Neural world models supporting flexible behaviour

Timothy Muller

Magdalen College

University of Oxford

Trinity Term 2018

DPhil Thesis

Abstract

A core component of intelligent behaviour is our ability to flexibly adapt to changes in the external world or internal goals. In order to achieve such flexible behaviour, it has been argued that humans and animals maintain sophisticated internal models – or cognitive maps – of the environment that monitor the state of the world and encode the relationships between objects. This thesis examines 1) how internal models may be monitored and updating in a changing world, and 2) the general principles under which internal models may be learnt, how they may be used for flexible and efficient planning, as well as how they may be generalised to new experiences.

In the first part of the thesis, we investigate the neural substrates of maintaining flexible behaviour in a changing environment. In order to maintain flexibility, the uncertainty in internal models guiding behaviour needs to be increased when the current behavioural strategy no longer accurately reflects the environment. We use functional MRI to measure an internal model representing participants' beliefs about the state of the world in medial orbitofrontal cortex. We show that changes in pupil diameter – thought to index the neuromodulator noradrenaline – predict increases in uncertainty in this internal model. Anterior cingulate cortex tracked the certainty in participants' beliefs and predicted changes in pupil diameter. Together these results provide evidence for neuromodulatory control of the flexibility of internal models in dynamic environments.

In the second part of the thesis, we investigate how internal models may be learnt, generalised to novel situations and used for flexible and efficient planning. Cells in the brain that fire when an animal is in certain spatial locations are an internal model of space. They also provide "ground truth" representations of a 2-dimensional environment allowing us to investigate general principles under which they may emerge. We show these cells can be thought of as representing the underlying structure or topology of the environment. We further show how these representations can be used for flexible and efficient planning of arbitrary structures, as well as how they may be generalised to novel situations allowing reuse of previously learnt knowledge. These results suggest the way we process space may be an instance of general coding algorithms responsible for organising knowledge of all different kinds.

Together, these studies demonstrate how neural world models may support flexible behaviours.

Acknowledgements

This DPhil could not have happened without a large community of amazing family, friends and supervisors.

First of all, thank you to my entire family scattered around the world, but in particular, Mum, Dad, Greg, Ellie, Jessie, Lula, Rona, and my incredible grandparents (this thesis is for them) – without all of your support, understanding, belief in me, and encouragement of curiosity none of this would have happened.

Thank you to Soph for somehow putting up with me, sharing the ups and the downs, for an unforgettable few years together in Oxford, and for pretending to be interested in neuroscience.

Thank you to the many people outside of the lab environment who have made this Oxford experience magical over the years. There are too many to exhaustively list but here are a few: John, Sinead, Mark, Santi, Amar, Steve, Sam; thanks to Sebas who has been an incredible friend on the journey with me from the very start and still is; Rauiridh for coding up the weirdest game in the world and late night meet ups at the KA; Harry and Kalina for being there from day one of the DPhil and sharing countless fun times; Yossi for agreeing that Magdalen is the best place to survive the zombie invasion; Daisy the deer in Magdalen for calm company while feeding her conkers in the final writing stretch. Of course thank you to my friendships outside of Oxford: my home friends, and my London gang for supporting and guiding me on endless adventures. Particular mention to Henry, my neuroscience partner in crime…

Thank you to the people at FMRIB who make it such an enjoyable place to work and scan: Sue, Mike, Jon, David, Dave and Duncan, as well as the many others that make it a great community. Thank you to the fantastically fun and smart bunch of people forming the comp neuro lab community that make this experience unforgettable. I can't believe how lucky I've been to work with these people: Alexa, Cam, Mona, Helen, Philipp, JamesB, Shirley, Yunzhe, Tom, Saad, Laurence, Kim and Zeb, and a special mention to the amazing newbies: Anna, Jacob, Veronika – I wish we could have overlapped for longer. I'm also grateful for the Christmas parties and summer fires at Jill's house, cosy cottage-based lab retreats with epically long, whiskey-fuelled science sessions, cosyne trips, hat games and cereal box games. Thank you also to Rogier for kindly helping with data collection whilst introducing me to some great beers in Nijmegen; Oiwi for attending Nando's Thursday

evening lectures with me way back; Hamed for his constant stream of jokes; Erie for inspiring me and being someone to look up to (both in default and non-default worlds); and a special mention to Emma, who I embarked on this journey with all those years ago and who has since been a constant source of friendship, laughter and insight.

I feel extremely lucky to have had the opportunity to not only share the DPhil experience, but also to work closely, with two truly great friends, James and Alon. Thank you to them for their patience, comradeship, and generosity. I've learned so much from both of them. Thank you for sharing this experience with me on all levels, in and out of work, and making the whole thing fun, even the bits that shouldn't be.

Finally I would like to thank my two incredible supervisors, Jill and Tim. I could not have asked for a better pair of neuroscience parents. I am grateful for the broad education I have received from them both, from how to code to how to think more like a scientist, via how to (sort of) ski (Figure 0). I am hugely grateful for the support and inspiration they have given me, as well as the freedom and encouragement to explore new things that I never thought I would have been able to. Being supervised by them has been a once in a lifetime experience. Thank you.



Figure 0 – My ski instructors. Jill, Tim, me, on a mountain in Utah, just before motor control failure.

Contents

1 INTRODUCTION 1

1.1 Learning, decision making and internal models in the brain 2

1.1.1 Behavioural evidence for internal models of the world 2

1.1.2 Multiple parallel decision making systems 3

1.1.3 Model-free vs. model-based decision making, states and internal models 5

1.1.4 Internal models in vmPFC/mOFC 8

1.2 Computational modelling of learning, decision making, and neural representation 12

1.2.1 Computational modelling of associative learning 12

1.2.2 Computational modelling of structural neural representations and flexible behaviours 17

1.2.3 Modelling vs. internal models 20

1.2.4 Summary 21

1.3 Maintaining flexibility in dynamic worlds 22

1.3.1 The problem of inflexibility 22

1.3.2 Behavioural flexibility and the ACC 25

1.3.3 Behavioural flexibility and the noradrenergic system 28

1.3.4 Parallels between the ACC and the NA system in behavioural flexibility 30

1.3.5 Work in this thesis related to behavioural flexibility in frontal cortex 31

1.4 Cognitive maps in the hippocampal-entorhinal system 33

1.4.1 Neural implementation of the cognitive map 34

1.4.2 Desirable properties of hippocampal-entorhinal representations 39

1.4.3 Theories for hippocampal-MTL function 47

1.4.4 Summary 52

1.4.5 Work in this thesis related to cognitive maps in the hippocampus and entorhinal cortex 53

2 NEURAL SUBSTRATES TRACKING BELIEF UNCERTAINTY 55

2.1 Abstract 55

2.2 Introduction 55

2.3 Methods **56**
2.3.1 Participants 56
2.3.2 Task 57
2.3.3 Behaviour 59
2.3.4 Bayesian learning model 60
2.3.5 fMRI acquisition and analysis 62
2.3.6 General Linear Models 63

2.4 Results **64**
2.4.1 Behaviour 64
2.4.2 Exploration and exploitation engage different neural substrates 65
2.4.3 Task variable-related activity explains differences between exploitation and exploration 65
2.4.4 Transitioning from exploitation to exploration engages specifically ACC 67
2.4.5 Increases in uncertainty of the current belief engage specifically ACC 67

2.5 Discussion **68**

3 CONTROL OF ENTROPY IN NEURAL MODELS OF ENVIRONMENTAL STATE **71**

3.1 Abstract **71**

3.2 Introduction **71**

3.3 Methods **74**
3.3.1 fMRI preprocessing and multivariate measure of representation strength 74
3.3.2 Region of interest definitions 76
3.3.3 Pupil acquisition and analysis 76
3.3.4 General Linear Models 76

3.4 Results **78**
3.4.1 Medial OFC represents probabilistic beliefs about the state of the environment 78
3.4.2 Noradrenaline as a candidate mechanism for increasing flexibility of belief representations 83
3.4.3 Anterior cingulate cortex predicts pupil change 84
3.4.4 Noteworthy analyses that didn't work (produced null results) 86

3.5 Discussion **86**

3.6 Supplementary data **89**

4 SPACE AS A STATE SPACE AND INTUITIVE PLANNING 91

4.1 Abstract 91

4.2 Introduction 91

4.3 Casting space under a state space framework 93

4.4 Grid cells as a PCA of 2D space 95

4.4.1 Covariance structure of 2D space 95

4.4.2 Grid cells as representing the topology of 2D space 97

4.4.3 Grid-like firing recovered as a PCA from real place fields 100

4.5 Intuitive planning 102

4.5.1 Global distances from local learning 102

4.5.2 Global navigation from local computation 107

4.6 Discussion 110

5 GENERALISATION OF STRUCTURAL KNOWLEDGE IN THE HIPPOCAMPAL ENTORHINAL SYSTEM 117

5.1 Abstract 117

5.2 Introduction 117

5.3 Conceptual framework 120

5.3.1 Problem: predicting the next time step in worlds with common structure 120

5.3.2 Explicit structural representations are necessary for generalisation 122

5.3.3 Structural knowledge is combined with sensory experiences in a conjunctive code 124

5.3.4 Hierarchical representations allow reuse of knowledge across space 127

5.3.5 Hebbian learning facilitates fast formation of conjunctive episodic memories 128

5.3.6 State transitions occur in the grid cells 129

5.3.7 Slow and fast learning across and within environments 130

5.3.8 Putting the components together 130

5.4 Model implementation 134

5.4.1 Agent architecture 134

5.4.2 Model learns spatial representations resembling those in the brain 137

5.4.3 One-shot learning and zero-shot inference 139

5.5 Empirical evidence for a preserved relationship between place and grid cells across environments **140**

5.5.1 Remapping experiment and experimental prediction 140

5.5.2 Preserved relationship between grid and place cells across environments 143

5.5.3 Grid and place cell remapping 144

5.5.4 Model-data correspondence 145

5.6 Discussion **146**

6 DISCUSSION **153**

6.1 Summary **153**

6.2 Discussion **154**

6.3 Conclusion **160**

7 APPENDIX **162**

7.1 Appendix 1 – Grid score and ideal grid rate map **162**

7.2 Appendix 2 – NeuroRisk **163**

8 REFERENCES **167**

1 Introduction

Our brains evolved to guide our actions to promote reproductive success. In order to act advantageously, the brain has to select appropriate actions from the vast number available. This may sounds trivial, but the world is complex and dynamic. Therefore decisions that yielded profitable outcomes in the past may no longer do so, and we constantly encounter new situations, requiring us to go beyond our previous experiences to make appropriate choices. In order to achieve reproductive success, therefore, our brains must have evolved the capacity to support flexible behaviour; that is, the ability to adapt to changes in the external world or internal goals. This kind of adaptive, flexible behaviour is supported by sophisticated internal models of the world that monitor the state of the environment and encode its structure, allowing us to make novel inferences and flexibly update plans.

This thesis investigates how the brain learns, represents, monitors, updates and uses such internal models to support flexible behaviour. More specifically, in the first part of this thesis we address the neural codes tracking whether the current internal model guiding behaviour is an accurate reflection of the environment, and how this information may be used to update or alter the use of internal models. In the second part we investigate what kinds of representations of the structure of the environment provide an efficient internal model that permits flexible choices, can be generalised to novel situations, and perform novel inferences. We use computational modelling approaches throughout, as well as functional imaging and pupillometry in the first part and some electrophysiological data analysis in the second.

In this introductory chapter, we first motivate the need for internal models by discussing early behavioural evidence for their existence, followed by discussion of a candidate neural implementation of an internal model. We then address the neural substrates of learning and decision making, highlighting how they are investigated in general. Following that we focus on the problem of maintaining flexibility in the face of a changing world, along with a few key neural systems important for such flexibility and relevant to the work in this thesis. We then turn our attention to the neural representations, information processing properties, and proposed functions of the hippocampal formation, which contains a well-characterised cognitive map – or internal model – of the world.

1.1 Learning, decision making and internal models in the brain

1.1.1 Behavioural evidence for internal models of the world

Two main schools of thought emerged in the first half of the twentieth century from the study of animal behaviour: the stimulus-response (S-R) theories and cognitive map theories. These were different theories proposed by the behaviourist psychologists that attempted to explain how with experience rats "tend to make fewer and fewer errors and to take less and less time" to reach a goal state from some start state, i.e. they learned how to perform some task. "All students agree as to the facts. They disagree, however, on theory and explanation" (Tolman, 1948).

The S-R theories proposed that instrumental behaviour reflected the emergence of an associative structure between stimuli and responses. In Thorndike's Law of Effect he proposed learning consists of the strengthening of associations between stimuli and responses by reward (Thorndike, 1911). This was exemplified by his study of cats in puzzle boxes whereby the cat had to press a lever to escape the box. Initially the cat would try various different responses until it pressed the lever and escaped the box. On subsequent trials the time to escape the box decreased as the cat learnt to press the lever. Thorndike proposed this was due to a strengthening of the association between the stimulus (box) and response (pressing the lever) by reward (escaping). This type of learning is known as instrumental conditioning. Also in this camp is classical or Pavlovian conditioning (Pavlov, 1927); the difference here being that the delivery of reward following the stimulus is not contingent on the animal's behaviour.

A second school of thought, most notably argued by Tolman, advocated the idea of an internal representation of the environment that can be flexibly used to make plans or novel inferences, such as taking shortcuts down previously untraveled routes. Tolman termed this internal representation a "cognitive map" (Tolman, 1948), which we use interchangeably with the term "internal model" in this thesis. These ideas were motivated in part by observations that could not be accounted for by the S-R theories. In his seminal paper, Tolman noted these observations fell in to five different categories (Tolman, 1948). We briefly explain what have become probably the two most well known: latent learning and spatial orientation. In latent learning experiments, animals learn the location of food in a maze environment faster if they have previously been able to explore the environment with no rewards present than rats that have not been previously exposed to the environment (Tolman and Honzik, 1930). Under the S-R theory, this should not occur since

the rats should only be able to learn from reinforcement by reward. In spatial orientation tasks, when the configuration of a maze is changed (but the food stays in the same spatial location), rats are able to take a novel route to the food location even when this involves going down a route never taken before (Figure 1.1) (Tolman et al., 1946). Similarly this is not possible under the S-R theory alone, as this new route (response) has never been reinforced.

These observations provide compelling evidence that rats learn an internal model of the structure of the maze, and can use this 'cognitive map' to support flexible behaviour, in this case to navigate to novel goal locations or to old goal locations via novel routes (Tolman, 1948).

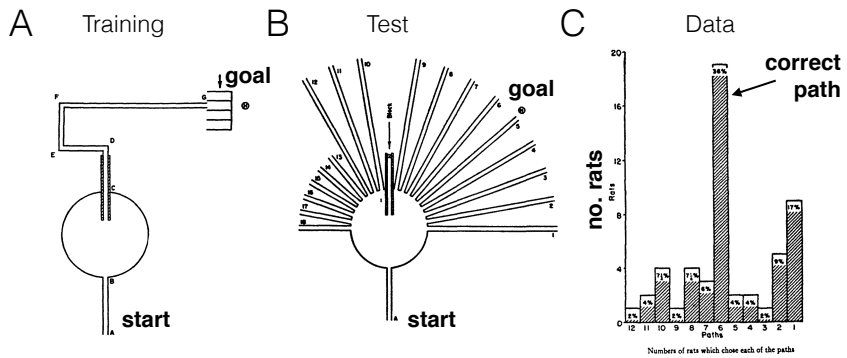


Figure 1.1 **Rats learn an internal model of the environment**. A) During training, the rats learned to navigate from the start to goal locations. B) After training, the structure of the environment was changed to the sun-burst arrangement with radial arms fanning out and the original arm blocked. The goal remained in the same place. C) The majority of the rats took the path that led to the goal (path 6), suggesting they had learnt the structure of the environment. Adapted from (Tolman, 1948).

1.1.2 Multiple parallel decision making systems

The classic behaviourists treated the brain as a black box, avoiding descriptions of observed phenomena that involved reference to specific underlying neural mechanism. Of course there is a lot to learn by investigating underlying neural substrates. For example, remarkably, despite their extremely different views on learning and decision-making, it turned out both camps of behaviourists were in some ways right.

By reviving an old place vs. response (map-based vs. S-R) learning task invented by Tolman that dissociates S-R and map-based decision making, and combining it with selective inactivation of brain regions, it was shown that these different learning and decision making processes operate in parallel in dissociable brain structures. In this task, a rat is trained to retrieve food from one arm of a cross maze. After training starting from one location, it is given a probe test in which the starting arm is at the opposite end of the maze (Figure 1.2).

The use of the response strategy (same left turn) indicates the learning was response-specific and inflexible, but use of the place strategy (right turn) shows the animal was able to use a map of the environment to flexibly select a response opposite to that which it had initially learned (Packard and McGaugh, 1996; Yin and Knowlton, 2006). After eight days of training, the majority of normal rats exhibited place responses, whereas after sixteen days of training they switched to a response strategy (Figure 1.2) (Packard and McGaugh, 1996). Therefore there is a shift in strategies controlling behaviour with extensive training.

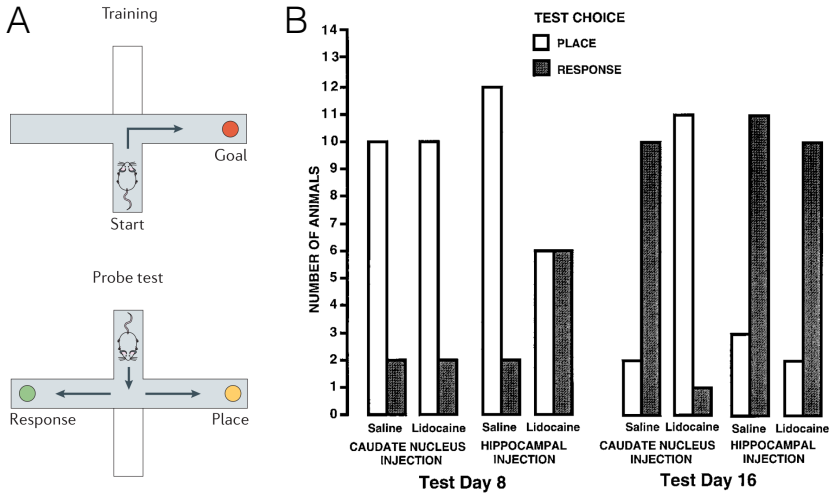


Figure 1.2 **Parallel learning systems in dissociable brain regions**. Place and response learning differentially effected by hippocampal and caudate inactivations. A) Cross maze experimental setup. B) Effects of training and hippocampal and caudate inactivation on strategy type (place or response). Detailed description of task and data is in the text. A adapted from (Yin and Knowlton, 2006); B adapted from (Packard and McGaugh, 1996).

However, inactivating (with lidocaine) either hippocampus or the caudate nucleus had dissociable effects on decision-making strategies. On day 8 probe trials, those rats with hippocampal inactivation failed to find the food altogether and those with caudate inactivation were unimpaired, exhibiting a place strategy as normal, suggesting a hippocampally mediated place but not response strategy was controlling behaviour. Rats with inactivated hippocampi on day 16, however, were unaffected, exhibiting a response strategy as normal, but those with inactivated caudate nuclei reverted back to a place strategy (Figure 1.2) (Packard and McGaugh, 1996). Hence, place and response strategies are mediated by dissociable brain regions – the hippocampus and caudate, respectively – and following a shift to a caudate-dependent response strategy, the hippocampal-based place strategy remains intact.

Although in the above we have discussed primarily spatial tasks, it should be briefly mentioned that a large body of non-spatial literature similarly found evidence for parallel learning and decision making systems. These studies revealed goal-directed (analogous to map-like) and habitual (analogous to S-R-like) behaviour in instrumental conditioning tasks (Balleine and Dickinson, 1998; Dickinson and Balleine, 1994), for example where an animal presses a lever for reward. Goal-directed behaviour is characterised by knowledge of the relationship between an action and its consequences as well as the desirability of those consequences (the outcome), whereas habitual behaviour is stamped in by past reinforcement and therefore insensitive to the current value of the outcome (Dolan and Dayan, 2013). Hence, by devaluing the outcome or degrading the stimulus-outcome contingency (both of which goal-directed but not habitual behaviour is sensitive to), it was demonstrated that both goal-directed and habitual behaviours exist in parallel (Balleine and Dickinson, 1998; Dickinson and Balleine, 1994). Importantly, this work similarly fuelled various lesion studies to localise these different behaviours to their neural substrates (Balleine and Dickinson, 1998), and was influential in the development of subsequent, more formal theories discussed in the next section.

From these various observations we can conclude that there are parallel learning and decision making systems controlling behaviour. It is worth noting that Tolman actually suggested there were multiple underlying learning systems (Tolman, 1949). He even noted that 'narrow learning' (S-R/response strategy) could be induced by 'an overdose of repetitions' (Tolman, 1948), as observed in the transition from place to response strategies with extensive training described above. The concept of multiple learning and decision making systems more broadly has proved essential throughout the study of behavioural control and learning, as well as the construction of AI systems (Daw et al., 2005; Kumaran et al., 2016; McClelland et al., 1995; Sutton and Barto, 1998).

1.1.3 Model-free vs. model-based decision making, states and internal models

Hence there is evidence of multiple learning and decision making systems in the brain. The core arguments made by the two different schools of behaviourist thought have persisted in various forms and continue to do so today (Dolan and Dayan, 2013). Rather than S-R and cognitive map theories, their descendants are now most commonly referred to as model-free and model-based, or habitual and goal-directed. More formal approaches to thinking about these different decision making strategies have helped crystallise their

various strengths and weaknesses as well as building of models to test their ability to explain behavioural and neural data (Daw et al., 2005, 2011; Sutton and Barto, 1998).

Model-free and model-based decision making strategies lie at the extreme ends of a continuum of strategies. A pure model-free decision maker attempts to effectively build a table with the values (expected future reward) it expects to receive when taking a particular action in a particular state (Daw et al., 2005; Sutton and Barto, 1998). It is therefore simply caching action values. This is contrasted with building an explicit internal model of the world, as in model-based learning. A model-based decision maker learns a model of the world – or a cognitive map – by learning a transition function, which provides a distribution over future states when a particular action is taken in a given state, i.e. $P(s'|s,a)$ where s and a denote states and actions, respectively. This can then be combined with the reward function (the expected reward at each state) to construct plans (Daw et al., 2005; Sutton and Barto, 1998). This can be done through a variety of approaches, for example tree search, in which the agent simulates paths through the environment then chooses the action that will give the highest reward.

These two approaches have opposite strengths and weaknesses. Because model-free simply caches action values, it is very computationally efficient at decision time as no search is required. However, it is inflexible to changes in the reward and transition structure of the environment: following environmental changes, it needs to relearn all the relevant values. Model-based on the other hand has the opposite pattern. Because it stores the full structure, it is entirely flexible following changes. However, this flexibility comes at the cost of having to plan through the structure, which can quickly become computationally intractable even when planning only moderately deeply. More intermediary decision making strategies (such as predictive/successor representations (Dayan, 1993)) have a more balanced set of strengths and weaknesses, trading some flexibility for computational efficiency. We explore this further in chapter 4.

We have referred to world models at various points already, although have not yet defined what they are. The definition of a world model is not exactly clear. The definition in the model-based RL literature (Sutton and Barto, 1998) is that which we gave above – $P(s'|s,a)$. This is a rather strict definition. By some definition, a retinal photoreceptor cell, or even one of its opsins, may be thought of as a rudimentary world model – some internal representation of the outside world. In this thesis, we are particularly interested in the types of world models that support flexible behaviour, and take the view that a world model should contain the necessary information to predict future outcomes or states and select

appropriate actions. Storing the environment transition dynamics ($P(s'|s,a)$) certainly satisfies this criterion; opsins certainly do not. However, there are likely other types of model that do satisfy this criterion that are not as strict as storing $P(s'|s,a)$. We therefore broaden our definition to include good state representations s.

The true underlying state is the current configuration of the world (Behrens et al., 2018). This is extremely high dimensional and impossible to represent. Agents therefore learn state representations that contain preferably only the information necessary for solving tasks, by discriminating between relevant dimensions and collapsing over irrelevant dimensions (Behrens et al., 2018). The resulting state representations should include the information necessary to predict outcomes and select appropriate actions (Wilson et al., 2014). More specifically, a good state representation should have a (likely low dimensional, latent) representation that is able to link current sensory information to policy or action information, and to (or via) internal knowledge about the structure of the task or problem at hand. A sensory encoder may transform sensory experiences into a compressed abstract representation of the aspects of the environment relevant to the current task. Embedded in the representation, therefore, is the information required (given an appropriate decoder) to generate the current sensory experiences relevant to the current task, as well as the information necessary (again given an appropriate decoder or controller) to select an optimal action under some policy to take the agent towards its goal. Furthermore, there could be a temporal component to this state representation such that transitions between states over time could occur in the latent state representation (which is efficient as would mitigate a need to, for example, generate or experience the sensory experience at each timestep when planning). The role of this temporal component is to predict the future. Therefore a good state representation will contain a compressed latent representation of relevant sensory experiences that allows a controller to select optimal actions, and may additionally contain the ability to predict future states.

Additionally, much of the relevant state information may not be directly observable from the current sensory experience, and therefore agents ought to build latent states to represent the different underlying states of the world in perceptually identical situations that are important for the current task (Wilson et al., 2014).

State representations indeed prove to be powerful models for flexibly solving tasks, even when the learning and decision making algorithms operating on them are model-free (Wilson et al., 2014), and, even when they are model-free, can exhibit behaviours that appear model-based (Akam et al., 2015). From the above definitions and discussion, we can

see how retinal photoreceptor cells clearly do not meet the criteria to qualify as a good state representation and therefore world model, whereas other representations that may not necessarily store the transition dynamics, such as compressed high-level (latent) representations, may well do. While we acknowledge the definitions here are rather loose, we hope they have broadly set the limits of the types of world model we are interested in in this thesis.

1.1.4 Internal models in vmPFC/mOFC

In this thesis we are interested in world models and their neural substrates. We focus on two loci of such world models: the orbitofrontal cortex (OFC) and the hippocampal formation. We discuss the hippocampal formation later in the introduction. Here we focus on the role of OFC in model-based decision making, specifically the medial OFC (mOFC), which we also refer to as ventromedial prefrontal cortex (vmPFC) interchangeably. We briefly highlight its importance in flexible decision making, and then turn to ideas suggesting it contains a world model in the form of state representations.

The frontal lobes have long been implicated in decision making since the famous case of Phineas Gage, who following frontal lobe damage underwent such drastic personality changes he "was no longer Gage" (Harlow, 1868). Although the extent of both his lesion and personality changes are not clear, more recent work has isolated the key role of vmPFC/mOFC. Lesions to this area in particular result in indecisiveness, odd decisions when they are made, and a lack of goal-directed behaviour, acting rather on immediate opportunities – suggestive of a lack of model-based control (Barrash et al., 2000; Bechara et al., 1994, 2000). These observations suggest this region plays a key role in value-based decision making. Indeed, a wealth of functional imaging research has implicated vmPFC/mOFC in value-guided decisions; it is engaged during both choice and feedback stages when participants are making decisions between options based on their value (Bartra et al., 2013), coding the difference between the chosen and unchosen option values at choice, i.e. the evidence in favour of selecting the option that was chosen (Boorman et al., 2009).

One remarkable feature of this brain region is the flexibility of its representations, or the identities of the options it is choosing between. In addition to representing the value of previously experienced options, it can represent and evaluate the value of imagined experiences constructed from memories (Barron et al., 2013), and accordingly is necessary for making decisions based on inferred values (Jones et al., 2012). It also combines different

types of information in order to make a decision (Behrens et al., 2008), and can even flip its frame of reference depending on what type of choice ought to be made (Nicolle et al., 2012). Such flexibility is expected from a region that represents or has access to internal models of the environment, and that represents the information relevant to solve the current task.

These observations, along with various functions ascribed to OFC, are consistent with view that OFC represent the current state in the task state space (Wilson et al., 2014), which indeed should encapsulate the relevant information for solving the task. Although there are numerous observations that motivate this view, we present two here to illustrate the point: reinforcer devaluation and reversal learning.

The hallmark of model-based choice is simulating the consequences of one's choices, and there is evidence to suggest this is dependent on the OFC (Daw et al., 2005; Izquierdo et al., 2004). A classic example of a model-based task is reinforcer devaluation. In this paradigm, animals are trained to associate cues or actions with an outcome. When this outcome is devalued outside of the task environment (e.g. by pairing its consumption with poison), selection of actions that were trained with the now devalued outcome are reduced at test time. This behaviour is evidence of the capacity to simulate the consequences of actions within some internal model of the task, which is critically reliant on an appropriate state representation. OFC lesions impair the ability to devalue the action, implying an inability to simulate its consequences (Izquierdo et al., 2004; Pickens et al., 2003), consistent with a lack of state representations (Wilson et al., 2014).

Potentially more convincing evidence comes from reversal learning paradigms. In a typical reversal learning paradigm, one of two stimuli will be probabilistically associated with reward. After animals learn this association (choose the high reward option), the contingencies are reversed such that the other stimulus is associated with reward. Both normal and OFC-lesioned animals learn initial task contingencies equally well (Butter, 1969). However, following reversals in contingencies, OFC-lesioned animals fail to adapt, taking even longer to learn the new contingencies than the original. Control animals adapt quickly, learning even faster than originally. These observations can be accounted for by OFC representing latent states of the task (Wilson et al., 2014). Control animals are able to adapt quickly since they can learn that there are two different states of the world (one for each set of contingencies) and quickly infer which is currently reflective of the true state of the environment. However, OFC-lesioned animals, unable to do this, have to unlearn their previous learning, which is why it takes them longer to adapt to reversals.

These observations, along with numerous others, are accounted for by a critical role of OFC in representing the current state in the task state space (Wilson et al., 2014) – a prediction that has recently been more explicitly tested. Humans performed a task that was defined by multiple states, which in turn were defined by not only what was currently on the screen but also the recent history. Despite the definition of the current state being divorced from the immediate sensory experience, the current location within the task state space could accurately be decoded from specifically vmPFC/mOFC using fMRI (Figure 1.3)(Schuck et al., 2016). This offers strong support for the proposition that OFC represents the current state and that it is able to flexibly represent information relevant for performing the current task.

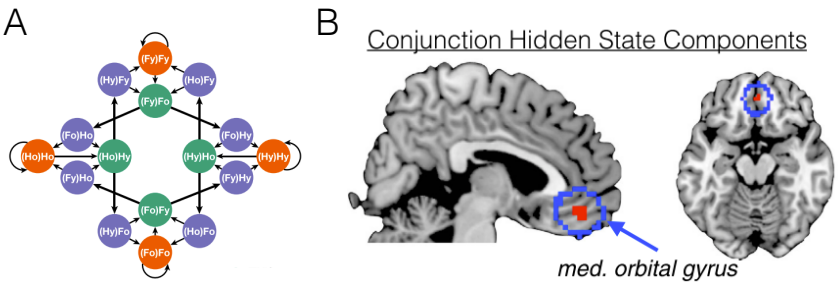


Figure 1.3 **The vmPFC/mOFC represents a cognitive map of task state space**. A) The underlying state space of a task performed by human participants. We do not describe the details of the task here, but rather illustrate the point that on any given trial the participant can be described as being in one of these sixteen states that define the task structure. B) Participants location in this state space can be accurately decoded from specifically vmPFC/mOFC using fMRI while they perform the task. Adapted from (Schuck et al., 2016).

Further evidence comes from single unit activity in rodents performing a task with multiple different blocks (defined by various task variables, such as location of reward) that each can be thought of as a different state of the task. This study demonstrated that OFC represents an abstract 'holistic' representation of the current block that cannot be explained by the sum of its specific task parts (Wikenheiser et al., 2017), as expected from an abstract state representation (Wilson et al., 2014). More specifically, when regressing out of neural activity all the features that together define a given block (location, flavour, magnitude of reward, etc.), the identity of a block can still be decoded from the residual neuronal activity in OFC. This suggests it represents something abstract about the structure of the task, i.e. the state representation of the block. Taken together, there is therefore converging

evidence that OFC represents the current location in a cognitive map, i.e. internal model, of the task space (Schuck et al., 2016; Wilson et al., 2014).

1.1.5 Summary

Hence we have seen behavioural evidence that animals construct internal models of their environment, and that these internal models exist alongside other, parallel learning and decision making systems. The extreme ends of these systems are entirely model-free and entirely model-based strategies. We have discussed what may constitute a world model, and suggested there may be softer definitions than explicitly storing all the transition dynamics. Finally we have seen how the OFC may represent the current location in an abstract task state space. Before digging in to literature that is more relevant to questions asked in this thesis, we first present a section highlighting how computational modelling in general has shed light on neural processing, in particular using the example of associative learning, and how it is beginning to offer insight in to more flexible behaviours as well as neural representation.

1.2 Computational modelling of learning, decision making, and neural representation

In this section, we look at some examples of computational modelling approaches to learning, decision making and neural representation, using them to highlight computational modelling more generally. Computational neuroscience attempts to formalise information processing and break it down in to its constituent components. A precise computational model allows one to predict animal behaviour and look in the brain for hidden variables proposed by the model, offering a more mechanistic description of neural function. Furthermore, computational descriptions are essential for simulating information processing to build artificial agents. Although this thesis is not principally about associative learning, we start by introducing some key computational models of associative learning since these offer some of the best examples of models explaining behaviour and neural activity, and more importantly, because we use these examples to bring in some key concepts, namely Bayes and model-based fMRI. We then turn to how more recent approaches are beginning to explain neural representations and flexible behaviours that are of interest to this thesis. In this section we only briefly highlight a couple of important examples to introduce key concepts, but do not dig deep in to the details.

1.2.1 Computational modelling of associative learning

Whilst we have seen that agents have multiple different strategies for making decisions, we have not discussed how they learn the appropriate information. For example, where do the values assigned to stimuli come from? We present models of candidate associative learning algorithms here. These examples are focused around value-based learning and decision making, where value is the expected reward (i.e. reward magnitude x probability). We start with the classic Rescorla-Wagner model and explore two of its generalisations that overcome some of its limitations.

1.2.1.1 Rescorla Wagner

Early formalisms of animal learning theory were put forward by Rescorla and Wagner (Rescorla and Wagner, 1972). The Rescorla-Wagner model offered an account for the learning in classical conditioning in which an animal learns the strength of association between a conditioned and unconditioned stimulus (CS and US, respectively). They key

principle is that animals update the strength of the CS-US association by learning from the errors in their prediction. The update of the CS-US association $V_{c(t)}$ of stimulus c at trial t is described by:

$$\Delta V_{c(t)} = \alpha_c \beta \left(R - \sum_s V_{s(t)} \right)$$

$$V_{c(t+1)} = V_{c(t)} + \Delta V_{c(t)}$$

where R is the reward on the current trial (positive or negative) and α and β are learning rates associated with the CS and US, respectively. The summation is to take into account the predictive strengths of all stimuli s. The prediction error is therefore the sum of the association strengths of all CS subtracted from the strength of the US (the difference between expected and received reward), and is used to update the CS-US association strength in proportion to the learning rate(s). This model proved useful in describing animal learning and was able to account for a variety of phenomena, such as blocking, wherein a CS that already fully predicts the US blocks the learning on another CS (Kamin, 1969).

The Rescorla-Wagner model has been generalised to account for some of its limitations. Specifically the Rescorla-Wagner model has two primary limitations. First, it artificially breaks time down into discrete trials and only predicts immediate reward. Real life does not consist of trials but rather occurs in continuous time or at least a temporal scale much more fine-grained than trials. Second, it is a point estimate and does not take into account uncertainty over other options. The first of these limitations is addressed by a generalisation of the Rescorla Wagner model to temporal difference (TD) learning. The second is addressed by a Bayesian treatment of updating of associations known as the Kalman filter.

1.2.1.2 Temporal difference learning and the firing of dopaminergic neurons

We briefly discuss TD learning here since it is of particular interest to neuroscientists. TD learning came from attempts to build artificial agents capable of learning from interactions with the environment in sequential decision problems, wherein the breaking down of time into discrete trials as in the Rescorla Wagner model is a particular problem (Sutton and Barto, 1998). TD learning takes into account time, since the values of the current state in TD learning are a prediction of the total discounted expected future reward V_t:

$$V_t = \mathrm{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k}\right]$$

where γ is a discount factor and r_t is the reward at time t. If we make various (Markov) assumptions that we do not go into here, we can arrive at the recursive Bellman equation:

$$V_t = r_t + \gamma \mathrm{E}[V_{t+1}]$$

From which the TD prediction error can be obtained (although we do not derive it here, for intuition, assume the above equation is true and subtract V_t from both sides – δ_t is 0 when the equation is true):

$$\delta_t = r_t + \gamma \mathrm{E}[V_{t+1}] - V_t$$

This prediction error is conceptually similar to that in the RW model. The difference is the introduction of the expectation of future reward; if the discount γ were 0, we are back at the RW model prediction error. By introducing this expectation we induce TD learning to take on prediction of long term expected future reward, not just immediate reward. It is noteworthy that the t for time in TD learning is not the same as that in RW (trials), but instead however time has been chosen to be broken up.

The reason we have been discussing TD learning is because it is one of the first and most well known demonstrations of formal descriptions offering a precise explanation of neural firing. The TD reward prediction error turns out to be a powerful descriptor of firing rates of dopaminergic neurons in the ventral tegmental area and substantia nigra pars compacta during classical conditioning. During initial learning between a CS and US, dopaminergic neurons fire to the unexpected presentation of the US. However, with learning, as the US becomes predicted by the CS, dopaminergic neurons no longer fire to the predicted US but rather to the surprising CS that now fully predicts the US (Figure 1.4). Unexpected omission of the US leads to a suppression of dopaminergic firing rates, as predicted by a negative TD reward prediction error (Figure 1.4) (Schultz et al., 1997). This striking correspondence between the formal model and neural firing has proved influential in inspiring computational approaches to interrogating neural function.

