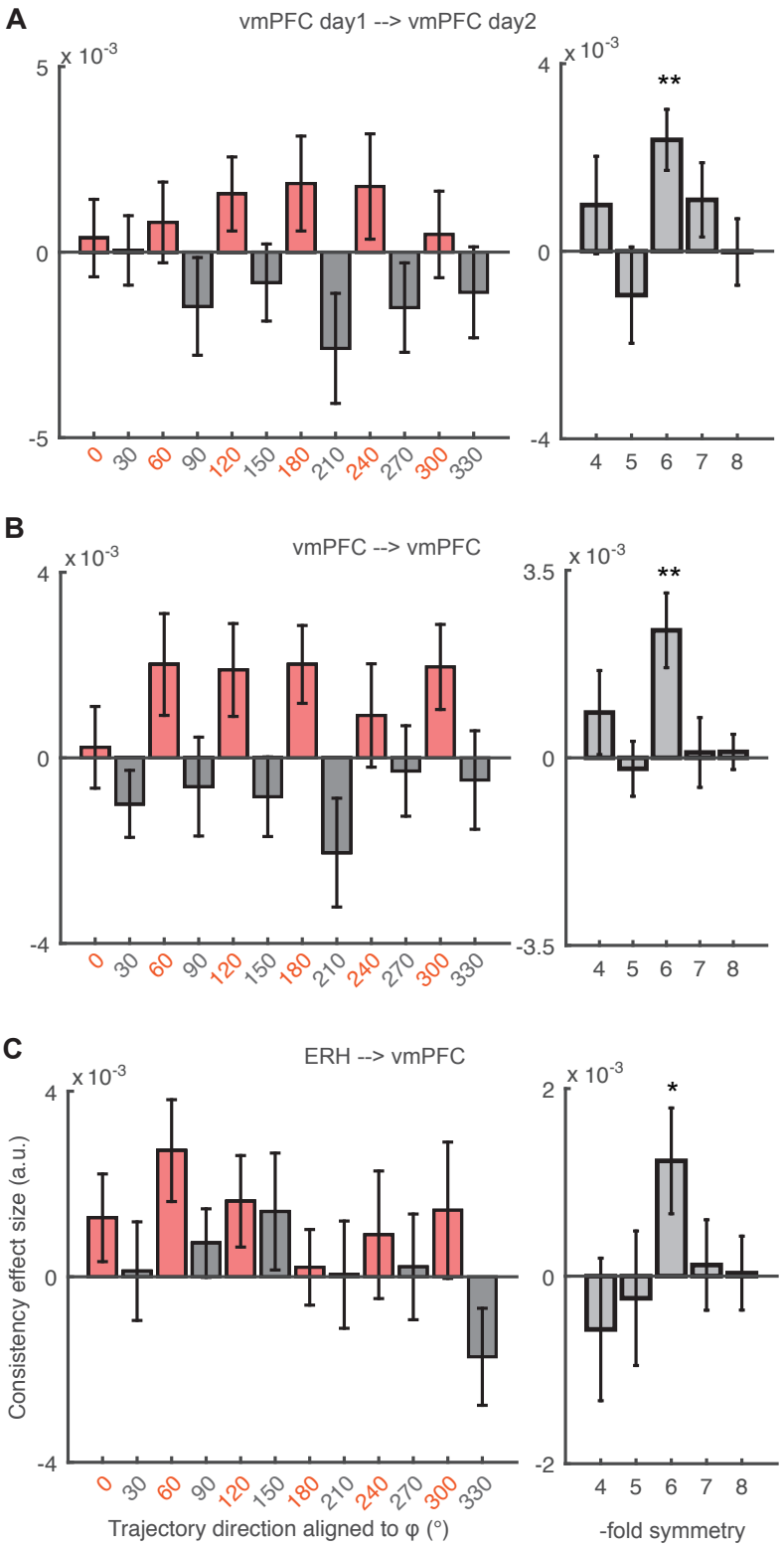


Fig. 7.8: Grid angle consistency between separate sessions acquired more than a week apart. (A) Cross-day consistency of the grid angle in vmPFC (left panel $t_{20} = 3.65$, **P < 0.01; right panel all P > 0.18 for control periodicities). **(B)** Within- and cross-day consistency in vmPFC (left panel $t_{20} = 3.41$, **P < 0.01; right panel all control P > 0.15). **(C)** Cross-region consistency between the ERH and vmPFC (left panel $t_{21} = 2.18$, *P < 0.05; right panel all control P > 0.46).

7.3.4 Replication of the results after a re-analysis

To rule-out possible confounding effects of motion on the hexagonal symmetry signals, we re-analyzed the data after excluding subjects that moved more than 1 voxel-diameter. We estimated motion using a range of measures: the average mean displacement, the total displacement, the average mean framewise displacement and the maximum framewise displacement.

The average mean displacement across participants was 0.4 ± 0.043 mm, thus less than 1 voxel-diameter (3 mm).

We found 7 subjects who had a total displacement > 3mm. This is a very conservative measure of subject movement as it penalizes many incremental movements that are well corrected by linear methods. Nevertheless, we tested using region of interest analyses whether our results survived the exclusion of these 7 subjects. Despite the inevitable loss of statistical power entailed by removal of data from more than 25% of the participants, we were able to confirm the majority of our findings. All results presented except for two (the correlation between behavior and hexagonal consistency in vmPFC and the hexagonal consistency vmPFC within-day) remained significant in this re-analysis (Table 7.1). We next focused on measures of rapid motion of the kind that cause a bigger problem for fMRI.

The average mean framewise displacement across participants was 0.031 ± 0.0045 mm, thus less than 3mm. Indeed, participants did not tend to make sudden head movements throughout scanning.

We found that 3 subjects made 1-2 rapid head movements in one scanning session, as assessed by the maximum framewise displacement (MaxFD). It is very unlikely that these events could have potentially affected the observed hexagonally symmetric signal because they did not occur frequently, with hexagonal periodicity, and they would have needed to

be precisely aligned with the grid angle. Despite this improbable situation, we tested again using region of interest analyses whether our results survived the removal of these 3 subjects. All the presented results remained significant in this re-analysis, except for one, which approached significance (the correlation between behavior and hexagonal signal modulation in vmPFC).

To sum up, we re-analyzed data after excluding subjects who moved more than 3mm in the scanner. Although the occasional results changed when we remove these subjects, in each case the vast majority of assessments survived. The statistical tests from these two re-analyses, one that is common within FSL and SPM analysis standards (MaxFD) and another more conservative measure (TD) are presented in Table 7.1.

Test	Presented in	Value including subjects after motion correction	Value excluding subjects with MaxFD > 3mm	Value excluding subjects with TD > 3mm
Correlation between behavior and hexagonal signal modulation in vmPFC	Fig. 7.5B	r=0.432, p=0.039	r=0.41, p=0.07*	r=0.4, p=0.09*
Correlation between behavior and hexagonal consistency in vmPFC	Fig. 7.6A, right	r=0.431, p=0.039	r=0.49, p=0.02	r=0.32, p=0.19**
Hexagonal consistency vmPFC, within-day	Fig. 7.6B, left	t(26)=2.61; p<0.05	t(23)=2.17; p<0.05	t(20)=1.54, p=0.13**
Control periodicities, vmPFC, within-day	Fig. 7.6B, right	All p>0.26	All p>0.3	All p>0.29
Hexagonal consistency ERH, within-day	Fig. 7.6C, left	t(27)=2.36; p<0.05	t(24)=2.3, p<0.05	t(21)=2.65; p<0.01
Control periodicities, ERH, within-day	Fig. 7.6C, right	All p>0.15	All p>0.18	All p>0.31
Hexagonal consistency vmPFC, cross-day	Fig. 7.8A, left	t(20)=3.65; p<0.01	t(17)=2.97; p<0.01	t(15)=2.77, p<0.05
Control periodicities, vmPFC, cross-day	Fig. 7.8A, right	All p>0.18	All p>0.3	All p>0.12
Hexagonal consistency vmPFC, within- and cross-day	Fig. 7.8B, left	t(20)=3.41, p<0.01	t(17)=2.66; p<0.05	t(15)=2.35, p<0.05
Control periodicities, vmPFC, within- and cross-day	Fig. 7.8B, right	All p>0.29	All p>0.6	All p>0.54
Hexagonal consistency, cross region ERH --> vmPFC	Fig. 7.8C, left	t(21)=2.04, p=0.053	t(18)=2.21; p<0.05	t(16)=2.32, p<0.05
Control periodicities, cross region ERH --> vmPFC	Fig. 7.8C, right	All p>0.8	All p>0.55	All p>0.25

Table 7.1: Ruling-out confounding effects of motion. Statistical tests from the data re-analyses after excluding participants who moved more than 3mm. *approaching significance, **no longer significant.

7.4 Discussion

Although the coarse nature of the fMRI signal urges caution in making conclusions at the level of neuronal codes, we have reported an unusually precise hexagonal modulation of the fMRI signal during non-spatial cognition. This hexagonally symmetric signal is a precise an unusual prediction for fMRI.

Its temporal waveform is different in every scan and every subject (Fig. 7.) and thus cannot be easily predicted by imaging artefacts. The trial-by-trial fMRI activity resembles a periodic wave with six peaks only if the angles of the trajectories the participants are running at are sorted in ascending order (modulo 60) from 0° to 360° (Fig. 7.A). However, participants did not move along trajectories in a sorted order, but in a random fashion and the signal looks random when the order of the trajectory angles is shuffled (Fig. 7.C). Moreover, the trial-to-trial fMRI activity is different from day-to-day in each participant because they navigated along various random angles in each day. These responses are also different from person-to-person because each subject has their own grid angle (Fig. 7.9).

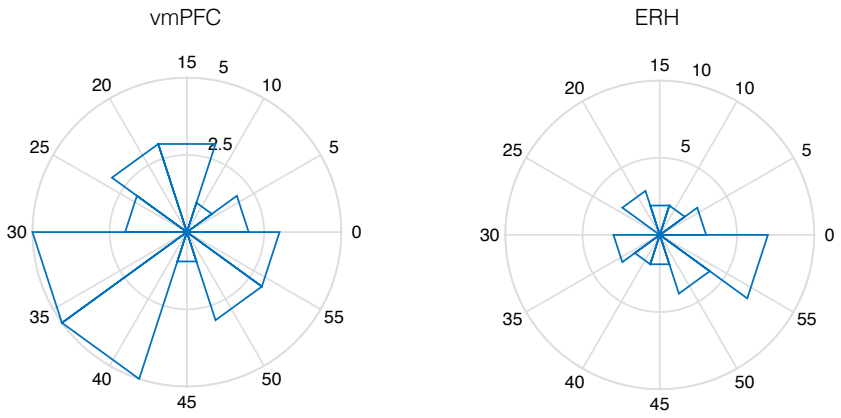


Fig. 7.9: Distribution of grid angles across subjects. Note that the distributions look different between the vmPFC and ERH ROIs however, within subjects, the grid angles between these ROIs are closer than expected by chance (see **Fig. 7.8**).

For example, if two participants have grid angles at 30° from each other, then trajectories that are aligned in one person, thus leading to fMRI responses with big amplitudes, will be misaligned in the other person, leading to fMRI responses with low amplitudes.

Therefore, potential confounds of this signal need not resemble a periodic wave with six peaks regardless of the order of the trials. Instead, the artifacts need to be consistently and significantly happening with a hexagonal periodicity that depends precisely on the order of the trajectories of moving in space, be aligned to one grid angle in one participant and to another grid angle in another participant. It is therefore unlikely that motion or physiological artifacts (such as breathing and heart rate) would cause the hexagonal pattern. For example, subjects would need to move more during aligned versus misaligned trajectories, according to the random order these trajectories are presented, which is different from day-to-day, and according to the orientation of the grid, which is different from person-to-person. The hexagonal pattern of fMRI activity is therefore difficult to explain with unspecific fMRI artefacts.

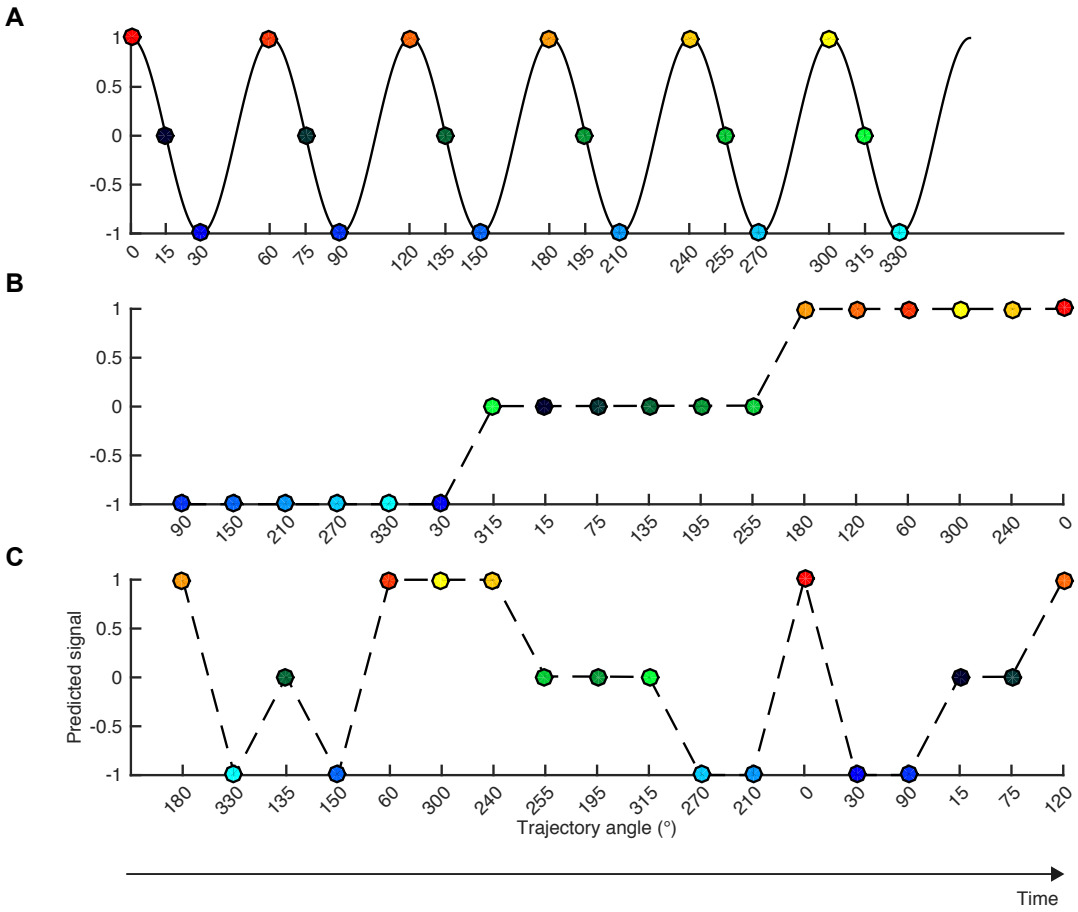


Fig. 7.10. Understanding the fMRI hexagonal symmetry signal. (A) The hexagonal symmetry pattern in the fMRI response becomes evident when the trajectory angles are sorted in an ascending order. **(B)** When the angles are ordered first as completely misaligned with the grid angle φ, then right in the middle between aligned and misaligned, and then completely aligned with the grid angle, the fMRI response will show a step-wise pattern, from a small response, to no change, and to a big response. **(C)** When shuffled randomly, the fMRI response also becomes random.

Moreover, when subjects perform our abstract cognitive task, this hexagonal signal exists in a set of brain regions markedly similar to those observed when subjects run in a virtual reality spatial environment (Fig. 7.), despite profound differences in the cognitive and perceptual demands of the two tasks.

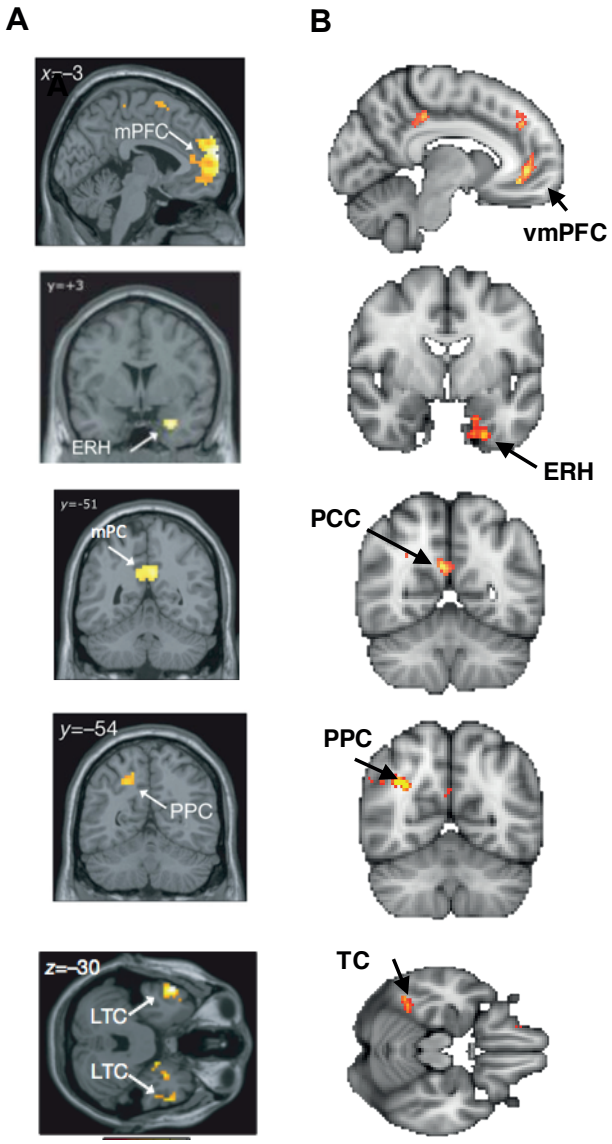


Fig. 7.11: The brain network for navigation in physical space (A) and conceptual space (B). Abbreviations: ventromedial prefrontal cortex (vmPFC), entorhinal cortex (ERH), posterior cingulate cortex (PCC), posterior parietal cortex (PCC), temporal cortex (TC). A is adapted with permission from Nature, Doeller et al, 2010.

The hexagonal grid is consistently oriented across sessions that are acquired both half an hour and more than a week apart. Together, this evidence suggests that gridlike codes that are known to underlie spatial navigation, and recently discovered in the temporal dimension (Kraus et al., 2015) and sound frequency dimension (Aronov et al., 2017), can

also be used to organize abstract knowledge of the type that is difficult to study in nonhuman species.

In the event that such conceptual grid cells can be recorded directly, it will be of interest to know whether they share relationships (such as relative phases) across spatial and conceptual tasks, suggesting that conceptual tasks can be solved by subconsciously mapping abstract dimensions onto preexisting spatial maps, or whether new organizations can emerge to represent conceptual problems. It will also be informative to study how such cells behave in conceptual problems that are not easily mapped onto continuous two-dimensional spaces (Stachenfeld et al., 2014; Dordek et al., 2016).

Chapter 8: A new experimental model for detecting grid-like coding as a function of location in space

We aimed to develop a new multivariate analysis method for detecting grid-like signals with fMRI that is more akin to those derived from electrophysiology recordings of grid cells, where the hexagonal firing pattern is estimated as a function of the subject's location in space. To achieve this goal, we needed two key ingredients: (i) a rich stimulus set that samples at high-resolution a well-learnt spatial environment (Dinoworld), and (ii) sub-millimeter fMRI that could capture fine biases in the preferred grid phase across the population of grid cells between different voxels. This chapter presents the behavioural part of this experiment, as well as a proof-of-concept showing that our novel analysis works on simulated data. The fMRI part of this experiment is presented in the next chapter.

8.1 Introduction

Grid cells are neurons important for the organization of knowledge. They show gradients in their anatomical connectivity and function along the long axis of the entorhinal cortex, which is oriented from dorsal to ventral in rodents and from posterior to anterior in humans (Barry et al., 2007; Brun et al., 2008b; Stensola et al., 2012; Navarro Schröder et al., 2015; Maass et al., 2015a). The scale of grid cells increases along the entorhinal axis and is thought to underlie a hierarchical organization of information (Strange et al., 2014). These gradients are paralleled by similar ones in the hippocampus and place cells, found in rodents (Kjelstrup et al., 2008; Strange et al., 2014; Collin et al., 2015; Brunec et al., 2017). Gradients in grid scale have not been investigated in humans, yet.

Measuring gradients in grid scale in humans is challenging because direct recordings in epilepsy patients along the entire axis of the entorhinal cortex are impractical, and current fMRI analysis methods can estimate the grid angle but not the grid scale (Doeller et al., 2010; Bellmund et al., 2016a). Nevertheless, fMRI has the advantage of recording the

entire brain simultaneously. Thus, with the right tools, it might be possible to measure gradients in grid scale in the human entorhinal cortex and how they map to the other parts of the brain.

Grid cells have several properties that could be leveraged to develop new experimental models to detect grid scales in human fMRI. First, their firing fields have hexagonal symmetry (Hafting et al., 2005). Second, grid cells are anatomically clustered into small modules that share the same grid angle and scale (Barry et al., 2007; Stensola et al., 2012). Third, grid scales cluster into even smaller modules that also share the same grid phase (Heys et al., 2014b; Gu et al., 2017). This means that small clusters of grid cells differ mostly in their grid phase. A sufficiently small fMRI voxel might not resolve individual clusters per se, however the distribution of clusters with different grid phases might not be perfectly uniform and most voxels might have a slightly higher number of clusters sensitive to a particular phase. Careful multivariate analyses might capture such heterogeneity, as it has been shown for orientation-sensitive cortical columns in the visual cortex (Haynes and Rees, 2006).

We hypothesized that humans have homologous gradients in grid scale along the posterior-anterior entorhinal axis, and they could be detected with careful computational analyses. To facilitate the detection of these gradients, we trained our participants extensively to learn a spatial environment at high resolution, such that the higher the resolution, the more likely it would be to detect grid cells with smaller grid scales. We then tested our novel analysis on simulated data. The next chapter will present the fMRI results of this experiment.

8.2 Methods

8.2.1 Participants and experimental design

Eight human participants reported to be in good health with no history of neurological or psychiatric disease and with normal or corrected-to-normal eye vision participated in this experiment at the Oxford Centre for Functional MRI of the Brain (FMRIB). Subjects were trained for hours outside the MRI scanner to play virtual reality (VR) games. Two participants only participated in some parts of the training because of motion sickness. All participants gave written informed consent and were paid for participating, as approved by the ethics committee of the University of Oxford (SOP number 004-V3).

We developed a battery of behavioural tasks for spatial memory and navigation in VR: the "Dinoworld Games". They were based on those routinely used for navigation in 3-dimensional environments, where participants are moving horizontally on a 2-dimensional virtual arena (Doeller et al., 2008b; Horner et al., 2016). Human participants were trained extensively to learn precisely the layout of a virtual environment and the location of ten objects hidden inside it, as well as how they related to each other. These tasks increased in difficulty over the course of training, and subjects could solve them by relying on different types of cues. This multimodal approach aimed to facilitate the formation of cognitive maps (Tolman, 1948b), which would allow the participants to locate themselves in space immediately, even while being immobile and teleported to different locations. Once the subjects could locate themselves in space quickly and precisely, they were ready to do the fMRI experiment (see next chapter). The reason we chose to use immobility and teleportation, rather than continuous navigation, was to minimize temporal autocorrelation between trials, which can be a confounding factor for multivariate analyses. Videos of the main training tasks can be found here: *goo.gl/vtzq7K*.

8.2.2 Virtual reality environment

We used Unity 5.5 software to present a first-person perspective view of a grassy plane surrounded by a circular wall, with a background of clouds and the sun, and peripheral landmarks such as a dinosaur, trees and buildings. The wall and the peripheral landmarks would change size as participants moved closer to or further away from them, therefore they could be used to estimate both allocentric and egocentric distances (Fig. 8.1). In contrast, the clouds and the sun were projected at infinity, thus they would not change in size depending on the participants' position, hence they only provided information about orientation but not location. Participants could navigate in this environment by pressing keyboard buttons to move forward and backward, as well as to turn left and right.



Fig. 8.1: VR environment. Example screenshots from the participants' point of view, showing how they can estimate egocentric distances between themselves and the landmarks. For example, they are closer to the dinosaur and further away from the well in the left screenshot, but closer to the well and further away from the dinosaur in the right screenshot. Participants can also estimate allocentric distances between landmarks. For example, the dinosaur is closer to the tree compared to the well.

8.2.3 Encode task

The aim of this task was to train participants to learn the locations of 10 objects hidden in the arena, by collecting them sequentially (Fig. 8.2A, B). The objects (a duck, an apple,

an ice-cream, a cupcake, a birthday cake, a basketball, a rugby ball, a bell, a bucket and a pair of sunglasses) were spread in random locations in the arena, that were kept consistent across the whole experiment in all participants. Subjects first familiarized themselves with the arena, and then they explored freely. One object at a time appeared in its correct location, and participants were instructed to collect it by moving over it, and to remember its location (Fig. 8.2C).

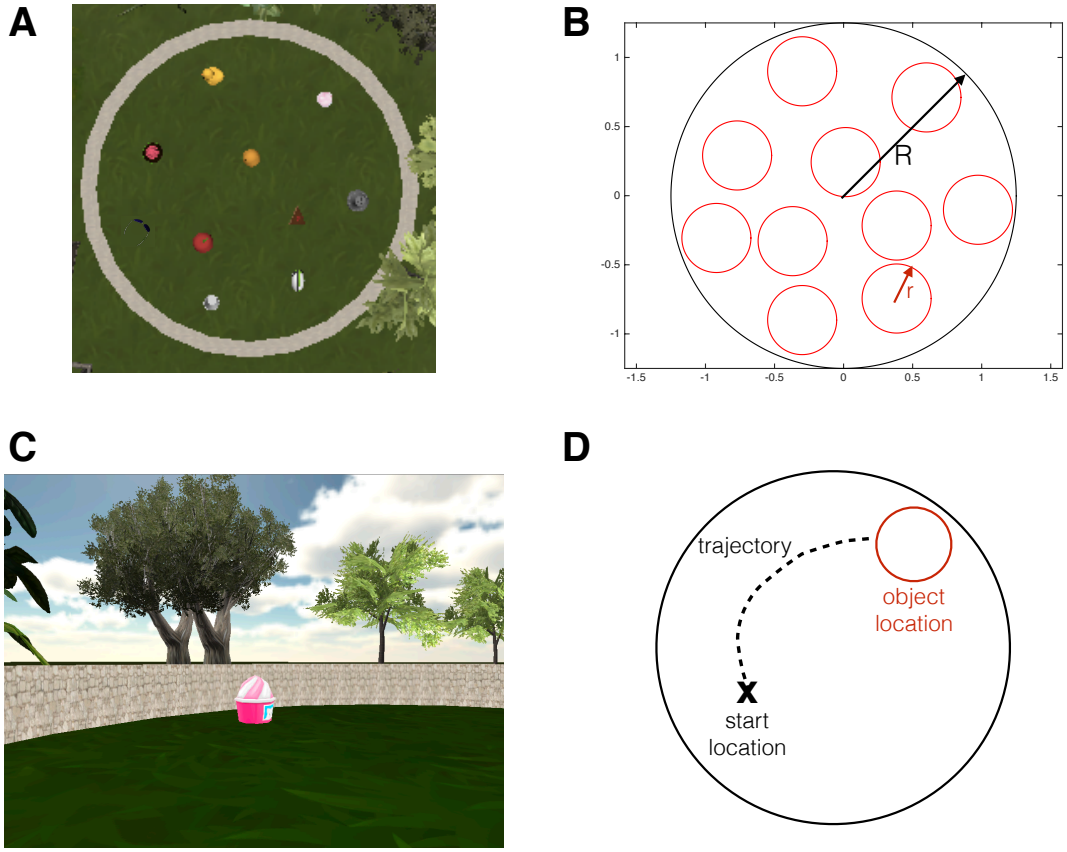


Fig. 8.2: Encode task. (A) Aerial view of the arena with all 10 objects made visible. **(B)** The radius (r) of the area covered by each object is 25% of the radius (R) of the whole arena. **(C)** Example trial where the participant needs to collect the ice-cream. **(D)** Example trajectory towards the ice-cream. Note that the trajectory is not straight.

8.2.4 Test and feedback task

Next, we trained and tested the spatial memory of the subjects by asking them to find specific objects from random start positions. We encouraged them to choose the direction

correctly and move towards the object in one straight trajectory. When subjects reached the desired end location they pressed a button and, as feedback, the object appeared in its correct location. A correct response meant that participants were inside the radius of the object's location, and this was signaled by the message "Brilliant". The subjects had to collect the object to move on to the next trial. We measured the error in their response by the distance between the end location and the centre of the object location (Fig. 8.3).

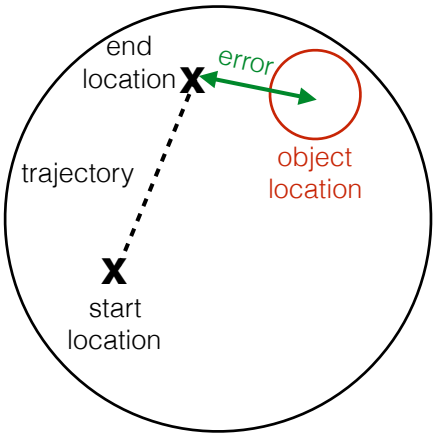


Fig. 8.3: Test and feedback task. Note that the trajectory is straight.

The reason we asked participants to move to the object in a single straight trajectory was to prevent them from using an image matching strategy. An example of this strategy would be to learn all 10 views from centre of the arena. Thus, a simple way to solve the task would be to first go from the random start location to the centre of the arena, then rotate to match the remembered view, and then navigate towards the object. However, this strategy is habitual and, unlike the cognitive map strategy, it ignores the rest of the environment.

We also tested some participants with two versions of this task. In one version, we polarized the environment by removing all landmarks apart from the dinosaur. Hence, when subjects were starting from a random location, they needed to first rotate and orient themselves relative to the dinosaur and clouds, and then rely on their memory of the

environment, and estimate distances using the walls and their own movement (path integration). In the other version, all landmarks were completely removed and subjects were only shown briefly where one object (the rugby ball) was located. Then, they could orient themselves using the clouds, and estimate distances using the wall and their own movement, thus relying more on path integration.

8.2.5 Teleport task

We aimed to further train participants' knowledge of the environment and object locations by teleporting them in random locations with random viewing directions. They needed to choose the closest object to them, which was given as a response option together with two other distractor options, and they received feedback at the end of each trial (Fig. 8.4A). Because we used a three-alternative forced choice task, chance performance was 33%. Correct performance on this task required participants to have very precise cognitive maps because they could not move or rotate, and often they needed to remember what landmarks lied behind them and estimate the distance to the wall behind them (Fig. 8.4B).

A version of this task without feedback was also used during scanning, because the analysis that we aimed to develop required the participants to be immobile not only physically in the scanner, but also as part of the task (see next chapter).

8.2.6 Draw map task

Finally, to assess if participants could recreate the environment from their memory, we asked them to position the individual landmarks around the arena and the objects inside the arena in their correct location.

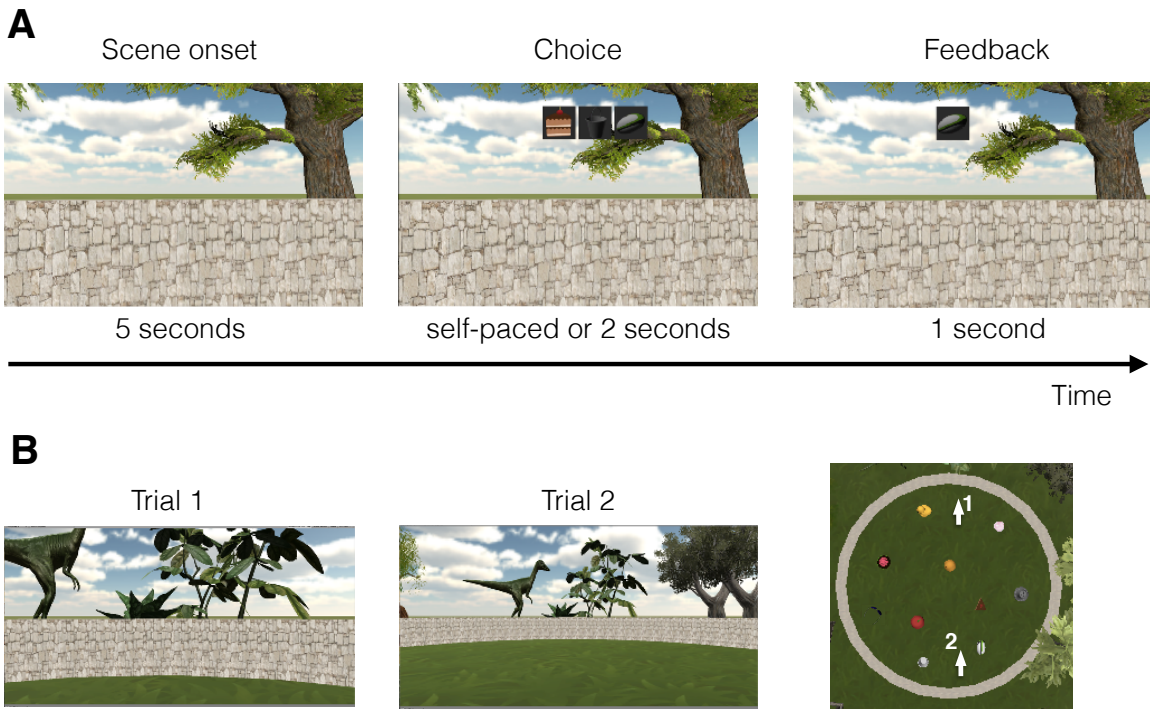


Fig. 8.4: Teleport task. **(A)** Example trial where the participants are shown a scene, then they need to make a choice which is followed by feedback. Here, the correct response is the rugby ball. **(B)** Two example trials shown from the participant's view and an aerial view with the corresponding locations and viewing directions marked by arrows. The viewing direction is the same in both trials, however the participant is moved further back in trial 2 compared to trial 1. Note that in trial 2, participants need to recall what the environment looks like behind them. Note also that the correct response in trial 2 is also the rugby ball like in (A), even if subjects are looking at the other side of the environment.

8.2.7 Simulations using ideal grid cells

We sampled 192 locations equally distributed in our Dinoworld environment (Fig. 8.5A). We then simulated the normalized firing rate (F) of ideal hexagonal grid cells (Fig. 8.5B) built from three identical planar waves propagating at 60° to each other (Derdikman and Knierim, 2014), in each of these locations:

$$F(x,y) = \sum_{l=1}^{3} \cos\left[\frac{2\pi}{R}(x - x_0)\sin\left(\frac{l\pi}{3} + \alpha\right) + \frac{2\pi}{R}(y - y_0)\cos\left(\frac{l\pi}{3} + \alpha\right)\right],$$

where x and y are the coordinates of each spatial location, R is the grid scale, x_0 and y_0 define the grid phase, and α is the grid angle in radians.

To simplify, we considered that each voxel was small enough to include a coherent population of grid cells with the same grid angle, grid scale and grid phase. We simulated 100 voxels. Separate voxels only differed in their grid phase (Fig. 8.5C), by analogy with the anatomical clusters of real grid cells (Heys et al., 2014a; Gu et al., 2017).

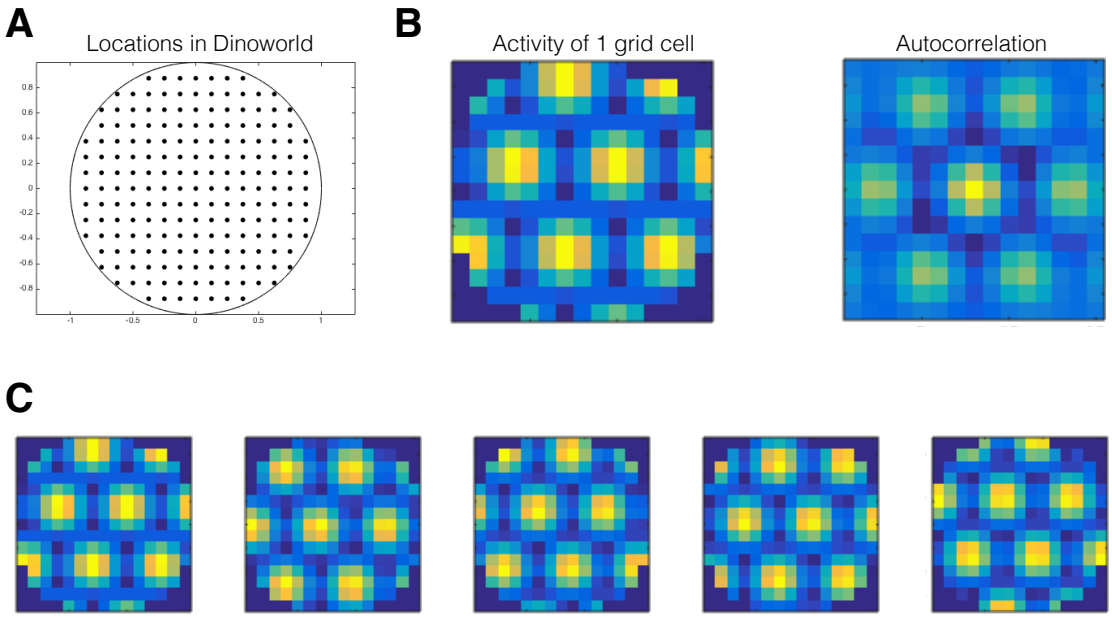


Fig. 8.5: Simulation of ideal grid cells. (A) Sampling locations in Dinoworld space. **(B)** Example activity of one grid cell and its autocorrelation showing hexagonal symmetry. Dark blue means no activity and yellow highest activity. **(C)** Other example grid cells with the same grid angle and grid scale like in (B), but with different phases. Note how the activity is shifted in space between different cells.

The aim of our novel analysis was to retrieve the hexagonal symmetry characteristic of grid cells by using a representational similarity analysis approach, such that patterns of brain activity across voxels are similar between locations that correspond to the firing field of a population of grid cells, and are dissimilar for the locations in between. Our novel analysis had 4 main steps:

(A) We correlated the activity patterns across voxels, between all 192 locations, and computed [1-correlation coefficient] to create a location by location representational *dissimilarity* matrix (RDM) (Fig. 8.6A). We used this measure because it is an intuitive

proxy of distance in neural space, hence of neural *similarity*. For example, if two patterns are highly correlated, their correlation coefficient is high. Hence, it makes sense for [1-correlation coefficient] to be small because these patterns are similar, hence their neural distance is small.

(B) We then used the similarity values for each row (seed location) and back-projected them in Dinoworld space, resulting in 192 maps (Fig. 8.6B_1). These maps are shifted relative to each other depending on which parts of Dinoworld space are correlated to the seed location (Fig. 8.6B_2).

(C) Because these maps were shifted relative to each other, we cropped and re-centred them (Fig. 8.6C).

(D) This allowed us to average them together. The resulting map was remarkably similar to the original autocorrelation map of the raw grid cell activity (Fig. 8.7).

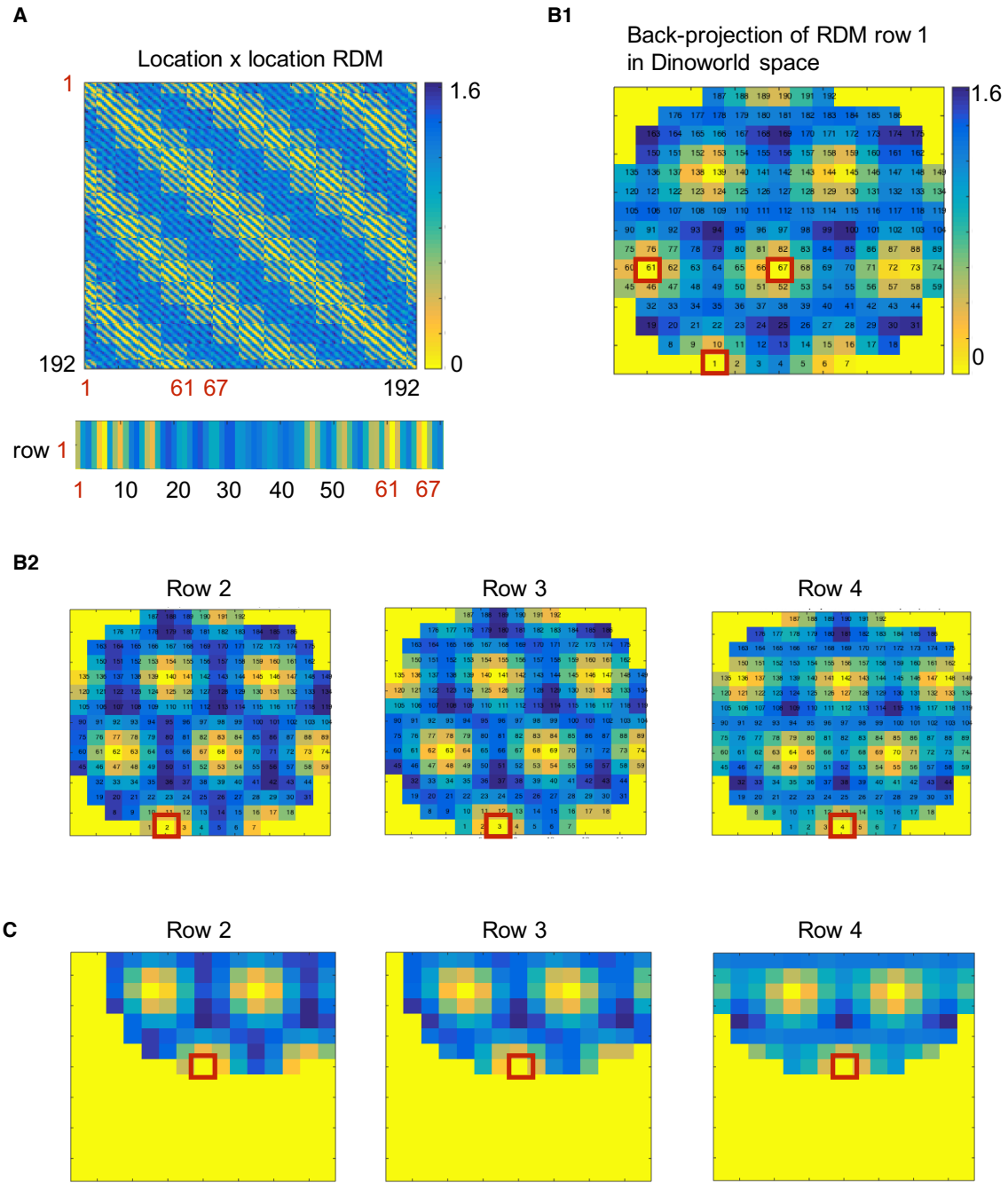


Fig. 8.6: Novel analysis applied on simulated grid cells. (A) Location by location RDM. For example, the row 1 represents the neural similarity between the voxel patterns for location 1 versus all 192 locations. Yellow means high similarity (small dissimilarity or neural distance). Note in the magnified row 1, shown underneath the matrix, how location 1 is very similar to locations 61 and 67. **(B1)** The similarity values for row 1 are back-projected into Dinoworld space. Note how locations 1, 61 and 67 are very similar (highlighted in red). **(B2)** We repeated this back-projection for all 192 rows, leading to 192 maps. Here, we illustrate the maps for rows 2, 3 and 4 and highlight the corresponding seed locations (row numbers) in red. Note how the maps are shifted relative to each other. **(C)** We cropped and re-centred the maps.

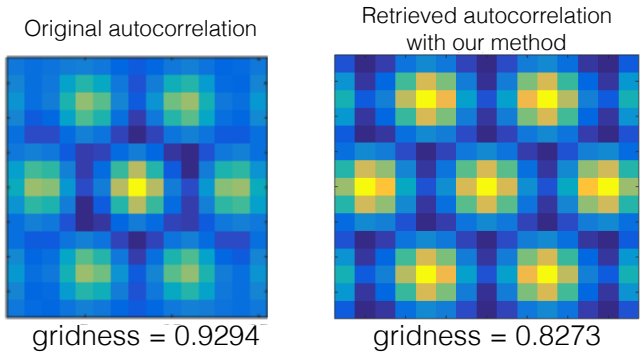


Fig. 8.7: Validation of our novel method on the raw grid cell activity. With our novel method, we retrieved the hexagonal symmetry characteristic of grid cells, at the same grid scale as our simulated raw data.

8.2.8 Gridness score

The next steps of this experiment aim to investigate hexagonal signals in real, fMRI data, using the same novel analysis described above. To determine to what degree the data has hexagonal periodicity, we could use the same gridness score that is conventionally used with electrophysiology data (Fig. 8.8). This score is calculated from the autocorrelogram retrieved with our new method (Fig. 8.7, right), defined as the lowest correlation between the original autocorrelogram and its rotations of 60° and 120° minus the highest correlation for 30°, 90° or 150°. Cells with gridness of 0 or greater are classified as grid cells (Sargolini et al., 2006; Doeller et al., 2010). If C_δ is the correlation for a specific rotation angle δ, then the gridness score is:

$$\text{Gridness} = \frac{1}{2}(C_{60} + C_{120}) - \frac{1}{3}(C_{30} + C_{90} + C_{150})$$

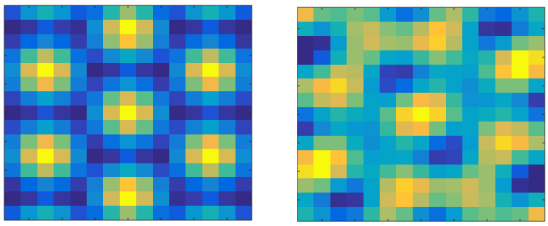


Fig. 8.8: Estimation of gridness. Left, an ideal, simulated, hexagonally periodic grid cell. Right, a simulated cell with disrupted periodicity. Cells with positive gridness scores are considered grid cells by convention.