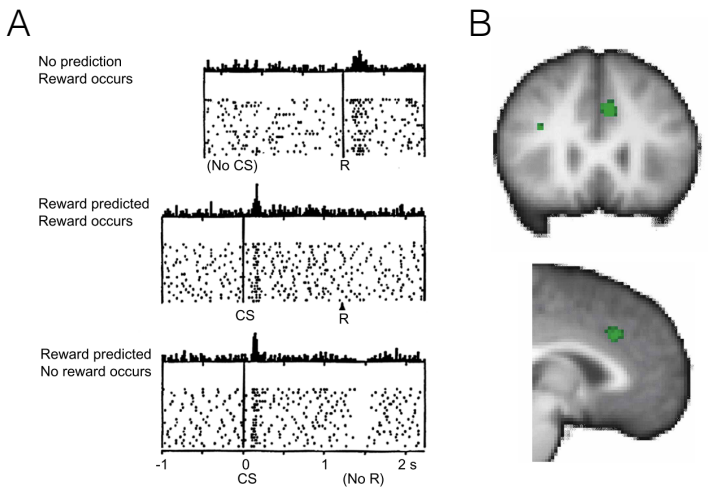


Figure 1.4 **Neural correlates of precise learning-related computational parameters**. A) The firing of dopaminergic neurons in the ventral tegmental area is predicted by the TD learning reward prediction error. Figure adapted from (Schultz et al., 1997). B) Activity in the ACC is predicted by the inferred volatility of the environment. Figure adapted from (Behrens et al., 2007).

1.2.1.3 Bayesian associative learning and neural correlates of environment stability

The other limitation of the Rescorla-Wagner model is that it does not take into account uncertainty. It is a point estimator; that is, it estimates a single value of the association strength, which is a point in the space of possible association strengths, rather than a probability distribution over all possible strengths. Representing the probability distribution allows the animal to take into account its uncertainty over the values. This uncertainty is an important piece of information as it informs how much one should learn from new observations. For example, in the Rescorla-Wagner model presented above, the learning rate is generally fixed. However, if an animal is uncertain of the true underlying state of the world, it should learn more than if it is certain. Animals should therefore flexibly adjust their learning to reflect their uncertainty in the state of the world.

In general, Bayesian learning models postulate that the learner represents beliefs and their associated uncertainty in the form of a posterior probability distribution over unknown parameters given data. The posterior is given by Bayes' rule:

$$P(unknowns|data) \propto P(data|unknowns)P(unknowns)$$

The unknowns here are the candidate parameters of the statistics of the environment (e.g. associative strengths). $P(data|unknowns)$ is the likelihood of observing the data given the parameters; $P(unknowns)$ is the prior probability of the parameters. Bayes rule specifies

how to optimally take into account uncertainty when updating beliefs upon observing new data.

Modelling the uncertainty over the belief of an associative strength permits adaptive learning rates. The Kalman filter is a Bayesian implementation of associative learning and computes the Kalman gain, which replaces the learning rate in the Rescorla Wagner equations. The Kalman gain grows with uncertainty resulting in the learner learning faster when it is uncertain. This allows it to account for various animal behavioural phenomena not explained by the Rescorla-Wagner model (Gershman, 2015).

Such a Bayesian treatment was applied to human learning in a reinforcement learning setting (Behrens et al., 2007). In this model, a probability distribution over reward is learned instead of a point estimate, allowing the learner to assess its ability to predict reward, since the width of this distribution corresponds to the amount of uncertainty of the learner. The width of this distribution determines how much the prior is updated and is analogous to the learning rate in the RW model. Sometimes the world may change faster than other times, i.e. is more volatile. In more volatile environments, we should learn faster and therefore this volatility should impact on how much our estimates of reward are updated. Accordingly, the volatility sets the width of the probability distribution over reward rates. The volatility itself is modelled as a probability distribution over volatilities; hence this model had another hierarchical layer (and in this sense can be thought of as a double Kalman filter), tracking the stability of the world.

This Bayesian approach was able to explain participants' choices better than classic reinforcement learning approaches similar to those described above, suggesting humans adapt their learning rates to the statistics of the environment (Behrens et al., 2007). Furthermore, the inferred volatility parameter predicted brain activity – as measured using functional MRI (fMRI) – in the anterior cingulate cortex, suggesting this brain region may be involved in regulating the degree to which beliefs are updated (Figure 1.4). Indeed, activity in this region predicted participants' learning rates. These results suggest this brain region may be involved in controlling flexible behaviour in the face of changing circumstances, which we discuss in more detail below. [1]

[1] To briefly complete the picture, we have discussed generalisations of Rescorla-Wagner to deal with its limitations. First to TD learning to predict expected discounted future reward to account for its limitation in the way it deals with time, and second to a Bayesian treatment to account for uncertainty. However, the TD model still does not

1.2.1.4 Model-based fMRI and localising computations

The finding that the volatility parameter predicts neural activity in ACC (Behrens et al., 2007) demonstrates that we can seek evidence for variables postulated by computational models using functional MRI. By adopting a computational approach, so called "model-based fMRI" allows us to characterise brain areas in terms of the computations they perform (O'Reilly et al., 2013a). We use model-based fMRI in chapters 2 and 3. Furthermore, rather than regressing model variables against univariate brain activity alone, we also explore whether we can explain trial by trial changes in multivariate signals, in an attempt to explain changes in the neural representation of internal models and characterise brain regions in terms of their representation.

1.2.2 Computational modelling of structural neural representations and flexible behaviours

We have presented a couple of examples of computational models that model associative learning and have variables that explain neural activity. These illustrate computational modelling in general, as well introduce Bayesian modelling and model-based fMRI, both of which we use in this thesis. We now briefly look at some more recent attempts to explain neural representations and flexible behaviours that are relevant to this thesis. We first explain how ideas from machine learning may offer insight in to the neural representations of grid cells, which we discuss further in this chapter, and then how artificial neural networks may exhibit flexible behaviours. As with the previous section, we keep these descriptions brief.

---

represent uncertainty and the Bayesian treatment is still concerned with prediction of immediate reward. There is a further generalisation to overcome both of these limitations by combining their strengths in the form of Kalman TD (Gershman, 2015), which we do not discuss here.

[2] We note here that another option is not to update the internal model *per se*, but rather disengage behaviour from it (Tervo et al., 2014).

1.2.2.1 Grid cells as a PCA of place cell activity.

We discuss grid cells in more detail later, but we briefly use them here to highlight how statistical modelling approaches have offered fresh insight in to the potential origins of their representation. Briefly, grid cells are spatial cells that have multiple firing fields that form a periodic triangular lattice over the environment (Hafting et al., 2005). Place cells on the other hand generally only have one (or a couple) of firing fields within a given environment (O'Keefe and Nadel, 1978). How place and grid cell representations depend on one another has been a topic of much interest, and, recently, ideas from machine learning have offered some insight.

Principle component analysis (PCA) is a statistical technique used to provide an efficient representation of the structure of data, and is therefore commonly used for dimensionality reduction. PCA provides a set of principle components that seek to maximise the variance of the data explained. These principle components are the eigenvectors of the covariance matrix of the data and those with larger corresponding eigenvalues explain more variance in the data. Hence PCA may be used to provide an efficient and low dimensional representation of data.

Recently it has been shown that grid cells may be understood as a PCA of place cell activity (Dordek et al., 2016). The activity of place cells as a function of time can be simulated by distributing (2D Gaussian) place cells throughout a virtual 2D environment and having an agent walk around the environment through the place fields. Performing non-negative PCA of this place cell activity by time matrix results in eigenvectors that are periodic over space, forming a triangular lattice as well as recreating various other properties of grid cells (Figure 1.5) (Dordek et al., 2016). Furthermore, these eigenvectors can be learnt in a simple neural network. This insight offers a fresh perspective on grid cells, suggesting they are an efficient representation for summarising the statistics of 2D-distributed place cells, and highlights how we may be able to gain insight in to the principles that give rise to neural representations. We explore these ideas in detail in chapters 4 and 5.

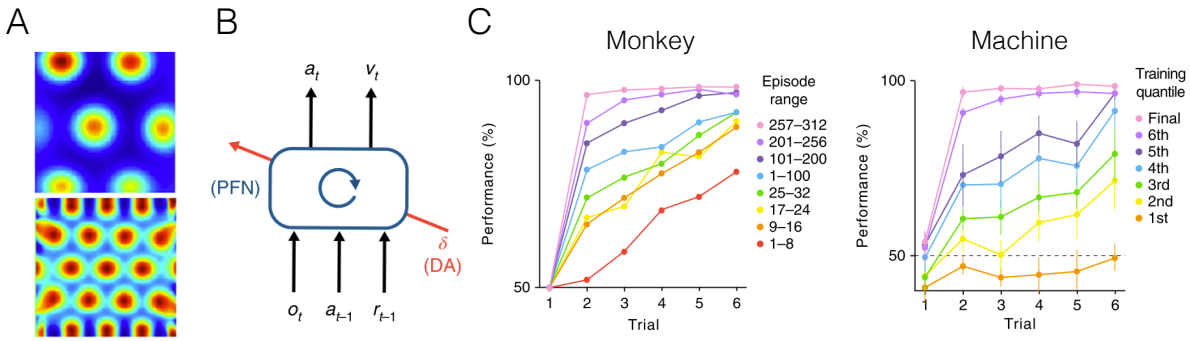


Figure 1.5 **Representations and flexible behaviours emerge from statistical models**. A) Triangular lattice patterns are extracted from a non-negative PCA of place cell activity. Firing rate map reveals grid-like firing (top), which is made more clear by in the 2D spatial autocorrelation (bottom). Figure adapted from (Dordek et al., 2016). B) Meta-RL agent architecture consists of a prefrontal (recurrent) network (PFN) with synaptic weights adjusted through an RL algorithm (driven by dopamine, DA). At each time step the agent receives past action *a* and reward *r* and outputs an action and state value estimate *v*. The learnt recurrent dynamics implement a second, fast learning algorithm that can solve the Harlow's task in one trial. C) Performance accuracy playing Harlow's task in Harlow's monkeys (Harlow, 1949) and meta-RL (Wang et al., 2018). With more training on the task, both learn to solve the task in fewer trials – eventually solving it in one when fully trained. Figure adapted from (Wang et al., 2018).

1.2.2.2 Flexible behaviour and brain-like representations in artificial neural networks

We now present a recent example demonstrating how flexible behaviours may be exhibited by artificial neural networks (ANNs). ANNs have a long history with neuroscience (Hassabis et al., 2017), all the way back to having been originally inspired by neurons in the brain. ANNs are powerful function approximators, and when combined with the reinforcement learning ideas discussed above have performed at super-human levels in a variety of tasks, including various games (Mnih et al., 2015; Silver et al., 2016). Furthermore, across a variety of domains, ANNs learn representations similar to those found in various regions of the brain when required to perform similar tasks as the corresponding brain region (e.g. (Banino et al., 2018; Kell et al., 2018; Khaligh-Razavi and Kriegeskorte, 2014; Mante et al., 2013; Singer et al., 2018; Sussillo et al., 2015; Yamins and DiCarlo, 2016)). However, ANNs do not generally exhibit the types of flexible behaviours we are interested in this thesis that humans perform effortlessly.

A key example of the types of flexible behaviour we are interested in are those arising from abstract knowledge of the structure of tasks. This is exemplified by Harlow's famous learning sets tasks (Harlow, 1949). When monkeys are tasked with choosing between two stimuli – one rewarded, one unrewarded – over multiple trials, they are able to learn to choose the rewarded stimulus over a few trials. This sounds simple, and indeed could be solved using the simple RL algorithms we mentioned above. However the key

insight of these experiments was to notice that the monkeys got better and better at performing this task with repeated experience, even though the different instantiations of the task used different stimuli each time. After extensive experience, the monkeys could learn to always pick the rewarded stimulus after observing only one trial (Figure 1.5). This ability requires some abstract representation of the task encoding the relationships between entities and events that generalises over specific stimuli.

This ability to learn how to learn, i.e. meta-learning, has recently been demonstrated in ANNs. When a recurrent neural network is supplied with the action it took and the reward it received on the previous trial, and is trained over many, many experiences using reinforcement learning (similar to TD learning mentioned above), it can learn to play Harlow's task (Wang et al., 2018). After much experience, the recurrent ANN learns weights that support activation dynamics that implement a second, fast learning algorithm that exploits the abstract structure of the task to allow the network to solve the task in one trial, like the well-trained monkeys (Figure 1.5). It is further noteworthy that when this same architecture is trained to perform the reward learning task described above in which volatility of the environment changes (Behrens et al., 2007), the ANN learns to flexibly adapt its learning rate depending on the volatility of the environment, similar to the humans (Wang et al., 2018).

This is of interest to neuroscientists for a few reasons: first, it showed how slow dopaminergic prediction error driven learning may train a more flexible and fast-learning recurrent learning algorithm in prefrontal cortex; second, it demonstrated that flexible behaviour exploiting underlying structure can be exhibited by ANNs; and third, it contained artificial neurons that resemble those found in the brain (Wang et al., 2018). In a similar spirit, in chapter 5 we explore how an ANN endowed with both slow and fast learning may be able to exhibit flexible behaviour and learn representations observed in the brain, in this case neurons in the hippocampal formation.

1.2.3 Modelling vs. internal models

Just a quick note on terminology: Decision-making may or may not involve internal models but can nonetheless be modelled. Therefore "model" has two different meanings in this thesis: 1) We can model some neural process with a computational model, and 2) A neural process itself may be a model of the external world. For example, we have used the term 'model-based' in two different contexts above. Model-based fMRI refers to looking for variables of some computational model in the brain, regardless of what this model is

modelling. Model-based decision making means that an agent is using some neural model of the external world to guide decisions. Neural models of the external world will be referred to as internal models or cognitive maps interchangeably throughout this thesis.

1.2.4 Summary

In this section we have highlighted a few key computational models. We have shown how associative learning models are able to account for animal behaviour and explain neural activity. In doing so, we have introduced Bayes and model-based fMRI that are both used in this thesis. We have also looked at how more recent statistical approaches may be able to explain the principles giving rise to neural representations and flexible behaviours that we are interested in in this thesis.

1.3 Maintaining flexibility in dynamic worlds

We have discussed the possible representation of internal models and that they can be used for decision making. However, an important consideration is how to keep these models or their use flexible in the face of a changing world. Internal models require updating if they no longer reflect the environment, or alternatively behaviour ought to be disengaged from the model. Such flexibility is the topic this section; we first highlight the problem of inflexibility, and then discuss some of the key neural systems that may be involved in maintaining flexible behaviour.

1.3.1 The problem of inflexibility

In order to make sense of their environment, organisms must infer the underlying causes or states of the world that generate the constant stream of observed sensory data. This inference is achieved by sequentially sampling data from the environment. As long as the environment is assumed to remain stable, certainty about its state increases as the number of data samples increases. This familiar notion is in fact an inevitable consequence of the mathematics of Bayesian inference – if beliefs about the environment are expressed as a posterior distribution over possible states, the entropy, or uncertainty, of this posterior decreases (tending towards a finite value) as the number of concordant data points increases (assuming there is no mechanism for driving up uncertainty over unsampled options). The entropy of a probability distribution over a discrete random variable X with possible values $\{x_1, \dots, x_n\}$ is given by:

$$H(X) = -\sum_{i}^{n} P(x_i)\log\left(P(x_i)\right)$$

which intuitively will be lower when there is more certainty, i.e. values x_i are nearer either 0 or 1.

In a dynamic environment, the tendency of uncertainty to decrease over time raises a problem of inflexibility – the better established one's model of the environment, the less flexible one's beliefs are (i.e. the less they can adapt to change). One of the simplest forms of flexibility is the ability to discard or forget old information. Therefore to counter the problem of inflexibility, theoretical accounts of inference that capture uncertainty must

include an explicit mechanism for driving up entropy – injecting uncertainty back into the model's state estimate to make rapid changes of belief possible. The simplest approach is to include a constant 'leak' term (Usher and McClelland, 2001) such that the model's belief distribution continuously decays towards a state of agnosticism (uniform belief distribution over possible states), i.e.:

$$prior_{t+1} \propto posterior_t + U(x_1, x_n)$$

where t denotes a trial number and $U(x_1, x_n)$ is a uniform distribution over the range of possible values, although there are more sophisticated approaches (Behrens et al., 2007; Nassar et al., 2012). As a simple demonstration, we can see how including such a leak allows a simple Bayesian learner to flexibly update its beliefs by discarding old information when attempting to learn the underlying mean of a noisy and dynamic data generation process (Figure 1.6). When the leak is removed, the learner learns well initially when the environment is stable but fails to accurately update its beliefs and appropriately increase its entropy when the underlying mean jumps to a new value, illustrating the importance of having some mechanism to permit flexibility.

We note here that the simple Bayesian learner discussed was not constructed to model a dynamic environment. Rather, it assumed observations were drawn from a stationary generative process. Adding the leak allowed it to forget its previous (now incorrect) inference of the mean and relearn a new mean. This is a simple way of dealing with non-stationarity, however it is not optimal. More sophisticated models can be constructed that explicitly assume a dynamic environment in the generative model. These models may allow uncertainty to increase over time even with concordant data points (for example, when exploiting a given option and not sampling the other options (Knox et al., 2012)). This results in exploration having an informational value, in order to decrease this uncertainty, and can be modelled (Knox et al., 2012). Under these more sophisticated models, the learner may be able to adapt to change without requiring something like a leak. Nonetheless, the simple demonstration mentioned above (Figure 1.6) shows that by increasing uncertainty in the model's beliefs, a learner is able to adapt to change (even when its assumptions do not explicitly assume a dynamic environment).

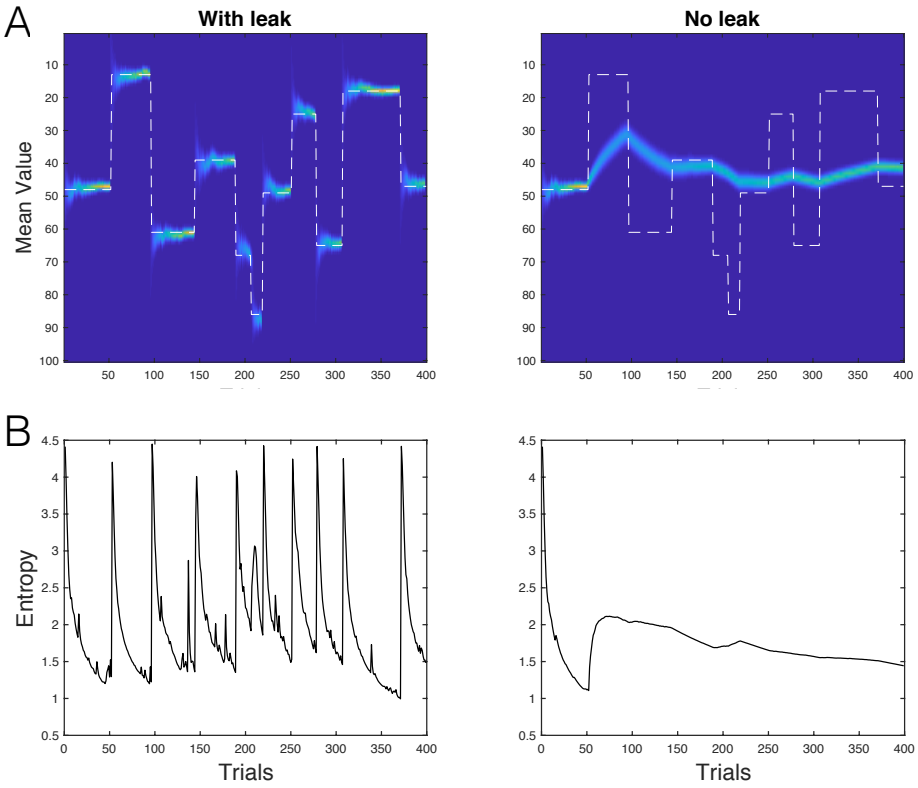


Figure 1.6 **The problem of inflexibility.** Learners need a mechanism to permit flexibility. A) A simple Bayesian model with a 'leak' that allows it to forget the past successfully tracks the mean of a noisy data generation process. However, the same Bayesian model but with the leak removed learns well initially when the environment is stable but is inflexible to changes in the environment, failing to accurately update its belief about the world. White line is the true underlying mean (which jumps randomly with a probability of 1/40 and takes a value between 0 and 100). The data observed by the learner is drawn from $N(mean, SD)$, where N is the normal distribution and $SD = 5$. The heatmap is the posterior probability distribution over mean values on each trial. B) Accordingly, the entropy of this posterior probability distribution increases appropriately following a change in the environment with the leak, but not without it.

Regardless of the exact mechanisms, the ability to adapt to change is crucial, and updating beliefs to reflect the environment is a key component of this. Brains therefore require some mechanism for flexibly updating internal models. This process requires (at least) two key components: First, it requires some mechanism for monitoring the quality of the current belief or tracking changes in the environment. Second, once the need for learning has been detected, the internal model(s) guiding decision making ought to be

updated or have their uncertainty increased to facilitate learning the new state of the world[2].

A range of evidence points to the ACC implementing computations responsible for both of these functions. However, despite this wealth of evidence, it is still not clear how uncertainty is modulated in internal models themselves (possibly represented in regions outside the ACC) – partly due to the challenge of measuring internal models representing beliefs as they are updated. We address this question in chapters 2 and 3, exploring the possibility that the neuromodulator noradrenaline mediates injection of uncertainty in to internal models. We now review relevant literature, starting with the monitoring of the quality of the current belief in ACC, followed by the possible role of the noradrenergic system.

1.3.2 Behavioural flexibility and the ACC

Previously, we discussed evidence for evaluation of options centred on vmPFC/mOFC. These signals have been typically reported during tasks in which subjects make decisions between two or so currently available options. However, what if the subject decides to switch away from performing this task altogether in search of better alternatives? There is converging evidence that the ACC plays a key role in such behavioural flexibility. For example, when human participants are given the option of making an evaluative choice such as those previously described, or searching for a better alternative, whereas vmPFC/mOFC activity reflects the evaluative choices as expected, the ACC activity reflects how valuable it would be to search for an alternative set of choices altogether (Kolling et al., 2012). In other words, whilst the vmPFC/mOFC is deciding which exact option to choose, the ACC is in parallel deciding whether or not to abandon these decisions altogether. This and the results discussed below suggest the ACC plays a key role in assessing the quality of the current strategy and adapting our learning and behaviour accordingly.

A first requirement for a region monitoring the current strategy is that it has access to the outcomes of choices. The ACC responds to outcomes, particularly when they are caused by one's actions (Walton et al., 2004). Additionally, such a region should respond to outcomes according to their importance for updating beliefs and guiding future behaviour. An instance when we should place higher weight on current outcomes, as opposed to

[2] We note here that another option is not to update the internal model *per se*, but rather disengage behaviour from it (Tervo et al., 2014).

relying on previous beliefs, is when the world is changing. Indeed, activity in the ACC at the time of outcome of a decision is greater when the world is changing more rapidly and predicts the degree to which that outcome will guide future choice (Behrens et al., 2007). These findings are suggestive that the ACC monitors the outcomes of our choices and uses them to adapt behaviour. Indeed, lesions to this area in macaques (Kennerley et al., 2006), or inactivation in rodents (Tervo et al., 2014), impair their ability to guide choices based on outcome history.

The ACC may not only be responsible for monitoring the current strategy, but relatedly could also be important in computing how valuable alternative strategies are. This is important information when decided whether to initiate the switch away from the current strategy in to exploration. Indeed, when macaques are foraging in an environment with patches that have depleting resources but costs for travelling between patches, firing of neurons in ACC code the overall value of switching away from exploiting the current patch to exploring alternatives, taking in to account all of the respective values and costs in the decision. When the firing rates of these neurons reach a threshold, the animal leaves the current patch to explore (Hayden et al., 2011). Similarly in humans, the ACC codes the value of searching for alternatives (Kolling et al., 2012). The ACC therefore contains the necessary information for deciding whether or not to abandon the current strategy in favour of alternative, consistent with its increased engagement in humans when making exploratory decisions (Daw et al., 2006) and shifting strategies (Schuck et al., 2015).

These behavioural adaptations are likely mediated via an updating of internal models. Indeed, ACC is specifically engaged when observations indicate the internal model guiding behaviour needs to be updated even when overt behaviour is identical (O'Reilly et al., 2013b), and its inactivation in mice disrupts model-based learning (Akam et al., 2017). Furthermore, when rats perform a probabilistic learning task wherein one of two options has a higher probability of delivering reward, they are able to maintain stable behaviour despite probabilistic outcomes. During this stable behavioural period, cellular recordings reveal population level activity in ACC is also stable across trials. After the high reward location switches, the rats adapt their behaviour accordingly after a period of evidence accumulation, abandoning the old belief in favour of exploring alternatives. Preceding their switch in strategy from exploitation to exploration, however, ACC activity undergoes an abrupt, widespread and coordinated remodelling of population-level activity (Figure 1.7) (Karlsson et al., 2012). Activity then returns to a new stable state as a new belief or strategy is formed. These tight temporal links between activity and behaviour together with the

other studies implicate the ACC in having a key role in monitoring and updating the current internal model or strategy.

A somewhat separate school of thought regarding ACC function suggests that it codes for the value of deploying cognitive control (Shenhav et al., 2013) or for choice difficulty (Shenhav et al., 2014). Cognitive control is defined as "our ability to flexibly adjust behaviour in accord with internally maintained goals and away from behaviours that are more automatic but distract from those goals" (Shenhav et al., 2016). Although there is debate (Kolling et al., 2016; Shenhav et al., 2016) in the literature regarding precisely how this theory fits with some of the others discussed above (e.g. (Kolling et al., 2012)), which we do not go in to here, it is clear from this definition that there is some common ground. Although the underlying computations may differ, the cognitive control perspective similarly places emphasis on the role of ACC in flexibly adapting behaviour.

Taken together, a general picture of the ACC emerges in monitoring the quality of the current strategy and facilitating behavioural flexibility. It can use its information regarding the quality of the current strategy to control the degree to which beliefs are updated by observations or to decide whether to abandon the current strategy altogether and search for alternatives.

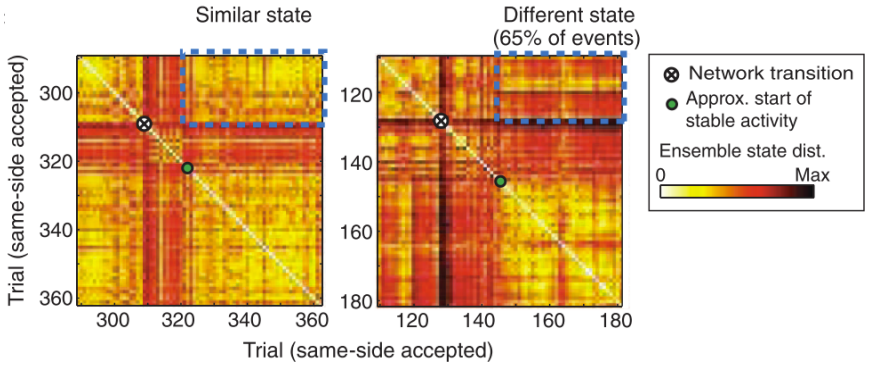


Figure 1.7 **Remodelling of ACC activity as beliefs are abandoned**. Trial by trial population similarity matrices of two example transitions reveal rapid and large changes in activity at 'network transitions', which precede a change in behavioural strategy. Before this when stable behaviour is present, and after stable behaviour has resumed again, the network activity is stable. In the majority of cases, the network returns to a different stable state despite the same strategy being implemented (right matrix). Yellow colours denote more similar population activity. Adapted from (Karlsson et al., 2012).

1.3.3 Behavioural flexibility and the noradrenergic system

Despite a wealth of evidence implicating the ACC in monitoring the current strategy, it is not clear what neural mechanisms facilitate flexibility in the internal models themselves. An intriguing possibility is that the requirement to update beliefs or disengage behaviour from an internal model is signalled to internal models or regions guiding behaviour, respectively, via a neuromodulatory route, in particular noradrenaline. Noradrenaline is a neuromodulator released from the brainstem nucleus locus coeruleus (LC), and its broad release in the brain can be indexed by pupil dilation (Ebitz and Platt, 2015; Joshi et al., 2016; Nassar et al., 2012). Converging theoretical and experimental work suggests that noradrenaline may play a key role in regulating behavioural flexibility.

Theoretical work suggests NA signals unexpected uncertainty (Yu and Dayan, 2005). Unexpected uncertainty is the uncertainty associated with not knowing the underlying statistics in the environment. This is contrasted with the expected uncertainty associated with known unreliability of predictive cues within a given context. Situations in which there is unexpected uncertainty are exactly those in which updating or disengaging from internal models is desired, since it is not known whether the current model is an accurate reflection of the statistics of the environment. It has been proposed that at a network level, NA may mediate this function via network resets: a reorganisation of networks guiding behaviour that permits rapid behavioural adaptation to a changing environment (Bouret and Sara, 2005).

This view is concordant with the suggestion that tonic NA release is associated with exploration, i.e. disengagement from the current task in favour of searching for alternatives (Aston-Jones and Cohen, 2005). Indeed, tonic NA release – as indexed by pupil diameter – is increased during exploratory relative to exploitative periods (Jepma and Nieuwenhuis, 2011). Potential causal evidence for this function of NA comes from the finding that tonic stimulation of LC causes disengagement from exploiting a patch; increasing LC tonic firing increased the likelihood of changing behaviour in rats performing a patch-foraging task (Kane et al., 2017). Furthermore, there is evidence to suggest that manipulating NA can alter internal exploration of options during decision making. Reducing NA renders decision makers more decisive (Coull et al., 1995), and it has been shown by measuring the neural representation of candidate options in rat hippocampus that this is related to reduced mental exploration of options before making a decision (Amemiya and Redish, 2016), suggesting NA also plays a role in balancing exploration and exploitation in internally

simulated as well as overt behaviours. Taken together these studies suggest that NA may have a broad role in disengaging from the current behaviour and increasing exploration.

One situation in which disengagement from the current task in favour of exploration is desirable is when the world has changed. One may therefore predict that NA release should reflect changes in the underlying state of the environment. Indeed, when humans are tasked with inferring the underlying mean (that jumps unpredictably) of a noisy process generating sensory observations, changes in pupil diameter positively correlate with the probability there has been a change in the underlying mean (Nassar et al., 2012). This study also showed that baseline pupil diameter reflected the unreliability of the current estimate of the mean, and not the noisiness of the data generating process (expected uncertainty), consistent with the increased pupil diameter during exploratory periods when there is similarly uncertainty over the underlying state of the environment (Jepma and Nieuwenhuis, 2011). Increases in both change point probability and uncertainty in the underlying mean are situations in which learning is desired, as they indicate the current belief is not a reflection of the true state. Accordingly, both change in pupil and baseline pupil diameter predicted the rate at which participants learnt from new observations. Remarkably, a task-independent manipulation of arousal (an unexpected tone) led to increased pupil diameter that predicted subsequent learning from new observations (Nassar et al., 2012). These findings are consistent with a role of NA in facilitating changes in behaviour and beliefs that follow unexpected changes in the environment by promoting both learning and information-seeking behaviours (Aston-Jones and Cohen, 2005; Bouret and Sara, 2005; Yu and Dayan, 2005).

The impact of NA on internal models and how this facilitates such learning and modulation of behaviour is not yet known, although recent research is beginning to uncover these processes. In the context of facilitating exploratory behaviours, there is evidence that it may operate by modulating the degree to which behaviour is guided by an internal model versus a stochastic behaviour generating process. When rats are trained to perform a task in which they will receive reward if they successfully counterpredict a virtual competitor, they use previous history of choices and an internal model to counterpredict the competitor (Tervo et al., 2014). However, when they are faced with a competitor they cannot defeat by counterprediction, they remarkably switch to a stochastic decision mode and behave randomly, by abandoning their internal model that they deem no longer sufficient for the task. Optogenetically driving noradrenergic input from LC to the ACC, or inhibiting ACC alone, causes this stochastic behavioural mode even when the animals are faced with

competitors they can counterpredict. Therefore, NA, in concert with the ACC, may control the degree to which behaviour is guided by an internal model and facilitate its abandonment when it is deemed no longer sufficient.

The impact of NA on internal models that facilitates learning is also not yet clear. However, some insight may be gained from a recent study in which the impact of NA on tuning of neurons in auditory cortex was measured (Martins and Froemke, 2015). Increasing NA in primary auditory cortex led to, amongst other things, a broadening of auditory neuron tuning curves to different sound frequencies (Figure 1.8). In other words, the selectivity of a given neuron was reduced. If we reconceptualise the tuning of a given neuron as a simple form of internal model, then NA can be thought of as having weakened the prior of the internal model. Such a state is desired to facilitate learning, and could be a candidate mechanism for facilitating learning in internal models.

Demonstrating such an effect on the neural representation of an internal model representing dynamic beliefs in more typical probabilistic learning tasks is challenging due to the requirement of measuring the neural model and NA simultaneously. This is likely at least partly why, despite much theoretical work proposing the impacts of NA on internal models, experimental work is relatively sparse. In chapter 3 we simultaneously measure an internal model, by decoding dynamic beliefs with fMRI, and NA, indirectly with pupil diameter, to explore the suggestion NA may reduce the selectivity of an internal model.

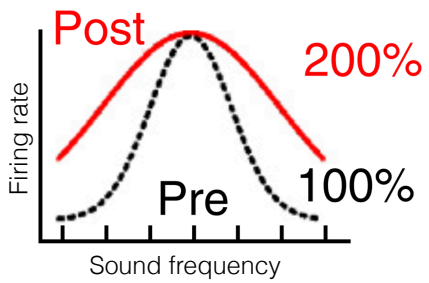


Figure 1.8 **Noradrenaline reduces selectivity of auditory neurons**. Increased NA in auditory cortex (red) gives rise to a broadening of the tuning curves (reduced selectivity) to sound frequency, compared to before increased NA (black). It is possible reducing selectivity of cortical representations is a general effect of NA. Figure adapted from (Martins and Froemke, 2015).

1.3.4 Parallels between the ACC and the NA system in behavioural flexibility

Across the various studies discussed above, a general pattern of NA function emerges as facilitating learning and behavioural adaptation. This is reminiscent of the ACC,

and indeed from the above discussions we can observe various parallels between them. These can be broadly summarised as both the ACC (Behrens et al., 2007; Daw et al., 2006; Karlsson et al., 2012; O'Reilly et al., 2013b; Tervo et al., 2014) and the NA system (Jepma and Nieuwenhuis, 2011; Kane et al., 2017; Nassar et al., 2012; Tervo et al., 2014) predicting the updating of beliefs and showing increased engagement during exploratory decisions, both important components of behavioural flexibility. These roles are consistent with each other and can be reconciled with monitoring the quality of the current strategy: if an agent has uncertainty over whether the current strategy or model guiding behaviour accurately reflects the world, it ought to learn more from observations (giving rise to a higher learning rate and updating of beliefs) and use the model less for guiding decisions (giving rise to exploratory choices) to discover the true state of the world.

These parallels suggest the NA system could be a candidate mechanism via which the computation of the requirement for behavioural flexibility in ACC may be broadcast to other brain regions, including those representing internal models. This would require appropriate connectivity between these systems. Indeed, anatomical work suggests ACC is one of the few cortical regions with descending control on NA release (Jodo et al., 1998). Furthermore, recent studies have shown that activity of ACC neurons predict pupil dilation and may temporally lead pupil-predicting LC activity (Ebitz and Platt, 2015; Joshi et al., 2016). However, despite the various parallels between these systems, evidence for their suspected (Aston-Jones and Cohen, 2005; Karlsson et al., 2012; Nassar et al., 2012) coordination during tasks requiring belief updating has been lacking. In chapter 3 we seek evidence for such coordination.

1.3.5 Work in this thesis related to behavioural flexibility in frontal cortex

We have therefore seen that flexible task representations exist in vmPFC/mOFC, whilst the ACC plays a key role in monitoring or updating of the current strategy. The responses of ACC bear similarity with those of the noradrenergic system, which may alter cortical representations in such a way as to facilitate model updating. In chapters 2 and 3 we address the possibility that the noradrenergic system may facilitate increases in uncertainty in internal models of the task. An important challenge to investigating this question is measuring a neural index of the internal model as participants perform the task to test predictors of changes in this model. We address this challenge using multivariate fMRI methods as human participants perform a probabilistic learning task in a dynamic environment. We show ACC tracks changes in the uncertainty of participants' beliefs and

predicts changes in pupil diameter. Changes in pupil diameter in turn predicted changes in uncertainty in a simultaneously measured internal model of the task decoded from vmPFC/mOFC.

1.4 Cognitive maps in the hippocampal-entorhinal system

Thus far we have focused on learning and decision making systems in the prefrontal cortex. Amongst other things we have seen that internal models of tasks are found in these regions and that these are used for flexible decision making. A largely separate literature is that of spatial cognition and memory, centred on the medial temporal lobe. Although these literatures have historically proceeded largely independently, it is becoming clear that interactions between these regions are important for both learning and decision making, as well as memory and spatial cognition.

Furthermore, although we have presented evidence for task representations in PFC, often it is difficult to interpret exactly what task variable a given neuron is coding for, or perhaps more importantly, identify the principles that give rise to their coding properties. One of the most fundamental tasks animals perform is that of moving around in space. Striking neural correlates of this task have been found in the hippocampal-entorhinal system (H-EC) and associated brain regions. This provides the opportunity to investigate the general principles that give rise to their coding properties, which in addition to understanding the organising principles of H-EC, may help shed further light on how task representations emerge in PFC, as well as the coordination between these brain regions to subserve diverse cognitive functions, including learning, decision making, memory and spatial cognition.

We therefore now turn our focus to introducing the H-EC system and how it may support cognitive maps. The next three sections can, broadly speaking, be viewed as describing cognitive maps and the H-EC system from the perspective of each of Marr's three levels (Marr and Poggio, 1976): implementational, algorithmic, and computational. We first describe empirical observations revealing striking neural correlates of a cognitive map that suggest its implementation in these brain regions. We then explain some of the desirable information processing properties of these representations and of the neuroanatomy and neurophysiology of this system, followed by outlining the few main theories of the computational functions of this system.

1.4.1 Neural implementation of the cognitive map

Although Tolman's original conception of the cognitive map was not restricted to space, but rather proposed it organised knowledge across multiple domains, the focus of research on cognitive maps has been in the spatial domain due to the discovery of cell types with precise spatial correlates in physical space. Here, we briefly introduce these cells and then highlight recent evidence calling in to question the specificity of these representations to physical space.

1.4.1.1 Place cells

The first neural support for the cognitive map came with the discovery of place cells in the rodent hippocampus (O'Keefe and Dostrovsky, 1971). Place cells are neurons that fire whenever an animal is at a specific location in an environment (Figure 1.9), often regardless of other factors such as the animal's direction or sensory cues (Muller et al., 1994; O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978). Within a given environment, many place cells will be active, together covering the spatial extent of the environment with overlap between their firing fields such that at any given location, a number of place cells will be active (O'Keefe and Nadel, 1978). The animal's location can therefore be predicted from across the population (Wilson and McNaughton, 1993b). Upon exposure to a novel environment, a new, independent (orthogonal) place cell code for that environment appears rapidly (termed place cell remapping), is stable within the environment, and reappears upon re-exposure to the environment (Muller and Kubie, 1987; Wilson and McNaughton, 1993b). The size of place cell fields increases along the dorsal to ventral axis of the hippocampus (Kjelstrup et al., 2008). Hence the hippocampal place cell code can be viewed as providing an internal map of space that is capable of mapping multiple different environments in orthogonal place codes.