8.2.9 Simulations using random noise

As a control, we also tested our method on random noise because we wanted to check if there was any bias in our analysis that might output hexagonally symmetric signals regardless of the nature of the input (Fig. 8.9). We found no pattern in the data, and the only correlation was between every location and itself (Fig. 8.9B and F; note the centre yellow dot). This shows that our method is unbiased.

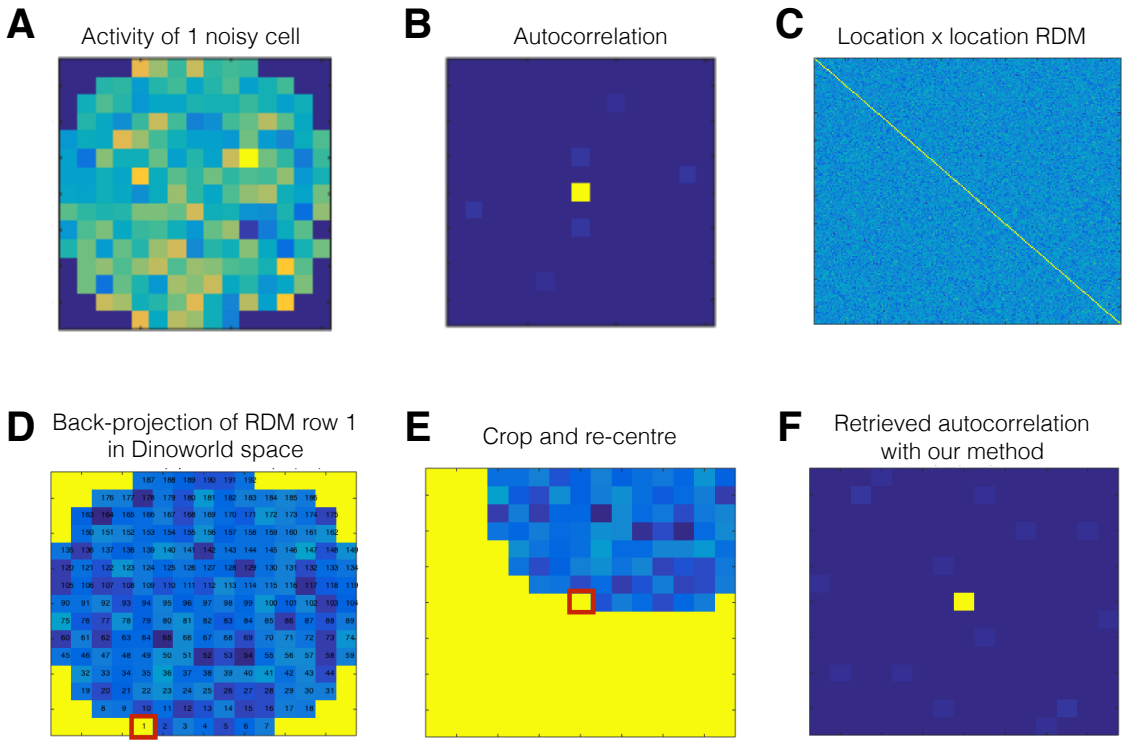


Fig. 8.9: Novel analysis applied on noisy data. (A) Example activity of one noisy cell in Dinoworld locations. **(B)** Autocorrelation of the noisy activity. **(C)** Location by location RDM. **(D)** Back-projecting the RDM into Dinoworld space. **(E)** Cropping and re-centering the back-projected maps. **(F)** Our analysis retrieves random noise when applied to noisy data.

8.2.10 Simulations using ideal place cells

Given that this new method estimates the firing activity of a neuron as a function of subject's location in space, then it should also be able to detect place cells. Consequently, we simulated ideal place cells and tested our method.

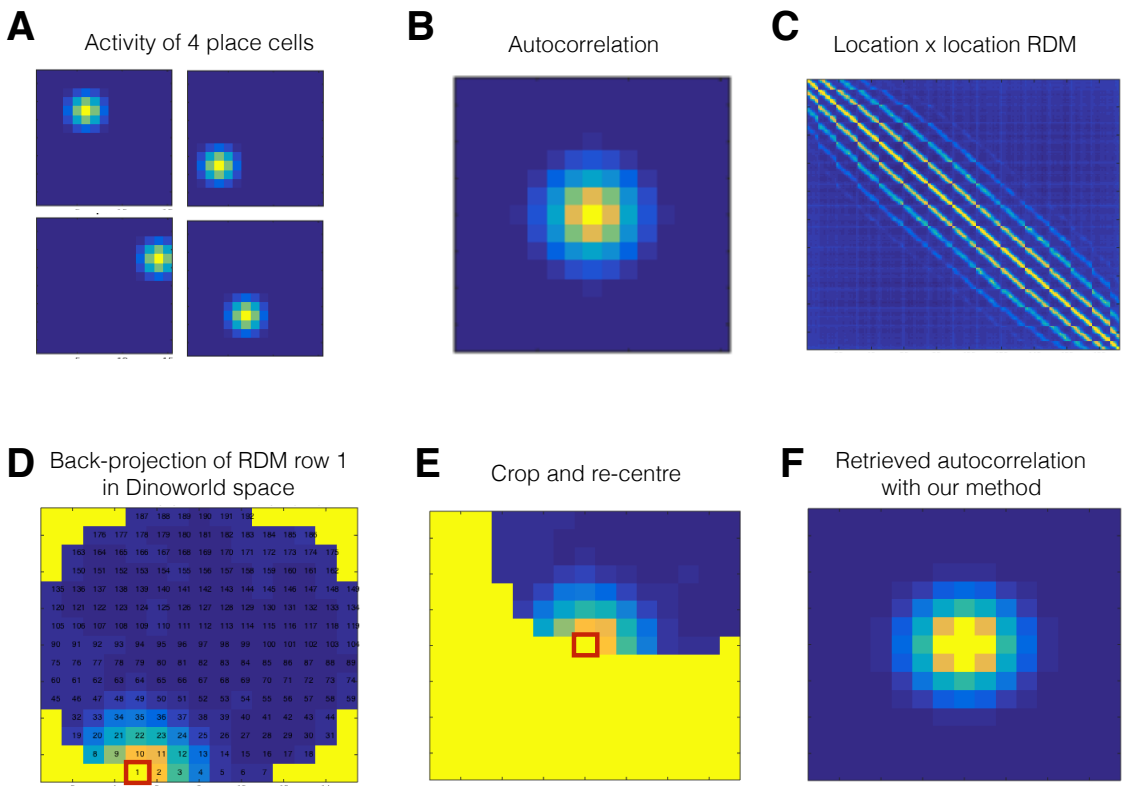


Fig. 8.10: Novel analysis applied on ideal place cells. Note the striking similarity between the original and the retrieved autocorrelation with our method.

8.3 Results

8.3.1 Encode task

Participants first familiarized themselves with the environment, by exploring most parts of space, as well as the boundaries, even when an object was presented to be collected. After the first session of training, subjects were navigating predominantly from one object to the next (Fig. 8.11).

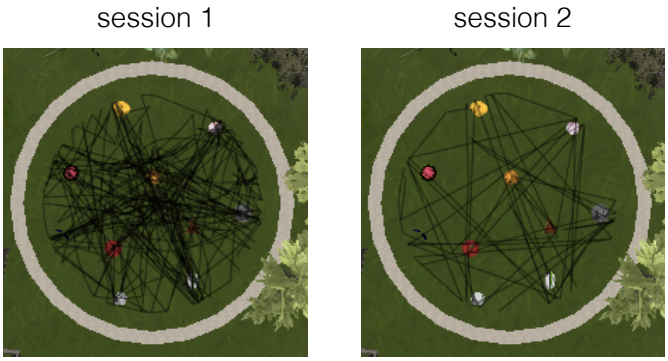


Fig. 8.11: Trajectories during the encode task in one representative subject.

8.3.2 Test and feedback task

The training regime was tailored to each participant until they reached optimal performance, therefore different subjects varied in the number of training sessions. Thus, we concatenated the data from all sessions in each subject, and then split it into five equal blocks (quantiles). We then computed the mean error (Fig. 8.3) for each block, and we averaged this measure across subjects. Following the training sessions with the *encode task*, subjects were already performing correctly on average in the *test and feedback task* during the first quantile. Their response location was within the radius of the object location and, as training progressed, the error decreased even more (Fig. 8.12A).

In the task versions where the only visible landmark was the dinosaur, most participants were within the radius of the object location (Fig. 8.12B). When all landmarks were completely removed, participants could locate the objects but with less precision (Fig. 8.12C). Note that the participants only did a few trials with these tasks because the goal was not to train them to navigate by relying mostly on motion cues, given that in the final task in the scanner participants were not navigating by movement, but by teleportation. The aim of these trials was simply to test if subjects had formed a mental map of the environment that could be accessed during navigation even in the absence of landmarks.

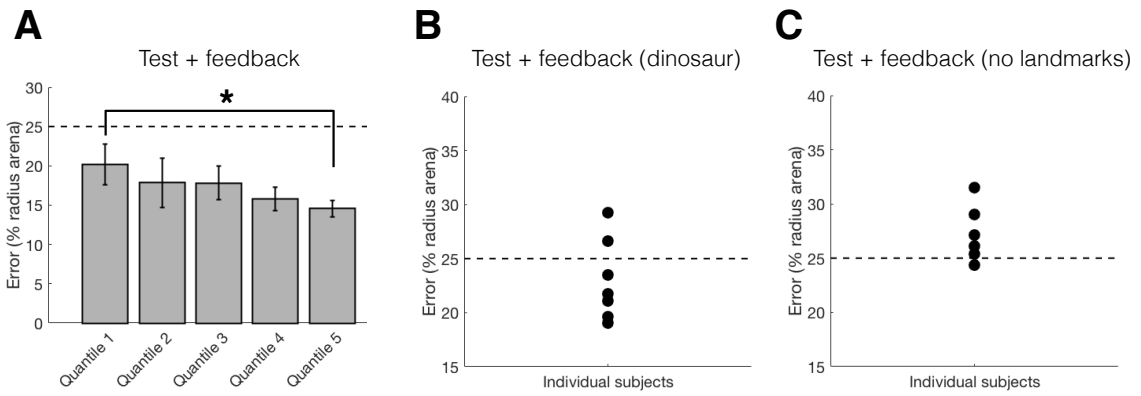


Fig. 8.12: Performance in the test and feedback tasks. (A) Subjects made smaller errors with additional training (quantile 1 vs quantile 5, t_7=2.5, * $p<0.05$). Note that correct performance means the error is within the radius of the object, which represents less than 25% of the radius of the arena (dashed line). **(B)** Performance of individual subjects when the only visible landmark is the dinosaur. **(C)** Performance of individual subjects when all landmarks are removed.

8.3.3 Teleport task

In the teleport task, we computed the performance as percentage of correct responses. The number of training sessions varied across participants. Overall, they continued to improve with training (Fig. 8.13), being able to tell apart objects that were very close to each other.

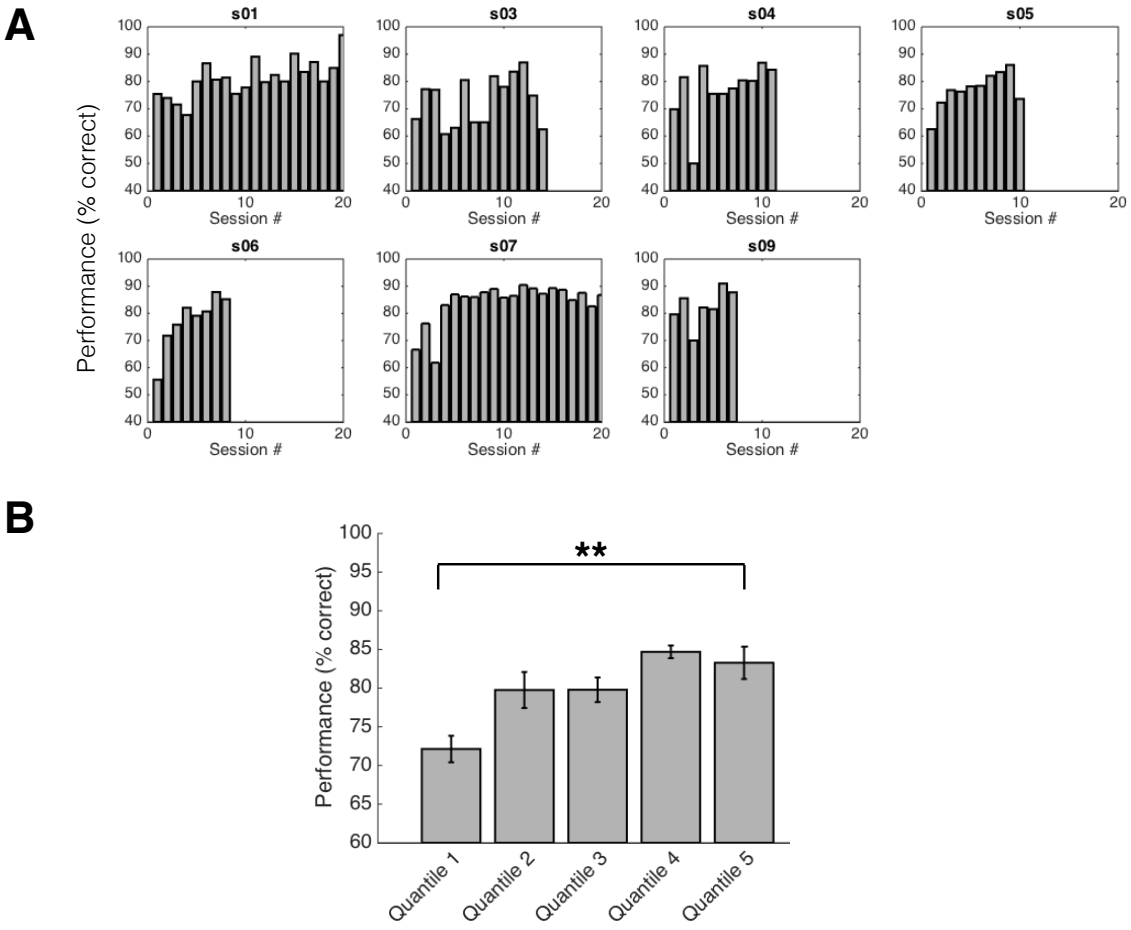


Fig. 8.13: Performance in the teleport task. (A) Performance of individual subjects. **(B)** Group results. The performance improved significantly over the course of training (quantile 5 vs quantile 1, t_6=5.25, ** p<0.005).

8.3.4 Grid scale, grid angle and place field size estimation

We simulated grid cells from different modules across the long axis of the entorhinal cortex, with increasing grid scales and consistent grid angles. Both the grid scales and angles could be estimated with our method (Fig. 8.14A). Similarly, we also simulated

different place cells along the long axis of the hippocampus, with increasing field sizes, and our method could retrieve this gradient, too (Fig. 8.14B).

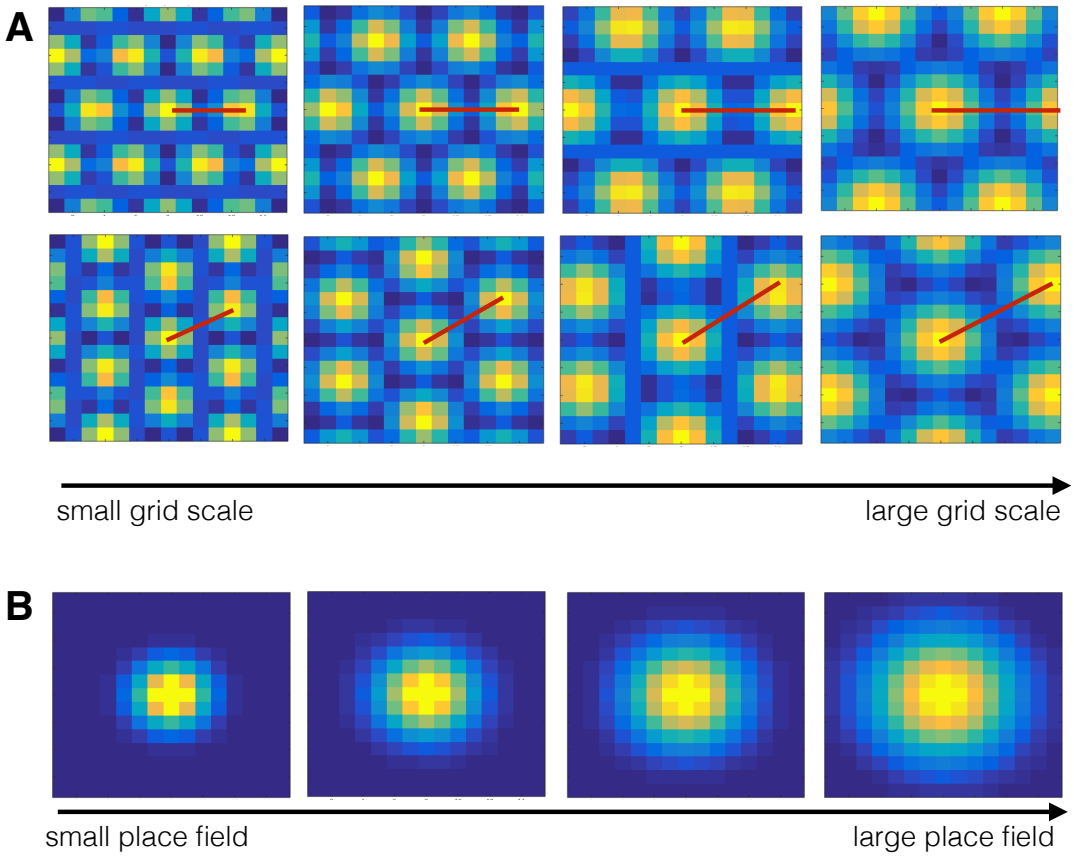


Fig. 8.14: Detection of grid cell and place cell gradients with our novel analysis. (A) Note how our method can detect the scale and orientation of the grid, as well as the size of the grid fields and **(B)** of the place fields.

8.4 Discussion

Here we show evidence that, after intensive behavioural training, humans can localize themselves in space with high accuracy even during immobility and teleportation. Moreover, we developed a new analysis method that detects simulated grid cells and place cells as a function of the subject's location in space, allowing the detection of gradients in grid scales and place field sizes. This raises the possibility that, with high resolution fMRI and careful analyses, it might be possible to detect similar gradients of grid cells and place cells in humans (see next chapter).

The “teleport task” has two main advantages. First, it can being used together with multivariate fMRI analyses (Kriegeskorte and Kievit, 2013) which require different trials close to each other in time to be independent from each other (see next chapter). Hence, tasks that require physical movement are not optimal for these analyses because the continuous nature of movement leads to confounding temporal correlations in the data. And second, knowing an environment in detail opens up the possibility to find a large range in the scales of spatial representations (Geva-Sagiv et al., 2015).

As shown in detail in Chapter 3, humans and other animals can use different strategies to navigate in space, with the cognitive map strategy being the most advanced one, and requiring the highest cognitive abilities (Doeller et al., 2012; Geva-Sagiv et al., 2015). A novel environment can be naturally learnt through free exploration, by combining two main types of information: self-motion cues derived from physical movement and visual cues coming from stationary landmarks (Ravassard et al., 2013). Once a neural representation of the environment is formed as a cognitive map, the behavior is thought to become less reliant on those cues. For example, there is extensive research on how animals can find their way in space while navigating in the absence of visual cues, such as during darkness, by relying on path integration (Foster and Wilson, 2006; McNaughton et al., 2006). However, evidence from research on replay and preplay (Wilson and McNaughton, 1994), as well as VR studies (Aronov and Tank, 2014; Horner et al., 2016; Bellmund et al., 2016a), suggests that animals and humans also know where they are in space when their movement is restricted in various ways. This implies that it should also be possible to locate oneself in space even in the absence of movement, by relying on visual cues, memory and imagination. Take for example the simple case where a healthy person wakes up in the morning and recognizes immediately where they are just by looking at the surrounding environment. In this behavioural experiment, we took advantage of VR

technologies to train participants to learn an environment by relying on different types of cues, and then we tested their knowledge of the map by making available to them only some cues. Our data strongly suggests that people solved the "teleport" task by accessing their cognitive maps of the environment, most likely using grid-cell and place-cell mechanisms.

The reason our VR arena was surrounded by many distinct landmarks was to ensure participants could learn it well, and would be able to rely on those cues in the version of the task where they could not move. Indeed, in cue-rich environments, visual information has a stronger contribution to the formation of cognitive maps, than self-movement information (Buzsáki and Moser, 2013). Interestingly, even if our tasks relied less on path integration compared to other spatial navigation studies (Doeller et al., 2010; Horner et al., 2016), our participants were still able to use their knowledge of the environment and switch to a path integration strategy when, in the absence of peripheral landmarks, it became the optimal strategy. Critically, once the environment was well-learnt, participants could generalize their knowledge to novel associations between objects and landmarks, such as those behind an object which were not seen during training when the object was approached to be collected. This function is the hallmark of cognitive maps (Tolman, 1948).

Using the same neural mechanisms during spatial movement and immobility has implications for studying non-invasively in humans many other non-spatial situations that do not require movement, such as abstract thinking and decision-making (Constantinescu and Behrens, 2016; Garvert et al., 2017; Kaplan et al., 2017). However, because grid cells and place cells are very well understood in the spatial domain (Burgess, 2014; Rowland et al., 2016), the most straightforward approach for developing this new experimental paradigm was to start by testing and validating it in physical space.

Our results strongly suggest that participants formed accurate cognitive maps of the spatial environment. Thus, the next steps are to test if this accuracy is sufficiently high to measure different grid scales using sophisticated analyses of human fMRI data, even during immobility.

Chapter 9: Developing a representational similarity analysis method to detect gradients in place-cell signals and grid-cell hexagonal symmetry in humans, using high-resolution fMRI

The previous chapter presented a behavioural method and neural simulations for detecting gradients in place-cell signals and grid-cell hexagonal symmetry with our novel analysis. This chapter asks whether we can apply this method on real fMRI data and shows the logistics of working with big fMRI data, and some preliminary results. We found that neural similarity increases from the posterior to anterior hippocampus in humans, which suggests increases in place field sizes along the dorso-ventral hippocampus in rodents. We also found hexagonally symmetric representational similarity patterns in the entorhinal, retrosplenial and parahippocampal cortices, as a function of viewing direction. These independent analyses suggest that place-like and grid-like signals may exist as participants perform our spatial task, in the absence of movement. Our novel analysis is still work-in-progress, and the next step is to use the developed pipelines and examine the data more carefully.

9.1 Introduction

Chapter 7 presented an encoding analysis method that takes advantage of several functional properties of grid cells that give rise to a hexagonally symmetric signal at the population level. This chapter describes a novel representational similarity analysis (RSA) method that, in addition, also takes advantage of several anatomical properties of grid cells, which are grouped into clusters of neurons with different functional properties. This new analysis aims to facilitate the detection of grid cell scales and place field sizes, therefore to allow the estimation of functional gradients across the postero-anterior axes of the entorhinal cortex (ERH) and the hippocampus (HPC), the human correspondence of the dorso-ventral axis in rodents.

RSA studies focus on pattern similarities across different experiment conditions, as a method to infer similarity between different neural representations. Thus, it is important to minimize potential factors that do not have neural origin, but could drive such similarities. For example, noise can drive autocorrelation in the data, unless carefully controlled and corrected for. Grid cells are most active when subjects run fast (Sargolini et al., 2006). However, a movement task would induce large temporal autocorrelation in the data, and this would be problematic for RSA. Consequently, a task that requires participants to localize themselves in space in the absence of movement would be optimal for RSA. Evidence from research on replay (Ólafsdóttir et al., 2016b; O'Neill et al., 2017), as well as from virtual reality studies (Aronov and Tank, 2014) shows that grid cells are active even when the subject's movement is restricted. In agreement with this idea, humans have hexagonal symmetry in the RSA patterns, as a function of imagined spatial direction in the absence of movement (Bellmund et al., 2016a).

The hippocampal formation is ~6 times bigger in humans than rats. The rat ERH is ~4 mm along the long axis (Stensola et al., 2012), whereas the human ERH is ~25 mm along its long axis (Maass et al., 2015b). Rats have approximately ~4 grid modules along the ERH long axis, of ~1.5 mm length, that share the same grid angle and scale (Stensola et al., 2012).This raises the possibility that similar modules of ~9 mm might also exist in humans. At a smaller scale than the grid modules, neurons in layer II of the rat ERH tend to form anatomical clusters of ~0.15 mm (Brecht et al., 2014; Ray et al., 2014; Ray and Brecht, 2016). Within a grid cell module in mice, a single cluster of ~0.12 mm has the same grid angle, phase and scale, such that different clusters differ only in their grid phase (Heys et al., 2014a; Gu et al., 2017). This suggests that humans might also have similar clusters of grid cells of ~0.7 mm, at least 6 times bigger than those found in mice. Place cells in mice

that code for the same spatial location also form clusters, of ~0.035 mm (Dombeck et al., 2010), which might correspond to ~0.2mm in humans.

We hypothesized that submillimeter fMRI voxels might be small enough to capture differences in the tuning of place cells and grid cell phases across the human brain (Fig. 9.1). We used an fMRI sequence to obtain crisp images with a resolution of 0.9 mm isotropic. To facilitate the detection of grid-cell and place-cell signals with RSA, we used a spatial navigation task where participants were teleported to random locations.

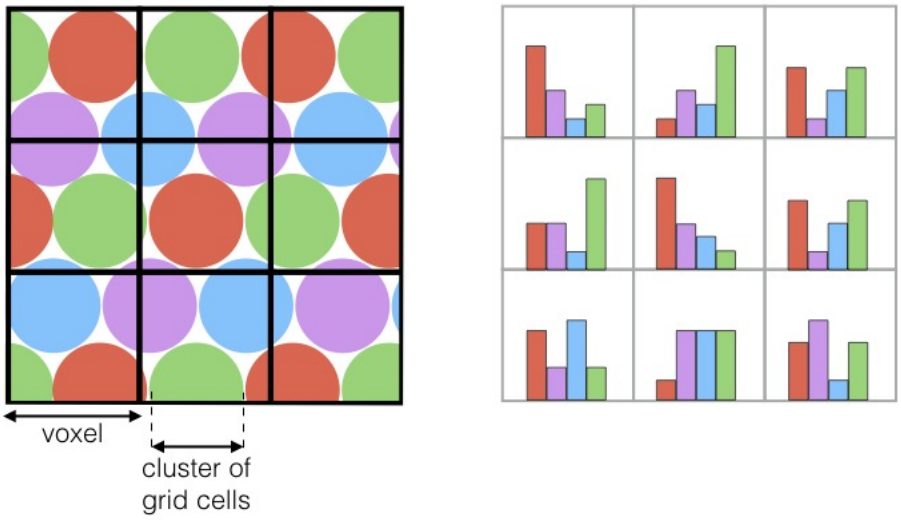


Fig. 9.1: Rationale of the experiment. Simulated anatomical topography within a module of grid cells in the ERH, captured by 9 fMRI voxels with a resolution of 0.9 mm. Each coloured disc represents an anatomical cluster of grid cells with the same scale, orientation and phase. The estimated size of these clusters in humans is ~0.7 mm. Note how clusters of similar phases have a hexagonal arrangement in anatomical space. Different colours represent clusters with different grid phases. These irregularities cause each voxel to sample a different proportion of neurons with different preferred phases, leading to potential biases in the phase preference of each voxel. If the voxels are small enough relative to the anatomical clusters of grid cells, then we might be able to use our novel analysis to look for hexagonal symmetry in humans, as a function of the participant's location in Dinoworld.

9.2 Methods

9.2.1 Participants and experimental design

Three subjects trained extensively to learn the layout of a virtual environment (see the previous chapter), played a computer game while their brains were being scanned with high-resolution fMRI in a 7T scanner, over the course of 6 - 8 fMRI scanning sessions, according to three piloted protocols. We acquired each fMRI session on a separate day, spanning over several weeks. We also acquired high-resolution anatomical data, on a separate day. All participants gave written informed consent and were paid for participating, as approved by the ethics committee of the University of Oxford (SOP number 004-V3).

This experimental approach where we acquire a lot of data from a few subjects comes with the advantage of keeping subject specific details that could allow us to use intricate analyses to identify important features in the data. In contrast, grouping data from many subjects together and making inferences at the group level (Chapter 7) is ideal for increasing signal to noise ratio, however it comes at the cost of smoothing out the data and erasing fine details that might be important. Thus, we chose to do statistics at the level of each single subject.

Subjects learnt to navigate in a virtual reality environment and find 10 hidden objects inside the environment. Once participants could locate themselves quickly in space, while being teleported and immobile in a random location with a random direction, they continued to do the part of the experiment in the scanner. During scanning, the subjects played a version of the *teleport task* described in detail in the previous chapter. In each trial, they saw a snapshot of the environment they were trained with and were instructed to imagine the closest object to them. In some trials, they had to choose one of three offered

objects: the correct object and two distractor objects. To prevent any confounding learning effects, subjects did not receive feedback in the scanner.

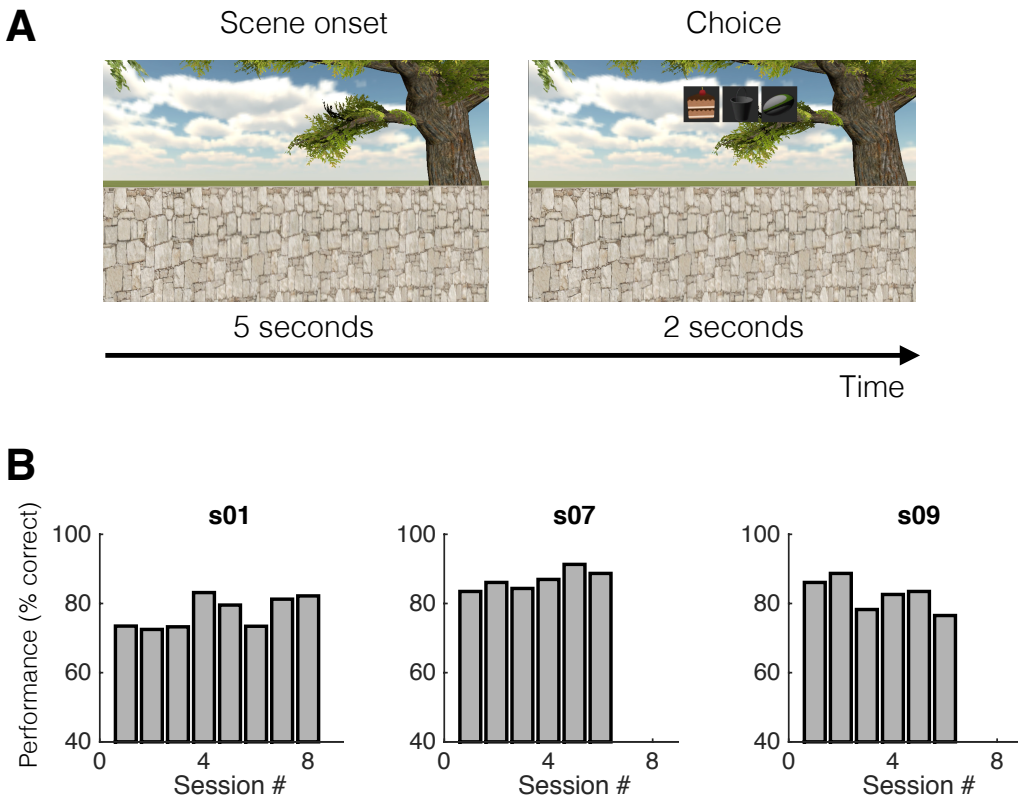


Fig. 9.2: Teleport task in the scanner. (A) Example trial where the participants are shown a scene (a location-angle pair) and they need to think which object is closest to them. This version of the teleport task was like the one participants were trained with outside of the scanner, with the difference that they were asked to make a choice only in 30% of the trials and did not receive feedback. **(B)** Behavioural performances for each individual subject and session. The average performances for each subject were: 77.3%, 86.8% and 82.6%.

High-resolution fMRI at high magnetic fields combined with RSA analyses and cross-validation require careful considerations when setting up the experimental design, such as having many stimulus repetitions that are counterbalanced between sessions and lead to high signal-to-noise ratio. Thus, we used a rich experimental design with 2304 different conditions (location-angle pairs), leading to 4 hours of scanning. Each experimental

condition was one of 192 locations in the environment, paired with one of 12 viewing directions sampled at every 30°.

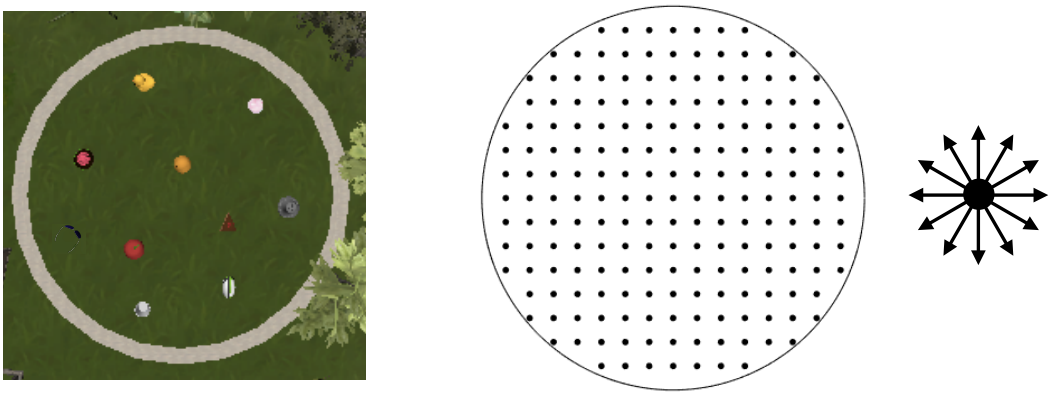


Fig. 9.3: Rich stimulus set. Aerial view of the spatial environment and distribution of all locations and viewing directions.

We piloted three different behavioural schedules, one for each subject. The first subject followed a schedule that used all conditions which were randomly allocated to different sessions. Because such a schedule was not optimized for RSA analyses, we will only illustrate anatomical data and raw timecourse analyses from this subject. The schedule of the second subject was the gold standard for RSA analyses because it included 6 identical repetitions of the same conditions, for each of the six sessions, in a shuffled order (Fig. 9.4A). However, this means it included only a subset of the conditions. The schedule of the third subject was an optimal tradeoff between RSA and our novel analysis because it used the entire stimulus set, that was carefully counterbalanced across the six sessions (Fig. 9.4B).

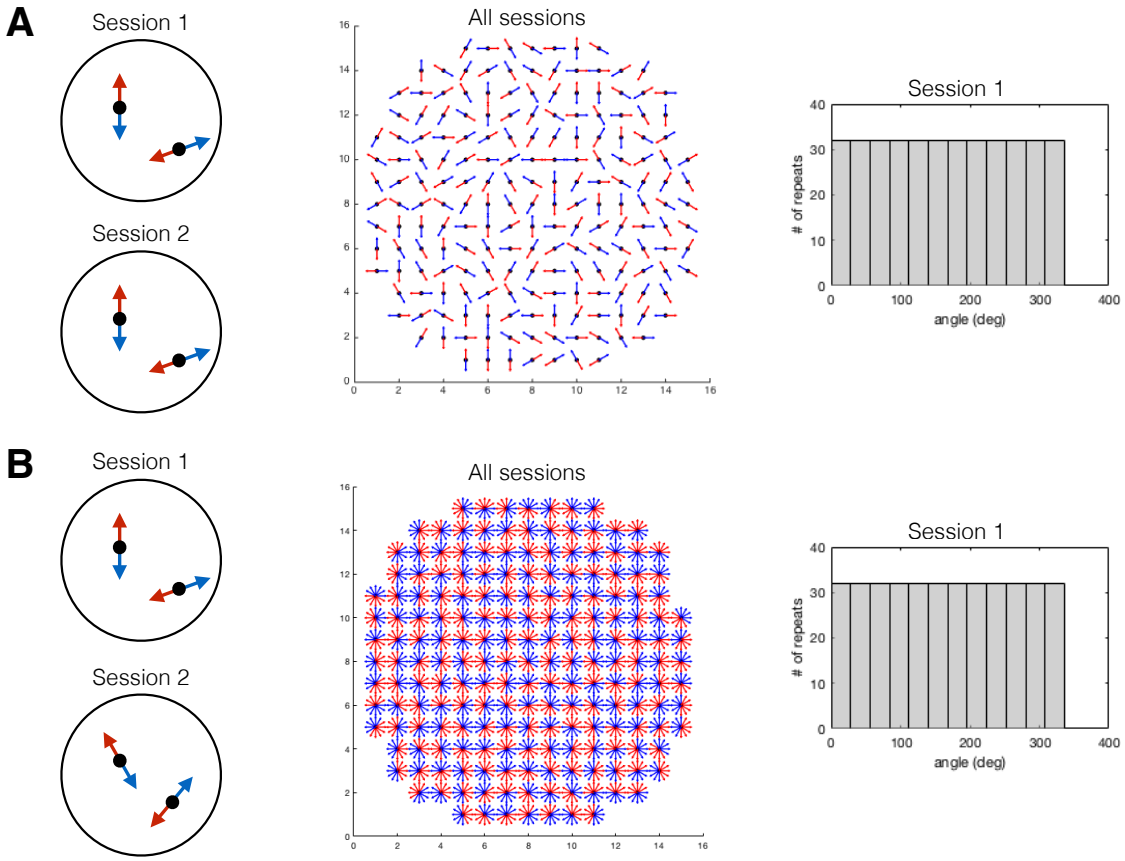


Fig. 9.4: Behavioural schedules optimized for RSA. (A) Two example sessions from the schedule optimized for RSA (subject s07). In each one of the six sessions, we had two repeats of each location, consisting of two unique angles arranged at 180° from each other to cancel out head direction effects. All sessions had identical conditions (location-angle pairs), however, to prevent any confounding order effects, the order of the conditions was shuffled randomly in each session. All angles were sampled uniformly in every session. **(B)** Two example sessions from the schedule created as a tradeoff between optimizing RSA and maximizing the use of our rich stimulus set, making the conditions between two consecutive sessions as similar as possible (subject s09). The angles in each consecutive session were rotated at 30° relative to the previous session. Thus, across all six sessions, each location was paired with all possible 12 angles. Moreover, in each session the angles were uniformly distributed across all possible trials. Therefore, two adjacent sessions had almost identical conditions, with the only difference being that each location-angle pair was rotated by 30°. Like in (A), the order of the conditions was shuffled randomly in each session.

9.2.2 MRI data acquisition

We acquired a set of images at high field 7 Tesla MRI (Siemens MAGNETOM) aiming to obtain submillimeter voxel sizes and high spatial and temporal signal to noise ratio (SNR) in the medial temporal lobe (Fig. 9.5).

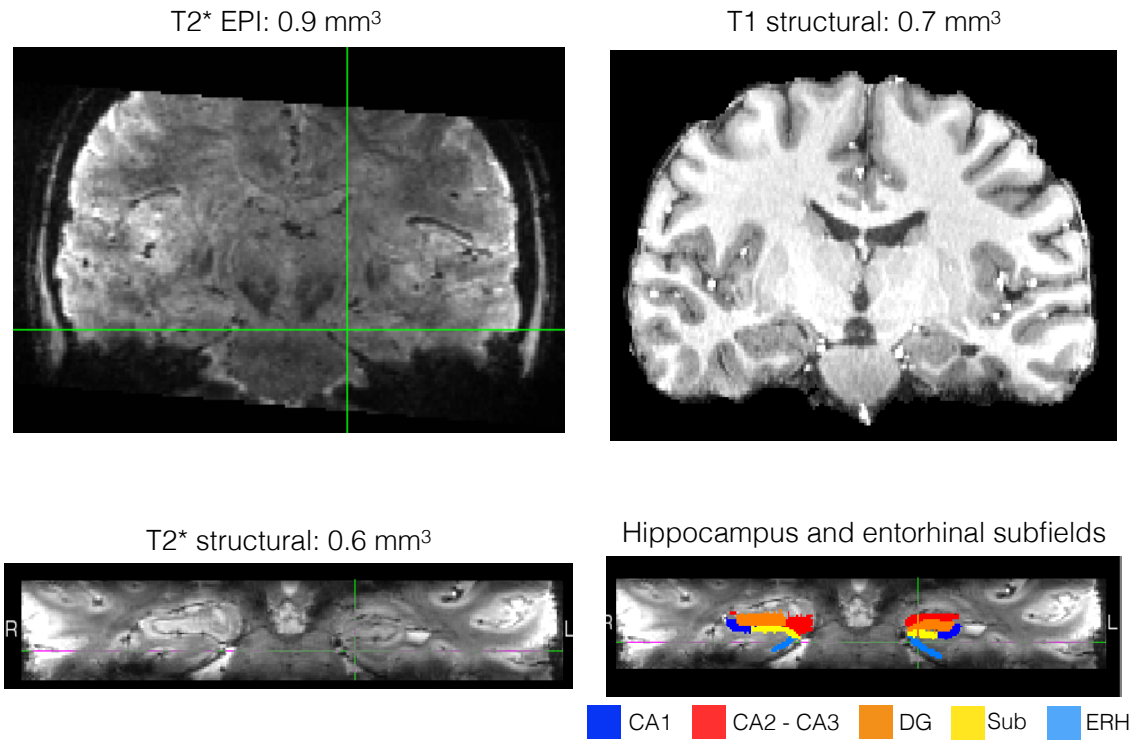


Fig. 9.5: Submillimeter fMRI data. Coronal slices of different MRI modalities, illustrating the cornu ammonis hippocampal subfields (CA1, CA2 and CA3), dentate gyrus (DG), subiculum (Sub) and entorhinal cortex (ERH).

We acquired blood-oxygenation-level-dependent T2*-weighted functional images using a three dimensional echo-planar imaging (3D EPI) pulse sequence (Poser et al., 2010) on a 32-channel surface coil. We chose the parameters of the EPI sequence carefully, aiming to find the optimal trade-off between high SNR in the medial temporal lobe, small voxel size, short repetition time (TR) and large brain coverage (Navarro Schröder et al., 2015). We chose a voxel size of 0.92 mm isotropic because we thought it might be small enough to allow RSA to capture variations in the distribution of grid phases across the population of grid cells (Fig. 9.1).

Small voxel sizes come at a cost of reduced the SNR, however this can be compensated by having shorter TRs and averaging across more data points for the same duration of the EPI session, which increase SNR. We therefore chose a TR of 2.94 s, which also allowed us to sample each trial of 5-7 seconds at multiple time points, giving us a good estimate of

the HRF. We used a field of view of 210 mm and we acquired 72 slices in each slab, with slice oversampling 22.2 %, leading to a partial, although large, brain coverage. The combination of short TR and large brain coverage was possible because we used parallel imaging with a GRAPPA factor of 4 in the phase-encode direction and a factor of 2 in the slice acquisition direction, which sped up the acquisition time of each 3D slab. We then used an echo time (TE) of 20 ms and a flip angle of 14° to obtain a sharp image contrast. Finally, to improve the signal in the medial temporal lobe, a region known to be susceptible to BOLD signal dropout, we tilted the slices at 30° relative to the antero-posterior axis (Weiskopf et al., 2006a; 2007). Stimulus presentation and subject button presses were registered and time-locked to the fMRI data.

The medial temporal lobe is also susceptible to signal distortion, therefore we aimed to minimize it by placing calcium dielectric pads over the head regions corresponding to the medial temporal lobe, bilaterally, during fMRI data acquisition. Moreover, we acquired separate field map images to further correct for distortion during offline analyses, with the following parameters: voxel size = 2 mm isotropic, TR = 620 ms, TE1 = 4.08 ms, TE2 = 5.1 ms, field of view = 192 mm, flip angle = 39°.

To facilitate registration between different imaging modalities, we collected an additional whole-brain T2*-weighted EPI image using identical image parameters as the partial-volume EPIs described above. Moreover, we also acquired a whole brain T1-weighted high resolution structural image, using a magnetization-prepared rapid gradient echo sequence (MPRAGE), with the following parameters: voxel size = 0.7 mm isotropic, TR = 2.2 s, TE = 2.96 ms, field of view = 224 mm, flip angle = 7°, GRAPPA acceleration factor = 2.