It is hard to overestimate how remarkable these cells are. Describing the activity of neurons in relation to how they represent the external world had only relatively recently been achieved by Hubel & Wiesel's demonstration of cells in visual cortex coding for orientation gratings (Hubel and Wiesel, 1959). This marked the beginning of investigation in to how neurons represent the external world that initially focused on sensory processing. Finding neural correlates of cognitive maps seemed far more elusive. Place cells were then discovered, which clearly have a relation to the external world since they code the animal's location. However, as just described, this coding of location is invariant to other factors such

as head direction, suggesting this is a highly processed and abstract representation that offers what is one of the most striking window in to high level cognition.

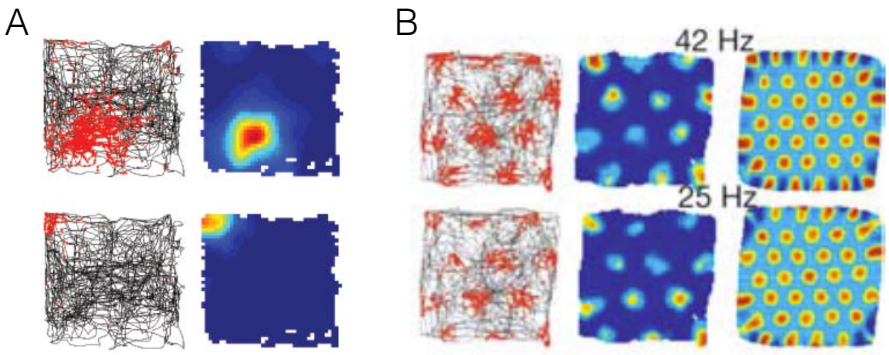


Figure 1.9 **Place cells and grid cells**. A) Place cells recorded in the hippocampus fire at a specific location in the environment. Left: black line denotes animals path and red dots are spikes; right: smoothed firing rate map. B) Grid cells recorded in MEC fire at multiple locations in the environment that form a triangular lattice. Same as A, with the addition of a 2D spatial autocorrelation (right). A and B adapted from (Fyhn et al., 2007) and (Hafting et al., 2005), respectively.

1.4.1.2 Grid cells

Attempts to understand the origins of this place cell code led to the discovery of grid cells in the neighbouring medial entorhinal cortex (MEC)(Hafting et al., 2005). Unlike place cells, which generally have only one (or a couple) of firing fields within a given environment, grid cells have many firing fields that form a triangular lattice spanning the entire environment (Figure 1.9). Topographically neighbouring cells share the same orientation and scale (distance between peaks) but can differ in their phases (Stensola et al., 2012). The grid cell scale increases along the dorsal to ventral axis of the MEC (Hafting et al., 2005; Stensola et al., 2012). Like place cells, a grid code is established rapidly upon entry to a novel environment and is invariant to environmental sensory cues (Hafting et al., 2005). It has therefore been suggested that grid cells "are part of a universal spatial metric" (McNaughton et al., 2006) providing an allocentric map of space.

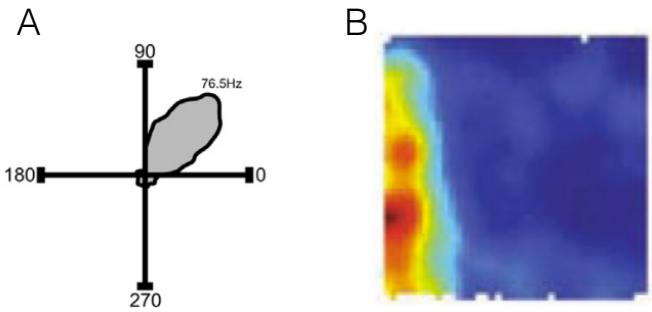


Figure 1.10 **Head direction cells and border cells**. A) Head direction cells recorded in MEC fire when the animal's head is facing in a particular direction. Polar plot illustrates firing rate binned in terms of the animal's head direction. B) Border cells also recorded in MEC fire along borders of the environment. A and B adapted from (Grieves and Jeffery, 2017) and (Solstad et al., 2008), respectively.

1.4.1.3 Other spatial cell types

In addition to place and grid cells, numerous other cell types each coding an important aspect of spatial cognition have been discovered (Grieves and Jeffery, 2017). In MEC, head direction cells, as the name implies, fire whenever the animal's head is facing a particular direction, regardless of the animal's location (Figure 1.10)(Taube et al., 1990), whilst speed cells code for the animal's speed (Kropff et al., 2015), together providing the necessary information to compute vectors in space. Also in MEC, border cells fire when the animal is next to an environmental border, irrespective of other sensory cues and exact spatial location (Figure 1.10)(Solstad et al., 2008), band cells form spatially periodic bands in the environment (Krupic et al., 2012), and object vector cells fire when the animal is a particular vector from an object of interest in the environment (regardless of the object identity) (Høydal et al., 2018). The presence of these various cells together with the grid cells suggests the MEC is ideally suited for computing transitions through space ($P(s'|s,a)$), indicating it is an internal world model. In hippocampus, in addition to the animal's current location, the location of other agents on the map is also encoded (Danjo et al., 2018; Omer et al., 2018), and reward locations are coded for by an apparently designated population of cells (Gauthier and Tank, 2018). Taken together, these various cell types across both regions code for different and important features of the spatial environment and together may constitute a map of physical space.

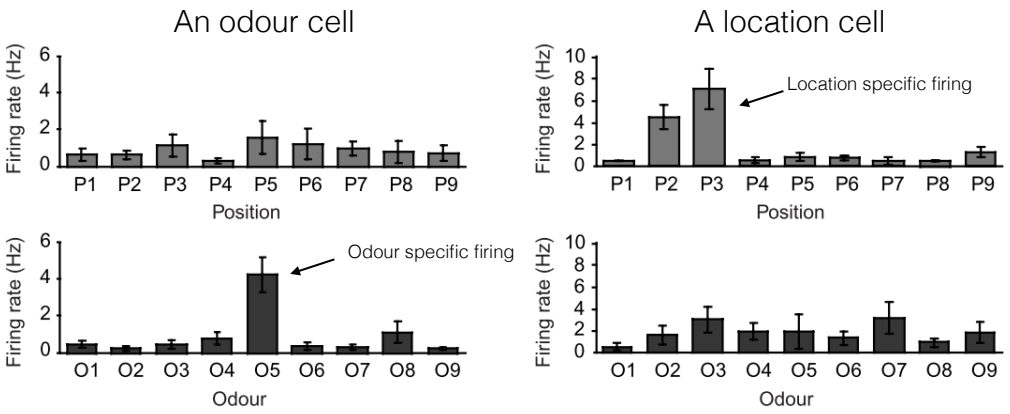


Figure 1.11 **Non-spatial coding in hippocampus**. In a task designed to enable dissociation of hippocampal neuron activity associated with different task components, such as odour and location, significant coding was found for all task variables, not just spatial. Here we show an example odour cell and an example location cell. P1-9 and O1-9 are the different locations and odours, respectively, used in this task. Figure adapted from (Wood et al., 1999).

1.4.1.4 Cognitive maps for non-spatial and abstract spaces

Although these remarkable cells convincingly demonstrate neural correlates of spatial features of the environment, attributing to them the function of simply mapping space does not capture the whole story. First of all, that hippocampal place cells are modulated by ongoing behaviour or task demands, such as which direction an animal is running down a maze (McNaughton et al., 1983), suggests they are not a veridical code for space *per se*, but rather represent location in the context of the task (Eichenbaum et al., 1999).

In addition to the modulation of spatial coding by other factors, it has also been demonstrated that the firing of hippocampal neurons is directly related to non-spatial stimuli and behaviour relevant to performing an ongoing task. For example, an elegant study explicitly dissociated spatial from non-spatial coding. Rats performed a simple non-match-to-sample task requiring them to dig for food if the odour in a cup was different to the odour on the previous trial (non-match; if the odour was the same they should not dig). Importantly, the location of the cup was changed on each trial and did not predict whether the odour was the same or not, allowing dissociation of odour coding from location coding. Neurons were found in hippocampus to code for all three pieces of relevant information: odour, the abstract non-match relationship, as well as location (Figure 1.11)(Wood et al., 1999). Some cells were selective for one variable, such as odour regardless of location, and others were selective for multiple, such as conjunctive position and odour coding.

Furthermore the prevalence of neurons selective for spatial and non-spatial information was comparable. Similarly, in humans, hippocampal cells have famously been found to fire to specific abstract non-spatial stimuli, such as Jennifer Aniston, or other given individuals, landmarks or objects (Quiroga et al., 2005). These studies, along with others (Eichenbaum et al., 1999), therefore provide strong evidence hippocampal neurons code for non-spatial information.

More recently it has been shown that non-spatial information with the same topology as spatial information is represented in the hippocampus and neighbouring structures in a manner analogous to space. In addition to space, another important dimension for organising experience is time, which has a 1-dimensional topology. Time cells have been found in rodent hippocampus, which each code for a particular moment relative to some behaviourally relevant event, such as starting to run on a treadmill, and their firing bears striking similarity to place cells coding for location on a topologically similar linear track (Kraus et al., 2013). Coding for time has also been found in MEC, but in this case rather than firing to a particular moment in time, firing as a function of time is periodic, analogous to spatial grid cells alongside them in MEC (Kraus et al., 2015). Similarly when rats are tasked with manipulating a sound along a continuous 1D frequency axis, place cell-like and periodic grid cell-like coding for sound frequency is observed in hippocampus and MEC, respectively (Aronov et al., 2017). Finally, when human participants perform a non-spatial task requiring them to observe stimuli morphing in two continuous dimensions (the lengths of a bird's neck and legs), that therefore has the same topology as 2D space, grid cell-like coding as is observed in spatial tasks (Doeller et al., 2010a) is observed in the fMRI signal in MEC and medial PFC, suggesting this entirely non-spatial task is represented in a grid-like code (Constantinescu et al., 2016).

Together, these findings demonstrate that spatial coding in the hippocampal formation is modulated by other non-spatial factors, that these regions represent non-spatial information independent of spatial information, and that the representation of non-spatial information in tasks topologically similar to real space appears to be in a manner analogous to the coding of spatial information.

1.4.1.5 Discrete maps

The demonstrations of coding of non-spatial information in a manner analogous to spatial information (Aronov et al., 2017; Constantinescu et al., 2016; Kraus et al., 2013, 2015) have necessarily focused on continuous topologies, since space is continuous.

However, the observations that place cells code for specific stimuli or behaviours (Quiroga et al., 2005; Wood et al., 1999) suggest hippocampal neurons may represent discrete information, such as locations in tasks that can be described by discrete states. Indeed, when humans passively transition through discrete state spaces that can be described as a set of stimuli on the nodes of structured graphs, the structure of these graphs can be recovered from the neural representation of the stimuli in both hippocampus and MEC (Garvert et al., 2017; Schapiro et al., 2013). These results are consistent with the hippocampal formation representing the structure of tasks of potentially arbitrary topologies, as would be expected of a system that supports cognitive maps responsible for organising information of all different kinds (Tolman, 1948).

1.4.2 Desirable properties of hippocampal-entorhinal representations

In addition to these regions containing representations consistent with the implementation of both spatial and non-spatial cognitive maps, they also have a number of neuroanatomical and neurophysiological features that suggest properties appealing from an information processing point of view. These appealing properties facilitate learning and efficient representation of the knowledge in a cognitive map, use of this knowledge for flexible behaviour, and the implementation of various computational functions attributed to this system (as discussed in the next section). We briefly run through some of these key properties, noting they could each have an entire review written about them alone. We also discuss desirable properties of the recorded representations. We start with properties specific to hippocampus, then neighbouring entorhinal cortex, followed by those common across both.

1.4.2.1 Well connected to bind disparate sources of information

Our everyday experiences are multimodal. Imagine walking through a garden on a sunny day; you'd feel the warmth, see and smell the flowers, hear the birds, you may be talking to a friend and having a positive emotional response. All of these features of the experience, as well as where and when it occurred, are processed in different cortical regions. In order to create a coherent memory of this experience, this information must somehow be bound together. The hippocampus is ideally placed to do this; it receives convergent afferents from nearly all cortical association areas, and these afferents are widely distributed on to the cell population in hippocampus (Figure 1.12)(Burwell et al.,

1995; Eichenbaum, 2004; Witter and Amaral, 2004). The high-level hippocampal representation in turn has return connections (via entorhinal cortex) to these cortical areas, allowing it to index and reactivate the representations. Hence, the hippocampus has access to high-level information processed in disparate parts of cortex providing it with the opportunity to bind it in to a coherent memory.

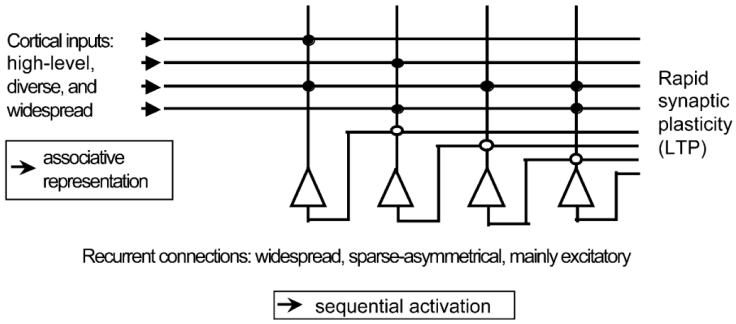


Figure 1.12 **Well-connected, recurrent associative networks with fast memory**. Schematic of these three desirable information processing properties of hippocampus, particularly true in CA3. Afferents from diverse cortical regions converge on principal neurons that have recurrent, asymmetrical connectivity. Rapid synaptic plasticity facilitates the formation of fast memories that can pattern complete and index the original cortical representation. Figure adapted from (Eichenbaum, 2004).

1.4.2.2 Recurrent associative networks and pattern completion

In order to bind this incoming information together, the intrinsic connectivity must also be appropriate: neurons representing one component of an experience must be connected to neurons representing another in order to form associations between them. The principle neurons in CA3 do just this: they send large numbers of projections to other CA3 principle neurons. These recurrent connections are broad across the CA3 population, sparse, and involve mostly excitatory synapses (Figure 1.12)(Eichenbaum, 2004; Treves and Rolls, 1994; Witter and Amaral, 2004). Once connections between neurons representing different components of a memory have been formed (see below), this associative network has the ability to recall the entire experience from a part of it. When reminded by your friend of the flowers in the garden, you may also recall the sound of the birds. This is facilitated by the recurrent, excitatory connections from CA3 neurons representing the flowers to those representing the bird song. These high level hippocampal neural representations then index, via other parts of cortex, the relevant cortical area representing

this mode of the experience (auditory cortex for bird song, for example) leading to the experience of recalling it.

1.4.2.3 Fast formation of associations

You only experienced your walk in the garden once yet can remember it. Such one shot learning requires the fast formation of associations between hippocampal neurons. During the experience of an event, CA3 neurons receiving inputs from (and indexing) the representations of disparate components of that event will be simultaneously activated. Their simultaneous activation combined with dense recurrent connectivity allows CA3 cells to establish Hebbian learning between each other, strengthening the associations between these representations (Figure 1.12)(Neves et al., 2008). Neurophysiologically, these hippocampal neurons are believed to undergo LTP, which is thought to mediate fast synaptic plasticity. This is consistent with the finding that mice without NDMA receptor plasticity – thought to mediate LTP – in CA3 are impaired in rapid learning of new spatial memories and cannot pattern complete from partial cues (Nakazawa et al., 2002, 2003). Fast formation of associations together with converging afferents and dense recurrent connectivity suggest the hippocampus can rapidly bind together different components of an experience to form a memory and that this memory can be retrieved from partial information.

1.4.2.4 Sparse representations and pattern separation

Many features of a walk in the garden will be common across garden walks; hence it is important to have mechanisms to prevent interference between memories of similar experiences. The pattern completion mechanisms just described encourage retrieval of memories from partial cues, making the representation of the partial cue more similar to the representation of the original holistic memory. This needs to be balanced with allowing representations of different memories to be different enough to avoid interference between them. Sparse codes in the hippocampus are thought to facilitate reduced overlap between representations and interference between memories, termed pattern separation (Rolls, 2013; Treves and Rolls, 1994). Sparsity is the proportion of neurons in a given brain region that are active in response to a given stimulus. Sparse coding, as in hippocampus, is when this proportion is relatively small (e.g. 1%), and is contrasted with densely distributed coded (e.g. 20%)(Kumaran et al., 2016). Two randomly selected sparse populations will have low overlap.

The dentate gyrus, thought to have among the sparsest codes in the brain, effectively orthogonalises incoming inputs from entorhinal cortex (which contains approximately five times fewer neurons), resulting in a sparse, pattern separated code. DG then selects a small subset of CA3 neurons that are also sparse and now pattern separated that then undergo autoassociative learning, as just described above (Kumaran et al., 2016; McClelland et al., 1995; Rolls, 2013; Treves and Rolls, 1994). Such orthogonalised codes are reminiscent of hippocampal remapping discussed above, whereby the hippocampal representation of one environment is orthogonal to that of another. Indeed, it has been proposed that the particularly pronounced remapping observed in CA3 may counteract memory interference in the pattern completion mechanisms that occur there (Colgin et al., 2008). Notably, in addition to reducing memory interference, sparsity also allows a greater number of representations for a given number of neurons, resulting in increased memory storage capacity (Treves and Rolls, 1991).

1.4.2.5 Invariant representations

As we have already mentioned above but is worth making explicit here, hippocampal representations can be invariant to lower level features. Place cells can be invariant to the animal's current direction and other cues (Muller et al., 1994; O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978). Additionally, the neurons found in human hippocampus that fire to specific abstract non-spatial stimuli, such as given individuals, objects or landmarks, were invariant to very different pictures of that individual or even modality by which that stimulus was presented (image vs. letters/word) (Quiroga et al., 2005). The authors conclude these observations suggest an invariant and sparse code containing abstract information important for memory (Quiroga et al., 2005). Such representations are desirable for memory since the abstract concept of an entity, rather than its low level features, can be bound to something else.

1.4.2.6 Grid cells form an efficient basis for 2D space: what's so special about hexagons?

In this section we have thus far focused on the desirable properties of the hippocampal code. We now turn to grid cells in neighbouring MEC. As described above, grid cells form a triangular lattice over the environment, and as we discuss below, grid cells form discrete modules of increasing grid scale along the dorsal to ventral axis on the MEC

(Stensola et al., 2012). Here we briefly discuss a few different but converging perspectives on why grid cells may form hierarchical triangular lattices.

From a functional perspective, it has been shown grid cells provide an efficient basis for navigation (Bush et al., 2015; Stemmler et al., 2015). Since a hierarchical triangular lattice is able to uniquely encode locations, it allows vectors to be computed between distant locations, even when these locations may be further apart than the largest grid scale (Bush et al., 2015). Such vector-based navigation allows planning that is efficient since it is not necessary to search through locations in the environment. Indeed it has recently been demonstrated that an artificial agent that has learnt grid-like representations is able to perform such vector-based navigation and use it to flexibly compute shortcuts via previously untraversed routes (Banino et al., 2018).

From an information theoretic perspective, the grid code is a highly efficient representation of spatial location. (Mathis et al., 2015) asked what type of lattice representation maximises the spatial resolution for a given population of neurons, resulting in the highest accuracy in decoding location from this population. They show how solving this problem can be mapped onto the problem of packing non-overlapping spheres (or circles in 2D), and that the optimal lattices for solving this problem are those with the highest packing ratio themselves, i.e. the densest lattices represent space most accurately, and these are hexagonal/triangular lattices. They also demonstrate the requirement for grid modules of different scales to resolve the ambiguity caused by periodicity at the smallest scale, while still preserving their high spatial resolution (Mathis et al., 2015). Hence across the hierarchical population of grid cells spatial information can be accurately decoded, even though within a single cell the representation may be similar in different parts of space, due to the periodicity. Similarly, a hierarchical triangular lattice was shown to minimise the number of neurons required to encode location at a given spatial resolution (Wei et al., 2015). Both of these studies therefore demonstrate that a hierarchical triangular lattice is optimal for decoding of spatial information from a population of neurons.

From a statistical perspective, a grid code emerges from a principle component analysis (PCA) of place cell activity (Dordek et al., 2016). As discussed previously, place cell activity as a function of time can be simulated by distributing place cells throughout a virtual 2D environment and having an agent walk around the environment through the place fields. Performing non-negative PCA of this place cell activity matrix results in eigenvectors that are periodic over space forming a triangular lattice, again with discrete modules with different grid scales (Dordek et al., 2016). This suggests grid cells are an efficient representation for

summarising the statistics of 2D distributed place cells. We discuss this further in detail in Chapter 4.

Each of these different theoretical perspectives arrives at a hierarchically organised triangular lattice – with grid module scale factors matching those observed in the brain – that forms an efficient combinatorial code of space that is optimal for encoding location and can be used for efficient navigation.

1.4.2.7 Generalisable representations

Although place cells provide a one-to-one mapping with location that is likely useful for memory, that the code remaps across environments such that there is no consistent relationship between place cells across environments suggests that any method used for navigation learnt on this place cell code would need to be relearnt in each new environment. Grid cells within a module on the other hand retain their relationships with each other across environments (Figure 1.13)(Fyhn et al., 2007), suggesting that navigation strategies learnt to operate on this grid code, or knowledge more generally, can be generalised to novel environments. Learning does not therefore need to start afresh. Grid cells are not the only cell type in MEC that appears to generalise (Behrens et al., 2018). The recently discovered object-vector cells generalise over different objects (Høydal et al., 2018), whereas their hippocampal counterparts do not, rather exhibiting more specific coding for objects (Figure 1.13)(Deshmukh and Knierim, 2013). MEC representations may therefore generalise over sensory particularities, rather representing knowledge about the underlying structure of the environment, such as the 2D-ness of space. We discuss this further in chapters 4 and 5.

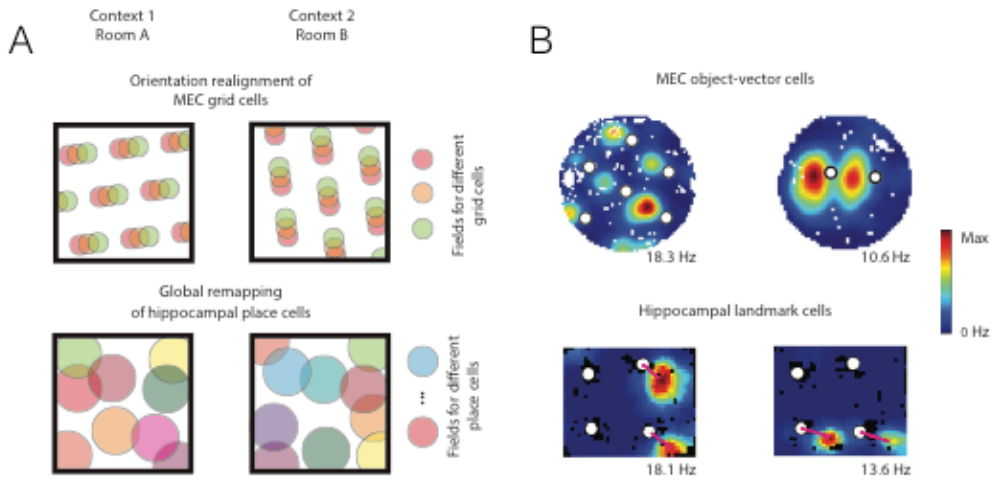


Figure 1.13 **Generalisable representations in MEC**. A) In spatial remapping experiments, MEC grid cells maintain a constant spatial phase structure (top), whereas hippocampal cells do not (bottom). In other words, the relationships between grid cells generalise across environments. B) MEC object-vector cells respond when an animal is a given vector from an object and generalise over object identities (top), whereas hippocampal cells are selective for object identity (bottom). White dots denote objects. Figure adapted from (Behrens et al., 2018); originals from B are from (Høydal et al., 2018) and (Deshmukh and Knierim, 2013).

1.4.2.8 Hierarchical representations

We now discuss some properties that are in common with hippocampus and MEC. As we have already mentioned, hippocampal place cells and entorhinal grid cells exhibit an increase in place field size or grid scale, respectively, with increasing distance along a dorsal to ventral axis of the hippocampal formation (Figure 1.14)(Hafting et al., 2005; Kjelstrup et al., 2008; Stensola et al., 2012). In the case of the grid cells, these increments in grid scale are discrete rather than continuous such that grid cells form discrete modules (Stensola et al., 2012). Place cell firing field sizes on the other hand appear to increase linearly along this axis (Kjelstrup et al., 2008). As we described above, the hierarchy observed in grid cells is essential for accurate encoding of location across the population of cells (Bush et al., 2015; Mathis et al., 2015; Wei et al., 2015). In addition to providing an efficient combinatorial code, hierarchical representations may also allow reuse of information over space. The relationships between high frequency (low scale) grid cells within a module will be true across different parts of space due to their periodicity. Therefore, due to the hierarchical representation, learning in one part of space could be repeatedly reused in different parts of space. We revisit these ideas in chapter 5.

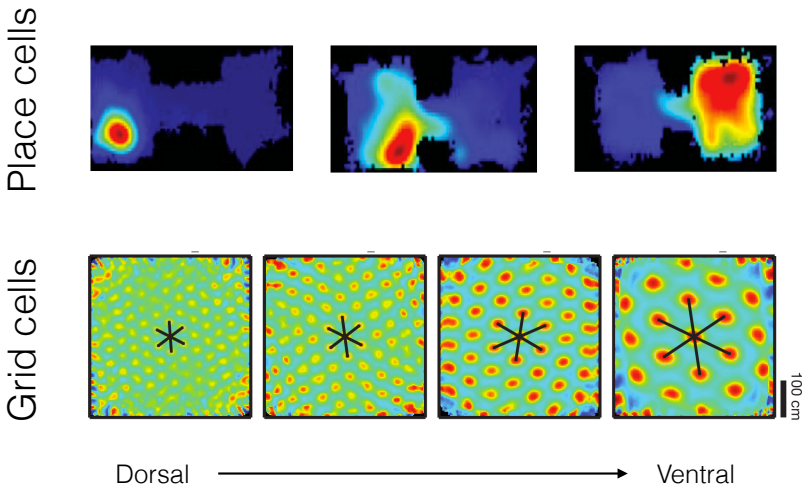


Figure 1.14 **Hierarchical representations**. Both place cells and grid cells show increases in size or scale more ventrally along the hippocampal formation. Place cells on a linear track show increasing field size (top) and grid cells in a square box show increases in grid scale (bottom). Notably the increases in grid scale are discrete, i.e. the grid cells cluster in to discrete modules that share similar scales. Adapted from (Eichenbaum, 2017a)(top) and (Stensola et al., 2012)(bottom).

1.4.2.9 Predictive/successor representations

Place cells, as we have discussed, are traditionally thought to represent the current location. Recently, it has been suggested that several features of place cell properties may be explained if a place cell is reconceptualised to encode successive or future spatial states (Stachenfeld et al., 2017), termed successor representations (Dayan, 1993). Representations predictive of future locations are useful since they offer rapid evaluation of possible futures (Behrens et al., 2018), useful for reinforcement learning (Stachenfeld et al., 2017). Furthermore, the eigendecomposition of these successor representations resemble grid cells (Stachenfeld et al., 2017). Hence it is possible that representations in the hippocampus are optimised for future prediction, to facilitate reinforcement learning.

1.4.2.10 Replay

A striking feature of the representations observed in these structures is that during offline periods such as sleep the cells that code for routes that were traversed during behaviour reactivate and replay the neural code of the route (Ólafsdóttir et al., 2015, 2016; Wilson and McNaughton, 1993a). These reactivation events occur during sharp wave ripples and, since they are not constrained by the statistics of the environment or behaviour, the sequences are played out much faster than the route they code for could be traversed,

potentially allowing many replay events. These events are predicted by theories of consolidation whereby the reactivation of memories is thought to help form associations in cortex and extract the statistical regularities across memories (McClelland et al., 1995). As well as reactivating previously traversed routes, these replay events also explore novel routes through the environment suggesting they may support preparation of future experiences (Ólafsdóttir et al., 2015).

1.4.3 Theories for hippocampal-MTL function

The neural representations and their various information processing properties in the hippocampal formation, along with neuropsychological evidence discussed below, have inspired numerous theories of the computational function of this system, and/or been predicted by them. Although some of these theories have been implicit in the discussions of the neural evidence above, we lay them out more explicitly here.

1.4.3.1 Complementary learning systems

The most influential computational theory about the relationship between hippocampus and cortex is complementary learning systems (CLS) (McClelland et al., 1995), which states the goal of this system is to perform fast learning of new memories in hippocampus, which are then slowly integrated with knowledge in neocortex. CLS has its roots in the ideas of Marr (Marr, 1971; Willshaw et al., 2015), who regarded the hippocampus as acting as a temporary store for rapidly formed memories that were then transferred to neocortex. Marr also laid out the core ideas for how these memories may be formed in recurrent networks with fast Hebbian learning and that these memories index representations in cortex (Marr, 1971; Willshaw et al., 2015).

This work inspired CLS (McClelland et al., 1995), which emphasised that there were two, complementary learning systems. In CLS, as Marr suggested, memories are first stored in sparse hippocampal codes by fast synaptic changes, and these changes support subsequent reinstatement of the memory in neocortex. Reinstatement of the memories in cortex (replay) trains this complementary, slow learning system to extract the common statistical structure across experiences, or memories – termed systems consolidation. This allows generalisation across experiences. The fast learning hippocampal system allows rapid learning of new experiences without disrupting this knowledge in cortex. Interleaved

reinstatement helps integrate these new experiences with old knowledge without disrupting it.

CLS was firmly ground in the information processing properties of this system, described above. Another guiding principle during the development of the theory was neuropsychological observations in patients with hippocampal lesions. Great insight in to the computational goals of a system can of course be gained by observing deficits following its damage. The exemplary patient was the famous HM who had bilateral lesions to the hippocampus and neighbouring temporal lobe structures that resulted in temporally graded retrograde amnesia, as well as profound anterograde amnesia, "forgetting the incidents of daily life as fast as they occur" (Scoville and Milner, 1957). The general picture across a range of observations appears to be that whilst hippocampally lesioned patients cannot form new memories or remember recent events, their memory of distant events (many years previously) and general ability to reason are retained (Squire et al., 1992). This was a key observation guiding the development of the consolidation process in CLS, which is able to account for this temporally graded amnesia since it proposes the transfer of knowledge to cortex is slow (McClelland et al., 1995). This differed from Marr's proposition that this transfer occurred over night during sleep to store the day's experiences when there was no incoming sensory information to interfere (Marr, 1971; Willshaw et al., 2015).

In the subsequent years, the CLS theory has faced a couple of critical empirical challenges, and has recently been updated to accommodate them (Kumaran et al., 2016). First, to accommodate for the observation that hippocampus is likely responsible for some forms of generalisation requiring inferences across new memories (Bunsey and Eichenbaum, 1996; Zeithamova and Preston, 2010) – not supported by sparse orthogonalised codes – by allowing recurrent activation of hippocampal traces via entorhinal cortex, which allows the system to discover higher order structure present within a set of related experiences (Kumaran and McClelland, 2012). Second, to accommodate for the observations that experiences can become hippocampally independent much faster than the original CLS theory predicted (Tse et al., 2007, 2011) by showing experiences that are consistent with previous knowledge represented in cortex can be consolidated much faster (McClelland, 2013; Saxe et al., 2013).

The ideas of Marr and subsequent CLS theory tied together many of the empirical observations and known properties of the hippocampal system to explain how it may mediate fast memory formation and interact with cortex, which extracts and consolidates knowledge. This synthesis of ideas has been highly influential in the development of

subsequent theories and explanation of data. An interesting potential hole left open in these ideas is how cortex then impacts on future hippocampal representations, placing new experiences and memories in the context of the knowledge represented in cortex. We discuss this further in chapter 5.

1.4.3.2 Declarative memory theory

The declarative memory theory of the hippocampal formation also builds on the ideas of Marr and seeks to explain the deficits observed in hippocampally lesioned patients and animals (Squire et al., 1992, 2004). The declarative view also identified the principle role of the hippocampus as rapidly forming associations that index cortical representations of the full memory, which gradually becomes independent of hippocampus. One of the main contributions of the declarative view was to identify multiple different types of memory and their differential dependence on the hippocampus. This was motivated by observations such as the retained ability to acquire a new skill or procedure in hippocampally lesioned humans, monkeys and rats (Cohen and Squire, 1980; Squire et al., 1992). This led to their development of a memory taxonomy, which distinguishes between declarative – the ability to bring to mind or declare the content of a memory – and non-declarative memory. Declarative memory can be further subdivided in to episodic and semantic memory, memory for specific events or facts, respectively. This is contrasted with non-declarative memory, which is associated with memories of skills and procedures. This taxonomy was useful in describing the effects of hippocampal lesions on memory; they were able to narrow the contribution of hippocampus to specifically declarative memory (Squire et al., 1992). The ability to use old and acquire new non-declarative memories is retained. The ability to acquire new declarative memories is lost entirely, as is the ability to recall recent memories. However, the ability to recall distant declarative memories is spared, since patients are able to exhibit apparently normal recollection of old facts and even events (Squire et al., 1992, 2004). Hence the declarative view offers a taxonomy of memory, proposing the crucial role of the hippocampus is in the formation and subsequent consolidation of specifically declarative memories, and is consistent with the mechanistic framework laid out by CLS (McClelland et al., 1995; Squire et al., 2004).

1.4.3.3 Spatial navigation theory

Whereas the declarative memory perspective began with the characterisation of amnesic patients, the spatial navigation view began with the discovery of place cells and was more focused on the cellular representations. The original spatial navigation view (O'Keefe and Nadel, 1978) proposed the hippocampus supported a map that was a Euclidean coordinate space dedicated to allocentric physical space where place cells index the location on the map, providing the basis for spatial navigation and memory. The hippocampal role in memory is to bind them to their spatial context. The periodic lattice of grid cells was said to provide a metric for space that supports path integration by self-motion cues – the mechanism by which location on the map is updated (McNaughton et al., 2006). A wealth of computational modelling explains how these cell types could support spatial processing and navigation (Bush et al., 2015; Hartley et al., 2013). Further cementing the hippocampal role in spatial navigation were studies demonstrating that lesions to hippocampus led to dramatically impaired performance in the Morris water maze, wherein rats escape from the water by remembering the location of a hidden platform and swimming to it, including by taking novel shortcuts (Morris et al., 1982).

There is therefore very convincing evidence both in terms of cellular representations and lesions that the hippocampus plays a crucial role in spatial navigation. However, the issue with at least the original spatial navigation theory is that it proposes the hippocampus is dedicated to spatial processing (O'Keefe and Nadel, 1978). It fails to account for a wide range of behavioural and neural data, such as the non-spatial cells described above (Wood et al., 1999), implicating the hippocampus in non-spatial tasks (Eichenbaum et al., 1999). Despite this, there are still relatively narrow focuses on the spatial processing properties of the hippocampus (Hartley et al., 2013) that do not account for its non-spatial properties, thus failing to offer a unified account of hippocampal function.

1.4.3.4 Relational theory

Research on declarative memory in humans and on spatial navigation in rodents perused largely independently, but recently there has been increasing interest in merging these fields (Eichenbaum and Cohen, 2014). The relational theory attempts to account for observations in both literatures, and can be viewed as an extension of the declarative memory theory. The relational theory proposes hippocampus as supporting declarative memory by providing a general relational processing mechanism across different tasks and modalities (Cohen and Eichenbaum, 1993; Eichenbaum and Cohen, 2014). Hippocampus

enables relational representations by not only forming memories but also linking memories via their common elements, creating a memory space (Figure 1.15)(Eichenbaum and Cohen, 2014; Eichenbaum et al., 1999). Linking memories allows them to be recombined and flexibly used to make novel inferences.

Relational processing is facilitated by three representational schemes: associative representations, sequential organisation, and relational networking (Eichenbaum, 2004). As we have discussed, associative representations are supported by fast learning in recurrent networks with converging inputs. The sequential organisation of memories, i.e. the ordering of their elements as they unfolded in time, can also be supported by the dynamics in these recurrent networks. In particular, the sparseness of the recurrent connectivity inherently produces asymmetric connections (since the likelihood of a reciprocal connection is low) resulting in unidirectional enhancement of the recurrent connectivity that therefore represents a sequence. Finally, these same mechanisms can also support linking distinct memory representations into a relational network. These networks include cells that receive no external inputs but become associated with a particular sequence or with overlapping sequences. When these cells are activated by separate episodes sharing overlapping features, they can allow the association of separate episodes that share features. Hence these relational networks can link related memories, allowing them to be recombined in novel ways and facilitating inferences (Eichenbaum, 2004).

The relational account of the hippocampus is able to account for observations that the other theories cannot. For example, the non-spatial odour cells discussed above (Wood et al., 1999) are entirely predicted by the relational theory but are a strong challenge to any proposition that the hippocampus is dedicated to space. Furthermore, whist in this study some cells coded for multiple task variables (e.g. odour and location) – as expected by cells representing their association; other cells coded for only one variable, such as odour regardless of location across different trials, or a particular phase of the trial, such as approaching a stimulus, regardless of its identity. These cells could serve as those that represent common features across episodes to link memories acquired in different experiences (Eichenbaum, 2004). Due to this linking of memories, the relational theory therefore naturally accounts for the ability to generalise across memories, explaining why hippocampally lesioned animals are impaired on transitive inference tasks (Bunsey and Eichenbaum, 1996; Dusek and Eichenbaum, 1997). Such generalisations across memories were not originally accounted for by CLS or the original declarative memory view. Indeed one of the updates to CLS mentioned above was to account for this (Kumaran et al., 2016).

The generalisation across memories in that case is facilitated by recurrent activation in a loop including entorhinal cortex (Kumaran and McClelland, 2012) rather than within hippocampus via linking neurons, as proposed in the relational theory.