Finally, to manually create individual anatomical regions of interest (ROIs), we acquired another type of structural image, with a better resolution and contrast, covering the medial

temporal lobe. Two T2*-weighted ultra-high-reolution images were acquired in opposing phase-encoding directions (left-to right and right-to-left), with the following parameters: voxel size = 0.6 mm isotropic, TR = 50 ms, TE = 25.7 ms, field of view = 192 mm (Cho et al., 2010; Hodgetts et al., 2017). Slices were aligned parallel to the long axis of the HPC (Maass et al., 2014). These two images were averaged together and used for subsequent segmentation protocols.

9.2.3 Pre-processing of functional images

We used FMRIB's Software library (FSL) (Smith et al., 2004b) for the preprocessing stage of the analysis. We segmented brain matter from nonbrain (Smith, 2002). We corrected functional data for motion artefacts, we removed low frequency signals using a high-pass filter at 1/100 Hz, and we prewhitened the data because the GLM works under the assumption that all time points are independent from each other (Woolrich et al., 2001). We also corrected for geometric distortions in the fMRI images using the acquired fieldmaps (Jenkinson, 2003). Because we aimed to do RSA analyses and keep fine details in the data, we did not smooth the fMRI data. Also, because our EPI sequence was in 3D, and all slices in the brain slab were acquired at the same time, repeatedly, we did not need to correct for slice timing. Moreover, to keep the high resolution of the original fMRI data, and to prevent spatial transformation and interpolation processes, we performed statistics in subject space. However, we registered the fMRI images to other image modalities to obtain transformation matrices that we could use later for moving anatomical ROIs into subject-specific space. Thus, we registered the high-resolution anatomical image using boundary-based reconstruction and then we normalized them into standard space (Montreal Neurological Institute – MNI152, voxel size 1 mm isotropic) using non-linear registration.

9.2.4 Segmentation of the hippocampal formation

We used the acquired high-resolution T2* anatomical image (0.6 mm isotropic) to manually segment the hippocampal subfields CA1, CA2, CA3, the dentate gyrus, the subiculum and the ERH using ITK-SNAP, based on a standard 7T protocol (Wisse et al., 2012). We then divided the CA1 into 8 subdivisions along the antero-posterior axis, aiming to look for a gradient in the size of place fields.

9.2.5 fMRI analysis of inter-voxel similarity along the hippocampal long axis

We first aimed to ascertain that our participants had place-like signals during immobility, using an analysis technique that works even at lower anatomical resolutions. We thus looked for neural markers of place cells in our data, using an analysis method that has been tested during a spatial navigation task in virtual reality (Brunec et al., 2017). We reasoned that if such neural signals also exist during our task that did not involve continuous movement, then their detection with our new analysis should be more likely. Briefly, we predicted that the fMRI pre-processed timecourses should be more similar to one another in the anterior HPC, where place fields should be larger, compared to the posterior HPC, where place fields should be smaller. This might be the case because place cells with larger fields should lead to a bigger overlap of the tuning curves of different voxels that are sensitive to spatial location. Thus, we calculated the inter-voxel similarity for all unique pairwise voxels within different ROIs along the long axis of the HPC, and then we averaged the resulting correlation coefficients (Fig. 9.6). We hypothesized that finding a gradient in inter-voxel similarity along the HPC is a good predictor for finding a gradient in place cell fields with our novel analysis.

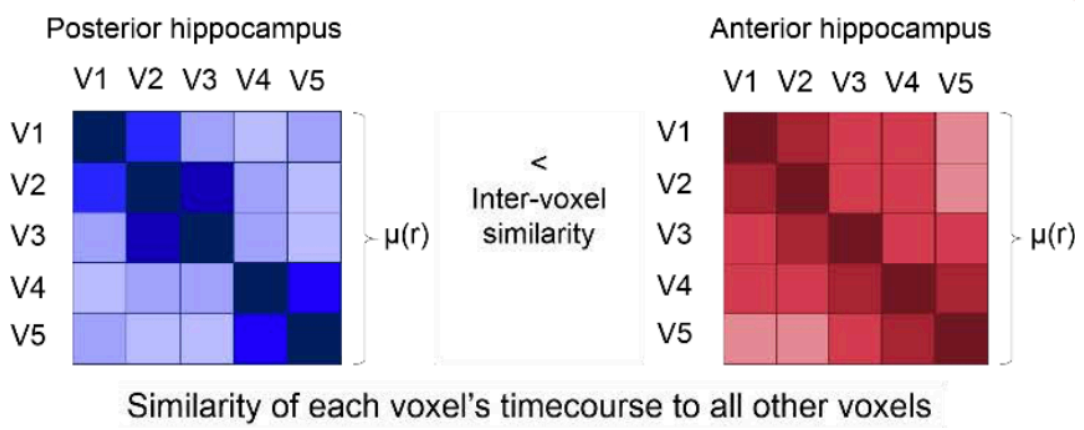


Fig. 9.6: Inter-voxel similarity analysis. We hypothesized that the similarity between the timecourses of the voxels is higher in the anterior compared to the posterior HPC in humans. Here, darker colours represent higher similarity, assessed by increased correlation coefficients. We computed the mean of all unique correlation coefficients μ(r), that is the data above the diagonal of the similarity matrix, for each ROI. This is akin to measures of functional connectivity, calculated between voxels within an ROI. Adapted from (Brunec et al., 2017).

9.2.6 fMRI analysis of head-direction and grid-cell signals modulated by spatial direction

We next aimed to ascertain that our participants also used grid-like mechanisms to solve our task during immobility, using an analysis technique that works even for bigger voxel sizes. Therefore, we carried out an analysis for detecting head-direction and grid-cell signals using RSA analyses (Nili et al., 2014), as a function of spatial direction (Fig. 9.7), that have already been validated during a spatial imagination task in the absence of movement (Bellmund et al., 2016a). Again, we reasoned that if we found grid-cell signals using this analysis on our teleport data, then it should also be possible to find them with our novel analysis that depends on the spatial location, and not spatial direction. This RSA that depends on spatial direction takes advantage of the property that grid cells have a common grid axis that is more uniformly distributed than the grid scale, therefore it is easier to detect at the population level with fMRI. Thus, we could carry it out after down-

sampling the voxel size of the pre-processed fMRI data at 3 mm isotropic, which had the advantage of increasing the signal to noise ratio.

First, we modelled the fMRI timeseries using a general linear model (GLM) in SPM12, grouping together all six sessions in each subject. In one session, we included 384 trialwise regressors for the main effect of each location-angle condition, during the scene onset periods. Because of the sensitivity of the blood oxygen level-dependent signal to motion, we also included six motion regressors obtained during the motion correction step of the pre-processing. Moreover, we modelled the main effect of each separate fMRI session in the GLM. All regressors were convolved with a canonical haemodynamic response function.

Second, the resulting beta coefficients corresponding to the 384 experimental conditions were entered in RSA searchlight analyses, using spheres with a diameter of 7 voxels (21 mm). This approach calculates pattern similarity values for search spheres centered on the current voxel, iteratively. We analyzed only search spheres including more than 30 voxels and the pattern similarity was calculated across all voxels within a sphere, separately for all possible pairs of conditions. We used [1 - Pearson correlation coefficient *r*] as a measure of neural distance because it yields small values for patterns that are highly correlated, and thus have a small neural distance and are very similar. The resulting representational dissimilarity matrices (data RDMs) were compared with two model RDMs (Fig. 9.7), using Spearman's rank correlation coefficient which was Fisher z-transformed (Nili et al., 2014). The model RDM for head-direction signals predicted increased neural similarity between similar directions (<= 30° apart) and the model RDM for grid cells predicted increased neural similarity for hexagonally symmetric directions (multiples of 60° apart). For each separate experimental session, the output of the searchlight analysis were two *r*-value maps corresponding to the two model RDMs. We entered these *r*-value maps across

sessions into a t-test, and presented the results as *p*-value maps in MNI space (uncorrected), separately for each model RDM.

Searchlight analyses have the benefit of identifying head-direction and grid-like signals across the whole brain, however they have some limitations and can misidentify a cluster as informative or fail to detect truly informative voxels (Etzel et al., 2013). Thus, we did not use them for statistical inference, but instead as a tool to identify informative brain regions that we could then use to create ROIs for subsequent unbiased tests.

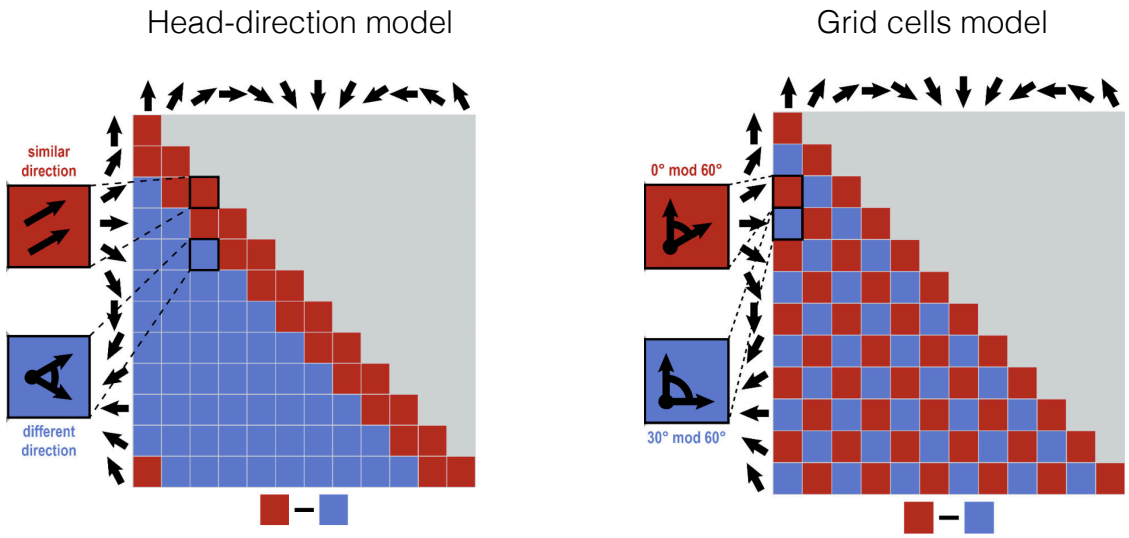


Fig. 9.7: RSA models for head-direction and grid-cell signals, as a function of spatial direction. In the head-direction model, two directions were considered similar if they had an absolute angular difference of 0° or 30°. In the grid cells model, two directions were considered similar if the reminder of their modulus 60° was 0°. For simplicity, the model RDMs are shown as a 12x12 matrices according to the spatial direction. However, for precision, during the analysis pipeline we used 384x384 trialwise RDMs, with an entry for each location-angle pair. Adapted with permission from (Bellmund et al., 2016a).

9.2.7 fMRI novel analysis

Next, we used the anatomical and searchlight-derived ROIs to look for place-like and grid-like signals as a function of subjects' location in space, with our novel RSA approach. The details of this analysis are presented in Chapter 8. Briefly, we aimed to retrieve the autocorrelation of populations of neurons that fire as a function of spatial location, such as place cells and grid cells. Similar to the autocorrelation of single-cell recordings of these neurons, we aimed to retrieve the size of place cell fields, and the scale and orientation of grid cells, which would allow us to map human analogues of functional gradients in the hippocampal formation, and other brain regions. We hoped this would be possible by using an RSA approach such that patterns of brain activity across voxels are similar between locations that correspond to the firing field of a population of neurons, and are dissimilar for the locations outside of the firing field.

Our novel analysis aims to address both goals of the previously described analyses (intervoxel-similarity analysis for place cells and RSA for grid cells as a function of spatial direction), in one go, because it shares common features with both of them (Fig. 9.8). It is similar to the inter-voxel similarity analysis because it aims to look for gradients in place cell activity, however it specifically targets the neural response to different Dinoworld locations. Thus, instead of correlating fMRI timecourses between voxels, we correlated the voxel patterns of beta values between specific experimental conditions. This would bring more direct evidence of place-cell activity. Our novel analysis is also similar to the RSA for grid cells as a function of spatial direction, however our experimental conditions depend on spatial coordinates of the subject's location in Dinoworld, as opposed to their viewing direction. This means that it should be able to detect not only grid-cell angles, but also grid cell scales and place-cell signals.

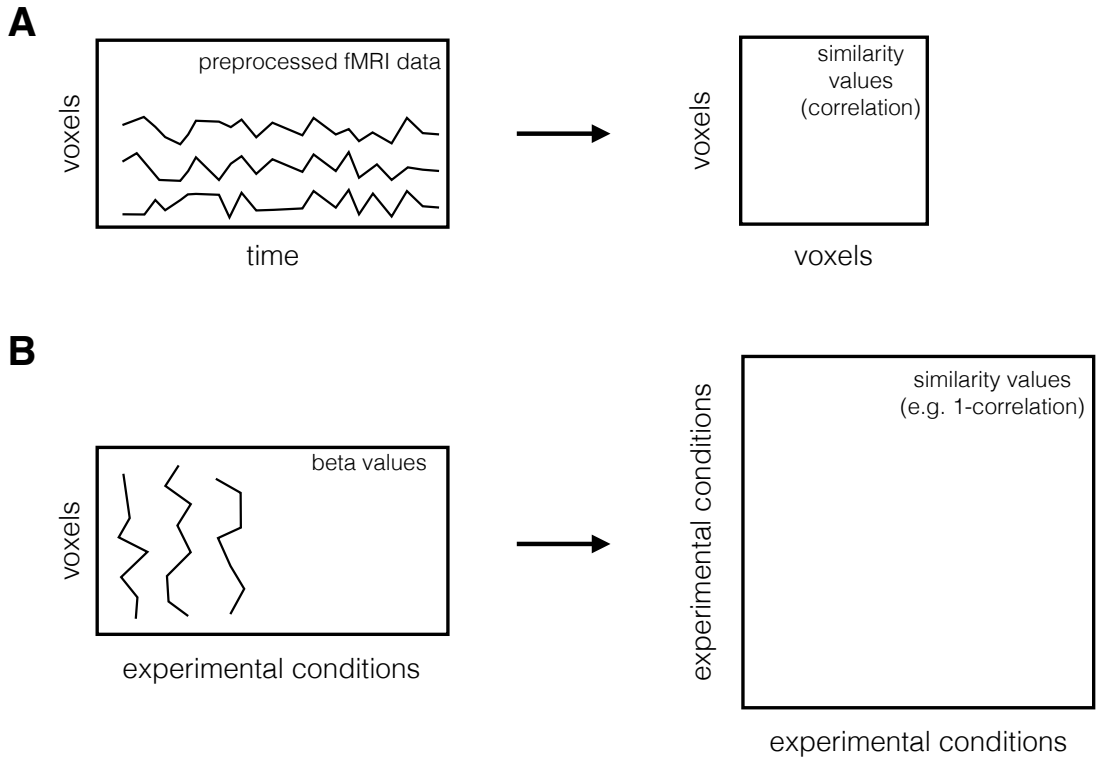


Fig. 9.8: Comparison between the three analyses used in this chapter. (A) Inter-voxel similarity analysis outputs a voxel – by – voxel similarity matrix. **(B)** RSA analyses output a condition – by – condition dissimilarity matrix, using different measures such as [1 – Pearson correlation coefficient], Euclidean distance, crossnobis distance. In the RSA for detecting grid-cell signals as a function of viewing direction, the experimental conditions are the viewing directions, whereas in the RSA for detecting grid-cell and place-cell signals as a function of spatial location, the conditions are the identity of the spatial locations.

To obtain our experimental conditions, we ran another GLM, again grouping together all six sessions in each subject. However, in one session, our regressors were the main effect of each of the 192 locations. Therefore, the onsets of the trials with the two viewing angles corresponding to the same location (red and blue arrows in Fig. 9.4) were grouped together into the same regressor. Again, we included motion regressors, we modelled the main effect of each separate fMRI session in the GLM, and all regressors were convolved with a canonical haemodynamic response function. We then ran our RSA on the 192 beta coefficients resulting from our GLM.

We used three distance measures for RSA: [1 – Pearson correlation coefficient], Euclidean and cross-validated Mahalanobis (crossnobis). The correlation distance is the most common used one in the literature, however a recent report showed that it is less reliable for data where the voxels are spatially correlated (Walther et al., 2016). Thus, we used the correlation distance for the simulations from Chapter 8 because our data were ideal. For example, the correlation distance between two vectors b_j and b_k, corresponding to the beta values for all voxels for conditions j and k, is:

$$d_{Pearson}(b_j, b_k) = \cos(\alpha)$$

where α is the angle between b_j and b_k.

However, we also used crossnobis on our real fMRI data because neighbouring voxels tend to have similar amount of noise (similar amounts of variance over time) and, hence, be correlated. Calculating the crossnobis distance involves two main steps. First, we calculated the Mahalanobis distance, which is the squared Euclidean distance that takes into account the voxel – by – voxel covariance matrix obtained from the residuals of the GLM. This is a form of spatial pre-whitening of the beta coefficients that downweights the noisy voxels by taking into account the spatial structure of the noise.

If the squared Euclidean distance is:

$$d^2_{Euclidean}(b_j, b_k) = (b_j - b_k) * (b_j - b_k)^T$$

And the pre-whitened beta values are:

$$b_j^* = b_j * \Sigma^{-1/2}$$

$$b_k^* = b_k * \Sigma^{-1/2}$$

where Σ is the noise covariance matrix. Then, the squared Mahalanobis distance is:

$$d^2_{Mahalanobis}(b_j^*, b_k^*) = (b_j^* - b_k^*) * (b_j^* - b_k^*)^T$$

Second, we cross-validated between different fMRI sessions, by computing the distance between conditions j and k separately for each possible pairs of sessions, and then averaging across all pairs:

$$d^2_{Crosnobis}\,(b_j^*, b_k^*) = (b_j^* - b_k^*)_{sessionA} * (b_j^* - b_k^*)^T{}_{sessionB}$$

9.2.8 Tackling challenges for big data experiments

Our data was big and it required many resources (Van Horn and Toga, 2014). We started with a big stimulus set (2304 conditions) and big fMRI data size (3 GB per experimental session, compressed). This led to intense training routines outside of the scanner (~10 hours), big reconstruction time from k-space to nifti-space (up to 1h after the end of the scanning session), big acquisition time (4 hours of scanning corresponding to 6 experimental sessions) and a big set of statistical images that was ~600 times larger than those of typical studies (Fig. 9.9). Such big data requires complex workflow. For example, the GLM and searchlight analyses require ~10^9 separate tests, one for each individual voxel in the brain (>10^6) and each individual condition (>2300). These computations create temporary files which expand the memory demands even more.

Our big data required a large team effort with the physics and analysis colleagues in FMRIB, to adapt the hardware and analysis pipelines to match the memory demands. For example, we modified the FEAT pipelines for data pre-processing to use cluster nodes of 500 GB RAM (a standard node has 10 GB of RAM). We also ran SPM and other Matlab scripts on the same cluster nodes, or on a separate computer with 1TB RAM. Moreover, we adapted our scripts to Matlab which has limitations in loading and saving big data (for example, our largest SPM.mat file was 19 GB, whereas in a typical experiment it is only

a few MB). Overall, we took several logistical factors into account when adapting and building a diverse combination of pipelines for data management, processing and complex analyses. Therefore, the results in the following section present some highlights of several analyses that are still ongoing.

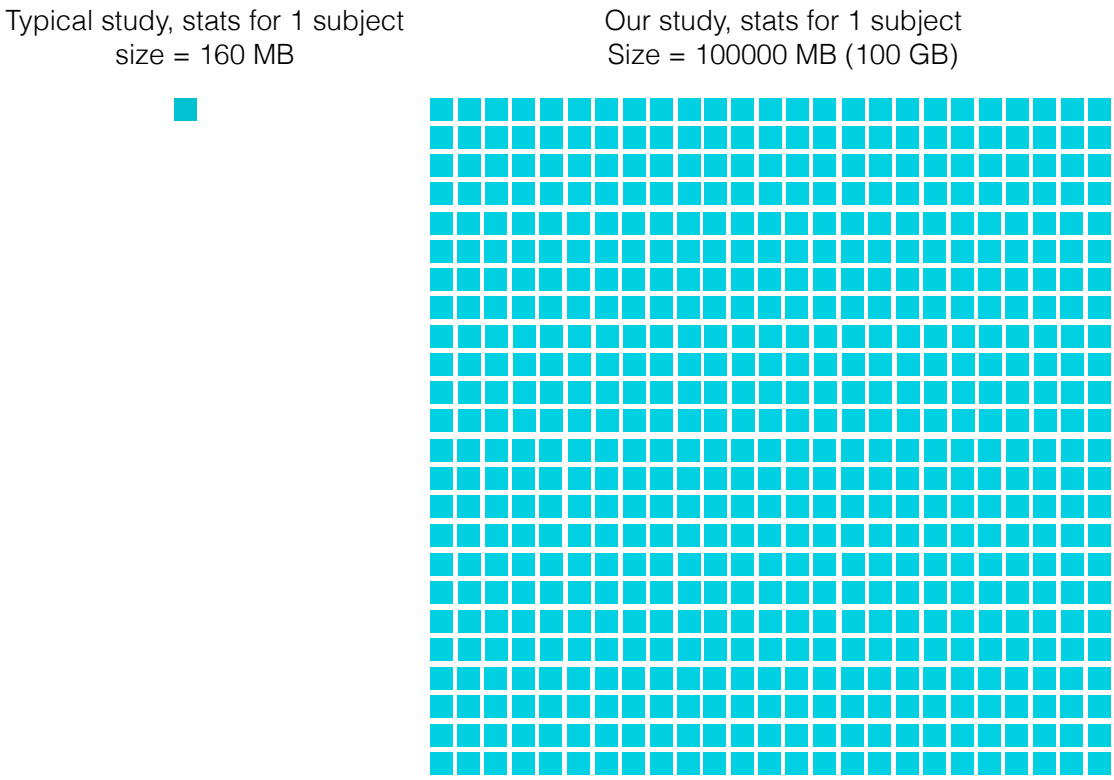


Fig. 9.9: The output of our GLM statistical analysis was 600 times bigger than the output of a typical study. The parameters for a typical study were taken from the fMRI data used in Chapter 7, for one 20-min session in one subject: 3 mm isotropic voxels covering the whole brain (64*64*45 voxels), ~400 volumes and ~10 regressors. When doing stats across all sessions in one subject, the summed size of all outputs (beta, contrast and t-stat images) is 160 MB. In the current study, a 40-min session in one subject had: 0.9 mm isotropic voxels with partial coverage of the brain (228*228*72), ~750 volumes and ~400 regressors. The statistical images add up to 100 GB across all sessions in one subject. The calculations are for outputs given by SPM because it performs GLM much quicker than FEAT, such that SPM takes ~3 days on our big data compared to FEAT which takes ~3 weeks. However, SPM uses uncompressed nifti data which means that a 3 GB fMRI data in its compressed form becomes 11 GB after decompression.

9.3 Results

9.3.1 Co-registration of high-resolution images

To test for functional gradients across the hippocampal formation, we needed to extract the 0.9 mm isotropic fMRI T2* data from the manually defined ROIs based on the 0.6 mm isotropic anatomical T2* data. This means that we needed to match carefully two images of the same modality, but with different resolution and field of view, to preserve the anatomical specificity. This involved several optimization steps, based on the transformation matrices created by FEAT during pre-processing, with various costs functions. Briefly, we obtained a warp field that transformed the segmented ROIs into the space of the fMRI data (Fig. 9.10, bottom). This warp field was, in turn, obtained from the concatenation of the following transformation matrices: (i) 0.6 mm anatomical image registered to the 0.9 mm whole-brain fMRI image and (ii) 0.9 mm whole-brain fMRI image registered to the 0.9 mm partial-field-of-view fMRI image. Because all these three images were of the same modality (T2*), we used linear registration (FSL FLIRT). Moreover, to obtain a second validation of the first step, we also registered the ROIs into the space of the 0.7 mm whole-brain anatomical image (Fig. 9.10, top). Again, this was possible by concatenating two transformation matrices: (i) same as above and (ii) whole-brain fMRI image registered to the whole-brain anatomical image. Step (ii) was achieved with FSL BBR, a non-linear transformation that uses the white-grey matter contrast to improve the registration process between T2* and T1 images. We inspected each image visually and carefully. Some example brain slices are shown in Fig. 9.10.

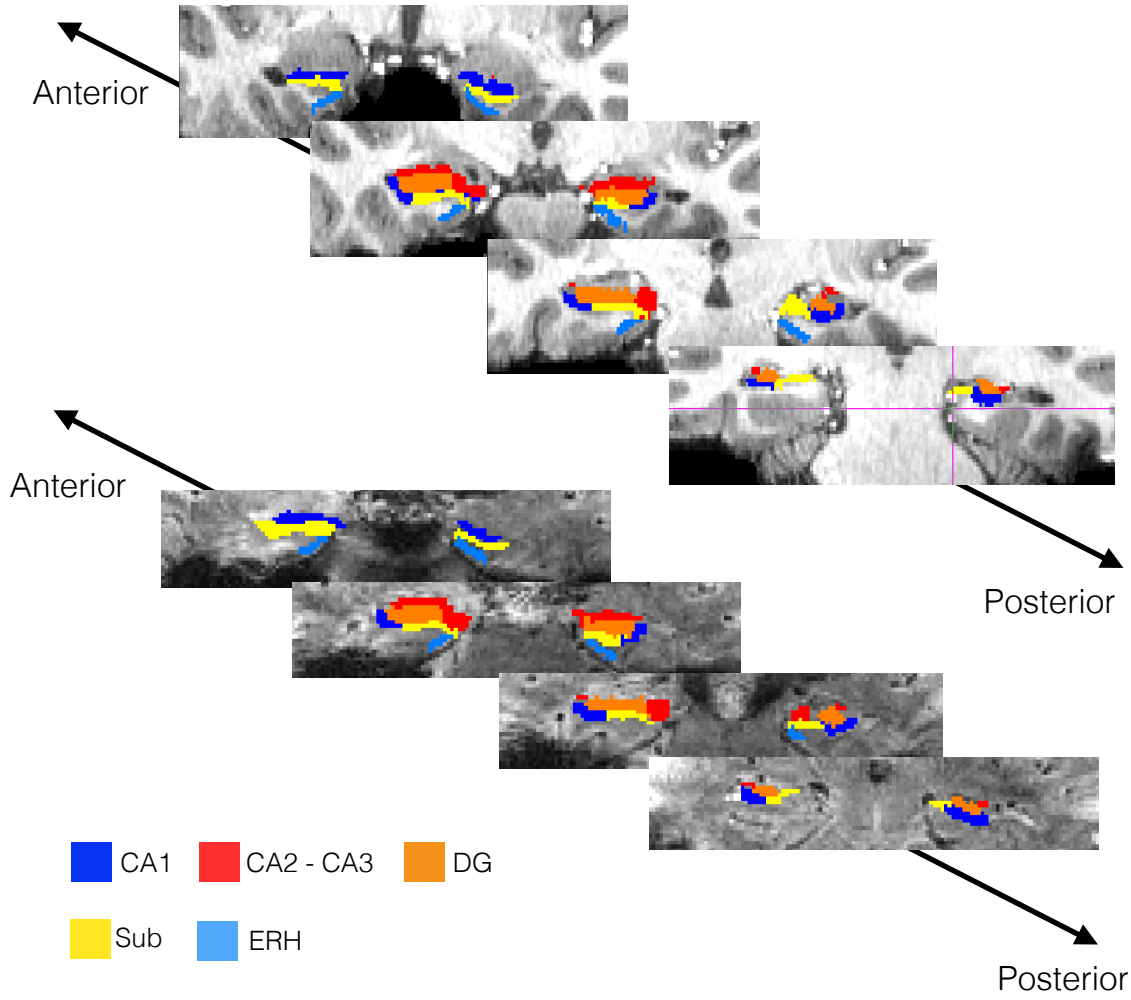


Fig. 9.10: Example registrations of the hippocampal subfields. Top, T1 structural space. Bottom, T2* functional space.

9.3.2 Gradients in inter-voxel similarity during immobility

Next, we used these ROIs to look for changes in inter-voxel similarity along the antero-posterior axis of CA1. To visualize such gradients at fine spatial resolution, we split the CA1 into 8 subdivisions, with inter-voxel similarity calculated for each division separately (Fig. 9.11A). Indeed, the similarity tended to decrease from anterior to posterior divisions. When separating data across a boundary approximately at the uncal apex, then the intervoxel similarity was significantly higher in the anterior compared to the posterior CA1

(Fig. 9.11B). This raises the possibility that humans, like rodents, also have a gradient in the size of place cell fields across the long axis of the HPC.

Nevertheless, this analysis does not control for several confounding factors that are common for any functional connectivity analysis of this kind. For example, it could be possible that the anterior HPC has more place cells than the posterior HPC, or the anterior HPC is more susceptible to motion and noise which could artificially induce spatial autocorrelations between voxels. Future analyses could rule-out some of these confounds by correcting more carefully for motion, for example by modeling breathing and pulse into the GLM, or using automated cleaning methods (Griffanti et al., 2014). These de-noising methods have not yet been officially developed for 3D EPI data, however we have developed some preliminary advanced motion correction methods as part of our big data pipelines. Overall, to reliably test for such gradients, we would need to test for similarities for each location in the spatial environment, which is the approach we are developing with our new method (for simulations on ideal place cells, see previous Chapter).

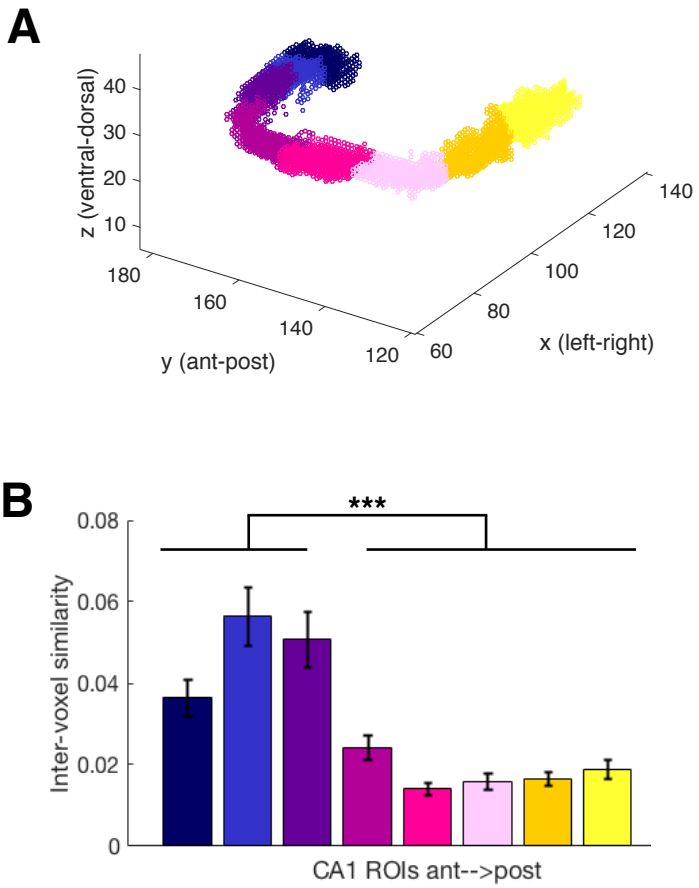


Fig. 9.11: Example gradient in inter-voxel similarity in CA1. (A) CA1 ROIs divided along the antero-posterior axis of the HPC. **(B)** Note the gradient in the corresponding inter-voxel similarity values for each ROI. The similarity is significantly higher in anterior compared to the posterior subdivisions of CA1 (*** $p<0.001$).

9.3.3 Head-direction and grid cell signals during immobility

Using searchlight analyses at the whole-brain level, we found a head-direction effect in the parahippocampal and the retrosplenial cortices, as well as a grid cell effect in the ERH (Fig. 9.12A). The anatomical location of these effects was remarkably similar to the one during imagined direction in space in the absence of movement (Bellmund et al., 2016b), and consistent with reports from the rodent literature of head-direction cells (Taube, 2007) and grid cells (Moser et al., 2014). We then focused on these regions to create ROIs and test for grid cell signals using an unbiased approach. Specifically, we tested for grid cell signals using the ROIs defined from the head-direction analysis. These two analyses were

independent because we excluded the conditions corresponding to the same head-direction condition (the diagonal of the models presented in Fig. 9.7). We then tested for grid cell signals using the ERH ROI obtained from the grid cell analysis from a completely different dataset.

To test with RSA for a grid-like signal as a function of viewing direction, at the ROI level, we calculated the difference in mean pattern similarity between the aligned (0° mod 60°) and misaligned (30° mod 60°) conditions for each session (Fig. 9.7 right). We then tested for *increased* pattern *similarity* (which is equivalent to a *decreased* pattern *dissimilarity*), by comparing this difference against 0, using a one-tailed t-test (Bellmund et al., 2016a). We found such significant hexagonally symmetric patterns, suggestive of grid cells, in the ERH, parahippocampal and retrosplenial cortices (Fig. 9.12B). Although these results are preliminary, they suggest that it is possible to measure grid-like signals in humans during our teleportation task.

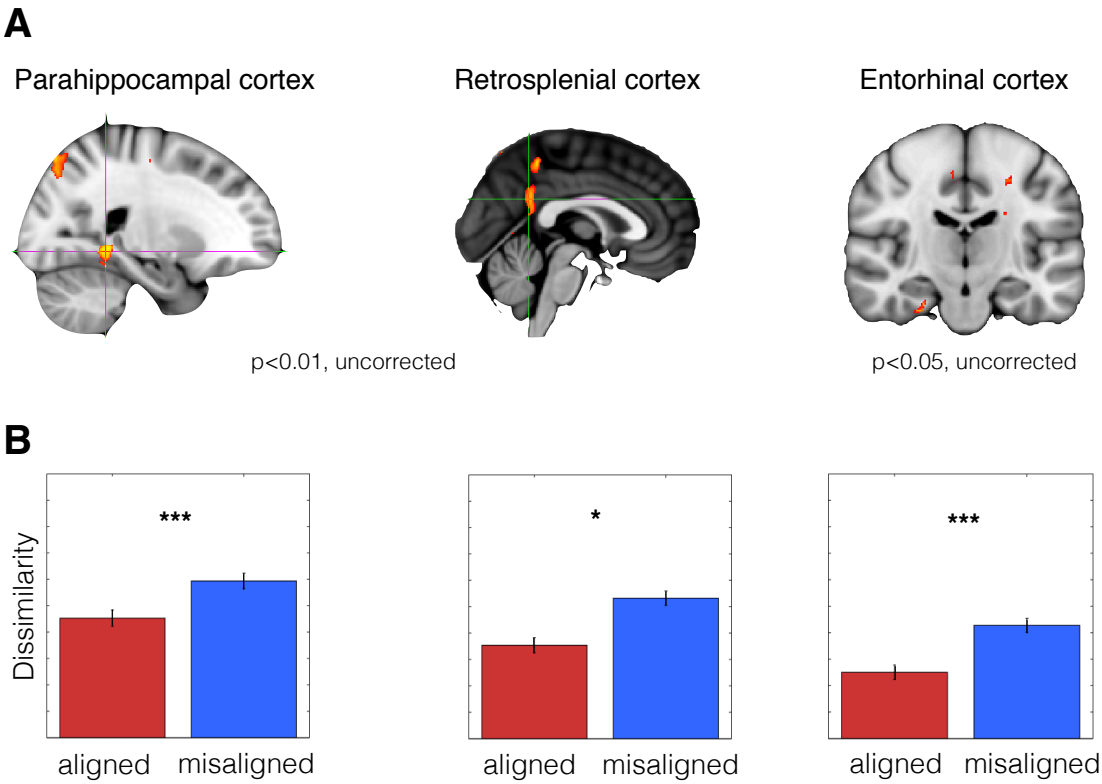


Fig. 9.12: Head-direction signals and grid-like hexagonal symmetry in one example subject. (A) Whole-brain searchlight results showing head-direction signals in the parahippocampal cortex and retrosplenial cortex, and grid-like signals in the ERH. **(B)** Significantly less dissimilarity (increased similarity) for spatial directions that are aligned (0° mod 60°) compared to those that are misaligned (30° mod 60°), in the ROIs from (A).

9.3.4 Novel analyses

We hoped the pattern of place cells and grid cells would pop up after applying our work-in-progress analysis. For example, we hoped we could see similar functional gradients in the place cell fields along ROIs defined as subdivisions of the HPC along the anterior-posterior axis, or grid cell scales across the ERH, as we obtained during our simulations. However, the results looked noisy, even after using the crossnobis distance. For example, we used the same ERH ROI from the searchlight analysis as in Fig. 9.12, where the RSA analyses as a function of viewing direction were significant. However, we did not find

hexagonal patterns with our analysis (Fig. 9.13). There are several possible explanations for these findings (see Discussion).

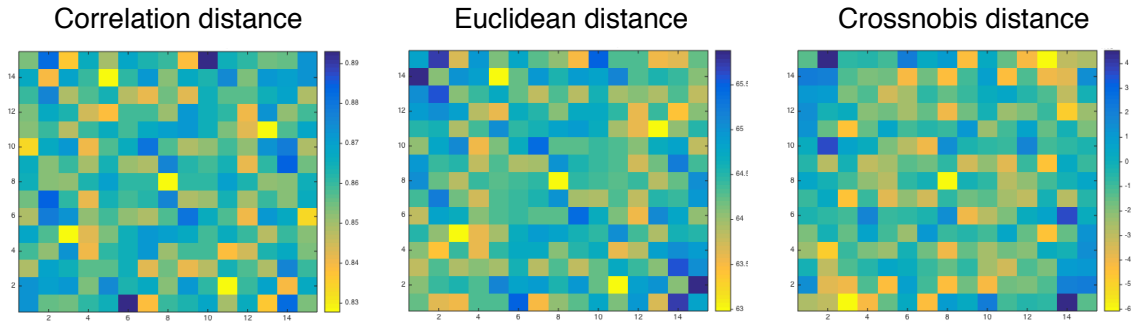


Fig. 9.13: Example results of our novel analysis on real fMRI data from the ERH. Autocorrelation results of our analysis, using three different distance measures. Note the absence of hexagonal symmetry. For visualization and scaling purposes, the centre point of the autocorrelation (which is always 0 because it corresponds to data correlated with itself, hence a distance 0) was made as NaN.

9.4 Discussion

This chapter highlights several challenges and opportunities of working with big fMRI data and raises the possibility that, with ongoing work, it might be possible to detect fine gradients in place-cell and grid-cell activity in humans. As big data require many resources, we piloted carefully different experimental designs and image sequences, that led to a rich dataset open for exploration. Preliminary data suggest that we can find correlates of place cell gradients and grid cell hexagonal symmetry at the population level in humans, using independent analyses. This is hopeful for future developments of our protocol.

One explanation for not finding positive results with our novel analysis is that our ROIs were too big relative to the size of the grid modules, therefore, our signal would be "contaminated" by grid cells with different grid scales, canceling out the hexagonal symmetry. Another explanation could be that subtle breathing or cardiac motion could

have dramatic influences on our small voxels, therefore in two participants we acquired simultaneous physiology monitoring. Although such motion correction pipelines have not yet been developed for 3D EPI data, we have some in-house code that we plan to test. Another possibility could be the cortex is folded and one ROI might include different grid modules that would become separated in space had the cortex been flattened out. Surface analyses would address this concern, and they also have the benefit of being less computationally demanding because the big number of voxels in our data would be reduced to a much smaller number of vertices.

Finally, if we continue to detect the grid-like signals with the RSA depending on viewing direction but not with our novel analysis depending of spatial location, then it might be possible that we need stronger signals for our analysis to work, as well. This could be achieved if we let the participants move in the scanner and account for the temporal autocorrelation induced by movement by cross-validating individual trials between completely independent runs. Thus, if the current crossnobis procedure is:

$$d^2_{Crosnobis}\,(b_j^*, b_k^*) = (b_j^* - b_k^*)_{sessionA} * (b_j^* - b_k^*)^T{}_{sessionB}$$

this would become:

$$d^2_{Crosnobis}\,(b_{j_sessionA}^*, b_{k_sessionB}^*) = (b_{j_sessionA}^* - b_{k_sessionB}^*) * (b_{j_sessionA}^* - b_{k_sessionB}^*)^T$$

Once the novel analysis works on real data during spatial navigation, we also plan to test it during other tasks that are more easily done in humans compared to animals, such as non-spatial tasks, which are often discrete rather than continuous, or see how these functional gradients relate to hierarchies of memories and abstract concepts. On a more computational note, we could also adapt our novel method to include principal component analysis (PCA) because a computational model has shown that the output of PCA on the

place cell – by – place cell covariance matrix are eigenvectors with hexagonal symmetry (Dordek et al., 2016).

Chapter 10: General discussion and future directions

This thesis investigated how the human brain represents abstract concepts and physical space into cognitive maps. Using functional magnetic resonance imaging (fMRI) and computational analyses, we found markers of grid cells and place cells that provide evidence of close relationships between spatial and non-spatial processing. This work has implications for understanding the mechanisms of advanced cognition in healthy humans and how they go awry during psychiatric disorders.

10.1 Developing new techniques for measuring grid-like and place-like representations in humans

In Chapters 6 and 7, we developed a battery of behavioural tasks that allowed humans to learn and navigate a non-spatial, conceptual environment, in a manner similar to spatial navigation. We could then measure neural activity with fMRI while participants navigated in this conceptual environment, and found the same grid-cell mechanism that is well-understood in the spatial domain. According to these results, individual grid cells recorded in rats in a two-dimensional spatial environment are also active during a one-dimensional sound frequency task (Aronov et al., 2017).

In Chapter 7, the hexagonally symmetric signal detected as a function of movement direction is predicted at the voxel level because grid cells have a dominant grid orientation (see section 3.2.2). In detail, early rodent studies showed that co-localized grid cells have largely the same grid orientation (Barry et al., 2007). Subsequent studies characterized in detail the grid angles at the level of individual entorhinal modules and found that the grid orientation can be mixed. (Stensola et al., 2012; Krupic et al., 2015; Stensola et al. 2015). However, grid orientation was more similar between cells with similar grid scale, and grid

cells with small scales outnumber those with large scales. Thus, these high frequency grid cells should dominate the signal at the neural population level, such that within a voxel there will be a preferred grid orientation.

In Chapters 8 and 9, we aimed to develop a novel analysis method that has the potential to detect gradients in place-cell and grid-cell coding, as shown by our simulations. This method required a big data approach, with a rich experimental design and high-resolution MRI data. Thus, we developed a novel experimental task that involved teleportation in the spatial domain and showed that participants could learn this task precisely. We also developed protocols and adapted existing pipelines to work on big data. We have preliminary evidence of a gradient across the long axis of the hippocampus, suggestive of the place cell gradients found in rodents. We also show preliminary evidence of grid-cell coding.

Using fMRI comes with the benefit of covering the whole brain. Thus, we found in Chapter 7 that grid-cell coding exists in a default-mode network, the same set of brain regions where this coding pattern and individually recorded cells were found during spatial navigation (Doeller et al., 2010; Jacobs et al., 2013). Although preliminary and with a more restricted brain coverage that did not include the prefrontal cortex, we replicated some of these brain regions using independent analyses in Chapter 9. Inter-species differences in anatomy might explain why animals use different cognitive strategies for simpler tasks, and why grid cells were not found outside of the hippocampal formation. This is interesting because the default-mode network is also used in humans during other higher cognition functions, such as valuation, memory and imagination, but also during restful mind wondering. These results suggest that, in humans, grid cells have evolved for organizing advanced cognition, which might be consolidated and updated during rest by replay and preplay mechanisms.

10.2 Open questions

Our work raises some open questions that have broad implications for memory and cognition that is not necessarily spatial.

How do place cells and grid cells organize information hierarchically? Humans have a remarkable ability to organize information hierarchically, and this could be tested easily at the behavioural level. However, the main challenge is on the technical and analysis side. Results from sensory cortices suggest that, technically, this is possible (see Chapter 4). Therefore, the next step is to improve the methods presented in Chapters 8-9 and investigate at the mesoscale level gradients of place and grid cell ensembles, which have been proposed to have a role in representing hierarchies.

How do grid cells code for higher dimensions? Computational simulations show that grid cells represent the dominant modes of variance in the place cell data using a method that could generalize over any number and type of dimensions (Dordek et al., 2016). This suggests that grid cells might represent very high dimensional worlds that are not necessarily spatial.

How do place and grid cells form predictive cognitive maps? Theoretical accounts show that many disparate properties of hippocampal place cells in rodent spatial tasks can be explained by the proposal that place cells encode predictions of future spatial states using the successor representation (Stachenfeld et al., 2016), which value each state depending on how close it is to a reward and how valuable the reward is. The successor representation can also account for fMRI representations in the hippocampal formation when humans navigate a discrete, non-spatial task (Garvert et al., 2017).

Overall, our experiments open the possibility to test in humans for novel roles of mechanisms discovered in animals, inspired by theoretical and computational work, with implications for understanding cognition during health and disease.

References

Aggleton JP (2012) Multiple anatomical systems embedded within the primate medial temporal lobe: implications for hippocampal function. Neurosci Biobehav Rev 36:1579–1596.

Aggleton JP, Wright NF, Vann SD, Saunders RC (2012) Medial temporal lobe projections to the retrosplenial cortex of the macaque monkey. Hippocampus 22:1883–1900.

Aghajan ZM, Acharya L, Moore JJ, Cushman JD, Vuong C, Mehta MR (2015) Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. Nat Neurosci 18:121–128.