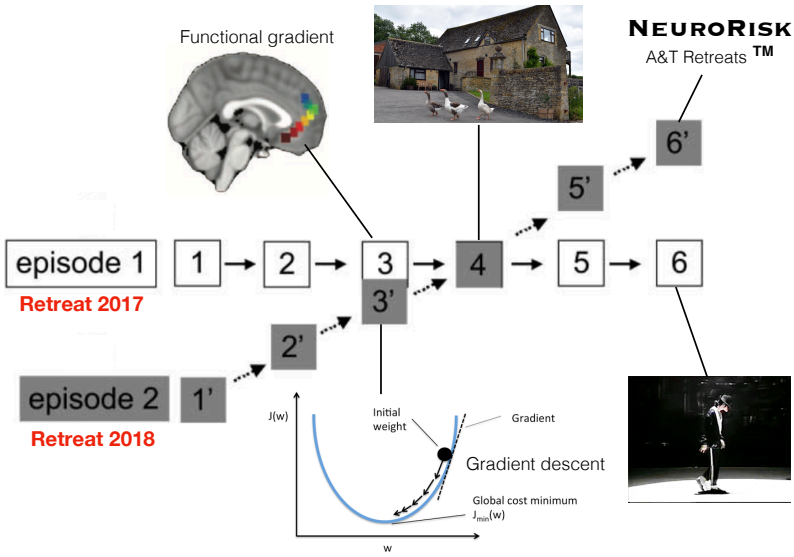


Figure 1.15 **Relational theory of hippocampal function**. A simple relational network is shown. Different memories (episodes) are organised in to a memory space that links episodes by their common elements, facilitating inferences among indirectly related events. For example, two different computational neuroscience lab retreats (episodes) partially overlap in that they shared some elements (the cottage fully overlapped and the word "gradient" partially overlapped). Other elements differed such as evening entertainment (see Appendix for details of NeuroRisk). These two memories therefore become linked by their common elements allowing their relationship to be inferred. Figure gratefully borrowed from Alexa Constantinescu, originally adapted from (Eichenbaum, 2004).

1.4.4 Summary

By approaching the hippocampal formation and its role in cognitive maps from (broadly speaking) Marr's three levels, we have seen these different theories of the computational goals of the hippocampus utilise its desirable information processing properties and explain the observed neural representations to different degrees. Across these theories a few general principles emerge. The neuroanatomy and neurophysiology of the hippocampus suggests a number of desirable information processing properties for forming fast memories. Specifically it is proposed the hippocampus is responsible for fast formation of declarative memories and that these memories index the original representation in cortex, which slowly extracts the statistics of memories. Memories can

become independent of hippocampus over time accounting for neuropsychological observations.

One of the most important differences between the theories is the primacy of spatial representation in the hippocampus. The spatial navigation view suggests the primary organising principle of hippocampal representations is space, upon which memories are formed, but struggles to account for non-spatial representations in hippocampus. The other views emphasise the critical function is declarative memory, of which spatial memory and its representation is a significant example. The relational theory in particular suggests the hippocampus is not dedicated to spatial cognition, "but organises experiences in memory, for which spatial mapping and navigation are both a metaphor for and a prominent application of relational memory organisation" (Eichenbaum, 2017b). It is therefore important to investigate the general principles under which the representations observed in a spatial context emerge, as this may tell us something general about the organisation of knowledge in these systems. We explore this in chapters 4 and 5.

An important component missing from the theories described above is how knowledge abstracted from memories and represented in cortex is generalised to novel experiences, placing them in context of previous learning. Furthermore, although there have been successful attempts to learn spatial representations in artificial agents (e.g. (Banino et al., 2018; Cueva and Wei, 2018)), they don't incorporate the majority of the properties of cognitive maps we described above. We explore this in chapter 5.

1.4.5 Work in this thesis related to cognitive maps in the hippocampus and entorhinal cortex

We take the view that the hippocampal formation is not dedicated to spatial cognition. Rather, in chapters 4 and 5 we explore the general principles that may give rise to the observed neural representations in the H-EC, as well as support flexible behaviour. In chapter 4, we show that by reconceptualising space as a state space, grid cell-like representations can be obtained as an efficient, low dimensional representation of 2D-distributed place cells, suggesting they summarise the statistics of 2D space. We then show that grid-like representations may be used to navigate without searching more than one step ahead, but retain flexibility to the goal location since the structure of the environment is stored in the grid cells. In chapter 5 we shall see that when the algorithmic (information processing) components described above are assembled together to achieve the computational goal of generalisation, an artificial neural network learns representations that resemble the neural implementation of the cognitive map.

Hence, taken together with the previous section of the introduction, in this thesis we address how internal models support a few different forms or key components of behavioural flexibility, one in each of the four main chapters: 1) Monitoring the current strategy and requirement for searching for alternatives in a changing environment, 2) Facilitating the updating of or disengagement from internal models of the current task, 3) Representations that allow efficient planning but retain flexibility to changes in goal location, and 4) Generalisation of previously learnt knowledge to new situations. Each of these is, in some way, concerned with the neural processes and representations that allow agents to exhibit flexible behaviour in response to changes in the environment or internal goals.

2 Neural substrates tracking belief uncertainty

2.1 Abstract

Adapting our behaviour to a changing world is a key component of flexible behaviour. This process requires the ability to monitor the quality of the current strategy in order to evaluate whether behavioural adaptation is required. Furthermore, uncertainty in internal models representing beliefs ought to be increased when these beliefs no longer accurately reflect the environment. This chapter is principally concerned with the former point – the monitoring of the current strategy, whilst the next chapter follows on from this one to addresses how uncertainty in internal models may be increased in light of realisation the current strategy is no longer adequate. In this chapter, we present a simple exploitation-exploration task designed to drive participants through periods of forming and revising beliefs about the inferred state of the environment, along with a Bayesian model to model their behaviour. Using functional MRI, we found a network of brain regions known to be involved in model-based choice centred on the medial orbitofrontal cortex and hippocampus to be engaged during exploitative behaviours. In contrast, we found the anterior cingulate cortex (ACC) amongst other regions to be more engaged during exploratory choices. Furthermore, specifically the ACC was engaged at the switch from exploitative to exploratory periods as well as during increases in the uncertainty of the current belief about the state of the world during stable exploitative periods. These results are consistent with a role of the ACC in monitoring the quality of the current strategy and initiation of exploratory behaviour, as well as behavioural flexibility more generally.

2.2 Introduction

Flexibly adapting our behaviour to a changing world is one of the core functions of the brain. This process requires updating internal models of the environment to allow accurate predictions that guide decision-making. As we discussed in the introduction, there are (at least) two key components to this updating process. First, it requires some mechanism for evaluating whether the current strategy is worth pursuing. Second, once the need for behavioural adaptation has been detected, the internal model(s) guiding decision making need to be updated to accurately reflect the environment.

This and the following chapter are together concerned with the neural substrates of such flexibility. We explore the monitoring of the current strategy as well as how this information may be used to update internal models guiding decision making via neuromodulatory systems. This chapter addresses the neural substrates underlying the monitoring of the reliability of the current strategy and "switching in" to states of exploring alternative courses of action. In the following chapter we ask how internal models may be updated in light of evidence suggesting the current strategy is unreliable, and attempt to link these two processes via neuromodulation.

In this chapter, we first present a simple probabilistic exploration-exploitation learning task designed to drive participants through periods of forming and revising beliefs about the inferred state of the environment. Reward contingencies are stable for a period of time resulting in periods of exploitation, followed by a change in reward contingencies and resultant exploratory behaviour to discover the new state of the environment. The probabilistic nature of the task means participants are constantly having to make inferences about the underlying state of the environment. This allows us to test for regions sensitive to changes in the certainty of their beliefs (inferred strategy reliability). Furthermore, as we will show in the next chapter, it allows us to measure an internal model representing these dynamic beliefs as well as investigate how this model may be updated.

We also present in this chapter behavioural analyses along with a Bayesian model to model participants' behaviour. We then present univariate brain analyses that reveal the different neural substrates of exploitation and exploration and speak to a role of ACC in monitoring the reliability of the current strategy, as expected from the literature discussed in the introductory chapter. It should be noted here that the primary goal of this experiment was to test for the mechanisms involved in updating the internal model, which are the focus of the next chapter. We did not design a task intended to distinguish between the various different proposed functions of the ACC, such as computing foraging value or cognitive control (Kolling et al., 2016; Shenhav et al., 2016).

2.3 Methods

2.3.1 Participants

Twenty-two healthy human volunteers participated in the experiment. Participant identities were anonymised for analyses. All participants were right-handed, aged between 20 and 31 (mean age: 26), and 8 were female. All participants had normal or corrected to

normal (with contact lenses) vision. Three participants had to be excluded because of technical difficulties with fMRI, and three had to be excluded from pupil analyses because of technical difficulties with eyetracking in the fMRI scanner. We estimated the required sample size a-priori. Based on a previously published result demonstrating an effect of model updating in the ACC (O'Reilly et al., 2013b), we performed a simple power calculation: a-priori sample size for a t-test. This effect from O'Reilly et al. (2013) had a Cohen's d of 3.68. For an anticipated effect size of Cohen's d = 3.68, desired power level of 0.9, and probability level of p = 0.001, the required total sample size for a two-tailed hypothesis was computed as 12 participants. The same calculation based on pupil effects (Nassar et al., 2012) (Cohen's d = 3.58) also yielded a sample size of 12. We collected data from a number of participants well in excess of the number required for the known size effects, and in line with other papers in the field. The study was approved by a local University of Oxford ethics committee (ref: MSD-IDREC-C1-2013-171), and all participants gave written informed consent.

2.3.2 Task

To establish a task environment in which participants were constantly making inferences about the current state of the environment, we used a 4-arm bandit task (Figure 2.1). At any given time, one of four response options had a high (70 or 90%) probability of delivering reward and the other three had a low (20%) probability of delivering reward, if selected. The high reward option switched after a variable number of trials (mean 20 ± SD 5) and independently of the participants' behaviour – adding to the unpredictability (due to the variable length of the exploratory period). When the high reward option switched, it was randomly selected with equal probability to have a 70 or 90% chance of reward delivery. Participants' task was to maximise reward, hence the optimal strategy was to always select the current high reward option. Due to the unpredictable length of the exploit period, participants therefore had to constantly infer the current state of the environment, i.e. which option was the current high-reward option. This simple task forced participants between periods of exploiting the high reward option and exploring for the new high reward option after it switched to a new option. Participants were explicitly instructed about the reward structure before beginning the task.

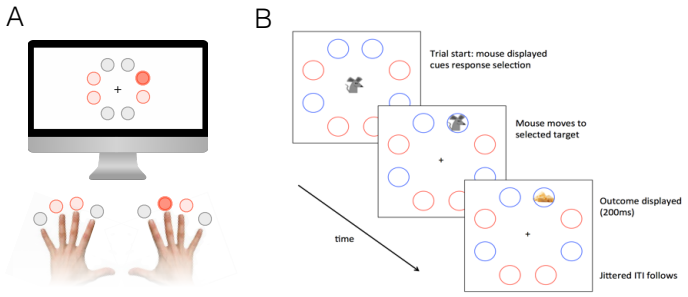


Figure 2.1 **Task to drive participants between forming and revising beliefs.** a) Participants chose freely between 4 available options on each trial. A total of 8 options were used in the experiment but only 4 were used in each block (indicated by option colours: red circles above the fingers denote the available options, and the dark red circle denotes the selected option). b) The task was implemented in a game where participants controlled a mouse seeking cheese or apples, counterbalanced across participants. Participants could select any of four available locations, denoted in this case by blue circles, again counterbalanced across participants.

Although there were only four options available for selection, there were eight options on the screen at any given time (Figure 2.1, see below for details and motivation). The four available options were distinguished from unavailable options by their colour (blue or red, counterbalanced across participants), and the available options changed in different runs of the task (participants completed four runs of the task – see below).

The task was embedded in a game in which participants controlled a mouse seeking cheese (or seeking apple, counterbalanced across participants, Figure 2.1). Following selection of an option, the mouse moved to the option and the outcome (reward, e.g. cheese, or no reward) was revealed and remained on the screen for 200ms. The next trial began soon after (jittered with a uniform distribution between 0 and 4 seconds; mean ITI of 3.44s). The short spacing between trials was necessary, both because we required large numbers of trials to drive participants between many periods of exploitation and exploration, and because we focused our analysis on a subset of trials. The close trial spacing was also designed to encourage a sense of consistent engagement with a particular strategy within exploit blocks.

Participants completed two sessions – one behavioural and pupillometry only, and one additionally with fMRI. Each session consisted of four runs of 200 trials (1600 trials total per participant). The available (blue/red) stimuli switched across runs (see below). Before the first session, participants were given detailed instructions of the task. These included information about underlying environment statistics: they were explicitly told the

probabilities of reward, although not the true underlying frequency at which the high reward probability jumped. Participants were told the probabilities so they would have a good understanding of the task when doing the fMRI session. Selecting one of the eight options corresponds to a button press for one of the eight fingers. Before starting the first session, participants were trained on the mappings between the options and the response buttons by doing 300 trials of a training task, whereby one option would be highlighted and they would be required to press the corresponding response button. After this, they performed the four runs of the task that comprised the behavioural session. The following day, participants completed the fMRI session, which similarly required participants to complete four runs of the task. The task and all instructions were the same across the two sessions. Participants were additionally instructed to centrally fixate the fixation cross throughout the whole experiment in both the behavioural and fMRI session.

The motive for having eight options but only four of which were available for selection was to test whether there was any difference in the neural encoding between options available for selection but not selected, and those not available for selection. The configuration of available options differed across the runs such that all options were available in two of the runs (to provide training and testing sets of the multivariate model). Therefore the unavailable options were intended to act as a baseline. Analyses (multivariate decoding analyses) testing for differences between available but not selected and unavailable options did not yield significant results (data not shown).

2.3.3 Behaviour

We classified trials as either exploitative or exploratory. We relied on as few assumptions or arbitrary cut-offs as possible when classifying the trials by using a simple heuristic. We defined the start of a period of exploitation as the first trial on which participants selected the true high reward probability option and received reward. The task provided a clear behavioural marker of the onset of exploration, which was defined as the first trial on which participants switched away from making the previously repeated exploitative selection.

We note here that some critical analyses, both in this chapter and next, focus on the exploit period. These analyses are restricted to only those trials in the middle of the exploit period that are at least 5 trials from the adjacent explore periods in either direction. This ensures that any observed variation in neural activity is due to changes in belief rather than choices. Furthermore, it ensures that any trials entering these analyses are those in which

behaviour is stable for >10 trials and it can be reasonably assumed that participants are exploiting. We state which analyses are focused on these trials when we discuss them.

2.3.4 Bayesian learning model

To generate trial-by-trial estimates of the certainty of participants' beliefs to regress against neural responses, we constructed a normative Bayesian learning model. The model reflected information communicated to participants – specifically, that one option had a high probability of reward and the remainders a low probability of reward. The model inferred the trial-wise probability that each of the four options was the high reward probability option, based on the data observed by each individual participant. For each trial, we calculated the entropy of the model's posterior probability distribution over options, based on evidence up to and including that trial. We used this as a proxy for the entropy in the participants' internal model, as presented in the next chapter. Entropy is low when the probability mass of the posterior is mostly over one option, i.e. participants are certain.

The model estimated the probability that each of four hypotheses $H_t \in \{1,2,3,4\}$ is correct – one hypothesis for each of the four locations $L_t \in \{1,2,3,4\}$ being the high reward-probability location. In each of these hypotheses, the three low probability locations were assigned p(reward)= 0.2, which was fixed. The other, high reward-probability location was assigned probability *P*, where *P* could take any value in the range 0.3 to 1. We allowed *P* to be more flexible to allow for possible differences in participants' estimates of probabilities, but constrained it to be greater than the default payoff of 0.2 by imposing a lower bound of p=0.3.

This information can be thought of as represented in a payout matrix – a matrix whose elements correspond to a probability of reward being delivered. The payout, or outcome, on trial *t,* O_t, was determined by a binomial distribution parameterised by *q*:

$$O_t \sim B(q)$$

Equation 2.1

where *q* was obtained from the payout matrix **M**. **M** depends on which location is the true high reward-probability location *H*, what the payout probability *P* at the high reward-probability location is, and which location *L* was selected on that trial.

The posterior probability that each location was the correct one, H, and the payout probability P takes some value p, was found by considering the payout probability q associated with the cell (h, L_t, p) in the payout matrix:

$$p(H_t = h \cap P_t = p|O_{1:t}, L_{1:t}) = p\left(q = \mathbf{M}_{h,l,p}\middle|O_{1:t}, L_{1:t}\right)$$

Equation 2.2

The posterior probability over q, $p\left(q = \mathbf{M}_{h,l,p}\middle|O_{1:t}, L_{1:t}\right)$, was given using Bayes' rule:

$$p\left(q = \mathbf{M}_{h,l,p}\middle|O_{1:t}, L_{1:t}\right) \propto p\left(O_t\middle|q = \mathbf{M}_{h,L_t,p}\right) \cdot p\left(q = \mathbf{M}_{h,L_t,p}\middle|O_{1:t-1}, L_{1:t-1}s\right)$$

Equation 2.3

The likelihood $p\left(O_t\middle|q = \mathbf{M}_{h,l,p}, L_t\right)$ was simply:

$$p\left(O_t\middle|q = \mathbf{M}_{h,L_t,p}\right) = \begin{cases} \mathbf{M}_{h,L_t,p} & \text{for } O_t = 1 \\ 1 - \mathbf{M}_{h,L_t,p} & \text{for } O_t = 0 \end{cases}$$

Equation 2.4

meaning that the likelihood distribution was a 2D matrix over

$$h = \{1,2,3,4\} \times p = \{0.3: 0.01: 1.0\}$$

The prior $p\left(q = \mathbf{M}_{h,L_t,p}\middle|O_{1:t-1}, L_{1:t-1}s\right)$ was none other than the probability of the values H=h and P=p, given all previous observations and the switch probability s:

$$p\left(q = \mathbf{M}_{h,l,p}\middle|O_{1:t-1}, L_{1:t-1}, s\right) = p(H_t = h \cap P_t = p|O_{1:t-1}, L_{1:t-1}, s)$$

Equation 2.5

This was obtained from the previous trial's posterior after accounting for the possibility of a switch in *H* with a leak. The probability of a switch *s* was fixed and modelled as 1/20. This reflects the true underlying mean frequency of a switch, but was unknown to participants (we could model this flexibly – however, once in the fMRI session, participants likely have a rough gauge of how frequently there is a switch due to having already completed 800 trials of the task in the behavioural session). The prior was therefore given by:

$$p(H_t = h \cap P_t = p|O_{1:t-1}, L_{1:t-1}, s) = [p(H_t = h \cap P_t = p|O_{1:t-1}, L_{1:t-1}, s) \cdot (1 - s)] + [1 \cdot s]$$

Equation 2.6

where $[1 \cdot s]$ effectively applies a uniform distribution over the previous trial's posterior, weighted by *s,* allowing the model to 'forget' past outcomes (otherwise the estimates in the posterior will just converge to the mean of all estimates).

To obtain the model entropy over Hypotheses (i.e. the spread of belief over which of the 4 possible locations was the high reward location), we marginalized over p:

$$p(H = h|O_{1:t}, L_{1:t}, s) = \sum_{p} p(H = h, P = p|O_{1:t}, L_{1:t}, s)$$

Equation 2.7

This gave the marginalised posterior shown in Figure 2.2C. We then applied the formula for Entropy:

$$ENT = -\sum_{h=1}^{4} p(H = h|O_{1:t-1}, L_{1:t}, s) \log\big(p(H = h|O_{1:t-1}, L_{1:t}, s)\big)$$

Equation 2.8

This was the equation that gave the entropy regressor used in analyses and shown in Figure 2.2B.

2.3.5 fMRI acquisition and analysis

In the fMRI session, 4 blocks, or runs, of functional data were acquired, with mean ~12 minutes each. Data were acquired on a 3T Siemens TIM Prisma, using a 64-channel head and neck coil. Functional scans were collected using a gradient echo-planar imaging (EPI) sequence, Multi-band 4, with TR = 1.35s, TE = 32ms, flip angle = 74 degrees, and voxel resolution of 2x2x2mm. Structural scans were acquired using a T1-weighted MP-RAGE sequence, 1x1x1mm voxels. All scans were of axial orientation angled to ACPC, covering the whole brain.

Preprocessing and analysis of fMRI data was performed using tools from fMRI Expert Analysis Tool (FEAT), part of FMRIB's Software Library (FSL). Data were preprocessed using the FEAT default options. Registration of EPI images to high-resolution structural images and to standard (MNI) space was performed using FMRIB's Linear and Non-Linear Registration Tool, respectively (FLIRT and FNIRT). For GLMs, the four runs of the task for each participant were pooled using fixed effects analyses, and then high-level analyses across participants were performed using mixed effects analyses. We locked each trial event to the time of outcome presentation and convolved the BOLD signal with the canonical HRF.

Multiple comparisons correction was performed with cluster mass-based permutation testing with the FSL function randomise. The clusters were thresholded voxelwise at $p=0.001$ and cluster masses surviving correction at $p<0.05$ are presented.

2.3.6 General Linear Models

We ran General Linear Models (GLMs), where the dependent variable was univariate brain activity. Analyses were carried out either on all trials or only included trials in the exploit period that were at least 5 trials from the adjacent explore periods. The GLMs with results provided in this chapter are as specified below:

GLM1 included the following regressors: Main effect exploit trials, main effect explore trials. Dependent variable: Univariate brain activity. A [1 -1] contrast tested for differences between exploitation and exploration, with positive activation indicating regions more active in exploitation than exploration, and vice versa for negative activations. All trials were included in the analysis. The goal of this analysis was to identify regions that were differentially active in exploitation and exploration phases of the task.

GLM2 included the following regressors: Main effect (bias; column of ones), entropy, whether reward was delivered at outcome, whether participants changed their choice on the subsequent trial, whether participants transitioned from exploitation to exploration on the subsequent trial (last exploit trial), whether participants transitioned from exploration to exploitation on the subsequent trial (first exploit trial). Dependent variable: univariate brain activity. All trials were included in the analysis. The goal of this analysis was to identify regions sensitive to certain task factors, as well as to test for regions with activity at the transitions between exploitation and exploration (over and above that explained by the other task regressors, such as switching options).

GLM3 included the following regressors: Exploit, explore, and all the regressors from GLM2 except the main effect regressor. Dependent variable: Univariate brain activity. A [1 -1] contrast tested for differences between exploitation and exploration. All trials included in the analysis. The goal of this analysis was to investigate whether much of the difference between exploitation and exploration was explained by the task factors in GLM2.

GLM4 included the following regressors: Main effect, change in entropy. Dependent variable: Univariate brain activity. Only exploit trials (>5 from adjacent explore phase) included in the analysis. The goal of this analysis was to identify regions in which activity reflected changes in the model entropy.

2.4 Results

2.4.1 Behaviour

Participants' behaviour alternated between exploitation and exploration – as defined in the Methods – as they formed and revised beliefs about the current state of the environment (Figure 2.2A). During exploitation, the high reward option was repeatedly selected without variation, whilst in the exploration period action selection was heterogeneous.

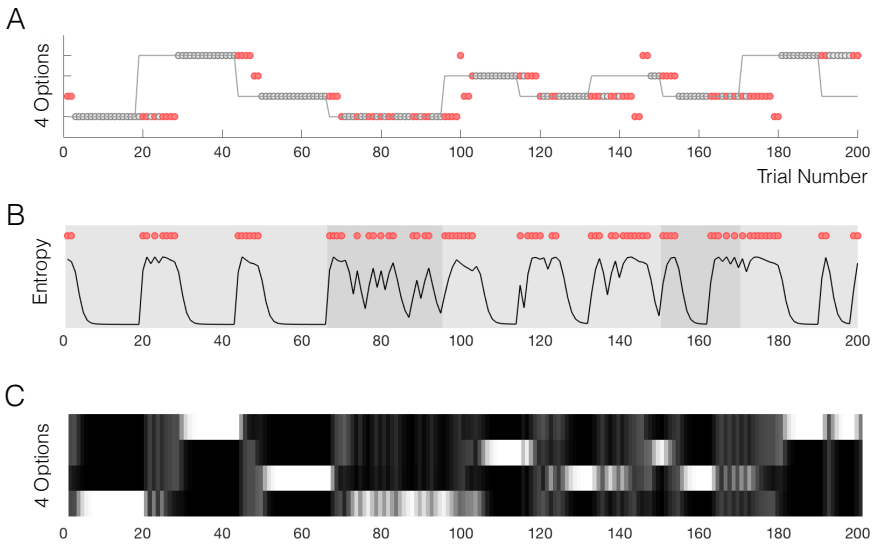


Figure 2.2 **Behaviour**. a) Example behaviour of a participant. Dot markers denote choices (location in y-dimension indicates which option was chosen) – grey circles are rewarded and red dots are unrewarded trials. The grey line indicates the true state of the world (high reward option), which the participant must infer. Note alternating phases of exploration and exploitation – the main analyses presented in the paper refer only to the exploitation phase in which there are no overt changes of action. b) Model entropy for the same run as in (a). Note that model entropy is low during exploit periods but increases following a reward omission (red dots are trials with reward omission). Background shading indicates the pay-out rate of the high reward option. During low pay-out (70%) exploit periods, indicated in dark grey, entropy tends to remain higher; during high pay-out periods (light grey) entropy reaches a floor during exploitation. c) State space of the model – 4 horizontal tracks represent the 4 possible states of the environment; shading indicates the probability assigned to each state by the model (light colours indicate high probability).

The possibility for voluntary exploration in the task created a constant pressure to evaluate whether the state of the environment had changed. The entropy of the estimated posterior of participants' beliefs (about the position of highest reward) was used as a measure of certainty – entropy is low when participants are certain. As expected, entropy was generally low in the exploit phase and high in the explore phase (Figure 2.2B,C). Within

exploitation phases, model entropy generally decreased over time as participants continued exploiting a single option, but varied due to probabilistic omissions of reward, and accordingly depended on the payout rate of the high reward option (70% or 90% reward) (Figure 2.2B,C).

2.4.2 Exploration and exploitation engage different neural substrates

The overall distribution of brain activity in explore and exploit phases of the task differed (GLM1, Figure 2.3). The explore phase engaged a frontal-parietal action network and the dorsomedial prefrontal cortex (dorsal anterior cingulate dACC and adjacent preSMA). The exploit phase engaged medial orbitofrontal cortex (mOFC) and the hippocampus, two regions known to be engaged in model-based choice (Boorman et al., 2016; Jones et al., 2012; Noonan et al., 2010; Wikenheiser et al., 2017). These results are consistent with previous reports of differences between exploitation and exploration that found a similar distribution of activity across the two phases (Daw et al., 2006; Kolling et al., 2012).

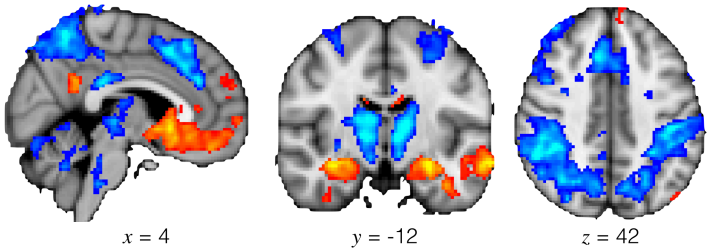


Figure 2.3 **Exploitation and exploration engage different brain regions**. Z-statistic map for the [1 -1] contrast in GLM1. Positive activation (hotter colours) denotes regions more active during exploitation than exploration, and vice versa for negative activation. Medial OFC and hippocampus were more active during exploitation, and a frontal-parietal action network was more active during exploration. Image thresholded at p<0.001; corrected for multiple comparisons.

2.4.3 Task variable-related activity explains differences between exploitation and exploration

Several task factors differed between explore and exploit phases. In addition to state uncertainty (model entropy) as mentioned above, the frequency of option choice switching and frequency of reward omission were also higher in explore periods. We first investigated whether modulation of these factors also modulated activity in similar regions to those engaged by exploitation and exploration, and second, tested whether this explained most of the variance differing between exploitation and exploration. A GLM containing

these task variables (amongst others; GLM2) indicated that response switching and high model entropy predicted activity throughout the frontal parietal network and dorsomedial frontal cortex (Figure 2.4). Reward omission predicted activity in dorsomedial frontal cortex only, and reward activated ventral striatum. Low model entropy predicted activation of medial OFC and hippocampus.

We next tested whether these factors explained much of the variance that drove the differences between exploitation and exploration. Adding these factors to a GLM testing differences between exploitation and exploration revealed that most of the difference between these states was accounted for by variance driven by these task factors (Figure 2.5). This suggests these variables capture a lot of the meaningful factors of variation in the task.

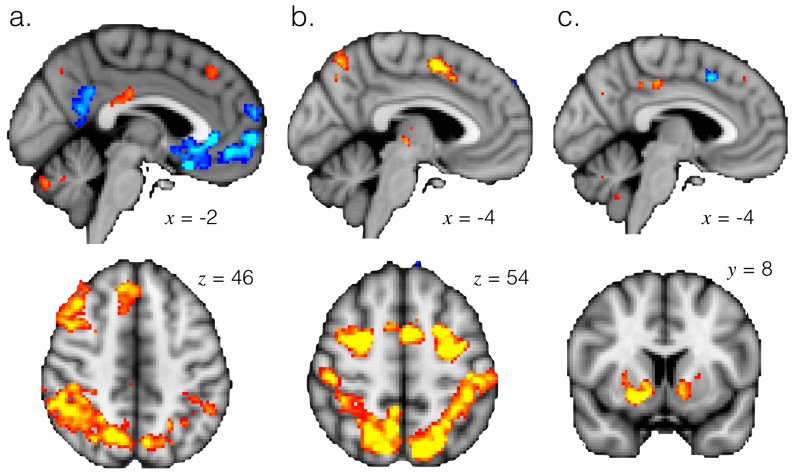


Figure 2.4 **Activation related to task factors**. Z-statistic maps for the entropy (a), whether changing response (b) and reward (c) regressors in GLM2. These variables explain variance in very similar brain regions to those in which variance is explained by differences between exploitation and exploration. Panels a and b Image thresholded at p<0.001; corrected for multiple comparisons. Panel c image thresholded voxelwise at p<0.001, uncorrected.

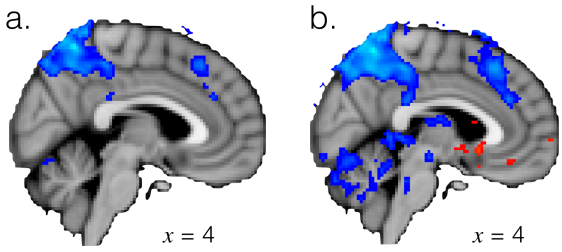


Figure 2.5 **A large amount of the difference between exploitation and exploration is captured by task factors**. Z-statistic map for the [1 -1] contrast in GLM3, testing for differences between exploitation and exploration having regressed out effects of the task factors reveals there is much less variance accounted for by differences between exploitation and exploration. a) Image thresholded voxelwise at p<0.001, uncorrected. b) Image thresholded voxelwise at a more liberal threshold of p<0.01, uncorrected.

2.4.4 Transitioning from exploitation to exploration engages specifically ACC

Given the remodelling of population dynamics in ACC during abandonment of a belief (Karlsson et al., 2012) and its role in regulating the degree to which an internal model guides behaviour (Tervo et al., 2014), we predicted that ACC would be engaged when participants switched from a state of exploitation to exploration. Indeed, we found the last trial of exploitation engaged specifically the ACC (GLM2)(Figure 2.6). The activity observed in ACC in this analysis is activity over and above that explained by changing response/option selection (included as one of the co-regressors in GLM2), which (by definition) occurs on the last trial of exploitation. Note that this effect is specific to exploration initiation, and was not observed when switching strategy from exploration to exploitation (Figure 2.6), suggesting a particular role of ACC in disengaging from the current strategy.

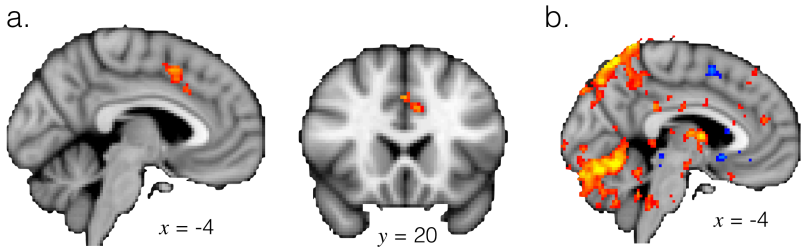


Figure 2.6 **ACC is engaged when transitioning from exploitation to exploration**. a) Z-statistic maps for the regressor for switching from exploitation to exploration, as in GLM2 (note this is variance explained by the regressor over and above that explained by the other regressors in GLM2, such as changing response and reward). Image thresholded at p<0.001; corrected for multiple comparisons. b) Z-statistic map at a more liberal threshold of p<0.05, voxelwise and uncorrected, for the regressor for switching from exploration to exploitation, showing this transition does not engage ACC. Hence ACC transition-related activity appears to be specific for switching from exploitation to exploration and not vice versa.

2.4.5 Increases in uncertainty of the current belief engage specifically ACC

Given the role of the ACC described above in monitoring the reliability of the current strategy, we predicted its activity would be explained by changes in the certainty of the current strategy. We tested this by asking where in the brain activity was explained by changes in model entropy induced by outcome observation – an estimate of the change in the participants' belief the current strategy is correct (GLM4). Importantly, we restricted this analysis to the exploitation period when the chosen option was by definition constant, by

only including exploit trials that are at least five trials away from adjacent explore periods (in either direction). By limiting our analysis to these core exploitation trials, we were able to dissociate uncertainty about the state of the environment from variability of action, and therefore could identify neural activity concerned with changes in participants' beliefs in the absence of changes in overt behaviour (since there is no choice variability). As predicted, difference in entropy induced by outcome observation in core exploit trials specifically activated ACC (Figure 2.7). This activation was such that an increase in entropy (uncertainty that the current strategy or state of the world is correct) predicted stronger engagement of ACC, as expected from a role of ACC in processing feedback that should indicate changes of belief or strategy are necessary (Behrens et al., 2007; Karlsson et al., 2012; O'Reilly et al., 2013b).

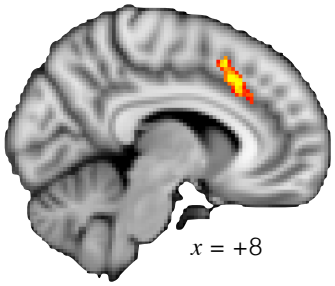


Figure 2.7 **ACC is engaged when uncertainty in the current belief increases**. Z-statistic maps for the difference in entropy during core exploitation trials only (those >5 trials from adjacent explore periods), showing ACC activity increases when uncertainty increases despite stable behaviour. Thresholded at p<0.001 and corrected for multiple comparisons.

2.5 Discussion

In this chapter we have presented a simple probabilistic learning task that drives participants between periods of forming and revising beliefs. We presented univariate analyses demonstrating engagement of different brain networks during exploratory and exploitative decisions; of particular interest in this thesis we find ACC and mOFC are more engaged during exploratory and exploitative choices, respectively. These differences appeared to be largely driven by task variables that we modelled, specifically: belief uncertainty (entropy), changing responses and reward outcome. Furthermore, specifically ACC was engaged when switching from exploitation to exploration as participants abandoned their beliefs, as well as during increases in the uncertainty of beliefs during periods of stable exploitative behaviour.

The univariate differences between exploitation and exploration are consistent with those previously observed (Daw et al., 2006). The regions we found to be more engaged during exploitative choices – the mOFC and hippocampus – are those thought to represent internal models of the task, consistent with such choices being guided by an internal model rather than random exploration. The regions more engaged during exploratory choices – the ACC and a frontoparietal action network – are those previously implicated in exploratory behaviour and increased learning (Behrens et al., 2007; Daw et al., 2006; Kolling et al., 2012; O'Reilly et al., 2013b).

Our ACC results are consistent with its proposed role in monitoring the reliability of the current strategy and updating or disengaging from internal models when a change in the current state of the world is detected. Engagement of ACC on the last trial of exploitation could reflect the abrupt, widespread and coordinated remodelling of population activity observed in rodent ACC as the current strategy is abandoned in favour of exploration (Karlsson et al., 2012). That this is activity over and above activity related to changing responses is consistent with the previous observation that when two cues required the same motor response, only the cue that signalled the requirement of updating internal models engaged ACC (O'Reilly et al., 2013b). The ACC activity in core exploit trials here reflects changes in model entropy. This indicates it is sensitive to changes in the uncertainty of beliefs, which can be interpreted as monitoring the quality of the current strategy and indicating the need for flexibility of behaviour or learning. This is consistent with the proposal that ACC signals reflect the value of exploring alternatives relative to staying with the current strategy (Hayden et al., 2011; Kolling et al., 2012) and track changes in the environment (Behrens et al., 2007). It must be noted, however, that our study was not designed to dissociate between neural correlates of monitoring and updating the current strategy, i.e. the agent's own policy and need for exploratory behaviour, from monitoring and updating the current state of the external world, independent of any requirement to change policy. Even though these two things are deeply related, they are conceptually different. In the current study, they are entirely correlated: when the world has changed, the policy ought to be updated. Whilst it is not clear whether activity in ACC in the present study is reflecting the updating of internal models or the drive for exploratory behaviour, it is likely both: it has been suggested that ACC activity at the time of choice reflects the value of exploring alternatives, and during learning, reflects the degree to which internal models must be updated (Kolling et al., 2016). We do not dissociate between choice and feedback due to the fast trial design in our task. Overall, our ACC results are consistent with a general

role in monitoring the quality of the current strategy and/or state of the world, and facilitating behavioural adaptation.

We note again that our task was not designed to interrogate the precise function of the ACC. In addition to not dissociating between a role in monitoring the current strategy and monitoring the state of the world (independent of strategy/policy), we also cannot and do not attempt to dissociate between other proposed functions, such as signalling foraging value or cognitive control (Kolling et al., 2016; Shenhav et al., 2016). The cognitive control perspective proposes that in a foraging context ACC signals choice difficulty (Shenhav et al., 2014), which predicts exactly the activations observed in our study. Indeed, the conditions in which we observe ACC activity are those in which choices are most difficult. Therefore, although the story presented here is rather one-sided, we do not dissociate between these theories. Although they differ on the precise underlying computations, we focus here on their common ground: both emphasise the importance of the ACC in behavioural flexibility.