Alme CB, Miao C, Jezek K, Treves A, Moser EI, Moser M-B (2014) Place cells in the hippocampus: eleven maps for eleven rooms. Proc Natl Acad Sci USA 111:18428–18435.

Amaral DG, Witter MP (1989) The three-dimensional organization of the hippocampal formation: a review of anatomical data. Neuroscience 31:571–591.

Aronov D, Nevers R, Tank DW (2017) Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. Nature 543:719–722.

Aronov D, Tank DW (2014) Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. Neuron 84:442–456.

Barron HC, Dolan RJ, Behrens TEJ (2013a) Online evaluation of novel choices by simultaneous representation of multiple memories. Nat Neurosci 16:1492–1498.

Barron HC, Garvert MM, Behrens TEJ (2016) Repetition suppression: a means to index neural representations using BOLD? Philos Trans R Soc Lond, B, Biol Sci 371:20150355.

Barry C, Burgess N (2014) Neural mechanisms of self-location. Curr Biol 24:R330–R339.

Barry C, Hayman R, Burgess N, Jeffery KJ (2007) Experience-dependent rescaling of entorhinal grids. Nat Neurosci 10:682–684.

Barry C, Lever C, Hayman R, Hartley T, Burton S, O'Keefe J, Jeffery K, Burgess N (2006) The boundary vector cell model of place cell firing and spatial memory. Rev Neurosci 17:71–97.

Beckmann M, Johansen-Berg H, Rushworth MFS (2009) Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. J Neurosci 29:1175–1190.

Behrens TEJ, Hunt LT, Rushworth MFS (2009) The computation of social behavior. Science 324:1160–1164.

Behrens TEJ, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CAM, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM, Matthews PM (2003) Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat Neurosci 6:750–757.

Bellmund JL, Deuker L, Navarro Schröder T, Doeller CF (2016a) Grid-cell representations in mental simulation. Elife 5:12897.

Benoit RG, Szpunar KK, Schacter DL (2014) Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. Proc Natl Acad Sci USA 111:16550–16555.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex 19:2767–2796.

Boccara CN, Sargolini F, Thoresen VH, Solstad T, Witter MP, Moser EI, Moser M-B (2010) Grid cells in pre- and parasubiculum. Nat Neurosci 13:987–994.

Bonnevie T, Dunn B, Fyhn M, Hafting T, Derdikman D, Kubie JL, Roudi Y, Moser EI, Moser M-B (2013) Grid cells require excitatory drive from the hippocampus. Nat Neurosci 16:309–317.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Brandon MP, Koenig J, Leutgeb JK, Leutgeb S (2014) New and distinct hippocampal place codes are generated in a new environment during septal inactivation. Neuron 82:789–796.

Brecht M, Ray S, Burgalossi A, Tang Q, Schmidt H, Naumann R (2014) An isomorphic mapping hypothesis of the grid representation. Philos Trans R Soc Lond, B, Biol Sci 369:20120521–20120521.

Brown TI, Carr VA, LaRocque KF, Favila SE, Gordon AM, Bowles B, Bailenson JN, Wagner AD (2016) Prospective representation of navigational goals in the human hippocampus. Science 352:1323–1326.

Brun VH, Leutgeb S, Wu H-Q, Schwarcz R, Witter MP, Moser EI, Moser M-B (2008a) Impaired spatial representation in CA1 after lesion of direct input from entorhinal cortex. Neuron 57:290–302.

Brun VH, Otnass MK, Molden S, Steffenach H-A, Witter MP, Moser M-B, Moser EI (2002) Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. Science 296:2243–2246.

Brun VH, Solstad T, Kjelstrup KB, Fyhn M, Witter MP, Moser EI, Moser M-B (2008b) Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. Hasselmo ME, Moser EI, Moser M-B, eds. Hippocampus 18:1200–1212.

Brunec IK, Bellana B, Ozubko JD, Man V, Robin J, Zhong-Xu L, Grady C, Rosenbaum SR, Winocur G, Barense MD, Moscovitch M (2017) Differential spatiotemporal representations along the hippocampal long axis in humans. bioRxiv.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci 1124:1–38.

Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. Proc Natl Acad Sci USA 108:11262–11267.

Burak Y, Fiete IR (2009) Accurate path integration in continuous attractor network models of grid cells. Sporns O, ed. PLoS Comput Biol 5:e1000291.

Burgess N (2014) The 2014 Nobel Prize in Physiology or Medicine: a spatial model for cognitive neuroscience. Neuron 84:1120–1125.

Bush D, Barry C, Manson D, Burgess N (2015) Using Grid Cells for Navigation. Neuron 87:507–520.

Buzsáki G, Leung LW, Vanderwolf CH (1983) Cellular bases of hippocampal EEG in the behaving rat. Brain Res 287:139–171.

Buzsáki G, Moser EI (2013) Memory, navigation and theta rhythm in the hippocampal-entorhinal system. Nat Neurosci 16:130–138.

Cacucci F, Lever C, Wills TJ, Burgess N, O'Keefe J (2004) Theta-modulated place-by-direction cells in the hippocampal formation in the rat. J Neurosci 24:8265–8277.

Canto CB, Wouterlood FG, Witter MP (2008) What does the anatomical organization of the entorhinal cortex tell us? Neural Plast 2008:381243–18.

Cao Q, Changlin M, Moser EI, Moser M-B (2016) Spatially periodic firing in grid cells requires local inhibition through parvalbumin interneurons.

Carmichael ST, Price JL (1995) Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol 363:642–664.

Carr MF, Jadhav SP, Frank LM (2011) Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nat Neurosci 14:147–153.

Chang L, Tsao DY (2017) The Code for Facial Identity in the Primate Brain. Cell 169:1013–1028.e1014.

Chen JL, Andermann ML, Keck T, Xu N-L, Ziv Y (2013) Imaging neuronal populations in behaving rodents: paradigms for studying neural circuits underlying behavior in the mammalian cortex. J Neurosci 33:17631–17640.

Cho Z-H, Han J-Y, Hwang S-I, Kim D-S, Kim K-N, Kim N-B, Kim SJ, Chi J-G, Park C-W, Kim Y-B (2010) Quantitative analysis of the hippocampus using images obtained from 7.0 T MRI. Neuroimage 49:2134–2140.

Clithero JA, Rangel A (2014) Informatic parcellation of the network involved in the computation of subjective value. Soc Cogn Affect Neurosci 9:1289–1302.

Collin SHP, Milivojevic B, Doeller CF (2015) Memory hierarchies map onto the hippocampal long axis in humans. Nat Neurosci 18:1562–1564.

Constantinescu AO, Behrens TEJ (2016) F1000 recommendation for "Extracting grid-cell characteristics from place cell inputs using non-negative principal component analysis." wwwf1000com.

Constantinescu AO, O'Reilly JX, Behrens TEJ (2016a) Organizing conceptual knowledge in humans with a gridlike code. Science 352:1464–1468.

Constantinescu AO, OReilly JX, Behrens TEJ (2016d) Supplementary video - Organizing conceptual knowledge with a grid-like code. Science 352:1464–1468 Available at: http://science.sciencemag.org/content/sci/suppl/2016/06/15/352.6292.1464.DC1/aaf0941s1.mov.

Couey JJ, Witoelar A, Zhang S-J, Zheng K, Ye J, Dunn B, Czajkowski R, Moser M-B, Moser EI, Roudi Y, Witter MP (2013) Recurrent inhibitory circuitry as a mechanism for grid formation. Nat Neurosci 16:318–324.

Craddock RC, Jbabdi S, Yan C-G, Vogelstein JT, Castellanos FX, Di Martino A, Kelly C, Heberlein K, Colcombe S, Milham MP (2013) Imaging human connectomes at the macroscale. Nat Methods 10:524–539.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

Davidson TJ, Kloosterman F, Wilson MA (2009) Hippocampal replay of extended experience. Neuron 63:497–507.

Davis T, Poldrack RA (2014) Quantifying the internal structure of categories using a neural typicality measure. Cereb Cortex 24:1720–1737.

de Lavilléon G, Lacroix MM, Rondi-Reig L, Benchenane K (2015) Explicit memory creation during sleep demonstrates a causal role of place cells in navigation. Nat Neurosci 18:493–495.

De Martino F, Moerel M, Ugurbil K, Goebel R, Yacoub E, Formisano E (2015a) Frequency preference and attention effects across cortical depths in the human primary auditory cortex. Proc Natl Acad Sci USA 112:16036–16041.

De Martino F, Moerel M, Xu J, van de Moortele P-F, Ugurbil K, Goebel R, Yacoub E, Formisano E (2015b) High-Resolution Mapping of Myeloarchitecture In Vivo: Localization of Auditory Areas in the Human Brain. Cereb Cortex 25:3394–3405.

De Martino F, Yacoub E, Kemper V, Moerel M, Uludag K, De Weerd P, Ugurbil K, Goebel R, Formisano E (2017) The impact of ultra-high field MRI on cognitive and computational neuroimaging. Neuroimage.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol 84:3072–3077.

Derdikman D, Knierim JJ (2014) Space, Time and Memory in the Hippocampal Formation. Springer.

Derdikman D, Moser EI (2010a) A manifold of spatial maps in the brain. Trends Cogn Sci (Regul Ed) 14:561–569.

Derdikman D, Moser M-B (2010b) A dual role for hippocampal replay. Neuron 65:582–584.

Diba K, Buzsáki G (2007) Forward and reverse hippocampal place-cell sequences during ripples. Nat Neurosci 10:1241–1242.

Dickson CT, Mena AR, Alonso A (1997) Electroresponsiveness of medial entorhinal cortex layer III neurons in vitro. Neuroscience 81:937–950.

Diedrichsen J, Kriegeskorte N (2017) Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. Cichy R, ed. PLoS Comput Biol 13:e1005508.

Doeller CF, Barry C, Burgess N (2010) Evidence for grid cells in a human memory network. Nature 463:657–661.

Doeller CF, Barry C, Burgess N (2012) From cells to systems: grids and boundaries in spatial memory. Neuroscientist 18:556–566.

Doeller CF, Burgess N (2008) Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. Proc Natl Acad Sci USA 105:5909–5914.

Doeller CF, King JA, Burgess N (2008a) Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. Proc Natl Acad Sci USA 105:5915–5920.

Doeller CF, King JA, Burgess N (2008b) Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. Proc Natl Acad Sci USA 105:5915–5920.

Dombeck DA, Harvey CD, Tian L, Looger LL, Tank DW (2010) Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. Nat Neurosci 13:1433–1440.

Domnisoru C, Kinkhabwala AA, Tank DW (2013) Membrane potential dynamics of grid cells. 495:199–204.

Donato F, Jacobsen RI, Moser M-B, Moser EI (2017) Stellate cells drive maturation of the entorhinal-hippocampal circuit. Science 355:eaai8178.

Dordek Y, Soudry D, Meir R, Derdikman D (2016) Extracting grid cell characteristics

from place cell inputs using non-negative principal component analysis. Elife 5:682.

Dragoi G, Tonegawa S (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. Nature 469:397–401.

Dupret D, O'Neill J, Pleydell-Bouverie B, Csicsvari J (2010) The reorganization and reactivation of hippocampal maps predict spatial memory performance. Nat Neurosci 13:995–1002.

Eichenbaum H (2014) Time cells in the hippocampus: a new dimension for mapping memories. Nat Rev Neurosci 15:732–744.

Eichenbaum H, Cohen NJ (2014) Can we reconcile the declarative memory and spatial navigation views on hippocampal function? Neuron 83:764–770.

Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003) Cellular networks underlying human spatial navigation. Nature 425:184–188.

Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci (Regul Ed) 12:388–396.

Etzel JA, Zacks JM, Braver TS (2013) Searchlight analysis: promise, pitfalls, and potential. Neuroimage 78:261–269.

Euston DR, Tatsuno M, McNaughton BL (2007) Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. Science 318:1147–1150.

Fiete IR, Burak Y, Brookings T (2008) What grid cells convey about rat location. J Neurosci 28:6858–6871.

Finkelstein A, Derdikman D, Rubin A, Foerster JN, Las L, Ulanovsky N (2015) Three-dimensional head-direction coding in the bat brain. Nature 517:159–164.

Fischl B et al. (2009) Predicting the location of entorhinal cortex from MRI. Neuroimage 47:8–17.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195–207.

Foster DJ, Wilson MA (2006) Reverse replay of behavioural sequences in hippocampal place cells during the awake state. 440:680–683.

Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8:700–711.

Fyhn M, Hafting T, Treves A, Moser M-B, Moser EI (2007) Hippocampal remapping and grid realignment in entorhinal cortex. Nature 446:190–194.

Fyhn M, Hafting T, Witter MP, Moser EI, Moser M-B (2008) Grid cells in mice. Hasselmo ME, Moser EI, Moser M-B, eds. Hippocampus 18:1230–1238.

Fyhn M, Molden S, Witter MP, Moser EI, Moser M-B (2004) Spatial representation in

the entorhinal cortex. Science 305:1258–1264.

Garvert MM, Dolan RJ, Behrens TE (2017) A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. Elife 6:1696.

Geva-Sagiv M, Las L, Yovel Y, Ulanovsky N (2015) Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation. Nat Rev Neurosci 16:94–108.

Goense J, Bohraus Y, Logothetis NK (2016) fMRI at High Spatial Resolution: Implications for BOLD-Models. Front Comput Neurosci 10:66.

Goense JBM, Logothetis NK (2006) Laminar specificity in monkey V1 using high-resolution SE-fMRI. Magn Reson Imaging 24:381–392.

Goense JBM, Logothetis NK (2008) Neurophysiology of the BOLD fMRI signal in awake monkeys. Curr Biol 18:631–640.

Grieves RM, Jeffery KJ (2017) The representation of space in the brain. Behav Processes 135:113–131.

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM (2014) ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage 95:232–247.

Gu Y, Kinkhabwala AA, Domnisoru C, Gauthier J, Tank DW (2017) Micro-organization of grid cells in layer II of the medial entorhinal cortex. Available at: www.cosyne.org.

Haacke EM, Bornw RW, Thompson MR, Venkatesan R (1999) *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley-Liss.

Haber SN, Behrens TEJ (2014) The neural network underlying incentive-based learning: implications for interpreting circuit disruptions in psychiatric disorders. Neuron 83:1019–1039.

Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. 436:801–806.

Haider B, Duque A, Hasenstaub AR, McCormick DA (2006) Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. J Neurosci 26:4535–4545.

Hales JB, Schlesiger MI, Leutgeb JK, Squire LR, Leutgeb S, Clark RE (2014) Medial entorhinal cortex lesions only partially disrupt hippocampal place cells and hippocampus-dependent place memory. Cell Rep 9:893–901.

Hall CN, Reynell C, Gesslein B, Hamilton NB, Mishra A, Sutherland BA, O'Farrell FM, Buchan AM, Lauritzen M, Attwell D (2014) Capillary pericytes regulate cerebral blood flow in health and disease. Nature 508:55–60.

Hamam BN, Kennedy TE, Alonso A, Amaral DG (2000) Morphological and electrophysiological characteristics of layer V neurons of the rat medial entorhinal cortex. J Comp Neurol 418:457–472.

Hardcastle K, Maheswaranathan N, Ganguli S, Giocomo LM (2017) A Multiplexed, Heterogeneous, and Adaptive Code for Navigation in Medial Entorhinal Cortex. Neuron 94:375–387.e377.

Harel N, Lin J, Moeller S, Ugurbil K, Yacoub E (2006) Combined imaging-histological study of cortical laminar specificity of fMRI signals. Neuroimage 29:879–887.

Hassabis D, Kumaran D, Vann SD, Maguire EA (2007) Patients with hippocampal amnesia cannot imagine new experiences. Proceedings of the National Academy of Sciences 104:1726–1731.

Hassabis D, Maguire EA (2007) Deconstructing episodic memory with construction. Trends Cogn Sci (Regul Ed) 11:299–306.

Haxby JV (2012) Multivariate pattern analysis of fMRI: the early beginnings. Neuroimage 62:852–855.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

Haynes J-D, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691.

Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523–534.

Heidemann RM, Ivanov D, Trampel R, Fasano F, Meyer H, Pfeuffer J, Turner R (2012) Isotropic submillimeter fMRI in the human brain at 7 T: combining reduced field-of-view imaging and partially parallel acquisitions. Magn Reson Med 68:1506–1516.

Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsáki G (2000) Intracellular features predicted by extracellular recordings in the hippocampus in vivo. J Neurophysiol 84:390–400.

Hevner RF, Wong-Riley MT (1992) Entorhinal cortex of the human, monkey, and rat: metabolic map as revealed by cytochrome oxidase. J Comp Neurol 326:451–469.

Heys JG, Rangarajan KV, Dombeck DA (2014a) The functional micro-organization of grid cells revealed by cellular-resolution imaging. Neuron 84:1079–1090.

Heys JG, Rangarajan KV, Dombeck DA (2014b) The functional micro-organization of grid cells revealed by cellular-resolution imaging. Neuron 84:1079–1090.

Hodgetts CJ, Voets NL, Thomas AG, Clare S, Lawrence AD, Graham KS (2017) Ultra-High-Field fMRI Reveals a Role for the Subiculum in Scene Perceptual Discrimination. J Neurosci 37:3150–3159.

Hoffmann MB, Stadler J, Kanowski M, Speck O (2009) Retinotopic mapping of the human visual cortex at a magnetic field strength of 7T. Clin Neurophysiol 120:108–116.

Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y (2013) Neural decoding of visual imagery during sleep. Science 340:639–642.

Horner AJ, Bisby JA, Bush D, Lin W-J, Burgess N (2015) Evidence for holistic episodic recollection via hippocampal pattern completion. Nat Commun 6:7462.

Horner AJ, Bisby JA, Zotow E, Bush D, Burgess N (2016) Grid-like Processing of Imagined Navigation. Curr Biol 26:842–847.

Howard JD, Gottfried JA, Tobler PN, Kahnt T (2015) Identity-specific coding of future rewards in the human orbitofrontal cortex. Proc Natl Acad Sci USA 112:5195–5200.

Huettel SA, Song AW, McCarthy G (2004) Functional magnetic resonance imaging. Sinauer Associates Sunderland.

HULL CL (1950) Behavior postulates and corollaries--1949. Psychol Rev 57:173–180.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. 532:453–458.

Insausti R, Amaral DG (2008) Entorhinal cortex of the monkey: IV. Topographical and laminar organization of cortical afferents. J Comp Neurol 509:608–641.

Jacob P-Y, Casali G, Spieser L, Page H, Overington D, Jeffery K (2017) An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. Nat Neurosci 20:173–175.

Jacobs J, Weidemann CT, Miller JF, Solway A, Burke JF, Wei X-X, Suthana N, Sperling MR, Sharan AD, Fried I, Kahana MJ (2013) Direct recordings of grid-like neuronal activity in human spatial navigation. Nat Neurosci 16:1188–1190.

Jbabdi S, Lehman JF, Haber SN, Behrens TE (2013) Human and monkey ventral prefrontal fibers use the same organizational principles to reach their targets: tracing versus tractography. J Neurosci 33:3190–3201.

Jenkinson M (2003) Fast, automated, N-dimensional phase-unwrapping algorithm. Magn Reson Med 49:193–197.

Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.

Jezek K, Henriksen EJ, Treves A, Moser EI, Moser M-B (2011) Theta-paced flickering between place-cell maps in the hippocampus. Nature 478:246–249.

Ji D, Wilson MA (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat Neurosci 10:100–107.

Jocham G, Brodersen KH, Constantinescu AO, Kahn MC, Ianni AM, Walton ME, Rushworth MFS, Behrens TEJ (2016) Reward-Guided Learning with and without Causal Attribution. Neuron 90:177–190.

Johansen-Berg H, Behrens TEJ (2006) Just pretty pictures? What diffusion tractography can add in clinical neuroscience. Curr Opin Neurol 19:379–385.

Johnson A, Redish AD (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. J Neurosci 27:12176–12189.

Jung MW, Wiener SI, McNaughton BL (1994) Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. J Neurosci 14:7347–7356.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kaplan R, Schuck NW, Doeller CF (2017) The Role of Mental Maps in Decision-Making. Trends Neurosci 40:256–259.

Kay K, Sosa M, Chung JE, Karlsson MP, Larkin MC, Frank LM (2016) A hippocampal network for spatial coding during immobility and sleep. Nature 531:185–190.

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352–355.

Kemp JM, Powell TP (1971) The structure of the caudate nucleus of the cat: light and electron microscopy. Philos Trans R Soc Lond, B, Biol Sci 262:383–401.

Kikkert S, Kolasinski J, Jbabdi S, Tracey I, Beckmann CF, Johansen-Berg H, Makin TR (2016) Revealing the neural fingerprints of a missing hand. Elife 5:649.

Kisler K, Nelson AR, Montagne A, Zlokovic BV (2017) Cerebral blood flow regulation and neurovascular dysfunction in Alzheimer disease. Nat Rev Neurosci 12:723.

Kitamura T, Ogawa SK, Roy DS, Okuyama T, Morrissey MD, Smith LM, Redondo RL, Tonegawa S (2017) Engrams and circuits crucial for systems consolidation of a memory. Science 356:73–78.

Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser M-B (2008) Finite scale of spatial representation in the hippocampus. Science 321:140–143.

Klein-Flugge MC, Hunt LT, Bach DR, Dolan RJ, Behrens TEJ (2011) Dissociable reward and timing signals in human midbrain and ventral striatum. Neuron 72:654-664.

Kobayashi Y, Amaral DG (2003) Macaque monkey retrosplenial cortex: II. Cortical afferents. J Comp Neurol 466:48–79.

Koenig J, Linder AN, Leutgeb JK, Leutgeb S (2011) The spatial periodicity of grid cells is not sustained during reduced theta oscillations. Science 332:592–595.

Kolasinski J, Makin TR, Logan JP, Jbabdi S, Clare S, Stagg CJ, Johansen-Berg H (2016) Perceptually relevant remapping of human somatotopy in 24 hours. Elife 5:3025.

Kondo H, Witter MP (2014) Topographic organization of orbitofrontal projections to the parahippocampal region in rats. J Comp Neurol 522:772–793.

Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D (2017) Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron 93:480–490.

Kraus BJ, Brandon MP, Robinson RJ, Connerney MA, Hasselmo ME, Eichenbaum H (2015) During Running in Place, Grid Cells Integrate Elapsed Time and Distance Run. Neuron 88:578–589.

Kraus BJ, Robinson RJ, White JA, Eichenbaum H, Hasselmo ME (2013) Hippocampal "time cells": time versus path integration. Neuron 78:1090–1101.

Kriegeskorte N (2015) Cross validation in brain imaging analysis. bioRxiv.

Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. Trends Cogn Sci (Regul Ed) 17:401–412.

Kriegeskorte N, Mur M, Bandettini P (2008a) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008c) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535–540.

Kropff E, Treves A (2008) The emergence of grid cells: Intelligent design or just adaptation? Hasselmo ME, Moser EI, Moser M-B, eds. Hippocampus 18:1256–1269.

Krupic J, Bauza M, Burton S, Barry C, O'Keefe J (2015) Grid cell symmetry is shaped by environmental geometry. Nature 518:232–235.

Krüger G, Kleinschmidt A, Frahm J (1996) Dynamic MRI sensitized to cerebral blood oxygenation and flow during sustained activation of human visual cortex. Magn Reson Med 35:797–800.

Kumaran D, Summerfield JJ, Hassabis D, Maguire EA (2009) Tracking the emergence of conceptual knowledge during human decision making. Neuron 63:889–901.

Kunz L, Schröder TN, Lee H, Montag C, Lachmann B, Sariyska R, Reuter M, Stirnberg R, Stöcker T, Messing-Floeter PC, Fell J, Doeller CF, Axmacher N (2015) Reduced grid-cell-like representations in adults at genetic risk for Alzheimer's disease. Science 350:430–433.

Langston RF, Ainge JA, Couey JJ, Canto CB, Bjerknes TL, Witter MP, Moser EI, Moser M-B (2010) Development of the spatial representation system in the rat. Science 328:1576–1580.

Lawrence SJD, Formisano E, Muckli L, de Lange FP (2017) Laminar fMRI: Applications for cognitive neuroscience. Neuroimage.

Lee JH, Durand R, Gradinaru V, Zhang F, Goshen I, Kim D-S, Fenno LE, Ramakrishnan C, Deisseroth K (2010) Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. 465:788–792.

Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci 23:3963–3971.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869–878.

Logothetis NK (2010) Bold claims for optogenetics. Nature 468:E3–4–discussionE4–5.

Logothetis NK, Guggenberger H, Peled S, Pauls J (1999) Functional imaging of the monkey brain. Nat Neurosci 2:555–562.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150–157.

Lovett-Barron M, Losonczy A (2013) Circuits supporting the grid. Nat Neurosci 16:255–257.

Low RJ, Gu Y, Tank DW (2014) Cellular resolution optical access to brain regions in fissures: imaging medial prefrontal cortex and grid cells in entorhinal cortex. Proc Natl Acad Sci USA 111:18739–18744.

Maass A, Berron D, Libby LA, Ranganath C, Düzel E (2015a) Functional subregions of the human entorhinal cortex. Elife 4:50.

Maass A, Schütze H, Speck O, Yonelinas A, Tempelmann C, Heinze H-J, Berron D, Cardenas-Blanco A, Brodersen KH, Stephan KE, Düzel E (2014) Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding. Nat Commun 5:5547.

MacDonald CJ, Lepage KQ, Eden UT, Eichenbaum H (2011) Hippocampal "time cells" bridge the gap in memory for discontiguous events. Neuron 71:737–749.

Mao D, Kandler S, McNaughton BL, Bonin V (2017) Sparse orthogonal population representation of spatial context in the retrosplenial cortex. Nat Commun 8:243.

Mars R, Sallet J, Rushworth MFS, Yeung N (2011) Neural basis of motivation and cognitive control. MIT Press.

Mathis A, Herz AVM, Stemmler M (2012) Optimal population codes for space: grid cells outperform place cells. Neural Comput 24:2280–2317.

McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser M-B (2006) Path integration and the neural basis of the 'cognitive map'. Nat Rev Neurosci 7:663–678.

Morris RG, Garrud P, Rawlins JN, O'Keefe J (1982) Place navigation impaired in rats with hippocampal lesions. Nature 297:681–683.

Moser EI, Roudi Y, Witter MP, Kentros C, Bonhoeffer T, Moser M-B (2014) Grid cells and cortical representation. Nat Rev Neurosci 15:466–481.

Muller RU, Kubie JL (1987) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J Neurosci 7:1951–1968.

Nakashiba T, Buhl DL, McHugh TJ, Tonegawa S (2009) Hippocampal CA3 output is crucial for ripple-associated reactivation and consolidation of memory. Neuron 62:781–787.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage 56:400–410.

Naselaris T, Olman CA, Stansbury DE, Ugurbil K, Gallant JL (2015) A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. Neuroimage 105:215–228.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63:902–915.

Nasr S, Polimeni JR, Tootell RBH (2016) Interdigitated Color- and Disparity-Selective Columns within Human Visual Cortical Areas V2 and V3. J Neurosci 36:1841–1857.

Navarro Schröder T, Haak KV, Zaragoza Jimenez NI, Beckmann CF, Doeller CF (2015) Functional topography of the human entorhinal cortex. Elife 4:252.

Navarro Schröder T, Towse BW, Burgess N, Barry C, Doeller CF (2017) Optimal decision making using grid cells under spatial uncertainty. bioRxiv.

Neubert F-X, Mars RB, Sallet J, Rushworth MFS (2015) Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. Proc Natl Acad Sci USA 112:E2695–E2704.

Nikolaou N, Meyer MP (2015) Lamination Speeds the Functional Development of Visual Circuits. Neuron 88:999–1013.

Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A toolbox for representational similarity analysis. Prlic A, ed. PLoS Comput Biol 10:e1003553.

Noonan MP, Chau B, Rushworth MF, Fellows LK (2017) Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision making in humans. J Neurosci:0692–17.

Noonan MP, Walton ME, Behrens TEJ, Sallet J, Buckley MJ, Rushworth MFS (2010)

Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proc Natl Acad Sci USA 107:20547–20552.

O'Keefe J (1976) Place units in the hippocampus of the freely moving rat. Exp Neurol 51:78–109.

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res 34:171–175.

O'Keefe J, Nadel L (1978) The Hippocampus as a Cognitive Map. Oxford Univ. Press.

O'Neill J, Boccara CN, Stella F, Schoenenberger P, Csicsvari J (2017) Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. Science 355:184–188.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proceedings of the National Academy of Sciences 89:5951–5955.

Oh SW et al. (2014) A mesoscale connectome of the mouse brain. Nature 508:207–214.

Olman CA, Davachi L, Inati S (2009) Distortion and signal loss in medial temporal lobe. García AV, ed. PLoS ONE 4:e8160.

Ongür D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cereb Cortex 10:206–219.

Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J (2011) A comparison of volume-based and surface-based multi-voxel pattern analysis. Neuroimage 56:593–600.

Ólafsdóttir HF, Barry C, Saleem AB, Hassabis D, Spiers HJ (2015) Hippocampal place cells construct reward related sequences through unexplored space. Elife 4:e06063.

Ólafsdóttir HF, Carpenter F, Barry C (2016a) Coordinated grid and place cell replay during rest. Nat Neurosci.

Ólafsdóttir HF, Carpenter F, Barry C, caswell (2017) Task demands predict a dynamic switch in the content of awake hippocampal replay. bioRxiv.

Parvizi J, Van Hoesen GW, Buckwalter J, Damasio A (2006) Neural connections of the posteromedial cortex in the macaque. Proceedings of the National Academy of Sciences 103:1563–1568.

Passingham RE, Stephan KE, Kötter R (2002) The anatomical basis of functional localization in the cortex. Nat Rev Neurosci 3:606–616.

Passingham RE, Wise SP (2012) The neurobiology of the prefrontal cortex. Oxford University Press.

Peelen MV, Downing PE (2007) Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. Trends Cogn Sci (Regul Ed) 11:4–5.

Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. Cereb Cortex 6:661–672.

Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP (2009) Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. Nat Neurosci 12:919–926.

Pérez-Escobar JA, Kornienko O, Latuske P, Kohler L, Allen K (2016) Visual landmarks sharpen grid cell metric and confer context specificity to neurons of the medial entorhinal cortex. Elife 5:534.

Poldrack RA (2007) Region of interest analysis for fMRI. Soc Cogn Affect Neurosci 2:67–70.

Poldrack RA, Mumford JA, Nichols TE (2011) Handbook of Functional MRI Data Analysis. Cambridge University Press.

Poser BA, Koopmans PJ, Witzel T, Wald LL, Barth M (2010) Three dimensional echo-planar imaging at 7 Tesla. Neuroimage 51:261–266.

Postelnicu G, Zollei L, Fischl B (2009) Combined volumetric and surface registration. IEEE Trans Med Imaging 28:508–522.

Quian Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10:173–185.

Quiroga RQ (2012) Concept cells: the building blocks of declarative memory functions. Nat Rev Neurosci 13:587–597.

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. 435:1102–1107.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13:669–683.

Raizada RDS, Grossberg S (2003) Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. Cereb Cortex 13:100–113.

Rajasethupathy P, Sankaran S, Marshel JH, Kim CK, Ferenczi E, Lee SY, Berndt A, Ramakrishnan C, Jaffe A, Lo M, Liston C, Deisseroth K (2015) Projections from neocortex mediate top-down control of memory retrieval. Nature 526:653–659.

Rakic P (2008) Confusing cortical columns. Proc Natl Acad Sci USA 105:12099–12100.

Ravassard P, Kees A, Willers B, Ho D, Aharoni DA, Cushman J, Aghajan ZM, Mehta MR (2013) Multisensory control of hippocampal spatiotemporal selectivity. Science 340:1342–1346.

Ray S, Brecht M (2016) Structural development and dorsoventral maturation of the

medial entorhinal cortex. Elife 5:175.

Ray S, Naumann R, Burgalossi A, Tang Q, Schmidt H, Brecht M (2014) Grid-layout and theta-modulation of layer 2 pyramidal neurons in medial entorhinal cortex. Science 343:891–896.

Remondes M, Schuman EM (2004) Role for a cortical input to hippocampal area CA1 in the consolidation of a long-term memory. Nature 431:699–703.

Richiardi J et al. (2015) BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. Science 348:1241–1244.

Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. Nature 497:585–590.

Robinson NTM, Priestley JB, Rueckemann JW, Garcia AD, Smeglin VA, Marino FA, Eichenbaum H (2017) Medial Entorhinal Cortex Selectively Supports Temporal Coding by Hippocampal Neurons. Neuron 94:677–688.e6.

Rotenberg A, Mayford M, Hawkins RD, Kandel ER, Muller RU (1996) Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. Cell 87:1351–1361.

Rowland DC, Roudi Y, Moser M-B, Moser EI (2016) Ten Years of Grid Cells. Annu Rev Neurosci 39:19–40.

Rudebeck PH, Murray EA (2014) The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. Neuron 84:1143–1156.

Rueckemann JW, Buffalo EA (2017) Neuroscience: Auditory landscape on the cognitive map. Nature 543:631–632.

Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex and reward-guided learning and decision-making. Neuron 70:1054–1069.

Saenz M, Langers DRM (2014) Tonotopic mapping of human auditory cortex. Hear Res 307:42–52.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014) Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90:449–468.

Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser M-B, Moser EI (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. Science 312:758–762.

Saxe R, Carey S, Kanwisher N (2004) Understanding other minds: linking developmental psychology and functional neuroimaging. Annu Rev Psychol 55:87–124.

Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK (2012) The future of memory: remembering, imagining, and the brain. Neuron 76:677–694.

Schapiro AC, Kustner LV, Turk-Browne NB (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. Curr Biol 22:1622–1627.

Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM (2013) Neural representations of events arise from temporal community structure. Nat Neurosci 16:486–492.

Schlesiger MI, Cannova CC, Boublil BL, Hales JB, Mankin EA, Brandon MP, Leutgeb JK, Leibold C, Leutgeb S (2015) The medial entorhinal cortex is necessary for temporal organization of hippocampal neuronal activity. Nat Neurosci 18:1123–1132.

Schmidt-Hieber C, Häusser M (2013) Cellular mechanisms of spatial navigation in the medial entorhinal cortex. Nat Neurosci 16:325–331.

Scholz J, Klein MC, Behrens TEJ, Johansen-Berg H (2009) Training induces changes in white-matter architecture. Nat Neurosci 12:1370–1371.

Schuck NW, Cai MB, Wilson RC, Niv Y (2016) Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. Neuron 91:1402–1412.

Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. 1957. American Psychiatric Publishing.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Sescousse G, Redouté J, Dreher J-C (2010) The architecture of reward value coding in the human orbitofrontal cortex. J Neurosci 30:13095–13104.

Sharp PE (1999) Complimentary roles for hippocampal versus subicular/entorhinal place cells in coding place, context, and events. Hippocampus 9:432–443.

Skaggs WE, McNaughton BL (1996) Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science 271:1870–1873.

Smaers JB, Gómez-Robles A, Parks AN, Sherwood CC (2017) Exceptional Evolutionary Expansion of Prefrontal Cortex in Great Apes and Humans. Curr Biol 27:714–720.

Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155.

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31:1487–1505.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004a) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl

1:S208–S219.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004b) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1:S208–S219.

Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, Van Essen DC, Feinberg DA, Yacoub ES, Ugurbil K (2012) Temporally-independent functional modes of spontaneous brain activity. Proc Natl Acad Sci USA 109:3131–3136.

Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TEJ, Glasser MF, Ugurbil K, Barch DM, Van Essen DC, Miller KL (2015) A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat Neurosci.

Solstad T, Boccara CN, Kropff E, Moser M-B, Moser EI (2008) Representation of geometric borders in the entorhinal cortex. Science 322:1865–1868.

Spiers HJ, Barry C (2015) Neural systems supporting navigation. Current Opinion in Behavioral Sciences 1:47–55.

Sporns O, Tononi G, Kötter R (2005) The human connectome: A structural description of the human brain. PLoS Comput Biol 1:e42.

Stachenfeld KL, Botvinick MM, Gershman SJ (2014) Design Principles of the Hippocampal Cognitive Map. NIPS Proceedings, Advances in Neural Information Processing Systems:2528–2536.

Stachenfeld KL, Botvinick MM, Gershman SJ (2016) The hippocampus as a predictive map. bioRxiv.

Stafford JM, Jarrett BR, Miranda-Dominguez O, Mills BD, Cain N, Mihalas S, Lahvis GP, Lattal KM, Mitchell SH, David SV, Fryer JD, Nigg JT, Fair DA (2014) Large-scale topology and the default mode network in the mouse connectome. Proc Natl Acad Sci USA 111:18745–18750.

Stemmler M, Mathis A, Herz AVM (2015a) Connecting multiple spatial scales to decode the population activity of grid cells. Sci Adv 1:e1500816–e1500816.

Stensola H, Stensola T, Solstad T, Frøland K, Moser M-B, Moser EI (2012) The entorhinal grid map is discretized. 492:72–78.

Stensola T, Stensola H, Moser M-B, Moser EI (2015) Shearing-induced asymmetry in entorhinal grid cells. 518:207–212.

Strange BA, Witter MP, Lein ES, Moser EI (2014) Functional organization of the hippocampal longitudinal axis. Nat Rev Neurosci 15:655–669.

Sugar J, Witter MP, van Strien NM, Cappaert NLM (2011) The retrosplenial cortex:

intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome. Front Neuroinform 5:7.

Sun C, Kitamura T, Yamamoto J, Martin J, Pignatelli M, Kitch LJ, Schnitzer MJ, Tonegawa S (2015) Distinct speed dependence of entorhinal island and ocean cells, including respective grid cells. Proc Natl Acad Sci USA 112:9466–9471.

Tang Q, Burgalossi A, Ebbesen CL, Ray S, Naumann R, Schmidt H, Spicher D, Brecht M (2014) Pyramidal and stellate cell specificity of grid and border representations in layer 2 of medial entorhinal cortex. Neuron 84:1191–1197.

Taube JS (2007) The head direction signal: origins and sensory-motor integration. Annu Rev Neurosci 30:181–207.

Taube JS, Muller RU, Ranck JB (1990a) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J Neurosci 10:420–435.

Taube JS, Muller RU, Ranck JB (1990b) Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. J Neurosci 10:436–447.

Terrazas A, Krause M, Lipa P, Gothard KM, Barnes CA, McNaughton BL (2005) Self-motion and the hippocampal spatial metric. J Neurosci 25:8085–8096.

Theysohn N, Qin S, Maderwald S, Poser BA, Theysohn JM, Ladd ME, Norris DG, Gizewski ER, Fernandez G, Tendolkar I (2013) Memory-related hippocampal activity can be measured robustly using FMRI at 7 tesla. J Neuroimaging 23:445–451.

Tolman EC (1948a) Cognitive maps in rats and men. Psychol Rev 55:189–208.

Turner R (2016) Uses, misuses, new uses and fundamental limitations of magnetic resonance imaging in cognitive science. Philos Trans R Soc Lond, B, Biol Sci 371:20150349.

Ugurbil K (2016) What is feasible with imaging human brain function and connectivity using functional magnetic resonance imaging. Philos Trans R Soc Lond, B, Biol Sci 371:20150361.

Ulanovsky N, Moss CF (2007) Hippocampal cellular and network activity in freely moving echolocating bats. Nat Neurosci 10:224–233.

Underwood E (2014) Nobel Prizes. Brain's GPS finds top honor. Science 346:149–149.

Van Horn JD, Toga AW (2014) Human neuroimaging as a “Big Data” science. Brain Imaging Behav 8:323–331.

van Strien NM, Cappaert NLM, Witter MP (2009) The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network. Nat Rev Neurosci 10:272–282.

Vann SD, Aggleton JP, Maguire EA (2009) What does the retrosplenial cortex do? Nat Rev Neurosci 10:792–802.

Vertes RP, Hoover WB, Szigeti-Buck K, Leranth C (2007) Nucleus reuniens of the midline thalamus: link between the medial prefrontal cortex and the hippocampus. Brain Res Bull 71:601–609.

Vincent JL, Patel GH, Fox MD, Snyder AZ, Baker JT, Van Essen DC, Zempel JM, Snyder LH, Corbetta M, Raichle ME (2007) Intrinsic functional architecture in the anaesthetized monkey brain. Nature 447:83–86.

Vogt BA, Vogt L, Laureys S (2006) Cytology and functionally correlated circuits of human posterior cingulate areas. Neuroimage 29:452–466.

Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J (2016) Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage 137:188–200.

Walton ME, Behrens TEJ, Buckley MJ, Rudebeck PH, Rushworth MFS (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron 65:927–939.

Walton ME, Devlin JT, Rushworth MFS (2004) Interactions between decision making and performance monitoring within prefrontal cortex. Nat Neurosci 7:1259–1265.

Wang G-Z, Belgard TG, Mao D, Chen L, Berto S, Preuss TM, Lu H, Geschwind DH, Konopka G (2015) Correspondence between Resting-State Activity and Brain Gene Expression. Neuron 88:659–666.

Weber B, Keller AL, Reichold J, Logothetis NK (2008) The microvascular system of the striate and extrastriate visual cortex of the macaque. Cereb Cortex 18:2318–2330.

Wehr M, Zador AM (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature 426:442–446.

Weiskopf N, Hutton C, Josephs O, Deichmann R (2006a) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. Neuroimage 33:493–504.

Weiskopf N, Hutton C, Josephs O, Deichmann R (2006b) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. Neuroimage 33:493–504.

Weiskopf N, Hutton C, Josephs O, Turner R, Deichmann R (2007) Optimized EPI for fMRI studies of the orbitofrontal cortex: compensation of susceptibility-induced gradients in the readout direction. MAGMA 20:39–49.

Westbrook C (2002) MRI at a glance. Blackwell Science.

Wheeler DW, White CM, Rees CL, Komendantov AO, Hamilton DJ, Ascoli GA (2015) Hippocampome.org: a knowledge base of neuron types in the rodent hippocampus. Elife 4:185.

Wikenheiser AM, Schoenbaum G (2016) Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. Nat Rev Neurosci 17:513–523.

Wills TJ, Barry C, Cacucci F (2012) The abrupt development of adult-like grid cell firing in the medial entorhinal cortex. Front Neural Circuits 6:21.

Wills TJ, Cacucci F, Burgess N, O'Keefe J (2010) Development of the hippocampal cognitive map in preweanling rats. Science 328:1573–1576.

Wilson MA, McNaughton BL (1994) Reactivation of hippocampal ensemble memories during sleep. Science 265:676–679.

Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81:267–279.

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

Winocur G, Moscovitch M, Rosenbaum RS, Sekeres M (2010) An investigation of the effects of hippocampal lesions in rats on pre- and postoperatively acquired spatial memory in a complex environment. Hippocampus 20:1350–1365.

Winter SS, Mehlman ML, Clark BJ, Taube JS (2015) Passive Transport Disrupts Grid Signals in the Parahippocampal Cortex. Curr Biol 25:2493–2502.

Wise SP (2008) Forward frontal fields: phylogeny and fundamental function. Trends Neurosci 31:599–608.

Wisse LEM, Gerritsen L, Zwanenburg JJM, Kuijf HJ, Luijten PR, Biessels GJ, Geerlings MI (2012) Subfields of the hippocampal formation at 7 T MRI: in vivo volumetric assessment. Neuroimage 61:1043–1049.

Witter MP (1993) Organization of the entorhinal-hippocampal system: a review of current anatomical data. Hippocampus 3 Spec No:33–44.

Woolrich MW, Ripley BD, Brady M, Smith SM (2001) Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14:1370–1386.

Yacoub E, Harel N, Ugurbil K (2008) High-field fMRI unveils orientation columns in humans. Proc Natl Acad Sci USA 105:10607–10612.

Yager LM, Garcia AF, Wunsch AM, Ferguson SM (2015) The ins and outs of the striatum: role in drug addiction. Neuroscience 301:529–541.

Yartsev MM, Ulanovsky N (2013) Representation of three-dimensional space in the hippocampus of flying bats. Science 340:367–372.

Yartsev MM, Witter MP, Ulanovsky N (2011) Grid cells without theta oscillations in the entorhinal cortex of bats. 479:103–107.

Yoon K, Lewallen S, Kinkhabwala AA, Tank DW, Fiete IR (2016) Grid Cell Responses in 1D Environments Assessed as Slices through a 2D Lattice. Neuron 89:1086–1099.

Yu X, Qian C, Chen D-Y, Dodd SJ, Koretsky AP (2014) Deciphering laminar-specific neural inputs with line-scanning fMRI. Nat Methods 11:55–58.

Zaldivar D, Logothetis NK, Goense J (2015) *Laminar Differences in Neural Activity During Positive and Negative Bold Conditions*.

Zeidman P, Maguire EA (2016) Anterior hippocampus: the anatomy of perception, imagination and episodic memory. Nat Rev Neurosci 17:173–182.

Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron 75:168–179.

Zhang S-J, Ye J, Miao C, Tsao A, Cerniauskas I, Ledergerber D, Moser M-B, Moser EI (2013) Optogenetic dissection of entorhinal-hippocampal functional connectivity. Science 340:1232627–1232627.

Zimmermann J, Goebel R, De Martino F, van de Moortele P-F, Feinberg D, Adriany G, Chaimow D, Shmuel A, Ugurbil K, Yacoub E (2011) Mapping the organization of axis of motion selective features in human area MT using high-field fMRI. Op de Beeck HP, ed. PLoS ONE 6:e28716.