The primary purpose of this study was to rather investigate how internal models are kept flexible. An intriguing possibility is that ACC acts in concert with the noradrenergic system, which, similar to ACC, has been ascribed numerous different roles that may be broadly summarised as facilitating behavioural flexibility. In the next chapter we explore the possibility that the ACC influences noradrenergic systems, and, by measuring a dynamic internal model, we explore whether noradrenergic systems may increase uncertainty in internal models.

3 Control of entropy in neural models of environmental state

3.1 Abstract

Humans and animals construct internal models of their environment in order to select appropriate courses of action. The representation of uncertainty about the current state of the environment is a key feature of these models that controls the rate of learning as well as directly affecting choice behaviour. To maintain flexibility, given that uncertainty naturally decreases over time, most theoretical inference models include a dedicated mechanism to drive up model uncertainty. In this chapter, we test the long-hypothesised role for noradrenaline in increasing the uncertainty, and thus flexibility, of neural models. Pupil diameter, a non-invasive proxy for noradrenaline release, predicted increases (but not decreases) in uncertainty in a neural state model encoded in medial orbitofrontal cortex, as measured using multivariate functional MRI. Furthermore, activity in the noradrenergic locus coeruleus and in anterior cingulate cortex predicted pupil diameter. These results provide evidence for top-down, neuromodulatory control of uncertainty in neural state models.

3.2 Introduction

In order to make sense of their environment, organisms must infer the underlying causes or states of the world that generate the constant stream of observed sensory data. As long as the environment remains stable, certainty about its state increases as the number of data samples increases. From a Bayesian perspective, if beliefs about the environment are expressed as a posterior distribution over possible states, the entropy, or uncertainty, of this posterior decreases with the number of concordant data points. This decrease in uncertainty raises the problem of inflexibility in dynamic environments: the better established one's model of the environment, the less flexible one's beliefs are.

As we mentioned in the introduction, to counter the problem of inflexibility, theoretical accounts of inference that capture uncertainty must include an explicit

mechanism for driving up entropy – injecting uncertainty back into the model's state estimate to make rapid changes of belief possible. The simplest approach is to include a constant 'leak' term (Usher and McClelland, 2001) such that the model's belief distribution continuously decays towards a state of agnosticism (uniform belief distribution over possible states).

In this case, the process by which model entropy is increased is theoretically separable from the process by which it is decreased. If the 'leak' were removed, the model entropy would still decrease with each new concordant observation, as this is a natural consequence of gaining information, but could not increase rapidly to accommodate environmental change. Such a model would be able to infer the state of a static environment successfully but could not efficiently adapt to change (as we showed with the simple demonstration in the introduction chapter). However, the question of whether there is a dedicated mechanism for increasing internal model entropy in the brain, as there is in theoretical models, has received little attention. Indeed in many experimental paradigms the processes for increasing and decreasing model entropy occur in parallel because the same information that signals a change in the environment also informs the observer about the new state of the environment. For example this is necessarily true in all reversal paradigms with two options and linked probabilities, the most common class of tasks used to probe inference. Hence the process of injecting uncertainty into the state estimate often occurs in parallel with inference about the new state that reduces uncertainty.

Furthermore, investigating mechanisms for increasing entropy in neural models is difficult due to the requirement of measuring the neural model representing beliefs. Internal models can be indexed behaviourally by measuring participants' beliefs, which are in turn indexed by analysing their choices (for example (Behrens et al., 2007; Nassar et al., 2012)). However, here we seek to directly identify predictors of changes to an internal model representing beliefs by measuring the neural model directly. This allows us to get closer to the neural representations of beliefs and the neural mechanisms of how they are updated (Boorman et al., 2016). In order to do this, it is necessary to obtain a trial-by-trial neural index of the internal model that provides a continuous measure of participants' dynamic beliefs. This has been impressively achieved before with fMRI in humans using repetition suppression, which allowed identification of neural learning signals that predicted changes to a neural internal model (Boorman et al., 2016). One downside to this repetition suppression approach is the large number of stimulus presentations required to probe the internal model, which would not be feasible in the current study.

We therefore turned to multivariate methods to provide a continuous trial-by-trial neural measure of the internal model. Typically, multivariate analyses are used to classify discrete categories (e.g. “this is an image of a dog”) or to determine the similarity between the patterns representing discrete categories. The issue with these typical methods is that participants’ beliefs in probabilistic learning tasks are not discrete, but continuous. Therefore in this chapter we obtain a continuous trial-by-trial neural measure of the internal model using multivariate analyses, by using a measure that is similar to the decoding accuracy on each trial, which we term representation strength. We were able to regress estimates of participants’ beliefs as well as changes in pupil diameter against this measure to investigate how beliefs may be represented and updated, respectively.

In this chapter we investigate the hypothesis that the brain has a mechanism for increasing entropy in internal models. Using the same task and data as the previous chapter, we first established that we could measure uncertainty in a neural representation of environmental state using multivariate analysis of functional MRI data. Then we identified predictors of increases in this uncertainty, in the absence of overt changes in behaviour. Participants formed a probabilistic model of the state of the environment that could be decoded from the mOFC. This was measured using multivariate analyses of the fMRI data wherein the decoder was trained to predict choice, but then we looked for regions where this representation of choice was predicted by the entropy derived from the Bayesian model. In other words, this internal model in mOFC represented participants’ certainty about the state of the environment in that the strength of the representation of the currently selected option, relative to other options, depended on the Bayesian model entropy. We defined our internal model in this way because we were specifically interested in regions that were tracking the dynamic and probabilistic underlying (hidden) state of the task (rather than simply representing, for example, actions). As well as measuring a neural model with fMRI, we simultaneously measured pupil dilation – thought to index noradrenaline release to cortex. We found trial-by-trial changes in pupil dilation explained changes in uncertainty in the mOFC neural model (i.e. how well choice could be decoded), providing evidence that the mechanism for driving up entropy in cortical models of the environment could be neuromodulatory. Furthermore, activity in the ACC explained the strength of the pupil dilation response, suggesting ACC may be involved in regulating these neuromodulatory processes.

3.3 Methods

Since this chapter is a different set of analyses on the same dataset as the previous, the participants and task are the same, and we used the same fMRI registration and multiple comparisons methods.

3.3.1 fMRI preprocessing and multivariate measure of representation strength

We generated whole-brain beta maps on each trial to obtain the activity in each voxel on every trial. As in the previous chapter, we locked each trial event to the time of outcome presentation and convolved the BOLD signal with the canonical HRF. These single-trial beta maps were then used as inputs in to multivariate analyses and testing for relationships between univariate brain activity and pupil diameter. Data was concatenated across the four runs for each participant before GLMs were run. These GLMs contained a bias term for each run separately to regress out any differences across runs (i.e. the GLM had four bias/main effect regressors, each with ones for one of the runs and zeros for the other three).

We seek a measure of representation strength, i.e. a trial-by-trial continuous measure indexing the internal neural model representing participants' beliefs. To obtain a measure of representation strength at each grey matter voxel, we ran a multivariate searchlight analysis over the whole-brain grey matter. We trained and tested the model on different runs of the task. The configuration of available options differed across the runs, such that each option appeared in two different runs. We therefore split the runs in to two pairs – each pair containing all eight options. We trained the model on one pair and tested on the other, and then repeated, training and testing on the converse pairs. This resulted in a trial-by-trial measure of representation strength.

The data used as input to the multivariate analyses are whole-brain, standard space beta maps for each trial. We then focus our analyses on the cortex using a grey matter mask. At each grey matter voxel, a set of voxels for that searchlight centre is defined by including all voxels within a radius of 7 voxels from the searchlight centre and that lie within the grey matter mask.

Before training and testing the multivariate model, at each searchlight we performed dimensionality reduction on the voxels and ensured a balanced training set. We performed dimensionality reduction to denoise the data (Mante et al., 2013) and to facilitate training the multivariate model (so that we didn't have many more dimensions

[voxels] than exemplars [trials]). We concatenated the data across runs, and then at each searchlight centre extracted an nVoxels x nTrials matrix. To perform the dimensionality reduction, we carried out principle components analysis on this matrix resulting in nVoxels principle components (PCs) – each a vector of length nVoxels. We then kept only the top 20 PCs. The eigenvalues associated with the remainder of the PCs were all near 0, suggesting we captured most of the variance in the top 20 PCs. On each trial, we then regressed these PCs against the activity across voxels for that trial. This resulted in a set of beta weights for each trial – a reduced-dimensionality (20 dimensions) representation of the activity on a given trial. These betas weights were taken as the dimensions to be used as predictors in fitting the model, and the selected option as the predicted variable (see below).

Because subjects had a free choice as to which option they selected, we had variable numbers of trials on which a given participant selected each option (i.e. an imbalanced training set). To avoid introducing bias in to the model to predict stimuli selected more often, we oversampled the training set such that there were equal numbers of trials of each stimulus type used to train the model. We did this by repeating the first n trials in the training set for each option until all classes had the same number of trials as the most selected option. This oversampling procedure successfully removed bias due to an imbalanced training set.

Having performed dimensionality reduction and balanced the training set, we trained a multinomial logistic regression model using the training set and then use this model to obtain probabilities each of the eight locations was selected on each trial of the test set. For each trial in the test data, we used the PC beta weights on that trial as input to the model to obtain the estimated probability that each option was the selected option. This is then repeated switching the training and test sets. The result of this analysis is a probability each of the eight options was selected on each trial of the task at each grey matter voxel.

To obtain a summary measure, we computed the ratio of the odds of classifying the trial as the chosen option, to the odds of classifying it as one of the offered but unchosen options (i.e. those options that were available for that run of the task but not selected on that trial). We took the logarithm of this ratio, providing us with a measure of the selectivity to the selected option of each searchlight on each trial, namely the log odds ratio, that is in theory normally distributed, although we used non-parametric statistics (permutation testing) to assess the significance of results at the whole brain level. This trial-by-trial scalar

measure was then used in subsequent regression analyses to identify regions in which it is explained by participants' belief entropy or baseline pupil diameter.

3.3.2 Region of interest definitions

The locus coeruleus ROI was defined as in recently published studies (Bär et al., 2016; Song et al., 2017). The LC complex was defined on the left (4 x 6 x 10 mm centred on MNI coordinates: x = -5, y = -34, z = -21) and right ROI (4 x 6 x 10 mm centred on MNI coordinates: x = 7, y = -34, z = -21). The ACC ROIs for the analyses to test whether ACC activity explained changes in representation strength in mOFC were defined by thesholding the activation maps for both GLM8 and GLM5 at p = 0.001 and p = 0.01 separately. Hence we tried the analysis with four different ROIs separately. For both the LC and ACC ROIs we extracted the timecourses by applying the inverse of each participant's registration to the mask to project the mask from standard space in to subject space and extracted the mean time series within this region of interest.

3.3.3 Pupil acquisition and analysis

We simultaneously measured pupil size while participants performed the task, in both the behavioural session and the fMRI session. We used an EyeLink eyetracker, SR Research to acquire pupil size data, sampled at 1000Hz.

We preprocessed the pupil size data using standard methods. We identified trials on which subjects blinked and removed the blinks by interpolation of values measured just before and just after the identified blinks. Trials on which the pupil measurement was lost were removed from the analyses. Finally, we removed the first 25 trials of each run of the task from analyses to avoid any luminance changes that may be caused by starting a new run of the task.

Pupil size in all analyses is normalised for each run by expressing it in terms of percentage difference from the overall mean pupil size on that run. Baseline pupil size is defined as mean (normalised) pupil size in the 20ms window preceding feedback.

3.3.4 General Linear Models

We then ran the following GLMs, where the dependent variable was either, multivariate representation strength, or change in baseline pupil diameter. Analyses were carried out either on all trials or only included trials in the exploit period that were at least 5

trials from the adjacent explore periods so that any variation is not due to changes in action. These GLMs are specified below:

GLM1 included the following regressors: Main effect (bias; column of ones), model entropy. Dependent variable: multivariate representation strength. Analysis carried out as a multivariate searchlight analysis. Only exploit trials (>5 trials from adjacent explore phase) included in the analysis. The goal of this analysis was to identify regions in which the representation strength varied with model entropy.

GLM2 was the same as GLM1 except instead of the entropy regressor used the probability (belief) assigned by the model that the currently exploited option is the high reward option. The goal of this analysis was to show representation strength in the region identified in GLM1 varied with the belief over the currently selected option (not just entropy, which captures the certainty of the belief).

GLM3 included the following regressors: Main effect, brain univariate activity in locus coeruleus (LC; trialwise activity from mean timeseries), difference in entropy following outcome between current trial and the previous (change in entropy), whether participants changed their choice on the subsequent trial, whether the current trial was the last trial of exploitation, and whole-brain mean univariate activity. Dependent variable: change in baseline pupil diameter between the current trial and the next. The goal of this analysis was to test whether univariate brain activity in LC explained changes in baseline pupil diameter.

GLM4 included the following regressors: Main effect, change in baseline pupil diameter between the current trial and the next, change in entropy. Dependent variable: change in multivariate representation strength in mOFC (and performed as a whole brain searchlight analysis). Analysis performed separately on only exploit trials (>5 from adjacent explore phase), and only explore trials. When this analysis was performed on explore trials, whether participants changed their choice was added as a regressor to the GLM. These analyses were performed separately on trials on which reward was delivered and omitted. The goal of this analysis was to test whether changes in pupil diameter explained changes in representation strength.

GLM5 included the following regressors: Main effect, brain univariate activity, difference in entropy following outcome between current trial and the previous, and whole-brain mean univariate activity. Dependent variable: change in baseline pupil diameter between the current trial and the next. Only exploit trials (>5 from adjacent explore phase) included in the analysis. (When this analysis was performed on all trials, whether participants changed their choice on the subsequent trial and whether the current trial was

the last trial of exploitation were added as regressors to the GLM.) The goal of this analysis was to identify brain regions in which univariate activity explained changes in baseline pupil diameter.

3.4 Results

In this chapter we present analyses addressing the question of how an internal model may be measured and updated. As in one of the analyses in the previous chapter, we focus our analyses to the core exploitation trials (at least five trials from the adjacent explore trials in either direction) so we can identify variation in neural activity due to changes in participants' beliefs rather than overt behaviour. All analyses presented are limited to this subset of trials unless otherwise stated.

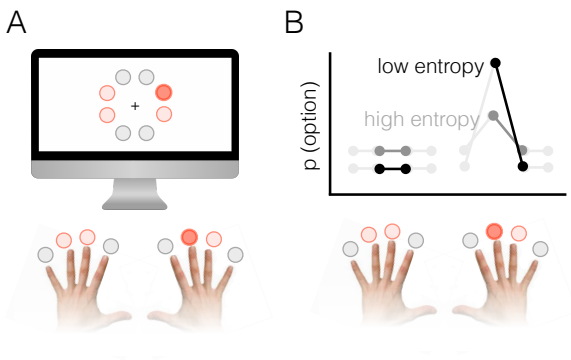


Figure 3.1 **Internal representation of the task.** A) Task schematic as shown in the previous chapter. B) Schematic of a probabilistic model of the state of the environment – the weighting of each option represents its probability of being the high reward option. When entropy is low, i.e. participants are certain, the weighting of the selected option is higher than when entropy is high.

3.4.1 Medial OFC represents probabilistic beliefs about the state of the environment

A probabilistic model of the current state of the world should comprise representations of candidate states, weighted by their probabilities. Recent studies have identified medial OFC as a likely site for such a probabilistic model in the context of model-based reinforcement learning (Chan et al., 2016; Schuck et al., 2016; Wilson et al., 2014). In the present task, such a probabilistic internal model could be thought of as a mixture model comprising representations of the possible options, wherein the strength of the representation of a given option on each trial is weighted by the belief that option is the high reward probability option (Figure 1b).

To identify regions that represented the task environment in this probabilistic

manner, we used a multivariate searchlight approach over the whole brain as follows: at each searchlight, in one half of the data (training set) we identified activity patterns characteristic of trials on which each option was selected. In the other half of the data (test set) we calculated the trial-wise probability of classifying as each option (we then swapped training and test sets; see Methods for details).

Simply searching for regions in which we could correctly classify the selected option, we were, as expected, able to decode at above-chance levels from a broad network of areas including motor and visual cortex (Figure 3.2a). A plot of the probability of decoding each action from visual cortex (Figure 3.2c, top) and motor cortex (Figure 3.2c, middle) reveals a representational structure that is sensitive to spatial and motor factors; the most likely options to be decoded, other than the actually selected option, are those corresponding to the chosen hand (and equivalently, the chosen visual hemifield).

However, the defining feature of a probabilistic model is that the relative weighting of the chosen option should depend on certainty, as shown in schematic Figure 3.1. To search for regions that had this property, we defined a trial-wise summary measure of representation strength as the ratio of the odds of classifying the trial as the chosen option, to the odds of classifying it as one of the unchosen options. This odds-ratio measure captured the extent to which the chosen option was represented more strongly than unchosen alternatives on each trial. Finally, we regressed trial-by-trial model entropy (as a proxy for the entropy of participants' beliefs; from the Bayesian model presented in the previous chapter) against this summary measure of representation strength (GLM1). This identified areas in which representation strength for the chosen option was higher when there was higher certainty that the chosen option was the high reward option.

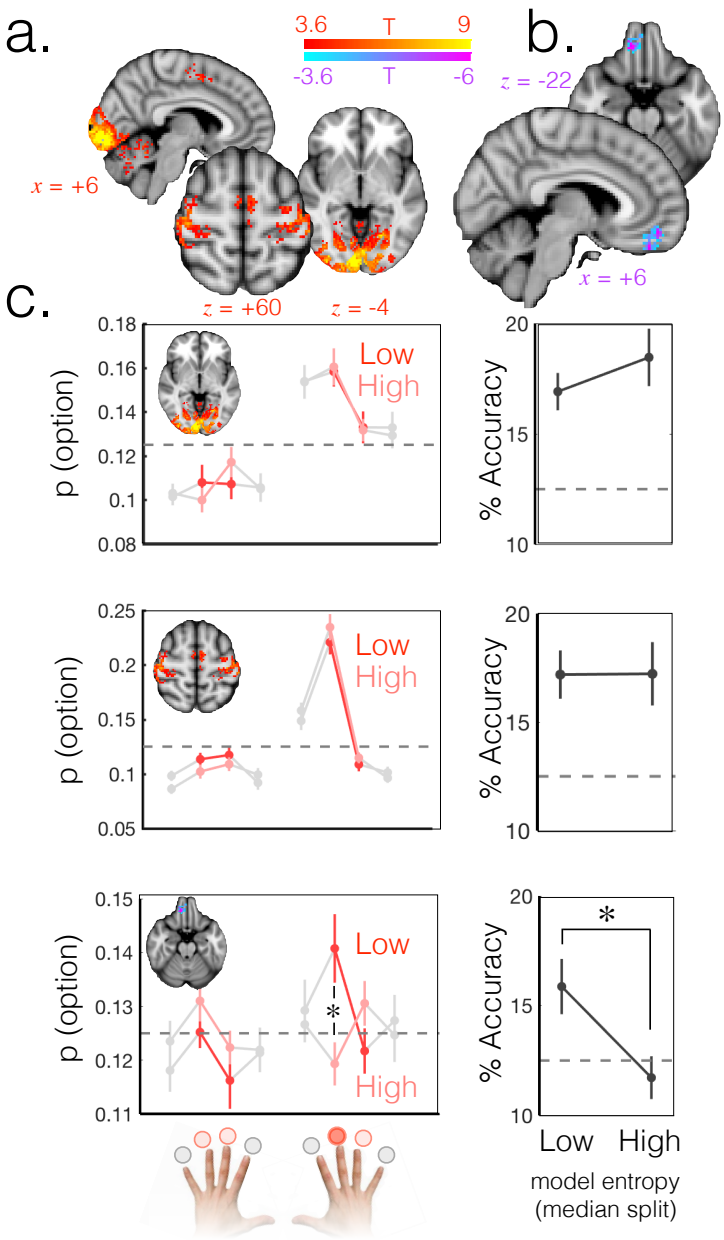


Figure 3.2 **Probabilistic beliefs represented in mOFC**. a) The currently selected option can be decoded above chance, as expected, in motor and visual cortex (t score map for above chance decoding of the chosen option across subjects; thresholded at p<0.001; corrected for multiple comparisons). **b)** Medial OFC represents the current state of the task probabilistically. Highlighted voxels are the centers of searchlights in which the representation strength for the chosen option was higher when model entropy was low (i.e. when participants have high certainty about the state of the environment; t score map for the effect of model entropy on representation strength; thresholded at p<0.001; corrected for multiple comparisons). **c)** Region of interest analyses demonstrating multivariate decoding patterns. Left column: the multivariate classifier probability that each option was selected; on trials when model entropy was low (dark red) and high (light red). Right column: decoding accuracy (i.e. proportion of trials when the option with the highest probability of having been selected by the classifier was indeed the option selected by the participant) in low and high entropy trials. All error bars are SEM. Top row and middle row: decoding in visual and motor cortices, respectively, is not sensitive to model entropy and errors in decoding tend to be to neighbouring options. Bottom row: decoding in mOFC is modulated by model entropy, such that decoding is higher when model entropy is low.

Using this approach we identified a region in mOFC (peak = 10,40,-22; Figure 3.2b) in which representation strength for the exploited option was higher when model entropy was lower (corrected for multiple comparisons using permutation testing; cluster mass p<0.05 corrected with a cluster forming threshold of p=0.001). This was the only cluster to survive multiple comparisons correction. This weighted representation of option preferences is consistent with a probabilistic internal model of the environment, whereby the currently selected option has a higher decoding probability when model entropy is low (Figure 3.1b and Figure 3.2c, bottom). It is distinct from an action/policy representation, in which the representation of options is not modulated by entropy, as was observed in visual and motor areas (Figure 3.2c, top and middle).

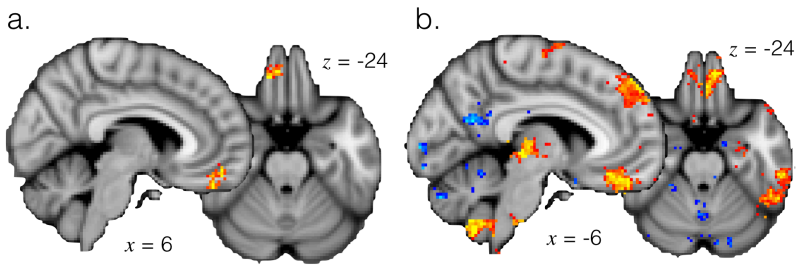


Figure 3.3 **mOFC effect from different measures.** Representation strength in mOFC is explained by probability assigned to the currently selected option, as well as the difference between high and low reward exploit periods. **a)** T-statistic map showing model probability (of the selected option being the high reward option) explains representation strength specifically in mOFC, as in GLM2. Image thresholded at p<0.001; corrected for multiple comparisons using cluster mass-based permutation testing: cluster forming threshold is voxelwise p<0.001, cluster corrected threshold p<0.05. **b)** T-statistic map showing regions where representation strength is higher in high (90%) vs. low (70%) reward exploit periods. Image thresholded voxelwise at p<0.05, uncorrected.

To probe the robustness of the mOFC result, we repeated the analysis using two measures other than the model entropy regressor to predict representation strength. Entropy is a summary measure of the uncertainty across options; low entropy could arise from any of the four options being represented more strongly than the others. To check that the result was specific to the current state of the environment, we repeated the above analysis but instead of model entropy, used the probability (belief) assigned by the model that the currently exploited option is the high reward option as the predictor variable (GLM2). This also produced a specific effect in the medial OFC (Figure 3.3). To confirm our result was not dependent on the particular learning model we used, we simply compared

representation strength during exploit periods with high and low payout rates. When the pay-out rate of the exploited, high reward option was 90% vs. 70%, representation strength was higher in mOFC (Figure 3.3).

We saw in the previous chapter that in addition to the medial OFC, the hippocampus was more active in exploit than explore periods and could perhaps be expected to contain a similar model of the environment (Eichenbaum, 2017c; Strange et al., 2005). In a searchlight ROI centered on the peak hippocampal voxel for this contrast (peak = 34,-14,-16), representation strength was significantly predicted by model entropy ($t_{(18)}$ = -2.11; p<0.05). However in the whole brain analysis described above (Figure 3.2b), this effect did not survive multiple comparisons correction and so should be interpreted with caution.

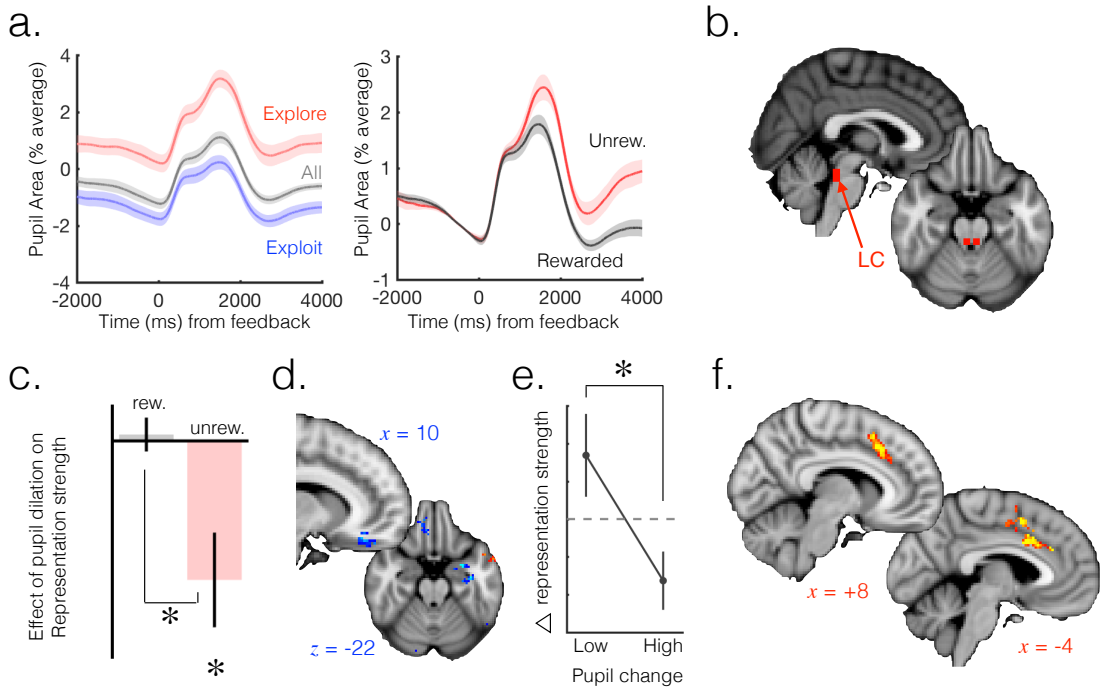


Figure 3.4 **Neuromodulatory systems as a candidate mechanism for increasing flexibility of belief representations. a)** Pupil size mean timecourses throughout a trial. Left: mean timecourses shown for all trials as well as trials split according to whether they were explore or exploit trials, revealing a larger pupil size in exploration than exploitation. Right: splitting trials according to whether the trial was rewarded or not reveals a sustained increase in pupil size following omission of reward. Shaded regions denote SEM. b) Locus coeruleus mask used to extract trialwise activity to regress against pupil size. c) Change in pupil diameter predicts changes in representation strength in mOFC on exploit trials. Beta weights for the effect of change in pupil diameter on change in representation strength are significantly below zero. This is true only on trials when reward was omitted. Error bars are SEM. d) Performing this analysis as a whole brain reveals, again, a relatively localized effect in mOFC (t score map shown, thresholded at p<0.01, uncorrected; analysis performed on unrewarded exploit trials). e) Median split on unrewarded exploit trials reveals that on trials when pupil change is high, change in representation strength is more negative than when pupil change is lower, in mOFC. Error bars are SEM; dashed line denotes 0 change (change in a.u.). f) Left: ACC region active when model entropy increased. Right: ACC region in which activity predicted pupil dilation, over and above the effect of increase in model entropy and mean brain activity. Both thresholded at p<0.001 and corrected for multiple comparisons.

3.4.2 Noradrenaline as a candidate mechanism for increasing flexibility of belief representations

Having confirmed that a probabilistic model of the state of the world is represented in the brain, we went on to ask how entropy in that model might be controlled. An intriguing possibility is that increases in neural model uncertainty are controlled via a neuromodulatory route. Noradrenaline, produced in the locus coeruleus (LC), is released very broadly in the cerebral cortex including the mOFC, and has been proposed as a candidate mechanism by which the level of uncertainty in neural models of the world may be controlled (Aston-Jones and Cohen, 2005; Bouret and Sara, 2005; Martins and Froemke, 2015; Nassar et al., 2012; Preuschoff et al., 2011; Yu and Dayan, 2005).

We used baseline pupil size (mean pupil area in the 20ms before outcome presentation, expressed as % signal change relative to the participant's mean pupil area over the task run) as a proxy measure for noradrenaline levels (Joshi et al., 2016). Trial-wise change in pupil dilation in our task was significantly predicted by activity in the LC across all trials (ROI defined as in Methods; Figure 3b; $t_{(15)} = 2.94$, p=0.01, GLM3: linear regression including mean activity across the grey matter as a co-regressor of no interest). Consistent with the hypothesis that noradrenaline levels are higher when uncertainty is higher, the baseline pupil size was larger in explore than exploit trials (Figure 3.4a, left), and showed a marked one trial increase at the transition from exploitation to exploration (see Supplementary Data at the end of this chapter), as observed in (Jepma and Nieuwenhuis, 2011). On a trial-by-trial basis, reward omission produced a sustained increase in pupil size following feedback (Figure 3.4a, right). Pupil changes could not be explained due to luminosity differences between the counterbalanced cheese and apple reward stimuli (see Supplementary Data at the end of this chapter).

We hypothesized that noradrenaline input to mOFC could be a mechanism by which uncertainty in the internal model is increased. To test this hypothesis, we carried out a linear regression (GLM4) in which the change in baseline pupil size was used to predict changes in representation strength in the mOFC (the region of interest shown in Figure 3.2b, representation strength extracted from a searchlight center defined on the peak; change in model entropy was included as a co-regressor of no interest).

This analysis indicated that changes in baseline pupil size indeed explained changes in representation strength in mOFC; as expected, this relationship was negative, such that increases in pupil size predicted decreases in representation strength. Interestingly, this

relationship was limited to unrewarded trials (the trials on which uncertainty increases; $t_{(15)}$=-2.94, p=0.01); within the subset of rewarded trials there was no relationship ($t_{(15)}$=0.37, p=0.71; difference between rewarded and unrewarded: $t_{(15)}$=2.82, p=0.01; Figure 3.4c). Although we focus here on the exploitation phase, we found this result replicated in the mOFC ROI when including only explore trials in the analysis (Figure 3.5; unrewarded: $t_{(15)}$=-2.38, p=0.03; rewarded: $t_{(15)}$=0.97, p=0.35; difference: $t_{(15)}$=-2.23, p=0.04). Performing the analysis on unrewarded exploit trials over the whole brain reveals a region relatively localized to mOFC (Figure 3.4d). To illustrate this result further, we took the unrewarded exploit trials and looked at the change in representation strength in the mOFC ROI on trials separated according to a median split on change in pupil diameter. As expected, we found change in representation strength was more negative when change in pupil dilation was higher (Figure 3.4e; $t_{(15)}$ = 2.72, p=0.02).

These results demonstrate that on trials when evidence against the current belief is observed (unrewarded trials), the change in uncertainty in the neural model is predicted by changes in pupil diameter, supporting the hypothesis noradrenaline increases uncertainty in neural models. The fact the relationship is specific to unrewarded trials suggests that the relationship between pupil dilation and representation strength in mOFC is directionally specific to the process of increasing, not decreasing, model entropy.

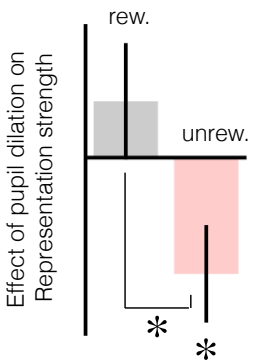


Figure 3.5 **Pupil change predicts representation strength in explore trials**. The relationship between change in baseline pupil diameter and change in representation strength in mOFC replicates in explore trials. The figure has the same notation as Figure 3.4c. When performing this regression (GLM4) on explore trials (whether participants changed their response on the subsequent trial was added as a coregressor).

3.4.3 Anterior cingulate cortex predicts pupil change

Having provided evidence that neural model uncertainty may be increased via neuromodulatory mechanisms, we next asked which brain regions may regulate changes in

neuromodulatory systems. Given the role ascribed to ACC in deciding whether to switch away from the current strategy, as well as our results in the previous chapter identifying ACC activity as tracking the entropy of the current strategy, we may predict ACC as a candidate region for regulating neuromodulatory systems that update internal models. Consistent with this, ACC is one of the few cortical regions reported to have an afferent influence on LC (Jodo et al., 1998), suggesting that if this region is responsible for computing changes in beliefs, it may be able to broadcast increases in uncertainty to other brain regions via a neuromodulatory route.

To probe this hypothesis we asked where in the brain univariate activity predicts changes in baseline pupil size. We conducted a whole brain GLM analysis, again focused on the exploit period, in which univariate brain activity at each voxel was the predictor variable and pupil size the dependent variable; change in model entropy was included as a co-regressor. Pupil dilation is globally correlated with neural gain (Eldar and Niv 2015), and we observed a strong correlation of pupil size with mean BOLD signal across the entire grey matter. When we controlled for this global effect by including mean activity across the grey matter as a co-regressor of no interest (GLM5), we observed a localized effect in ACC such that ACC activity on a given trial predicted the change in baseline pupil size from that trial to the next (Figure 3.4d, right; more activity in the brain associated with a greater increase in pupil dilation; corrected for multiple comparisons using permutation testing; $p<0.05$ at cluster forming threshold of $p=0.001$). Although we focus on the exploit period, this result is also present when conducting the analysis across all trials (Figure 3.6).

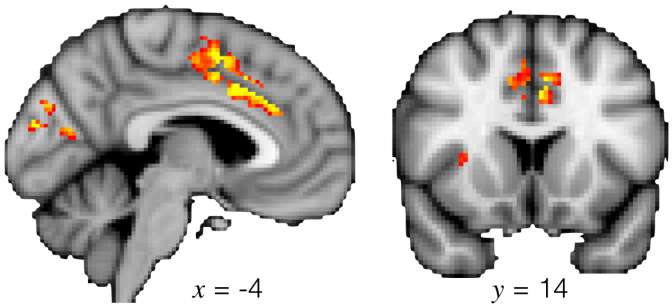


Figure 3.6 **ACC activity explains changes in baseline pupil size across all trials**. T-statistic map showing where univariate brain activity explains changes in baseline pupil size, as in GLM5. Analysis performed in a grey matter mask, hence the streak in the activation. Similar to Figure 3.4f in the main text except across all trials. Image thresholded at $p<0.001$; corrected for multiple comparisons.

3.4.4 Noteworthy analyses that didn't work (produced null results)

Given that ACC activity predicts pupil dilation and pupil dilation predicts changes in representation strength in mOFC, we tested whether there was a direct relationship between ACC activity and changes in representation strength in mOFC. However, there was no evidence for such a relationship (activity extracted from an ROI defined on peak of either the ACC effect tracking changes in belief uncertainty from the previous chapter, or the effect predicting pupil change, did not have a significant relationship with change in representation strength in mOFC; $p>0.1$ for all trial subsets tested). We further regressed change in representation strength in the mOFC against whole brain univariate activity to identify regions that predicted changes to it. We tried this for various different subsets of trials and co-regressors. These analyses also revealed no significant results. The lack of significance in these analyses may reflect the fact that both brain measures have relatively low SNR compared to the pupillometry data, to which both were significantly related.

3.5 Discussion

Models of the environment that capture the uncertainty of beliefs naturally become less entropic as the number of observations on which they are based increases; for them to become more entropic without a similar number of observations of the new world state requires an external input or leak. Such a source of entropy is necessary to make rapid changes of belief possible.

We identified patterns of activity indicative of a probabilistic model of the environment in the medial OFC. The hallmark of this probabilistic model is that it has stronger representation strength for the currently selection option when model entropy is low. This is distinct from sensory and motoric representations in visual and motor cortex that are not sensitive to model entropy. Changes in representation strength in medial OFC, over and above those predicted by our task model, were explained by the magnitude of pupil dilation, a proxy measure for noradrenaline release. Pupil dilation was in turn predicted by activity in the ACC – a region also active when model entropy increased (as we saw in the previous chapter).

That an internal model of the task is represented in mOFC is consistent with the proposal that this region represents the current location in task state space (Wilson et al., 2014), and in particular represents information that is hidden to participants but required to solve the task (Chan et al., 2016; Schuck et al., 2016). Our mOFC representation likely reflects the inferred current latent task state. Our work extends this task state space

proposal to include the representation of the strength of dynamic beliefs in a probabilistic task in which participants infer the hidden option-reward contingencies that define their location in the state space.

The fact that mOFC is the locus of the probabilistic model in the current task is probably due to the choice of task, as mOFC plays a key role in representing stimulus-action-outcome contingencies during model-based choice (Jones et al., 2012) and represents latent task states (Wilson et al., 2014). Different stimulus or task domains could rely on probabilistic models represented in other, task relevant parts of cortex, in particular if the task states are fully observable (Wilson et al., 2014). Furthermore, the exact nature of the representation we are interrogating in mOFC is an interesting avenue for future research. For example, it is possible the representation we observe here reflects the online representation of option-reward contingencies required to guide choice – for example, for online comparison – and is not where the option-reward contingencies are stored offline, which may be in another region such as hippocampus (Boorman et al., 2016) or striatum (Packard and Knowlton, 2002). The hippocampus showed some limited evidence of probabilistic option representations in the current study. The striatum also plays a key role in representing action-outcome associations, and indeed mechanisms for maintaining flexibility within these representations have been proposed (Franklin and Frank, 2015). Regardless of its exact nature, we are able to use the probabilistic model of the task decoded from mOFC as an index of the neural representation of participants' beliefs, i.e. a measure of an internal model.

We next sought evidence for a role of neuromodulatory systems in increasing uncertainty in such an internal model. Theoretical work has ascribed NA various functions relating to the modification of internal models (changing gain (Aston-Jones and Cohen, 2005), network resets (Bouret and Sara, 2005), and signalling unexpected uncertainty (Preuschoff et al., 2011; Yu and Dayan, 2005)). However, evidence for the impact of NA on internal models has been relatively sparse, partly due to the requirement of measuring the internal model. As we discussed in the introduction, increasing NA results in broader – less selective – tuning curves of auditory neurons (akin to a network reset) (Martins and Froemke, 2015). Such an impact on representations is what is required for NA to implement its proposed role in belief updating (Nassar et al., 2012). By simultaneously measuring pupil size and an internal model in mOFC, we observed changes in pupil diameter predict increases in uncertainty in an internal model, demonstrating a potential analog of this rodent work in humans performing complex decision-making tasks, and suggesting a broad

role of NA in reducing selectivity of cortical representations.

We then asked which brain regions may be involved in regulating the pupil-linked neuromodulatory system. As shown in the previous chapter, we found ACC tracked changes in the certainty of beliefs, suggesting it as a potential candidate. Consistent with this, ACC is one of the few cortical regions with descending control on NA release (Jodo et al., 1998) and activity of ACC neurons predicts pupil dilation and LC activity (Ebitz and Platt, 2015; Joshi et al., 2016). However, whilst, as we noted in the introduction, both ACC and NA have been implicated in behavioural flexibility, evidence for their suspected (Aston-Jones and Cohen, 2005; Karlsson et al., 2012; Nassar et al., 2012) coordination during tasks requiring belief updating has been lacking. Here we found ACC activity explained changes in baseline pupil dilation, over and above the effect of changes in uncertainty and mean brain activity. These results provide evidence for the NA system as a candidate mechanism via which the computation of uncertainty of beliefs in ACC may be broadcast to internal models.

An alternative explanation for the relationship between ACC and NA is that their activity is associated, not with a change in beliefs, but with a change in the drive to produce exploratory behavioural output. The circumstances in which model entropy should increase are the same circumstances that should drive participants towards exploratory (random) behavior. Indeed, increasing NA release to the ACC has been associated with the output of stochastic behavior (Tervo et al., 2014). However, an account in which ACC activity and NA release only track a drive to produce stochastic behavioural output, or to explore, cannot account for the association between pupil dilation and the increase in model uncertainty recorded in mOFC, in the absence of any behavioural change. Further, the animal work described above (Martins and Froemke, 2015) provides causal evidence that NA reduces selectivity of cortical representations, supporting the proposed role for NA in reducing the strength of beliefs, in addition to its role of increasing stochastic behaviour (Tervo et al., 2014). This hypothesis can be summarized in terms of economic models of choice as arguing that NA input to ACC alters the temperature parameter of a softmax choice model whereas NA input to internal models (mOFC) alters the option values, resulting in behavioural stochasticity (Tervo et al., 2014) and entropic beliefs (present results), respectively.

We started by asking whether there was a mechanism for increasing uncertainty in neural models. The results presented, in which increases in belief uncertainty engage ACC, which explains changes in pupil size, that in turn explain changes to the selectivity of internal models following evidence the underlying state of the world has changed (reward omission), provide a candidate mechanism. These results suggest different neural mechanisms may be

engaged when forming as compared to revising beliefs, and that the NA system may play a crucial role in implementing flexibility of internal representations during decision-making in humans.

3.6 Supplementary data

Task-related pupil size changes are not explained by outcome stimulus type

We checked our results could not be explained by differences in luminosity between outcomes signalling reward vs. no reward. We counterbalanced the stimuli signalling rewarding vs. non-rewarding outcomes across participants: half the participants received a cheese stimulus as the rewarding outcome and apple stimulus as the unrewarding outcome, and the other half of participants received the converse. We analyse the pupil data from the behavioural session and split the participants according to the stimulus outcomes they receive. We analyse the sixteen participants that entered pupil-brain analyses (ten received cheese as a reward, and six received apples as reward). We demonstrate that although noisier due to reduced power, pupil dilations to task factors are qualitatively the same when either the cheese or the apple was the rewarding outcome. We demonstrate this with two analyses (Figure 3.7). First, by performing an analysis looking at the baseline pupil size on trials around transitions from exploitation to exploration. We observed a large increase in baseline pupil size when transitioning in to exploration, as has been previously observed (Jepma and Nieuwenhuis, 2011), when analysing data from all participants, or splitting them according to either stimulus outcome identity. Second, by constructing a GLM with regressors: reward, changing response on the next trial, switching in to exploration, as well as a main effect regressor, and performing a timeseries analysis with this GLM on all data and on data split according to stimulus outcome identities. In both sets of analyses we found the results to be qualitatively the same (e.g. the directions of effects are the same) regardless of how the data is split. This demonstrates task factor-related changes in pupil size are not explained by differences in the low-level identity of stimuli signalling outcomes.

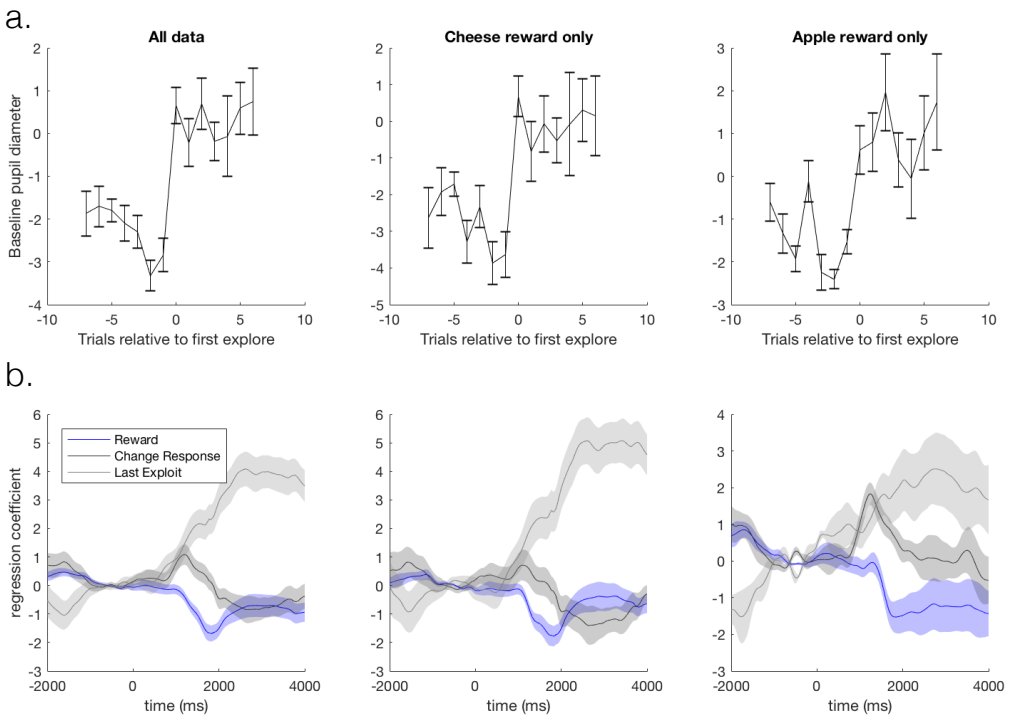


Figure 3.7 **Task-related pupil size changes are not explained by outcome stimulus type**. a) Mean baseline pupil size (again pupil size expressed as % of mean of that session) as a function of trials around a transition from exploitation to exploration. A marked one trial increase in pupil size was observed as participants transitioned from exploitation to exploration. Although noisier due to reduced power, this result was true when data was split according to either outcome identity type (middle and right panels). Error bars are SEM. b) Mean beta weights for the effect of the regressors in the GLM on pupil size, plotted as a function of time relative to outcome delivery. The data on each trial was normalised before performing the regression by demeaning the one second preceding outcome delivery. A constriction following reward delivery, a small dilation on trials on which participants changed their choice on the subsequent trial, and a large dilation when switching from exploitation in to exploration was observed. Again it can be observed that the results are qualitatively similar regardless of outcome identity type (middle and right panels). Shaded regions are SEM.

4 Space as a state space and intuitive planning

4.1 Abstract

It is proposed that a cognitive map encoding the relationships between entities in the environment supports the ability to flexibly navigate the world. Place cells and grid cells provide evidence for such a map in a spatial context and support flexible and efficient mechanisms for navigating in 2D physical space. Emerging evidence demonstrates analogous cells code for non-spatial information. This suggests these cells may not represent physical space *per se*, but rather something more general such as the topology of the environment. The neural coding principles that give rise to place and grid cells may therefore also support the arbitrary topology, non-spatial domains that cognitive maps were also originally proposed to encode. However, the grid cell-based mechanisms for navigation in physical 2D space do not generalise to arbitrary topologies. In this chapter, by reconsidering space more generally as a state space, we show grid-like firing patterns can be obtained as an efficient representation (principle component analysis) of the structure of 2D-distributed place cells, suggesting grid cells may represent the topology of the state space rather than physical space *per se*. We then show how grid-like codes contain not only local information about the structure of the environment but also knowledge of approximate distances between every pair of states (locations) in the environment, allowing efficient and flexible navigation of arbitrary topology environments – referred to as intuitive planning. This chapter is joint work carried out with Alon Baram.

4.2 Introduction

The cognitive map is defined as an internal representation encoding the relationships between entities in the world (Tolman, 1948). In the context of spatial cognition, the neural underpinnings of such a map have been identified, most notably in the form of place and grid cells (Hafting et al., 2005; O'Keefe and Nadel, 1978). More recently, encoding of information via a grid-like code has been identified in non-spatial tasks (Aronov et al., 2017; Constantinescu et al., 2016), suggesting neural coding of spatial and non-spatial information may be unified under a general framework (Behrens et al., 2018). Unifying under a general framework would offer insight in to the common principles giving rise to

spatial and non-spatial, arbitrary topology representations, as well as how cognitive maps may organise knowledge of all different kinds.

To understand the nature of such a generalised framework, we present a theoretical analysis of how place and grid cell receptive fields combine to provide an efficient representation of the cognitive map in the spatial domain, and how such grid fields may arise through experience. A key insight is that grid-like receptive fields arise as consequence of the structure of the (spatial) environment they represent. If correct, this theoretical account makes predictions for how grid-like representations of non-spatial cognitive maps would differ, depending on the structure of the cognitive environment.

To that end, in this and the next chapter we recast spatial cognition under the reinforcement learning – or state space – framework (Behrens et al., 2018; Dordek et al., 2016; Gustafson and Daw, 2011; Stachenfeld et al., 2017). We emphasise the *state space* component since the key contribution of the RL framework here is reconceptualising space as a state space, rather than optimising reward. As we will see below, this is achieved if the notion of a place cell is generalised to encode a state in an RL world, i.e. considering locations as states (Stachenfeld et al., 2017).

By examining neural coding under this framework, it is possible to ask whether spatial representations observed in the brain may be explained as representing the structure of 2D space. Doing so has accounted for a variety of features of spatial representations (Stachenfeld et al., 2017). For example, place cell representations may be better explained as coding future or successive states rather than simply the current state, which explains, for example, the observation that place cells skew away from goal locations in the direction of approach (as the agent is likely to approach goals) (Stachenfeld et al., 2017). Of particular relevance to us, which we explore further below, is the observation that grid cell firing may be thought of as the eigenspectrum of local place cell transitions (Dordek et al., 2016; Stachenfeld et al., 2017). This observation suggests we may be able to understand grid cell firing as encoding the relationships between place cells and therefore summarising the statistics or structure of 2D space.

Storing the structure of the environment, i.e. representing an internal model of the relationships between entities in the environment, has obvious advantages, such as facilitating flexible planning or inference of unobserved relationships/links. However, it is not immediately clear what the advantages are of two representations of the structure of the environment, one in the place cell population and one in grid cell population. Both encode a map of the external world, but they do so in different ways. Place cells encode the

map locally since their firing is spatially localised to a specific position. Grid cells encode the map globally since they each fire in multiple locations (and therefore do not represent a specific location), although across the population exact location is encoded. Recently, it has been demonstrated that global (grid) representations may be advantageous as they can be used to efficiently solve navigation problems since much of the necessary computation is already solved in the representation. Grid representations can allow an agent to directly compute the vector between two locations in the environment without requiring computationally expensive searches through place cells (Banino et al., 2018; Bush et al., 2015; Stemmler et al., 2015), endowing the agent with a "sense of direction" for intuitively navigating the map. However, cognitive maps encoding non-spatial knowledge will have arbitrary topologies and therefore such vector-based navigation that relies on 2D grid cells will fail to navigate these maps. It is therefore important to establish how grid-like representations may be used to flexibly and efficiently navigate arbitrary topology worlds as well as spatial worlds.

In this chapter we explore how casting spatial problems under a reinforcement learning or state space framework may lead to insights in to how representations emerge, as well as how these representations may be used to support flexible navigation in arbitrary topology worlds. First, we briefly explain what it means to cast spatial problems under an RL/state space framework. We then demonstrate how grid-like representations may be explained in this context as summarising the statistics of 2D space. To do this we reproduce the core result described in (Dordek et al., 2016) showing grid cells emerge as a PCA of place cells, and explore this further in a couple of minor ways. Finally, unlike previous work focusing on the spatial domain (Bush et al., 2015; Stemmler et al., 2015), we show how global grid-like codes may be used to flexibly and efficiently solve navigation problems in arbitrary topology worlds as well as spatial worlds.

4.3 Casting space under a state space framework

We briefly describe how spatial coding can be formalised and unified with non-spatial coding under a state space framework by considering locations as states. Imagine a simple linear track. We can define the different locations along the track as different states, such that a state indexes a given location (Figure 4.1). If we ignore the effects of the animal's policy, the neighbour relationships of states is defined by and represented in a states by

states transition T (or adjacency[3] A) matrix. Element T_{ij} is the probability of transitioning from state i to j ($P(j|i) = T_{ij}$) in the next time step, and hence any non-zero elements must be neighbour states. This matrix therefore represents the structure of the environment.

As we will see below, these matrices may be represented using grid-like codes, and are useful as they can provide distributions over future states or be used to compute distance matrices that encode the approximate distances between every pair of states in the world, allowing efficient planning.

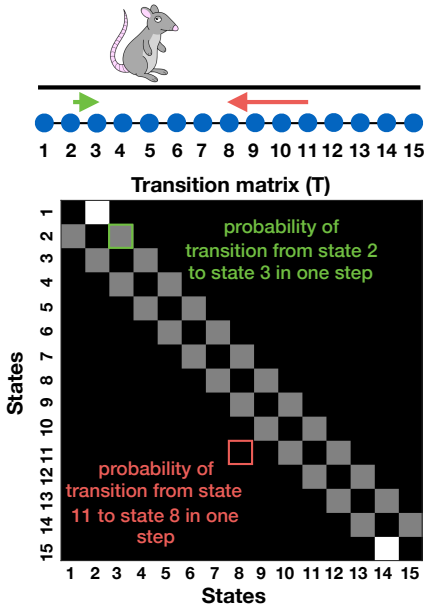


Figure 4.1 **Casting space as a state space**. Spatial cognition can be reconceptualised under a state space framework by generalising the notion of a place cell to encode a state. For example, the locations on a linear track (top) can be defined as different states (blue dots). The relationships between these states can then be represented in a states by states transition matrix T such that T_{ij} is the probability of transitioning from state i to j. Figure adapted from (Behrens et al., 2018).

---

[3] The difference between transition and adjacency matrices is a somewhat subtle but important one. Both define the structure of the environment, but their elements have different meanings. Element A_{ij} of an adjacency matrix is either 1 or 0 depending on whether it is or isn't, respectively, possible to transition from state i to j (A represents a graph and $A_{ij} = 1$ denotes the presence of an edge between vertices i and j). Element T_{ij} on the other hand, is the probability of transitioning from state i to j ($P(j|i) = T_{ij}$), and hence the rows of T must sum to 1. When, for example, computing powers of A or T, the resulting matrices have different meanings. A_{ij}^n is the number of walks of length n from state i to j, whereas T_{ij}^n is the probability of occupying state j n steps after occupying state i.

4.4 Grid cells as a PCA of 2D space

With this general framework in mind, we now ask whether such structures may be represented in the brain, and what form this representation may take. In the above section, we treated space as a discretised state space wherein each state indexed a discrete location (Figure 4.1). However, in the brain, we know that within an environment place cells represent locations and do so with a smooth change in firing field that can be approximated by a Gaussian (Dordek et al., 2016) (or other distributions, such as exponential (Stachenfeld et al., 2017)). Such a representation is continuous rather than discrete. We therefore investigate matrices that may represent the structure of continuous space, somewhat analogous to the adjacency matrix, and whether these may be represented in the brain.

4.4.1 Covariance structure of 2D space

We take as a starting assumption that place cells index locations in space. More specifically, we simulate a given place cell as a Gaussian indexing a single location in space. We note here that, as is discussed briefly and the end of this chapter and a focus of the next chapter, we do not believe place cells are so simple, but rather are crucial for combining structural representations with sensory representations. For the purposes of the PCA simulations, however, we simply provide place cell representations.

The structure of continuous 2D space, i.e. the relationship between one place (cell) and another, can be represented in the covariance matrix of place cell activity. In a place cell by place cell covariance matrix, element (i, j) is the covariance between place cell i and place cell j. Place cells will covary with one another if they fire at similar times, which will occur when they are indexing physically close locations such that their firing fields overlap. Therefore, under a random walk, the covariance matrix can be thought of as representing something analogous to adjacency, and hence the structure of the environment.

Following (Dordek et al., 2016), we can construct such a matrix by randomly distributing place cells in an environment, having an agent take a (random) walk through the environment to generate a place cells by time matrix, and then taking the place cell covariance of this matrix. We treat place cells simply as Gaussian distributions, and randomly (uniformly) distribute them over the environment (Figure 4.2). The firing rate r of place cell i at time t, when the agent is at location $\boldsymbol{X}(t)$, is given by a 2D Gaussian distribution:

$$r_i^t(\boldsymbol{X}(t)) = \exp\left(-\frac{(\boldsymbol{X}(t) - \boldsymbol{C}_i)^2}{2\sigma_i^2}\right)$$

Equation 4.1

where $\boldsymbol{C}_i$ is the i^{th} place field's centre and σ_i its width (σ_i assumed to be the same in both dimensions). The environment is a square box, with periodic boundary conditions to prevent boundary effects (i.e. it is a square torus). The locations of the agent $\boldsymbol{X}(t)$ are determined by the agent's walk over the environment. The walk is random (e.g. there is no directed exploration) but with a bias such that it is more likely to continue in roughly the current direction and does not take sharp turns (Figure 4.2). As in (Dordek et al., 2016), the agent's movement is defined by:

$$D^{t+1} = modulo(D^t + \omega Z, 2\pi)$$
$$x^{t+1} = x^t + vcos(D^{t+1})$$
$$y^{t+1} = y^t + vsin(D^{t+1})$$

Equation 4.2

where D^t is the current direction angle, ω is the angular velocity, Z is $N(0,1)$ where N is the standard normal Gaussian distribution, v is the linear velocity, and (x^t, y^t) is the location of the agent at timestep t. Z induces randomness in the agent's trajectory but also a bias to continue travelling in the same direction (as the mean of N is 0). As with the place cell rate maps, boundary conditions for the agent's trajectory were periodic.

Simulating the agent walking around the environment through place cell firing fields results in a place cells by time activity matrix (Figure 4.2). Notably, this matrix has therefore been constructed by observing only local information, i.e. the overlapping place cell receptive fields at a given time point. The agent has never observed the environment's overall structure. From this activity matrix, we obtain a place cell by place cell covariance matrix, where the covariance between two place cells is:

$$cov(r_i, r_j) = \mathrm{E}\big[(r_i - \mathrm{E}[r_i])(r_j - \mathrm{E}[r_j])\big]$$

Equation 4.3

where $\mathrm{E}[r_i]$ is the expected value or mean of r_i. Since this matrix contains information about the neighbour relationships between places (place cells), it describes the structure of the environment (Figure 4.2). In this instance the statistics of 2D space.

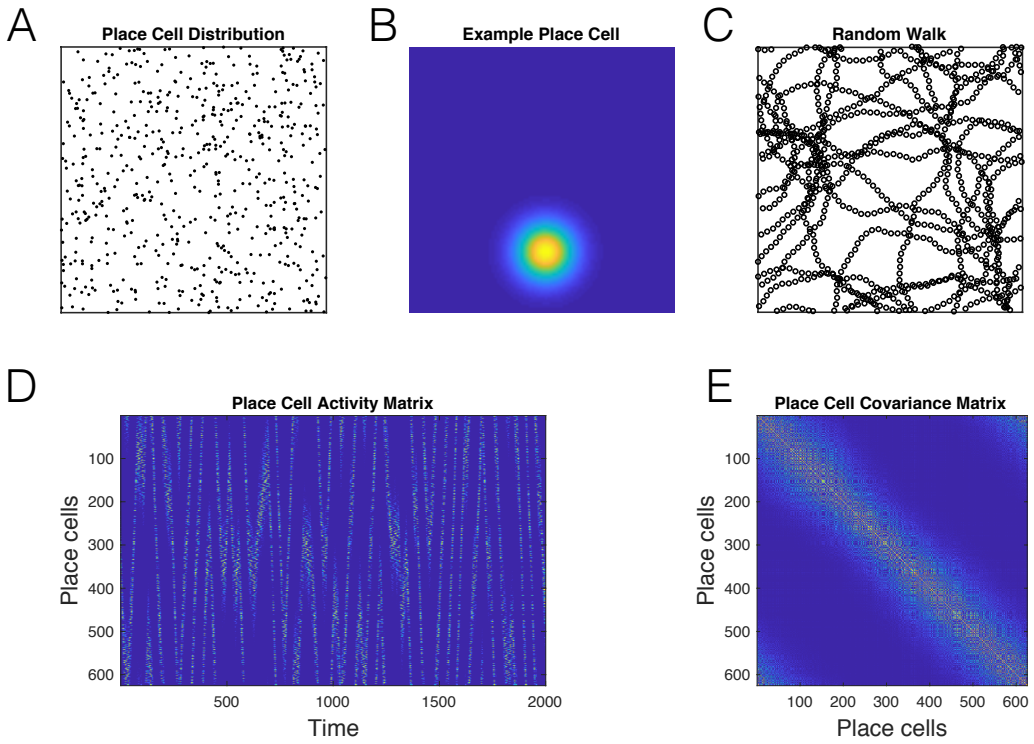


Figure 4.2 **Place cell covariance arises from local observations**. A) Uniform distribution of place cells in a square environment (black dots denote place cell centres). B) An example place cell rate map. C) First 1000 timesteps of a random walk through the environment. Periodic boundary conditions can be observed. D) Activity of each place cell as a function of time (first 2000 timesteps shown). E) Place cell by place cell covariance matrix of the place cell activity matrix. This matrix represents the structure of the square toroidal 2D space. Notably, we have arrived at this matrix without ever having observed the structure, only local place cell activity over time.

4.4.2 Grid cells as representing the topology of 2D space

We now ask whether grid cell activity may be explained as representing such a matrix. An efficient representation of the structure of data is derived from principle component analysis (PCA), a common method for dimensionality reduction. The principle components seek to maximise the variance of the data explained in orthogonal (eigen)vectors of the covariance matrix of the data. Eigenvectors associated with higher eigenvalues explain more variance. We take the eigendecomposition of the covariance matrix resulting in nNeurons eigenvectors and eigenvalues. We note that (Dordek et al., 2016) demonstrate a method for learning eigenvectors online in a simple neural network, which we do not do here.

As in (Dordek et al., 2016), we show these vectors have grid-like properties, suggesting grid cells represent the structure of 2D space. If we think of each grid cell as

being associated with a given eigenvector, that eigenvector is represented in the weights between place cells and the grid cell. We can then generate a rate map across the environment for each grid cell by projecting the eigenvectors onto the place cells:

$$G(x,y)_k = \sum_{p=1}^{n} \varphi_k^p \exp\left(-\frac{\left(x-c_x^p\right)^2}{2\sigma_x^2} - \frac{\left(y-c_y^p\right)^2}{2\sigma_y^2}\right)$$

Equation 4.4

where $G(x,y)_k$ is the firing rate of grid cell k at location (x,y). φ_k is the kth eigenvector and φ_k^p corresponds to the element of that eigenvector associated with place cell p, i.e. φ_k^p is the weight between place cell p and grid cell k. c_x^p and c_y^p are the (x,y) location of the centre of place cell p. The grid cell firing rate at location (x,y) is therefore the dot product between the vector of place cell activations at location (x,y) and the place to grid cell weights (φ_k). The resulting grid cell rate maps are grid-like, in that they are periodic over space (Figure 4.3). Plotting the place to grid cell weights for a given grid cell (i.e. giving the grid cell receptive fields in terms of place cells directly) also reveals this periodicity (Figure 4.4).

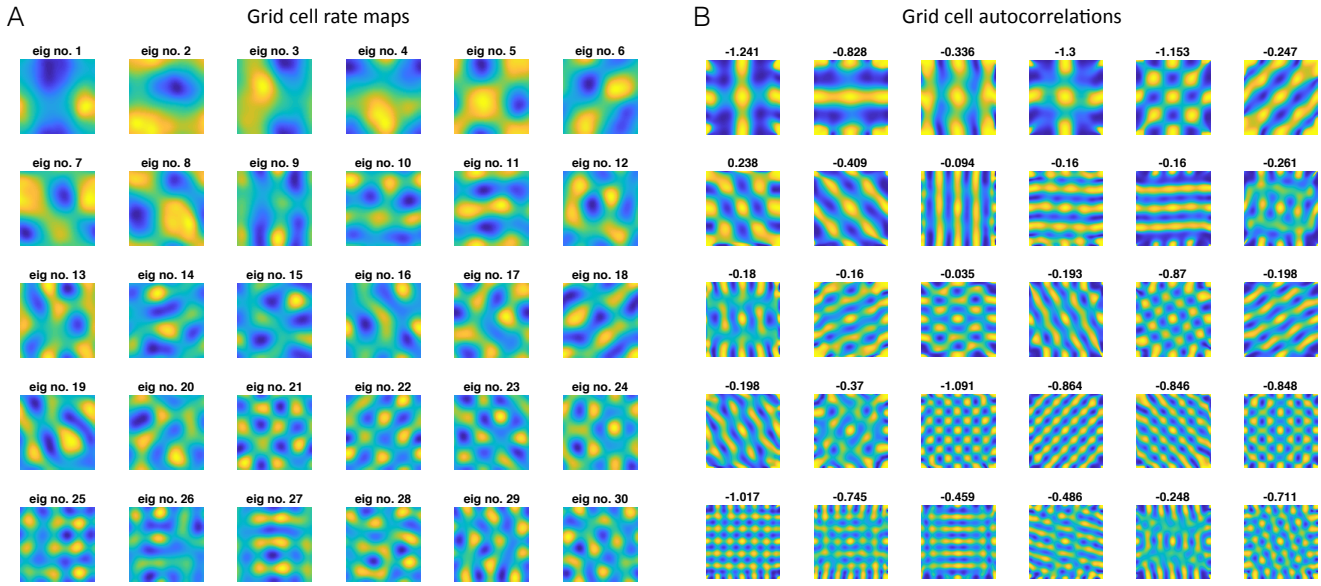


Figure 4.3 **Grid firing fields forming a square lattice emerge from a square environment**. A) Grid cell rate maps corresponding to the first 32 eigenvectors, and their autocorrelation (B). 2D spatial autocorrelogram panels in B correspond to equivalent panel in A. The value above each panel in B is the (hexagonal) grid score. Values >0 indicate a hexagonal grid cell; hence the square environment has yielded very few firing fields forming a triangular lattice. Grid scores are computed as in (Barry et al., 2012). They are explained further in Appendix 1, but the relevant details are that the higher the grid score the more 'griddy' it is, and grid scores >0 indicate it is more grid-like than expected by chance (Barry et al., 2012).

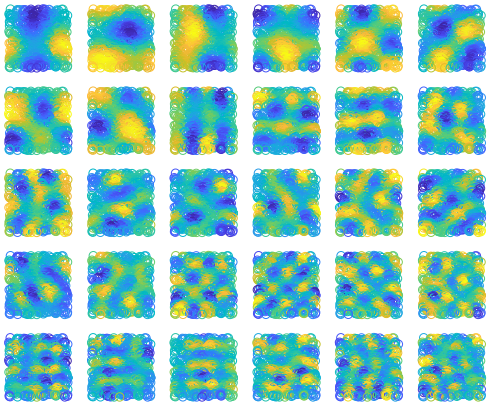


Figure 4.4 **Place to grid cell weights**. The place cell weights to each grid cell plotted at their corresponding place cell centre.

This periodic grid-like firing forms a square lattice over the environment, rather than a triangular lattice as is classically observed in the brain (Hafting et al., 2005). (Dordek et al., 2016) found that when placing a non-negativity constraint on the PCA, or on the place to grid cell weights, the resulting rate maps formed a triangular lattice. The non-negativity constraint was justified by the observation that place to grid cell weights are excitatory. Here, we ask whether the non-negativity constraint is required for a triangular lattice, or whether other conditions may be sufficient. We showed, holding all other settings constant, a triangular lattice is also obtained when the environment is defined by hexagonal rather than square boundary conditions, without a non-negativity constraint (Figure 4.5 & Figure 4.6 & Figure 4.7). This suggests non-negativity is not an absolute requirement to obtain a triangular lattice and that boundary conditions play an important part in defining grid fields.

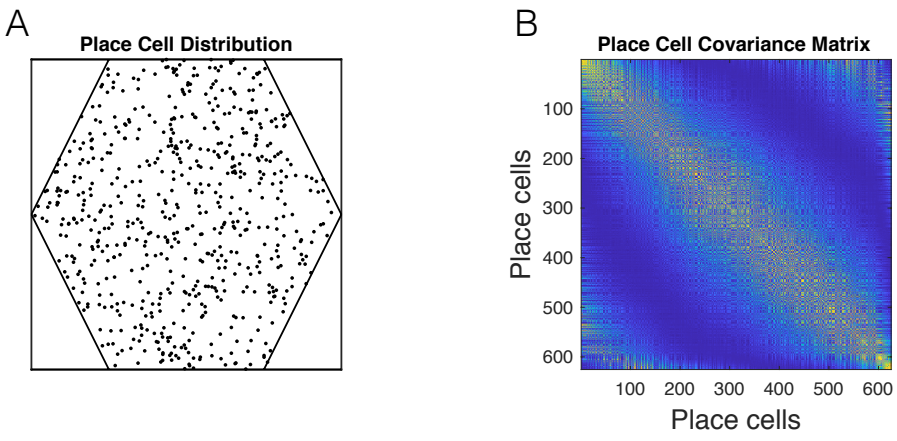


Figure 4.5 **Hexagonal environment**. A) Uniformly distributed place cells in a world with hexagonal boundary conditions. B) The covariance matrix of place cell activity in the hexagonal world.

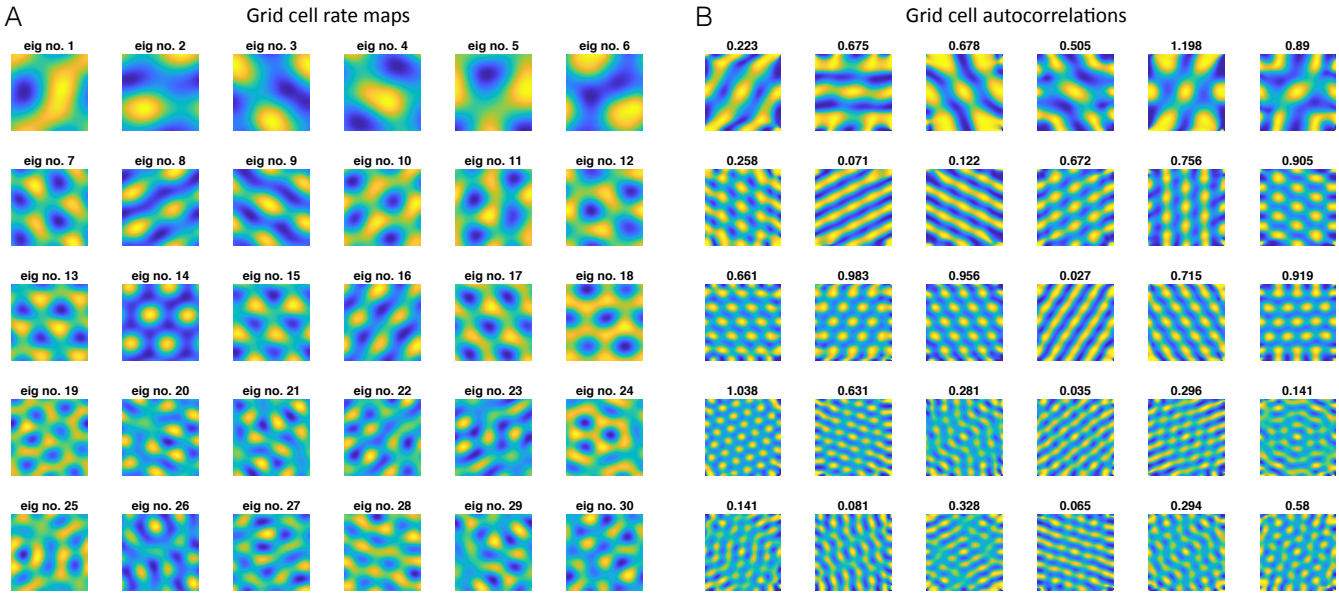


Figure 4.6 **Grid firing fields forming a triangular lattice emerge from a hexagonal environment**. A) and B) are analogous to those in Figure 4.3. Most grid scores have value >0 indicating grid cells emerging in a hexagonal environment form a triangular lattice.

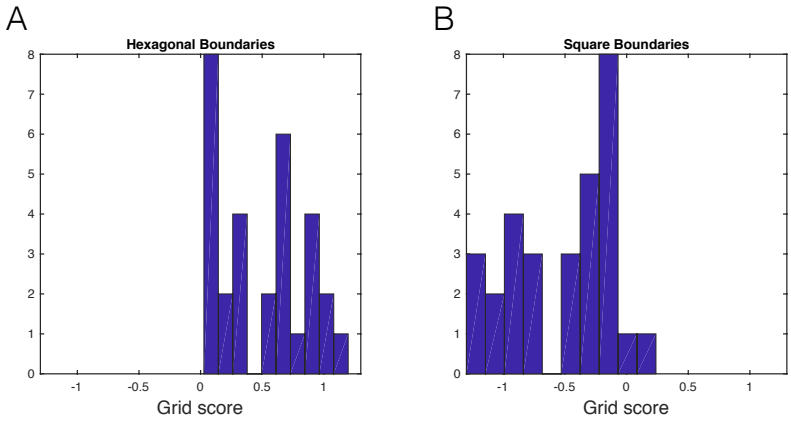


Figure 4.7 **Histograms of grid scores in hexagonal and square environments**. A) and B) are histograms of grid scores in environments with hexagonal and square boundary conditions, respectively, indicating that the hexagonal but not square environment gives rise to grid cells forming a triangular lattice.

4.4.3 Grid-like firing recovered as a PCA from real place fields

To determine whether this analysis works in real neural data where the agent's experience of the environment is naturalistic, and for some fun, we can try running PCA on

real rather than simulated place cell data to see if we obtain grid-like representations. Rather than randomly distributed Gaussians, we used real place cell rate maps obtained from (Chen et al., 2018) to define place cell activity at any given location. Following the same subsequent steps as above (instead of Equation 4.4, we used $G(x,y)_k = \sum_{p=1}^{n} \varphi_k^p \, P(x,y)_p$, where $P(x,y)_p$ is the firing rate of place cell p at location (x,y)), we obtained periodic grid-like rate maps, some of which formed a triangular lattice (Figure 4.8). If we randomly permute the elements of each eigenvector φ_k (place to grid cell weights) when computing the grid cell rate maps, we abolish this periodicity (Figure 4.8 & Figure 4.9), demonstrating that the periodic grid-like receptive fields are dependent on the relationships between place cells. Hence, we can successfully obtain grid-like representations from a PCA of real place cell data.

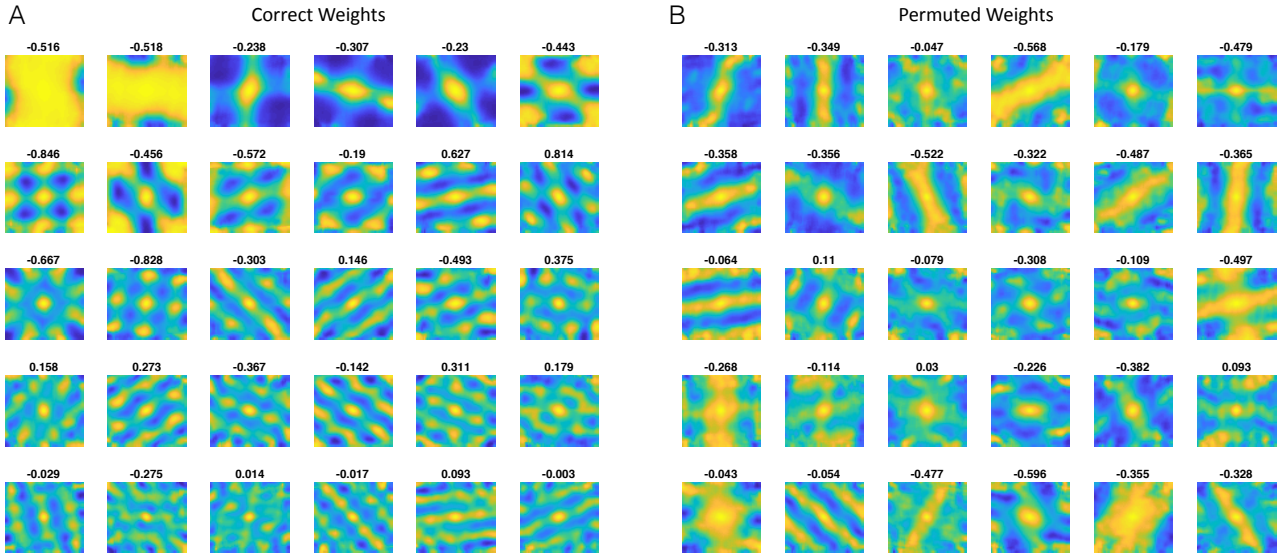


Figure 4.8 **Grid-like rate maps emerge from a PCA from real place cell rate maps**. A) Autocorrelation of grid cell rate maps obtained from PCA of place cell activity generated by walking through real rather than simulated place cell rate maps. B) Same as A but having permuted the eigenvector elements (place to grid cell weights).

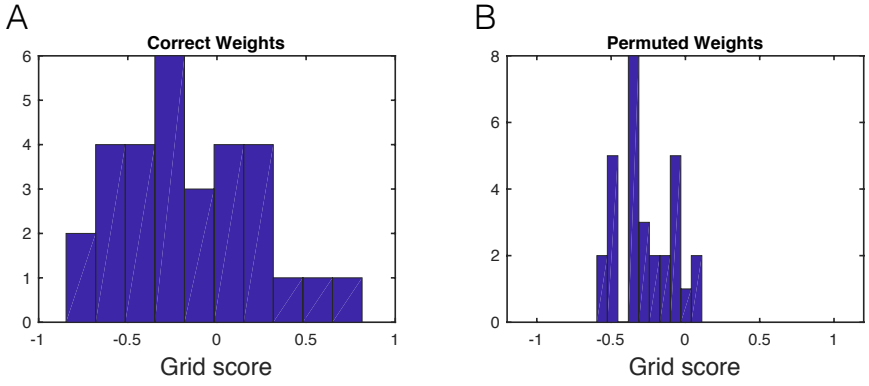


Figure 4.9 **Histograms of grid scores with correct and permuted place to grid cells weights**. More grid cells with grid scores greater than zero with correct weights (A) than permuted weights (B).

In this section, by demonstrating grid-like firing emerges from a PCA of 2D distributed place cell activity, we show grid cells may contain the necessary information to represent the structure of 2D environments. We note also that the mechanisms used to arrive at these grid-like firing fields, i.e. PCA of cell activity, are general, and not restricted to spatial domains, opening up the possibility that grid-like cells may represent structure in other, non-spatial domains.

4.5 Intuitive planning

In the previous section we demonstrated that we can obtain global representations from local learning, i.e. cells with periodic triangular lattice representations spanning the entire environment from only local observations (overlapping place fields). Here we explore how global representations can be used to provide distributions over future states and approximate distances between every pair of states in the world without search. Similar to vector-based navigation in a spatial context (Banino et al., 2018; Bush et al., 2015; Stemmler et al., 2015), these approximate distances provide a "sense of direction" – knowledge of distances without having to search – allowing "intuitive" navigation of the map, a process we refer to as intuitive planning. Because, unlike those for specialised for space, these representations are not restricted to the spatial domain, we show how global grid representations can be used for computationally efficient navigation to any point in an arbitrary topology state space.

4.5.1 Global distances from local learning

In order to demonstrate this we return to discrete state spaces wherein each state indexes a location on a discretised state space. We choose the adjacency matrix A over the transition matrix T since we borrow measures from spectral graph theory that are derived from A. A defines the relationships between states, such that element A_{ij} encodes the weight of the edge between state i and j. We can compute the distribution over future states n timesteps in the future by computing A^n (Figure 4.10). Element A_{ij}^n is the number of

walks[4] of length n from states i to j. As A^n is simply a power of A, all matrices A^n share the same eigenvectors of A:

$$A^n = \varphi \lambda^n \varphi^{-1}$$

Equation 4.5

where the matrices φ and λ are the eigenvectors and eigenvalues of A, respectively. It is notable that the eigenvectors of A representing a 2D box are periodic (Figure 4.11).

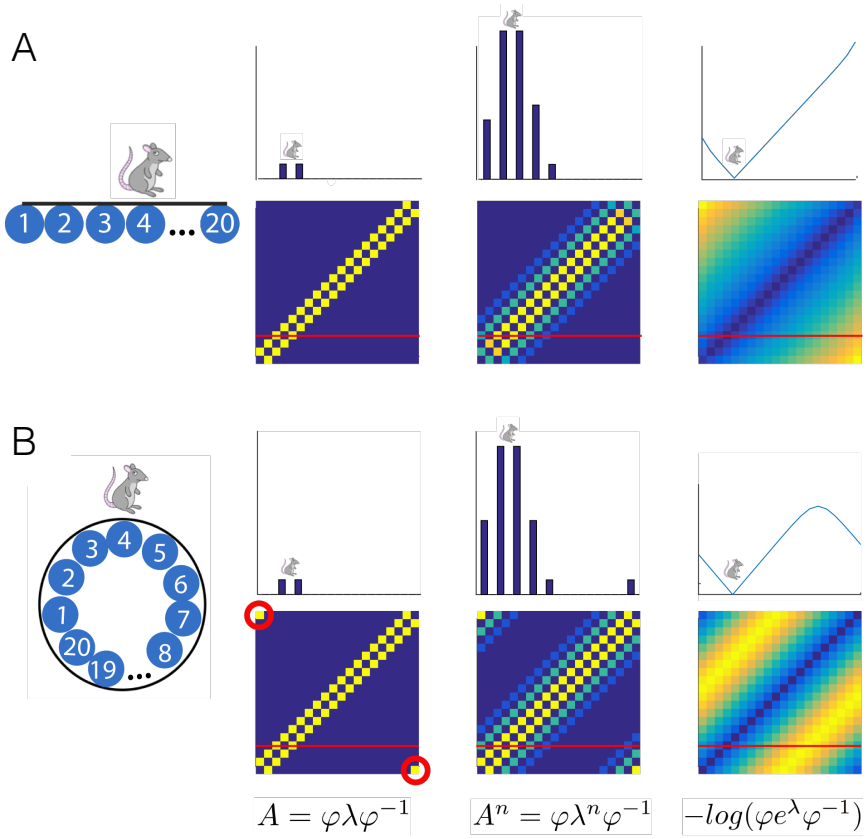


Figure 4.10 **Construction of local and global knowledge from locally-learnt eigenvectors.** A) A two-way linear track. For the three columns on the right, bottom is a matrix and top is a slice through that matrix (red line) at state four. From left to right: Schematic of the environment with a mouse occupying state four; the associated adjacency matrix A; A^n, where $n = 5$, and therefore A_{ij} is the number of possible walks of length n from state i to state j; $log(C)$, where C_{ij} is the communicability from state i to state j (communicability defined in the text). B) The same format as a for a two-way circular track. Red circles in the adjacency matrix mark the new adjacency between the ends of the linear track in a turning it in to a circular track. Bottom: Crucially, not only the matrix A containing local knowledge, but also A^n and $log(C)$ containing global knowledge can be constructed from the eigenvectors φ and eigenvalues λ of the matrix A. The local knowledge in state four is identical across the two environments (compare the slices through the adjacency matrices in A and B) but the global

[4] A walk between two nodes in a graph may revisit nodes and links along the way, whereas a path is a sequence of different nodes.

knowledge, obtained from the eigenvectors of A, is very different (compare slices in A^n and $log(C)$ across A and B).

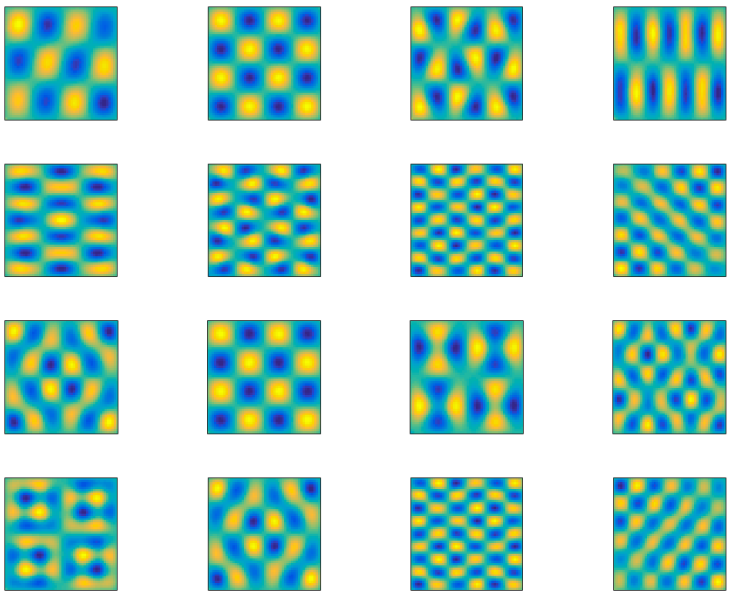


Figure 4.11 **Example eigenvectors of the adjacency matrix of a simple two-dimensional world are periodic.** Eigenvectors of the adjacency matrix A of a discrete-state 2D box projected back in to the space of the box have periodic waveforms. The permitted transitions between states in A were in the up/down and left/right directions and the boundaries were a square box, hence the square lattice.

Not only can we obtain distributions over future states from A, we can also obtain matrices of approximate distance metrics that encode the distances between all pairs of states, and therefore contain all the necessary information for planning the (approximately) shortest path to a goal. Because any summation of $B_i A^{n_i}$ (where n_i is an integer and B_i is an arbitrary scalar) will share the eigenvectors of A, we can compute (approximate) distance metrics that are a weighted sum of future states easily from the eigenvectors of A. For example, consider the following two weighted sums of future states:

$$\sum_{k=0}^{\infty} \frac{A^k}{k!} = e^A = \varphi e^{\lambda} \varphi^{-1}$$

and

$$\sum_{k=0}^{\infty} \gamma^k A^k = (I - \gamma A)^{-1} = \varphi(I - \gamma\lambda)^{-1}\varphi^{-1}$$

Equation 4.6

where γ is a discount factor, $\gamma < \frac{1}{\lambda_{max}}$ and λ_{max} is the largest eigenvalue of A in modulus (Aprahamian et al., 2016; Benzi and Klymko, 2013; Estrada et al., 2009). The infinite sums in the above equations converge to the matrix exponential (e^A) and resolvent ($(I - \gamma A)^{-1}$), respectively.

The dominant contribution to element (i, j) of the resulting matrices in each of these sums is the shortest path between i and j. This is because the contribution of longer paths is dramatically downweighted due to the factorial or exponential decay. Hence element (i, j) will have a high value if states i and j are closer and therefore contains information closely related to the shortest distance between the states, which is the necessary information for navigation. Since longer paths do contribute something, these measures do not necessarily vary monotonically with distance (see below).

In graph theory, these two summations (Equation 4.6) are known as the communicability between nodes of a graph (Estrada and Higham, 2010). The communicability between nodes i and j (C_{ij}, where C is the communicability matrix) measures the information flow between nodes i and j. C_{ij} will be higher if information flows more easily between nodes i and j. Communicability was proposed as a generalisation of the concept of the shortest path to include longer paths that may be relevant for information flow between pairs of nodes (Estrada et al., 2009). Notably, the resolvent summation is closely related to the successor representation in reinforcement learning (Dayan, 1993), which is computed from T rather than A.

Critically for our purposes, these summations – and therefore approximate distance metrics – can be easily computed from the eigenvectors and eigenvalues of A. It is not necessary to simulate through all paths in the environment. In order to obtain these summations, one can simply recombine the eigenvectors of A with different weights. Instead of the original eigenvalues λ, λ^n will give the distribution over states n steps in the future, e^λ will give the matrix exponent, and $(I - \gamma\lambda)^{-1}$ will give the resolvent. Furthermore, these reweightings are inexpensive to implement because the number of reweightings required is equal to the number of eigenvectors (grid cells), rather than the number of relationships that can be computed (which, assuming all eigenvectors/values are retained, in this case is number of eigenvectors/values squared). Hence approximate distance measures may be efficiently computed from grid-like cells.

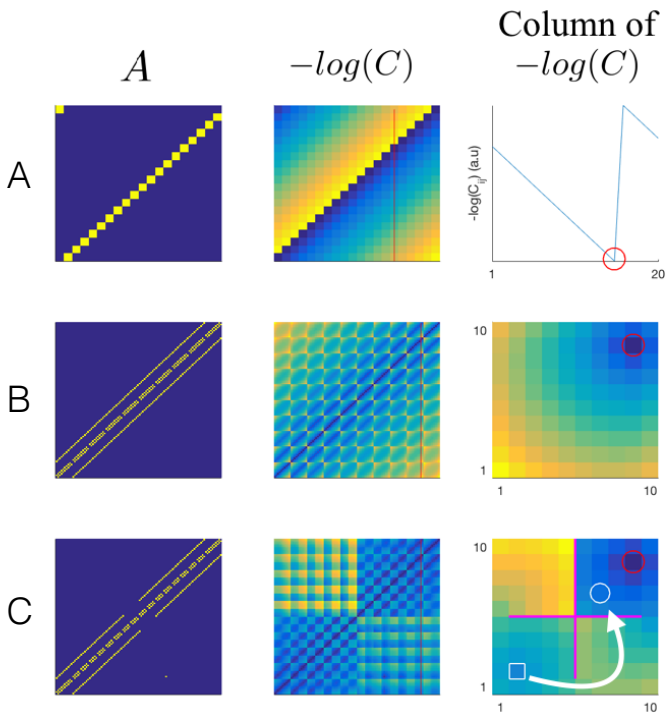


Figure 4.12 **Communicability measures in arbitrary topologies**. A) A one-way linear track. Left: The adjacency matrix A. Middle: the matrix $-log(C)$ where element C_{ij} is the communicability between states i and j. Red line marks the column of a target state (note this is different to Figure 4.10 where it is the row of the current state). Right: This column represented on the space defined by A. Red circle indicates target state. B) A two dimensional 10x10 box. The same format as A except the target column of $-log(C)$ is represented as a heatmap on a two-dimensional environment with warmer colours corresponding to higher values. C) A box with barriers and a wormhole connecting states which are not adjacent in a two-dimensional space. Same format as B. States separated by a magenta line are not adjacent. White squares and circles mark entrance and exit gates of a wormhole, respectively: the states are adjacent yet distant in the regular two-dimensional topology. The conventions used in this figure will be used in subsequent figures.

We emphasise that since A can represent any graph, these measures work for arbitrary topology environments. They reflect approximate distance measures in environments that cannot be mapped on to normal 2D maps, such as those containing wormholes linking distant parts of the map, as well as environments with one-way links between states (Figure 4.12). It is important approximate distances can be computed in arbitrary topologies since non-spatial cognitive maps representing knowledge will likely take arbitrary forms.

Both communicability measures can be approximately linearly related to the true distance, as defined by the shortest number of links between states. This is achieved by taking the $-log$ of the communicability, a measure we refer to as intuitive distance (Figure 4.13). As mentioned above, the relationship between intuitive distance and true distance is not only dependent on the length of the shortest path but also on the (further downweighted) longer paths. When the set of walks between a pair of states can be obtained from the set of walks between another pair of states with the same true distance, there is no variance in the intuitive distance for a given true distance. In these cases, intuitive distance will monotonically strictly increase with true distance. An example of such a case is a two-way, one-dimensional ring (Figure 4.13). However in environments where there are many short, yet not necessarily shortest, paths between states, the intuitive distance may differ between pairs of states with the same true distance. In such a case intuitive distance may not monotonically increase with true distance. Hence these approximate distance measures can give the impression a state is closer than it is, which predicts failures to navigate the shortest path between states (see below).

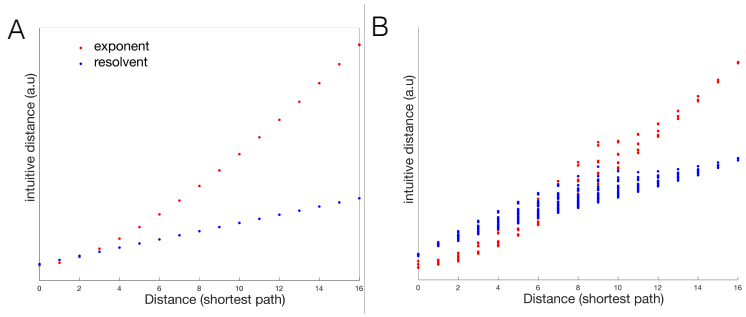


Figure 4.13 **Approximately linear relationship between intuitive distance and true distance**. Intuitive distances between pairs of states are approximately linearly related to the true distance, as defined by the length of the shortest path. A) Intuitive distance reliably increases with true distance on a two way, one-dimensional track (the same as that reported in Figure 4.10). B) Spread of intuitive distances for a given true distance in a 10x10 two-dimensional box (the same as that reported in Figure 4.12), because communicability is influenced by the number of routes between states.

4.5.2 Global navigation from local computation

One advantage of the representation proposed is that it is possible to calculate the shortest route between two states (locations), i.e. to navigate the environment, without a costly sequential search of all possible routes (such as by simulation of possible paths). Given we have access to all pairwise approximate distances between states, we can navigate between any pair of states. Given neighbour (N) and target (T) states, as well as the

eigenvectors and eigenvalues, the agent can compute without search which available neighbour state is closest to the target state. The agent can do this by computing for the target state T and each neighbour state N:

$$C_{NT} = \sum_{v=1}^{n} \varphi_v(N)\varphi_v(T)e^{\lambda_v} \quad \text{or} \quad C_{NT} = \sum_{v=1}^{n} \varphi_v(N)\varphi_v(T)\,(1-\gamma\lambda_v)^{-1}$$

Equation 4.7

where $\varphi_v(N)$ is the Nth element of the vth eigenvector – associated with eigenvalue λ_v – of A, and n is the number of eigenvectors. The agent then simply proceeds to the neighbour state with the maximal C_{NT}, or equivalently minimal intuitive distance $-\log(C_{NT})$. The agent therefore effectively performs ascent on communicability with the target state or descent on intuitive distance, which finds trajectories that are approximately optimal without any tree search. Since the structure of the environment is stored, the agent can flexibly reroute to a new target state, for example if goal locations or internal goals change. Further, parallel neural circuitry can naturally implement the summations to compute C_{NT}.

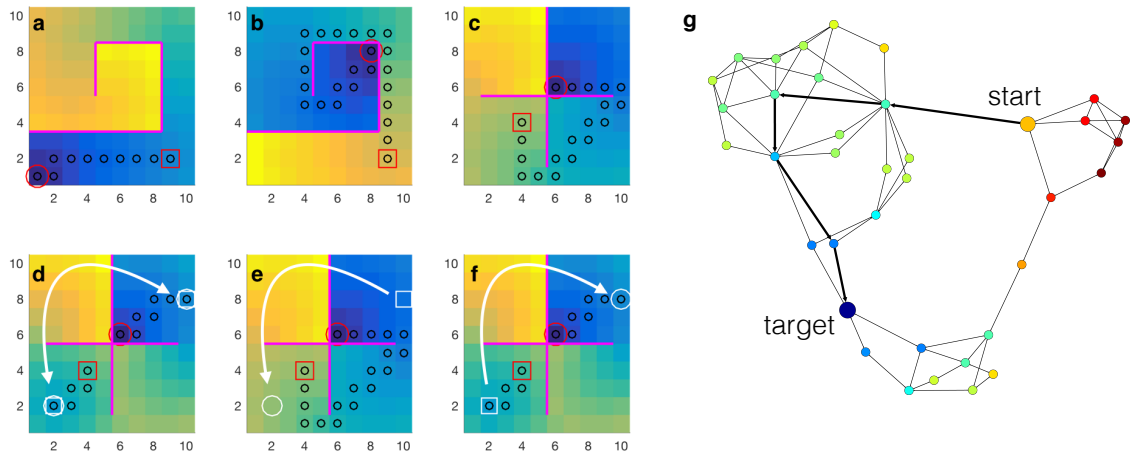


Figure 4.14 **Navigation in arbitrary topology worlds**. Successful navigation from start (red square) to target (red circle) states in two-dimensional worlds with barriers and differing start and target states (a-c). Introducing wormholes that connect disparate parts of the map and therefore warp it rendering it no longer described by a simple map does not affect the ability to successfully navigate to the target (d-f). Nor does the introduction of a one-way wormhole (e-f), resulting in complex eigenvalues and eigenvectors. In a-f navigation simulations, state transitions were permitted in the up/down and left/right directions (unless, for example, blocked by a barrier), and black circles denote the agent's path. g) Successful navigation in a graph of arbitrary topology. Colouring of nodes denotes intuitive distance (hotter is larger distance).

We constructed several discrete-state mazes and demonstrate intuitive planning can successfully navigate the shortest route to the target in several 2D, undirected mazes (Figure 4.14). The shortest route – defined by the minimum of transitions between start and target state – is successfully navigated using both the matrix exponential and the resolvent. Non-spatial cognitive maps, however, will likely have arbitrary topologies. We show the algorithm generalises to arbitrary topologies as its performance is not affected by the insertion of a wormhole, rendering it no longer topologically 2D (Figure 4.14). The algorithm also successfully navigated to the target in randomly generated, unweighted graphs, provided there was a path to the target (e.g. Figure 4.14).

Since intuitive distance is only approximately related to true distance (Figure 4.13), it is possible to construct mazes where the agent successfully navigates to the target, but does not take the shortest route. This is achieved by associating states that are not part of the shortest route with multiple longer routes to the target, which contribute to the communicability measure. Hence their communicability with the target can become higher than that of states on the shortest route, and the agent will erroneously choose these longer-route states. Such planning errors form potential experimental predictions. For example, an animal may select a slightly longer route to a goal if the first state of the route has higher communicability with the target than the first state of a shorter route (Figure 4.15).

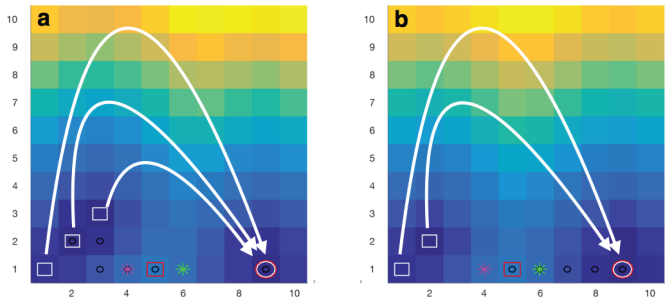


Figure 4.15 **Failure to navigate the shortest path**. Introducing multiple wormholes that cause multiple longer routes to the target to pass through a nearby state (magenta star) that is not in the direction of the shortest route (green star) causes the agent to take a longer route to the target (a). With fewer such wormholes, the shortest route is taken (b).

4.6 Discussion

In this chapter we have: first, shown how space can be conceptualised under a reinforcement learning or state space framework by considering locations as states, second, demonstrated that grid-like firing can be obtained from the PCA of 2D-distributed place cell activity by following the implementation of (Dordek et al., 2016), and third, presented a method for flexibly and efficiently navigating arbitrary topology environments using grid-like eigenvector codes. We first briefly discuss grid cells as a PCA and intuitive planning separately, and then several points that relate to both.

Grid cells as a PCA

Our implementation of grid cells as a PCA of place cells followed that of (Dordek et al., 2016). We do not claim to have done something novel here, but do note a couple of subtle extensions. First, in their implementation they required a non-negativity constraint on the eigenvectors (place to grid cell weights) in order to obtain a triangular lattice; without the non-negativity constraint a square lattice emerged. We showed that a non-negativity constraint is not necessary and hexagonal boundary conditions are sufficient to obtain a triangular lattice. Second, we showed grid-like rate maps emerged when using real place cell data to define the place cell rate maps.

The lack of dependence of the place cell covariance on the agent's policy is a potential weakness of the above simulations. As discussed above, the relation between one place (cell) and another is represented in the covariance matrix of the place cells' activity, such that the covariance between two place cells will be high if their firing fields overlap, i.e. they index physically close locations. The place cell covariance matrix can therefore be thought of as representing something analogous to adjacency in the sense that adjacent locations will have high covariance between their corresponding place cells. Whilst adjacency would provide a veridical representation of the structure of the environment, it is often more useful for an agent to represent where it is most likely to be in the future given its current state, which is additionally dependent on the agent's policy. For example, the agent may be more likely to travel into the firing field of place cell i from place cell j than vice versa, and this could be useful to know. Such representations would be analogous to the transition matrix. In the simulations above and those of (Dordek et al., 2016) the place cell covariances are independent of policy because the covariance is defined purely on overlapping firing fields and is not dependent on, for example, where the agent has just come from (because there is no temporal component to the place cell activity). Therefore,

assuming equal sampling of the environment, the agent could be teleported around the environment randomly (akin to shuffling the columns of the place cells by time activity matrix) and the covariance matrix would remain the same. Additionally, covariance matrices are symmetric and therefore do not take into account asymmetries in directions of travel. Whilst during development it may be useful to sample entire environments (maybe this is why babies crawl all over a room) to build more veridical, policy independent representations of the environment to act as bases for representation of similar environments encountered in the future, it is additionally important to acquire policy dependent representations. These could be obtained by, for example, having temporal lags in the place cell correlations, or taking in to account asymmetries with asymmetric STDP. Policy dependent representations have recently been investigated in the context of the successor representation under an agent's policy, which explains features of place cell representations such as skewing of the place field dependent upon the direction commonly travelled by the agent (Stachenfeld et al., 2017).

The vectors resulting from PCA in general are an orthogonal basis set, in which the leading eigenvectors (those associated with the highest eigenvalues) explain the maximal variance in the data. They therefore provide an efficient, low dimensional representation of the structure of data, suggesting grid cells are an efficient representation of space. This is consistent with observations such as a grid code requiring the fewest number of neurons in order to be able to decode location (Wei et al., 2015). We note here that the underlying computations that give rise to grid cells need not be a perfect PCA, but that the resemblance between grid codes and PCA suggest they are extracting similar statistics.

The implications of grid cells resembling a PCA of 2D space are quite deep. It implies they are summarising the underlying statistics or structure of 2D space. Instead of representing the low-level features of the environment, such as the sensory experiences occupying the space, they rather represent its topology; how states relate to one another. Therefore embedded in this representation is the necessary information to navigate the state space. Rules that are true in worlds with a certain topology, such as the shortest route between location A and location B in 2D is a straight line, can be extracted from representations of the topology of the environment, allowing efficient computation (Bush et al., 2015; Stemmler et al., 2015). One particular advantage of such representations rendering them extremely powerful is that, because they are separated from the representation of the entities they encode the relations between (e.g. place cells), they can

be generalised to novel situations sharing structure – allowing reuse of previously learnt knowledge. This is discussed again below and explored in the next chapter.

Intuitive planning

Hence the underlying structure of the environment may be represented in a grid-like code containing information about how states relate to each other, thereby allowing efficient navigation of the environment. Such efficient navigation has already been shown in 2D space (Bush et al., 2015; Stemmler et al., 2015). By recognising that global, grid-like codes could be used to easily compute a distance measure between any pair of states in the world, we presented a method for flexibly and efficiently navigating arbitrary topologies.

To place this kind of planning in context, it lies somewhere on the continuum that exists between model-free and model-based decision making systems, mentioned in the introductory chapter. The former is computationally efficient but inflexible and the latter is maximally flexible but inefficient. Intuitive planning, or planning based on predictive (successor) representations generally, trades some flexibility for efficiency. As discussed in the introduction, this comes from decomposing the value function in to a successor matrix and a reward vector, allowing the agent to be flexible to changes in reward, unlike model-free systems. Because it contains predictive representations that are a weighted sum of future states, expensive searches though states are not necessary since state representations already contain information about future states, unlike model-based systems. This comes at a cost to flexibility since a change in the transition structure of the environment is slow to incorporate in to the predictive representation.

There is some evidence for predictive representations in the brain. Behaviourally, that the value function is decomposed in to successor representations and reward vectors predicts some degree of choice behaviour is accounted for by flexibility to changes in reward but not transition structure, which has recently been shown in humans (Momennejad et al., 2017). Neurally, certain features of place cell responses are accounted for by considering them as predictive representations rather than indexing only the current location (Stachenfeld et al., 2017). Neural similarity measures reveal the structure of discrete state-space, non-spatial tasks can be recovered from hippocampus and entorhinal cortex (Garvert et al., 2017; Schapiro et al., 2013; Stachenfeld et al., 2017). These representations may be better explained by communicability measures that are weighted sums of future states than the adjacency matrix (Garvert et al., 2017), providing evidence the representations required for intuitive planning-like navigation may exist in the brain in non-spatial tasks.

A potential advantage of intuitive planning is that it allows place cells to retain their local scope – i.e. to only code the current location. The recent proposition that place cells are predictive representations accounts for a variety of experimental observations (Stachenfeld et al., 2017), however a potential downside is that they therefore lose their local coding, which may have implications for, for example, precise credit assignment. Further, these predictive representations emerge slowly (Stachenfeld et al., 2017) and therefore may be mediated by different mechanisms than those that give rise to immediate hippocampal codes upon entry to a new environment. Intuitive planning on the other hand does not require place cells to undergo slow learning of which future states they are predictive of.

We note that we do not actually believe intuitive planning as we presented here is implemented in the brain. There are some obvious issues. For example, it is unclear what mechanism would allow for switching between the relative weightings of the grid cells (e.g. λ to e^{λ}). The main purpose of intuitive planning was to illustrate the utility of global representations and how grid cells may, in theory, be able to navigate arbitrary topology spaces without search.

Global codes from local learning

An important point relating to grid cells as eigenvectors is that these global representations can arise from only local learning. In the context of grid cells emerging as a PCA of place cell activity, the simulated agent never has access to the overall structure of the environment, it only ever observes local (one-step) transitions through the environment. Nonetheless the resulting grid cells are global (periodic over the entire space). Similarly, the eigenvectors used in intuitive planning are derived from the adjacency matrix which describes only local (neighbour) relationships and can therefore also be constructed locally. These considerations are important because a real world agent only ever has access to its local experiences, yet is able to extract the underlying structure in the environment, represent it globally (e.g. grid cells), and can use this structure to exhibit sophisticated, flexible behaviours, such as navigating without search (Bush et al., 2015).

Non-spatial cognitive maps

Importantly, both framing grid cells as a PCA and intuitive planning are not specific to the spatial domain or 2D worlds. Indeed, intuitive planning was motivated by the need to navigate non-spatial, arbitrary topology cognitive maps not navigable by vector-based

navigation supported by 2D grid cells. PCA is a general method that is of course not specific to 2D data, or the identity of the dimensions. Hence, the grid cells resulting as PCA of 2D space presented above are agnostic to the identity of the two dimensions. They could, for example, have been the leg and neck length of birds (Constantinescu et al., 2016). Therefore, as a result of casting space as a state space, everything presented in this chapter is part of a general framework that can handle state spaces of arbitrary topology. Understanding the structure of tasks or causality – which also depend on the hippocampal formation (Duncan et al., 2018; Dusek and Eichenbaum, 1997; Eichenbaum, 2004) – could be viewed just as different statistics. That this framework can explain spatial representations suggests it is possible there is not much special about space *per se*, but rather spatial representations emerge as an instance of more general principles that can unify spatial and non-spatial coding (Behrens et al., 2018).

Experimental predictions

The theoretical work outlined in this chapter makes numerous experimental predictions. For example, if grid cells resemble the eigenvectors of covariance or adjacency matrices (above, and in (Dordek et al., 2016; Stachenfeld et al., 2017)), it should be possible to construct tasks with specific (non-spatial) structures and observe cells in the brain coding the (non-hexagonal) eigenvectors of the matrices defining the task structure. Such predictions are being testing in ongoing experiments in the lab. Predictions arising from intuitive planning include: unlike in classic tree search planning (Johnson and Redish, 2007), intuitive planning does not require a preplay through place cells leading to a goal. Relatedly, in intuitive planning the distance to a goal does not influence the difficulty of the next decision, unlike in tree search. We therefore predict for distant goals, when tree search is difficult or impossible, the distance to the goal will not predict reaction times. Finally, as described above, when there are many routes to a goal that pass through a nearby state, but the shortest route passes through a different state, intuitive distance and true distance can dissociate. This predicts that under these circumstances, humans or animals will take sub-optimal, longer routes to a goal.

Representation over computation

Using such structural representations to efficiently plan without search, as in intuitive planning, speaks to a more general point that having an appropriate representation can mitigate expensive computations (Akam et al., 2015). As discussed above, 2D grid cells

similarly provide a representation that supports navigation not requiring search through locations (Banino et al., 2018; Bush et al., 2015; Stemmler et al., 2015). There are other types of representation that are useful for mitigating computation, such as latent state representations and hierarchical representations (Akam et al., 2015; Botvinick, 2012; Solway et al., 2014). A model-free agent (that does not plan) is able to approximate the behaviour of a model-based agent (that plans through a model of the world) by building appropriate latent state representations (Akam et al., 2015). Hence appropriate representations can dramatically reduce computational costs at decision time. This makes sense given how difficult it is to plan with tree search: How often do we really plan deeply? We would far rather rely on some more efficient representation of the state space, such as hierarchies, or fall back on (sometimes sub-optimal) heuristics (Huys et al., 2015). Favouring appropriate representations over computation shifts the burden from expensive online planning to offline building of appropriate world models – consistent with offline replay of experiences thought to consolidate knowledge (Wilson and McNaughton, 1993a). Learning representations that mitigate expensive computation may be a general principle optimised over evolution.

Generalisation

Representing the underlying structure of the environment becomes particularly useful when we can generalise this structure to new situations to exploit previous learning. Whilst it is possible to learn within an environment how to navigate through the lower-level features of that environment (e.g. learning which specific objects occupying the space lead to which), this knowledge will not generalise well to new environments where the relationships between the specific objects may not be preserved. The advantage of representing the underlying topology is that this representation and the algorithms operating on it (such as planning or navigation algorithms) can be reused in situations where the underlying statistics or structure of the state space are preserved – we can immediately navigate in rooms we have never seen before, even if their arrangement of sensory experiences differs from any past experience. Additionally, learning the structure of the world occurs over many different experiences. It is therefore critical we are able to generalise structural representations across different environments.

How may we do this? As alluded to above, that the structural representations in grid cells presented in this chapter are separated from the states they encode the relationships between. There is therefore a separate neural population coding for structural relationships,

which we refer to as an *explicit* representation of structure. Because the representation is explicit, it may be reused in novel situations. Explicit representations are contrasted with encoding the structure in the connections between representations of objects, wherein there is not a separate population coding for the structure, and therefore this knowledge cannot be generalised.

Once we have an explicit structural representation, we are then faced with the problem of how to combine it with novel experiences to allow generalisation. The structural representations in the setups described above are learnt from or used in only one environment, e.g. one set of place cells, or adjacency matrix. In order to generalise the structure, it would require the inputs to the grid cells, i.e. place cells or states, to be realigned to match the old arrangement of place cells or states on which the structure was learnt. This alignment problem would not be easily solvable without having to search through the different combinations of place cell arrangements.

Notably, we have not discussed the role of place cells at all in this chapter. Rather they were simply provided. Far from their sole role being just to provide representations of specific locations (or states) from which to learn grid cells, we view their function as critical to closing the loop and combining structural representations (in grid-like cells) with novel sensory experiences, i.e. forming a bridge between old knowledge and new situations. We propose place cells are a conjunctive code between explicit structural representations and representations of sensory experiences, allowing previously learnt structure to organise new experiences and memories to facilitate generalisation. This, along with the rest of a framework for generalising structural knowledge in the hippocampal-entorhinal system, is the topic of the next chapter.

5 Generalisation of structural knowledge in the hippocampal entorhinal system

5.1 Abstract

A central feature of flexible adaptation to the environment is our ability to generalise our knowledge to novel situations. This allows previously learnt structure to be exploited to solve tasks in novel situations differing in their particularities. In this chapter, we present a framework that proposes that in order to generalise structural knowledge, the representations of the structure of the world, i.e. how entities in the world relate to each other, need to be separated from the representations of the entities themselves. Under this framework, structural information represented in grid-like cells and sensory information represented in high-level sensory cortices are combined to form a conjunctive representation in hippocampal place cells. This conjunctive code allows generalisation and can form fast episodic memories of what happened where. Hierarchical representations offer efficient representations of space and allow reuse of knowledge across space improving learning efficiency, as well as facilitate memory retrieval. We implement this proposal in an artificial neural network tasked with predicting the next sensory experience – a task that benefits from learning generalisable structural representations. The network learns grid and place cell representations and generalises structural knowledge across environments. Furthermore, our framework makes the prediction that, in order to generalise knowledge, the relationship between grid cells and place cells should be preserved across environments. We experimentally support this prediction by analysing single-unit data from a rodent remapping experiment. This chapter is joint work carried out with James Whittington.

5.2 Introduction

A core component of intelligent and flexible behaviour is generalisation. This allows previously learnt structure to be exploited to solve tasks in novel situations differing in their particularities. Flexibly reusing our understanding of the relationships between entities in the environment when encountering novel situations is clearly advantageous, as it means we do not have to learn afresh. When one enters a new room that consists of elements

arranged in a way unlike anything they have seen before, they can still understand the basic rules of 2D space and how they apply to this new room. When encountering a new social group, one draws parallels between its structure and those previously learnt, allowing the implications of inferring the dominant agent to be immediately understood (Behrens et al., 2018). Although such generalisations are often so effortless we don't even realise we are doing them, their neural mechanisms remain a mystery.

There is evidence to suggest the hippocampal formation may perform generalisation. Our ability to generalise knowledge to novel situations is thought to rely, at least in many instances, on abstract, conceptual knowledge. Concepts are formed through abstraction, and by capturing the similarities between entities in the world provide principles for organising new experiences in the context of previous knowledge (Kumaran et al., 2009). Critically, conceptual knowledge is thought to rely on the hippocampal formation (Constantinescu et al., 2016; Kumaran et al., 2009; Mack et al., 2016), and the hippocampus in particular is thought to be important for their generalisation to new situations (Kumaran et al., 2009).

The hippocampal formation is additionally known to be important for various other functions, such as memory, causal and transitive reasoning, structure learning, social cognition, and spatial cognition (Hassabis et al., 2007; McKenzie et al., 2014; Van Der Meer et al., 2012; Ólafsdóttir et al., 2015; Tavares et al., 2015). The principles relating these various functions and their neural implementation are not yet known. However, learning, representing and generalising abstract, conceptual or structural knowledge is a domain-general function, and may offer a unifying account for these various different roles of the hippocampal formation.

We now have a reasonably good characterisation of the representations in the hippocampal formation, most notably in the form of place cells and grid cells, and so can ask how these may support generalisation and investigate the principles under which these ground truth representations emerge. We propose grid cells offer a generalisable structural code. As discussed in the previous chapter, given grid cells resemble a PCA of place cell activity, i.e. the relationships between place cells (Dordek et al., 2016; Stachenfeld et al., 2017), and given grid cell-like firing has been observed in non-spatial tasks topologically similar to space (Aronov et al., 2017; Constantinescu et al., 2016), suggests they represent the topology of the state space. Furthermore, grid cells maintain their relationships with each other when the animal is in a structurally similar but different environment (Fyhn et al., 2007), suggesting information embedded in the grid code, i.e. structural knowledge, is

preserved across environments. This is in contrast to place cells, which remap across environments (Bostock et al., 1991; Fyhn et al., 2007; Leutgeb et al., 2005) such that place cells that were active next to each other in one environment most likely won't be in a second. We propose place cells may offer a conjunctive code between the structural representation in grid cells (in MEC) and sensory representations in LEC that allows for generalisation to new experiences. Consistent with this is the observation that place cell activity is modulated by the sensory environment as well as location (Komorowski et al., 2009; Wood et al., 1999). A conjunctive representation would predict place cell remapping is not random but rather place cells remap to locations consistent with the conjunction. Hence we propose grid cells represent the statistics of place cells, and use these representations to form a conjunctive code in the place cells allowing generalisation of the statistics to new situations.

Grid cells representing the statistics of place cells has obvious parallels with the highly influential complementary learning systems (CLS) theory, whereby statistics of memories stored in hippocampus are extracted by cortex (Kumaran et al., 2016; McClelland et al., 1995). CLS posits that hippocampus binds entities represented in disparate parts of cortex in to a single episode that is then stored. The statistics of these memories are slowly extracted by cortex, which learns common patterns across events allowing for generalisation. However, CLS hasn't closed the loop. It does not explain how cortex then constrains future hippocampal representations in order to place new episodes and memories in the context of previous experience.

In this chapter, we present a framework in which the structure of world, i.e. how the entities in the world relate to each other, is slowly extracted by grid cells and separated from the representations of the entities themselves – forming abstract conceptual knowledge. We close the loop in the sense that these representations are combined with high-level sensory representations in a conjunctive code in hippocampal place cells that allows generalisation of knowledge to novel experiences and memories. This conjunctive hippocampal code is endowed with Hebbian learning that allows fast formation of episodic memories. Hierarchical representations provide an efficient representation of space, reuse of knowledge across space, and facilitate recall of appropriate memories when recalling which sensory experiences happened where. We implement this proposal in an artificial neural network tasked with predicting the next sensory observation when walking on multiple 2D graphs, where each vertex is associated with a sensory experience but the arrangement of these sensory experiences differs across environments. To make accurate

predictions, the agent must learn the underlying hidden structure and generalise this across environments. Whilst fast Hebbian learning allows for fast formation of episodic memories within an environment, gradient descent learns to extract the statistics of these memories over multiple different environments, resulting in the learning of grid cells. We focus on 2D graphs so we can compare these learnt representations to the ground truth spatial representations measured in the brain. We empirically test a prediction of our model in data from a remapping experiment (Barry et al., 2012), demonstrating that place and grid cell relationships are preserved across environments despite place cell remapping.

In this chapter, we first outline the conceptual framework, motivating each of the key components mentioned above with reference to both their requirement to solve the problem of generalising structural knowledge and to the neuroscientific literature. Since this framework is difficult to explain, this section takes up the majority of this chapter. We then present the implementation of this proposal in an artificial neural network. We do not present details in this section, since this was work carried out by James Whittington. We then present analyses of single unit data to test the prediction that there is a preserved relationship between place and grid cells across environments.

5.3 Conceptual framework

In order to understand how this framework is able to generalise structural knowledge, several key components need to be simultaneously put together. Here we define the problem and run through each of the key components separately. Each is motivated by its requirement to solve some part of the computational problem of generalisation and also by neuroscientific considerations such as grounding it in data or offering mechanistic explanations for neural phenomena. We then put them all together by illustrating the key concepts in a 1D linear track environment.

5.3.1 Problem: predicting the next time step in worlds with common structure

Here we describe a task environment that encourages an agent to learn abstract, generalisable structural knowledge. In this environment, an agent is ultimately tasked with predicting the next sensory observation, and this is facilitated with an understanding of the structure of the environment. The agent walks around on multiple different discrete state environments, the structure of which are defined by underlying graphs that share the same statistical structure (e.g. 2D-ness). A sensory experience (e.g. a banana) is placed at each

node of the graph, randomly sampled from a set of candidate sensory experiences (Figure 5.1). The agent's task is to predict the sensory experience at the next time step. Learning to predict the next time step is a very powerful form of learning as, although technically unsupervised learning, it effectively offers large amounts of free supervision.

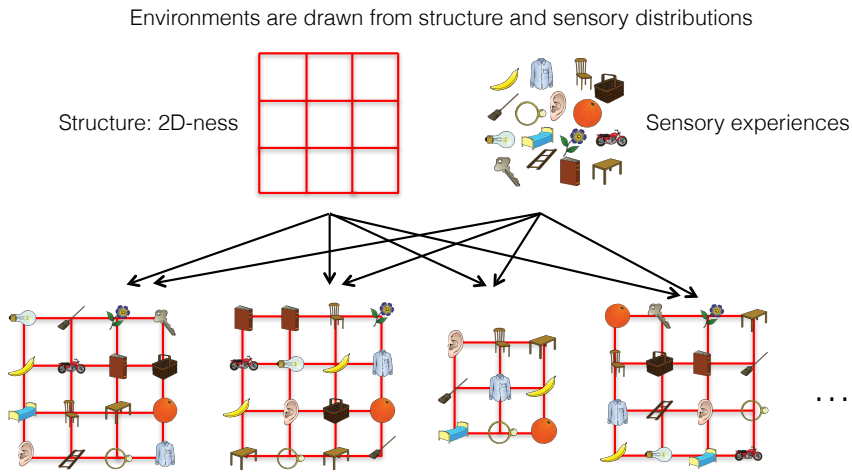


Figure 5.1 **Environments that share structure but differ in their particularities**. A given environment is made up of a structure and a sensory component. The structure component describes the underlying graph, in the present work 4-way connected 2D graphs. The sensory component describes which sensory experience drawn from a set of possible sensory experiences is present at each node of the graph. The sensory experiences for a given environment are drawn at random, resulting in environments that share structure but differ in the arrangement of their sensory experiences.

Learning the underlying statistics of the environment will enhance the agent's ability to predict sensory experiences. This is because the structure of the environment constrains what is possible. In 2D worlds, for example, we know that if point A and point B are close to each other, as are B and C, then A and C must also be close to each other. Or that if we travel equal distances east, south, west, north, we are back where we started (Figure 5.2). Basic rules such as these impose constraints on inference of what can be where, allowing the agent to, for example, infer unobserved links and as a consequence predict sensory experiences at previously visited nodes when approaching them over an unobserved link (Figure 5.2). In the present study we use 2D graphs so we can compare learnt representations to the well-characterised neuronal representations of space, although we note the framework is general to arbitrary graph structures.

Crucially, because the agent is transferred between different environments in which the sensory experiences at the nodes differ (Figure 5.1), it must learn a structural representation that can be generalised across environments that differ in their sensory particularities. The basic rules of 2D space just mentioned are true regardless of which room

we are in. Rather than learn afresh upon encountering a new environment, we wish to reuse this previously learnt structural knowledge where the statistics are true to facilitate immediate understanding. Hence this task environment encourages learning of generalisable structural representations.

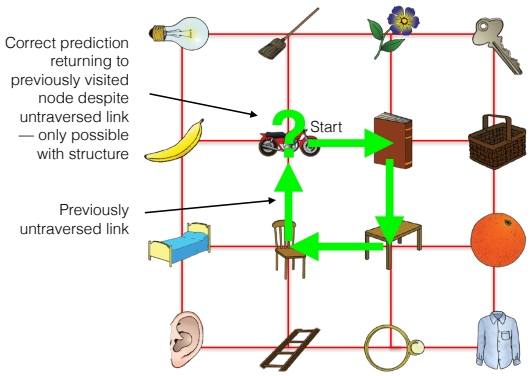


Figure 5.2 **Inference of previously untraversed links**. Understanding the basic rules of space places constraints on what can be where, allowing an agent to infer previously unseen links. Starting at the motorbike and going east, south, west, north takes the agent back to the motorbike. Only with an understanding of the rules of space can the agent realise it is back where it started and predict the experience of a motorbike.

5.3.2 Explicit structural representations are necessary for generalisation

We propose that in order to generalise structural knowledge across environments sharing structure but differing in their sensory particularities, the representation of the structure needs to be separated from that of sensory experiences. In such explicit representations the structure is represented in a separate population of neurons to that which represents the sensory experiences. This is contrasted with an alternative way of representing structure by encoding it between the representations of the sensory experiences themselves (i.e. in their connection weights) and remembering which specific sensory experiences are next to which (the banana is next to the table, which is next to…). While with sufficient learning this would be sufficient for prediction within an environment, the knowledge of structure would not generalise to new environments since these relationships are no longer true (the banana is no longer next to the table), and therefore algorithms such as planning or navigation would have to be learnt afresh in every novel environment. This is not to say that algorithms such as planning require generalisable representations in order to work (as they could be learnt on non-generalisable transition representations). Rather if they are learnt on top of generalisable representations, then they

can be reused quickly in new environments that share structure (e.g. 2Dness of space and how to plan through it), as this structure does not have to be relearnt, allowing immediate planning, for example. Furthermore, without learning and generalising abstract structure, it cannot be exploited to constrain inference: unobserved links cannot be inferred; rather all relationships need to be observed.

These problems can be overcome if the structure is represented devoid from the sensory experiences that it encodes the relations between. This allows the agent to represent its location in an abstract space that generalises across environments. As in (Manns and Eichenbaum, 2006), we propose that representations of the structure are in MEC and those of the sensorium are in LEC (Figure 5.3). Hence the representation of the environment has been factorised (separated) in to structural and sensory components, i.e.:

$$P(r_{sensory}, r_{structure}) = P(r_{sensory})P(r_{structure})$$

Equation 5.1

where the probability is over the spiking (r) of each neuron in the representation (Behrens et al., 2018). Factorisation allows the factorised components to be recombined in novel ways to previously unseen conjunctions, allowing generalisation (Figure 5.3)(Behrens et al., 2018; Eslami et al., 2018). Therefore, knowledge learnt in the abstract space – such as the fundamental constraints imposed by 2D space or other information useful for planning such as "the shortest route between two locations in 2D is a straight line" – can be reused across environments, or sensory experiences. An agent can therefore, for example, use previously learnt constraints to infer unobserved links in a novel environment, dramatically improving its predictive capabilities of sensory experiences.

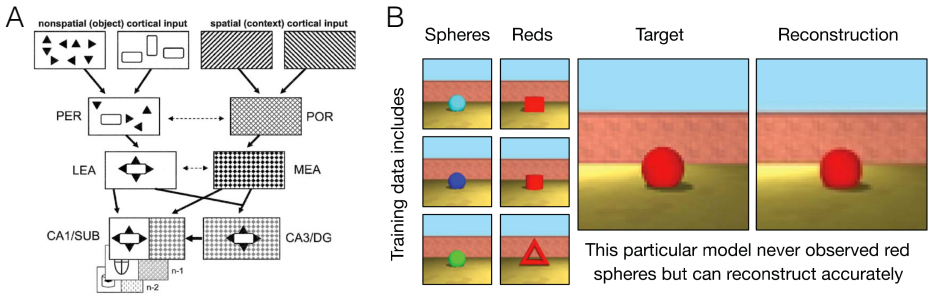


Figure 5.3 **Explicit structural representations.** Separating (factorising) structural from sensory representations allows generalisation of structure across sensory experiences. A) Structural or spatial representations are proposed to be in MEC (/MEA) and object representations in LEC (/LEA), and combined in hippocampus. Figure adapted from (Manns and Eichenbaum, 2006). B) Factorised representations facilitate generalisations to previously unseen conjunctions. Factorising the representation of object properties (colour and shape) in an ANN trained to predict scenes allowed it to represent and generate objects that are a previously unseen conjunction of these object properties

(red spheres), demonstrating generalisation. Rather than object features such as colour and shape, we factorise sensory experiences from structure. Figure adapted from (Eslami et al., 2018).

We propose grid cells are such an explicit, generalisable structural representation for 2D space. As discussed in the previous chapter, given a PCA of 2D-distributed place cells results in grid-like coding suggests grid cells are summarising the statistics of 2D space (Dordek et al., 2016; Stachenfeld et al., 2017). Given grid cells within a module retain their relative phase relationships across different environments suggests this code generalises (Fyhn et al., 2007). Other cells in MEC, such as border cells (Solstad et al., 2008), band cells (Krupic et al., 2012), object-vector cells (Høydal et al., 2018) and head direction cells (Taube et al., 1990), can also be thought of as coding important features of a 2D environment or an agent's movement through it. Hence MEC may represent the information necessary for understanding the structure of a 2D world (Behrens et al., 2018). In contrast, LEC contains representations of the sensory features of the environment, such as objects within it (Manns and Eichenbaum, 2006), and lesions to this area impair recognition of object identity but not location (James et al., 2018).

Hence we view explicit representations as crucial to generalising structural knowledge to novel permutations of sensory experience and propose that grid cells in the brain are such a representation for 2D space. The problem now becomes how to combine the structure with sensory experiences.

5.3.3 Structural knowledge is combined with sensory experiences in a conjunctive code

We propose that structural and sensory information are combined in a conjunctive code in hippocampus, allowing for generalisation. In other words, the factorised structure and sensory components are combined in a hippocampal conjunctive code. In this proposal, many hippocampal cells are consistent with a given location in an abstract structure and many are consistent with a given sensory experience. However only a subset of cells are consistent with both. MEC (structure/grid cells) proposes (provides excitatory drive to) candidate hippocampal cells consistent with the estimated current location in the structure. The location in the structure is coded for by the phase of the grid cells (see below). LEC (sensory representations) proposes candidate hippocampal cells consistent with the current sensory experience. Only the subset of hippocampal cells consistent with both the structure and sensory experience will be activated (Equation 5.1). The hippocampal population

therefore codes the unique structure by sensorium conjunction that fully describes the current experience.

Notably this hippocampal representation can be formed rapidly upon entering a new environment by these cortical regions proposing place cells, 'slotting in' the current sensory experience with previously learnt knowledge and offering immediate understanding of the environment. Because the place cells that are active are those consistent with the cortical representations and in turn index these cortical representations, knowledge coded in cortex (such as how place cells relate to each other in the grids) is embedded in the hippocampal representation. The hippocampal representation also has the ability to bind together all of this disparate information in to a unified thing for storage – a memory (see below).

There is evidence to support hippocampal cells as a conjunctive code. In addition to coding for space, hippocampal neurons have been found to code for sensory events, particularly in conjunction with spatial information (Komorowski et al., 2009; Wood et al., 1999). More recently, it has been observed that in fact a higher proportion of CA3 neurons code for both sensory information *and* spatial information than either alone (Figure 5.4)(Tsao et al., 2018). These results strongly point to the hippocampus as the locus of combining what and where information to form a conjunctive representation of the current experience.

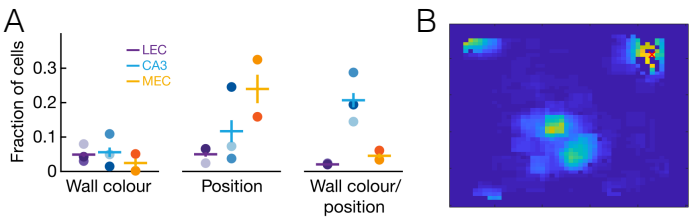


Figure 5.4 **Place cells as a conjunctive code**. A) As predicted by hippocampus as a conjunctive code, a higher proportion of cells are sensitive to both position and the sensory environment than either alone. Figure adapted from (Tsao et al., 2018). B) A conjunctive code is able to account for the multiple firing fields within a single environment often observed for a given place cell. Place cell taken from dataset analysed below (Barry et al., 2012).

The hippocampal conjunctive code as we propose here is able to offer mechanistic explanations for a variety of observed phenomena. Place cells will fire where their conjunction is true. Because we represent space hierarchically in order to repeatedly reuse

high frequency statistics across space (see below), high frequency place cells (those associated with high frequency grid cells) will be consistent with the structure in multiple locations in the environment (wherever the associated grid phase occurs). Therefore if the associated sensory experience is repeated at the same grid phase in a different part of the environment, the place cell will have multiple firing fields within an environment, as is observed experimentally (Figure 5.4)(Fenton et al., 2008). Similarly, again when combined with hierarchical representations, it also predicts repetition of place cells in multicompartment environments such that a given place cell may be active in the analogous location (structure) of multiple compartments, as is observed experimentally (Derdikman et al., 2009; Grieves et al., 2016; Kim and Maguire, 2018; Spiers et al., 2015).

More notably, a conjunctive code predicts a given neuron's firing field may move to a different location in a different environment due to different sensory experiences, and therefore naturally accounts for hippocampal remapping. Hippocampal neurons remapping in this way to locations where their conjunction is true is a significant departure from most other accounts of remapping that suggest it is random (Monaco and Abbott, 2011). This offers an experimental prediction: given a place cell should remap across environments to a location where the conjunction is true, this predicts that the relationships between place and grid cells should be preserved across environments despite hippocampal remapping. We analysed single unit data to test this prediction and provide evidence for it later in this chapter. Interestingly, it is supposedly surprisingly difficult to induce remapping experimentally: even in different boxes of the same dimensions, it requires introduction of significant changes in the sensory experience, such as textures, visual appearance and odours (Barry et al., 2012), as predicted by our framework.

The conjunctive code further explains the complete hippocampal remapping when the sensory modality is changed between echolocation and vision in bats flying around in a single enclosure (Geva-Sagiv et al., 2016); despite preserved structure, the sensory experience (and therefore conjunction) is different. It may also explain place field sizes: their firing fields may be smaller than their associated consistent grid phase depending on the extent of the sensory experience, offering a potential explanation for why there are not obviously discrete place cell modules (Kjelstrup et al., 2008) as there are with grid cells (Stensola et al., 2012).

Hence structural information is combined with sensory information in a conjunctive place cell code that allows for generalisation and formation of a unique code of the

experience for episodic memory (below), as well as accounting for neuroscientific phenomena, most notably remapping.

5.3.4 Hierarchical representations allow reuse of knowledge across space

We wish to generalise knowledge across space. Since the rules of 2D space are true in any part of space, rather than relearn them in every part of space, it is dramatically more efficient to reuse learning across space. This greatly reduces the number of place to grid cell weights that need to be learnt. Within the same environment, high frequency information can be reused in different parts of space because, within a module, the relationships between the high frequency (small scale) grid cells will be true in the different parts of space (Figure 5.5). Since these high frequency cells cannot disambiguate between low frequency parts of the environment (e.g. between the locations at the two arrows in Figure 5.5), we require a hierarchical representation to retain the ability to precisely encode location, by using lower frequency information that disambiguates between locations coded for identically at higher frequencies. These different hierarchical grid scales then make predictions to corresponding hierarchical representations in the place cell code. Hierarchical place cell representations facilitate memory retrieval (see below).

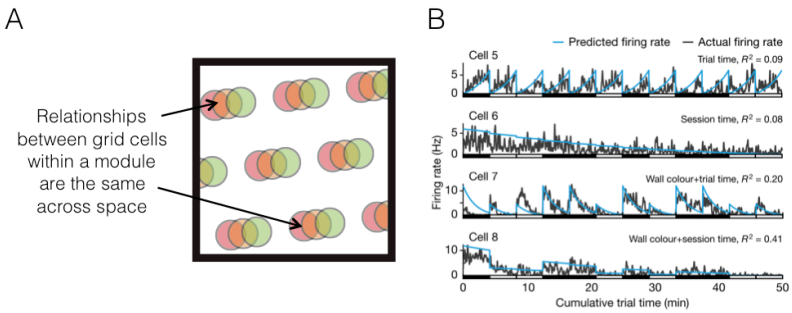


Figure 5.5 **Hierarchical representations**. A) The rules of 2D space are true across space; therefore high frequency information encoded in grid cells can be reused across space. The relationships between the green, pink and red grid cells are true in multiple parts of the environments, meaning this knowledge does not have to be relearnt in every part of space. B) Exponential temporal decays have recently been found in LEC, as predicted by our framework, since we use differing temporal smoothing factors to generate the hierarchy. Figure adapted from (Tsao et al., 2018).

We encourage hierarchical representations to emerge in our framework by temporally smoothing experience. While the agent only ever observed instantaneous sensory input, these sensory observations were then smoothed with different exponential decays to generate different degrees of temporal smoothing, corresponding to different

hierarchical levels. Lower frequency information allows the memories that form in hippocampus to contain information like “the banana is next to the table”, which allows it to be disambiguated from another banana in the environment that is next to a different object. Therefore, even if the same sensory experience appears in multiple locations within one environment, the surrounding stimuli, and therefore lower frequency place cells, will likely be different, allowing precise coding of the experience across frequencies and accurate memory retrieval.

Different hierarchical scales in the sensory input encourage different hierarchical scales to emerge in the place cells and grid cells. This results in a hierarchical grid representation that in addition to facilitating generalisation provides an efficient combinatorial code of space. The grid structural prediction is therefore at all frequencies, which allows the place cell memory to be recalled to make an accurate prediction of what specific observation is at the current state (see below).

Hierarchical representations of space are well known to exist along the dorsal-ventral axis of the hippocampal-entorhinal system. Grid cells form discrete modules along this axis with increasing grid scales more ventrally (Stensola et al., 2012) (as we saw in the Introduction). Similarly, hippocampal place fields also increase in size along this axis (Kjelstrup et al., 2008). Although we made the assumption that there were exponential temporal decays in the representations of sensory experiences in LEC, exactly this prediction has recently been confirmed (Figure 5.5)(Tsao et al., 2018). This may be produced in part from differing timescales across cortical representations (Murray et al., 2014). Hence there is empirical support for hierarchical representations as we model them in hippocampus, MEC and LEC.

5.3.5 Hebbian learning facilitates fast formation of conjunctive episodic memories

In order to form fast episodic memories in the conjunctive code that rapidly bind sensory experiences to their location in the abstract structure, combining what and where information, we require a form of fast learning. This allows storage of episodic memories (conjunctive codes) during inference of the place cell code from the structural and sensory experiences, and allows their retrieval during generation (prediction of the next time step) when returning to that location. This fast memory comes in the form of a simple recurrent neural network in the place cell code with a Hebbian learning rule on the weights between place cells, which allows for fast updating of the weights to form one-shot episodic memories.

The recurrent network combined with the hierarchical representations facilitates disambiguation of which sensory experience should be expected at the next time step when performing prediction. This is because the grid prediction of the place cells will activate all place cells consistent with the structure, regardless of the sensory experience they code for, i.e. all high frequency place cells consistent with the structure activated. In order to make an accurate prediction this needs to be narrowed down to just one sensory experience. This occurs in the dynamics of the recurrent neural network that utilises lower frequency information to keep active only place cells associated with the correct sensory experience. The high frequency place cells then make a prediction of the sensory experience.

These proposals fit well with classic theories of hippocampal function in memory. As outlines in the Introduction, the hippocampus is famously thought responsible for fast, episodic memory (Squire et al., 2004). Neurophysiologically, it supports LTP classically thought to mediate fast formation of memories, and neuroanatomically, it has recurrent attractor networks appropriate for retrieving memories from partial information (pattern completion) (discussed in the introduction).

5.3.6 State transitions occur in the grid cells

In order to predict what will occur at the next time step, the agent needs to make a prediction of where in the structure it will be at the next time step. These predictive state transitions occur through the grid cells because as we have discussed, grid cells are a natural basis for 2D navigation (Bush et al., 2015; Stemmler et al., 2015). The predicted grid code at the next time step is then used to predict the place cell code, which retrieves the memory of what previously happened at that location. It is desirable that the transitions occur in the space that is generalised so that any algorithms that operate on this information such as planning can also be generalised. In order for these state transitions to occur in the structural representation, it requires information concerning the agent's actions so that it can compute future states, which we provide as input to the grid code.

Alongside grid cells in MEC are cells carrying exactly this information: head direction cells (Taube et al., 1990), which as the name implies are tuned to a specific direction of the animal's head, regardless of where it is in the environment. Additionally, speed cells (Kropff et al., 2015) code the animal's speed, which is similarly useful for computing state transitions. Hence (and related to the section on explicit structural representations), the MEC appears to contain all the information necessary for coding the structure of a 2D world and vectors or transitions through it (Behrens et al., 2018).

5.3.7 Slow and fast learning across and within environments

There are therefore two types of learning in this system. The model combines Hebbian learning allowing for fast learning of episodic memories within a given environment with gradient descent, which slowly learns to extract the statistics of these memories over many experiences across many different environments. This has some parallels with meta-reinforcement learning as explained in the Introduction (Wang et al., 2018), in the sense that learning a generalisable representation encoding task structure facilitates learning within a new instance of the task. Both approaches result in neural networks exhibiting flexible behaviour.

5.3.8 Putting the components together

We now briefly put these components together and then run through them with an example. During inference, the agent observes a sensory experience causing excitatory drive to all place cells coding for that sensory experience. Simultaneously, the explicit structural representation in grid cells provides excitatory drive to all place cells consistent with the inferred current location in state space. Only the subset of place cells that are consistent with the conjunction of structure and sensory experience will be activated. Fast Hebbian learning allows this conjunctive code combining what and where to be stored in a memory of that experience.

A hierarchical representation allows for generalisation across space, such that high frequency information can be repeatedly reused across space, dramatically reducing the amount of learning required. Lower frequency information offers precise coding of location in a combinatorial code and disambiguation of which high frequency experience occurred at a given location during memory retrieval.

During generation, state transitions occur in the grid cells to predict the grid code at the next time step. This grid code in turn drives place cells that are consistent with it. From this place cell code, the recurrent dynamics in the Hebbian attractor combined with hierarchical representations allow recall of exactly what happened where (i.e. the previous sensory experiences at that location/grid code). From this refined place cell code, a prediction is made of the sensory experience at the next time step.

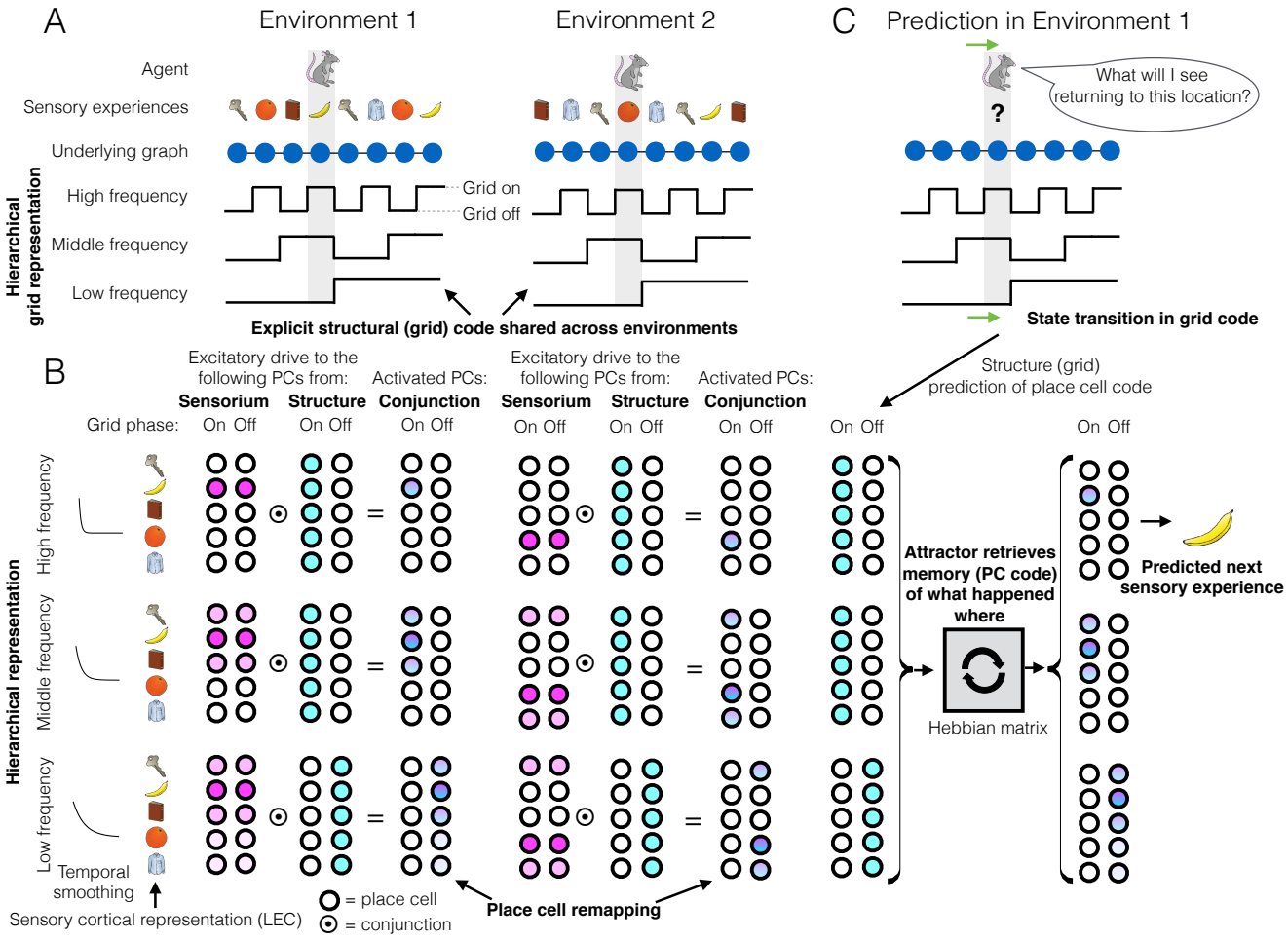


Figure 5.6 **Schematic of the framework and example representations in 1D**. A) A 1D environment (linear track) with different sensory experiences across two environments. The underlying structure of the environments and explicit, hierarchical structural representation are shared across environments. B) Place cell representations when the agent is at the shaded region in A. A given place cell is consistent with a sensory experience and a grid phase at the corresponding level of the hierarchy. Magenta and cyan denote place cells consistent with the sensory experience and structure, respectively, at the shaded region, and hence receive excitatory drive from the corresponding representations (in LEC and MEC). During inference, only those consistent with both – their conjunction – will become active. Because sensory experiences differ across environments, the place cells remap across environments. Hierarchical representations are created by different temporal decays on sensory experience. C) In order to predict sensory observations during generation, the agent must be endowed with fast (episodic) memory. These memories are formed from the conjunctive representations (as in A). When returning to a previously visited location, these memories are retrieved from the grid prediction of the place code, allowing prediction of the sensory experience at the next time step. A more thorough walk through of this example is in the main text.

5.3.8.1 Example in 1D (following Figure 5.6)

We illustrate the key components and example representations with the simple example of an agent on a 1D linear track (Figure 5.6). The following text walks through

Figure 5.6. **A)** An agent walks around a linear track in which it only ever observes sensory experiences. The linear track is discretised in to 8 states, on each of which there is one of 5 possible sensory experiences (e.g. banana, book…). The structure of this environment can be efficiently represented in a hierarchical grid code with three frequencies that uniquely codes location across the population (for illustrative purposes the grid cells are either on or off in their activations). Only one grid cell phase is shown; 'Grid off' could be 'Grid on' for another grid cell of the same frequency but phase shifted (phase shifted grid cells within a frequency/module are observed in the brain). The relationships between a set of grid cells within a module (e.g. grid cell 1 is on when 2 is off and 3 is somewhere in the middle) that are useful for computing transitions, etc., will be true across space – allowing repeated reuse and generalisation of high frequency knowledge across space. Because the structural representation is explicit – devoid from the sensory experiences, we see that if in a second environment sharing the underlying structure we randomly reassign a different set of sensory experiences to each state (Environment 2), the same structural code can still be shared across environments.

B) Place cells can be thought of as a conjunction of structure and sensory experiences. In this schematic, each module has 10 place cells denoted by a 5x2 grid of circles. The rows and columns of this grid denote the sensory and structural information that a given place cell is associated with, respectively. Within a given module, a subset of place cells will be consistent with a given grid phase (e.g. "on"). If the agent is in a location in the environment at this grid phase, the associated place cells will receive excitatory drive from the grid cells. A (partially overlapping) subset of place cells will similarly be consistent with a given sensory experience (e.g. "banana"). For example, when the agent is in the shaded location in Environment 1 in A, in the population of 10 place cells associated with the high frequency, all of the "Grid phase on"-consistent cells are driven by the associated high frequency grid cell (in MEC) and all of the "banana"-consistent cells are driven by the high frequency banana representation (in LEC). However, only the subset of place cells that is consistent with *both* will become active – i.e. the structure by sensory experience conjunction. In this example, this is only one high frequency place cell that codes "banana-grid on".

However, we can see that the "banana-grid on" conjunction is also true at the far right state of this linear track in Environment 1, where this place cell will also be active, and therefore it does not uniquely code the location. The solution to this problem is to have a hierarchical representation, whereby across the place cell population experience is coded

uniquely: i.e. at the far right state of the environment low frequency information would differ due to differing neighbour sensory experiences. We can obtain conjunctive place cells at each level of the hierarchy, giving us a complete place cell code that can uniquely represent different experiences. The different frequencies / levels of the hierarchy are produced by temporally smoothing the sensory experiences with differing decays.

Notably, upon entering a new environment in which sensory experiences are randomly reassigned (Environment 2), we observe a different place cell code – i.e. remapping. Remapping here is clearly not random; rather the place cells have maintained their relationships with the grid cells and hence have structural information embedded in their representation.

C) The agent's task is to predict the next sensory observation. It can carry out a state transition in the grid cells and predict the grid code expected in one time step from its current grid code and information about its actions (e.g. head direction). It can then use this grid code (predicted location information) to drive place cells consistent with it. In order to predict the next observation, it must remember the exact place cell code that previously occurred at this location (grid code). This is achieved with fast episodic memories and an attractor neural network. After the conjunctive code is generated as we just described, the weights between the place cells are updated in a Hebbian fashion, allowing formation of fast, episodic memories of what happened where. These weights (forming the memory) allow the specific place cell code that occurred at a given location to be recalled given the grid prediction of the place cells (location information). This occurs in an attractor network with the Hebbian weights that utilises the hierarchical information for retrieval of the specific memory. Once the recurrent network has finished iterating, the refined place cell code is used to make a (high frequency) prediction of the next sensory experience (e.g. "banana").

We note that since the structure (grid) representation in A) is identical across the two environments, it is not clear from this figure how to disambiguate which memory (environment 1 or 2) to retrieve from the grid prediction at the highlighted location. There are two ways this can be dealt with: first, have a new Hebbian matrix for each environment, or second, add another level to the hierarchy acting as a context signalling which environment the agent is in. We also note that whilst in this schematic we have a separate sensory representation for each sensory experience (e.g. banana), the sensory representations could themselves be combinatorial.

5.3.8.2 Summary

This system solves the computational problem of generalising structural knowledge to novel environments and forming fast episodic memories of these experiences. In doing so it may offer a unifying account of a variety of computational roles of the hippocampal formation, including memory, generalisation, inference and spatial cognition. Furthermore it is able to offer mechanistic accounts for a wide range of neural data (e.g. grid cells, conjunctive coding, remapping, hierarchy, fast memory, etc.) and makes novel experimental predictions (e.g. preserved relationships between place and grid cells across environments, further conjunctive predictions, representations in non-spatial tasks). As in the previous chapter, there is nothing special about space here. We focus on 2D graphs so we can compare representations learnt by a neural network model of this framework to ground truth representations recorded in real brains. That representations emerge resembling those for space in the brain suggests such spatial representations may emerge as an instance of more general coding mechanisms such as those presented here.

5.4 Model implementation

5.4.1 Agent architecture

We implement our conceptual framework in an artificial neural network tasked with predicting the next sensory experience when walking around on multiple 2D graphs, as described above. In this subsection we present the agent architecture (Figure 5.7) along with other information regarding the environment and training. We do not run through the details here because, as mentioned above, this work was carried out by James Whittington (Whittington et al., 2018). We just present the key components here to help aid interpretation of the model results presented in this section and to ground the conceptual framework just outline in the previous section.

The agent walks passively (actions are selected for it) around multiple 2D graphs. Within a given environment, as discussed in the previous section, at each vertex of the graph, a sensory experience is drawn at random from a set of candidate sensory experiences (Figure 5.1). Across environments, therefore, the same set of sensory experiences is used, but in different arrangements. The sensory experiences come in the form of a one-hot vector x_t, and the agent's task is to predict the sensory experiences as it moves on the

graph. This task is achieved by learning a generalisable structural representation that can exploit the structure of the graph to infer missing links and retrieve memories of what sensory experiences previously occurred at which locations on the graph within the environment.

Generation

Grid cells: During generation, to predict the sensory observation at the next time step, the agent must predict where it will be at the next time step. These state transitions occur in the grid cells g_t where t is the current time step, such that the the grid code on the next time step is a function of the grid code on the current time step and the action taken u_t: $g_t = f(g_{t-1}, u_t)$. This is analogous to path integration.

Place cells: This grid representation is an abstract representation of where the agent is in the state space and does not contain information about the sensory experience expected at that location. In order to make a prediction of the sensory experience, the conjunctive place cell code p_t that was previously inferred at that location (see below) needs to be retrieved. This is achieved in an attractor neural network $h_t^\tau = f(h_t^{\tau-1} + M_{t-1}\, h_t^{\tau-1})$, where τ is the iteration of the attractor, M_t is a Hebbian matrix containing the weights between place cells (that encode the memories – see below), with input $h_t^0 = g_t$ and the final output $h_t^{final} = p_t$. A given grid cell in the grid prediction g_t provides input to multiple place cells h_t^0, each of which may be consistent with a different sensory experience. Therefore there is no unique sensory information in the grid prediction and it alone cannot predict the sensory experience: the attractor must be used to refine the place cell code such that it retrieves the unique structure by sensorium conjunction.

Data: From this conjunctive place cell code, a prediction of the sensory experience x_t can be made $x_t = f(p_t)$. This is a categorical distribution to classify the next sensory experience. We note that this prediction is made only at the highest frequency (see below), which is the only frequency the agent observes (the lower frequencies are generated within the agent – see below). The data comes in the form of a one-hot vector representing the different possible sensory experiences.

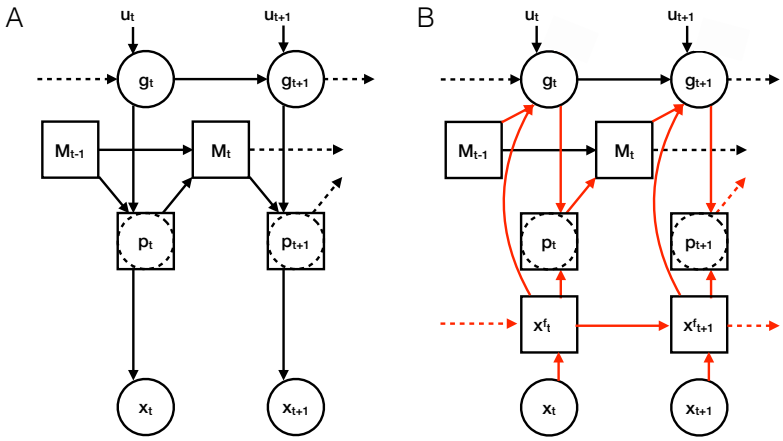


Figure 5.7 **Agent architecture**. A) Generative model, and B) inference architecture. x_t, g_t, p_t, u_t, and M_t denote, respectively, the sensory experiences, grid cells, place cells, action taken, and Hebbian memory matrix at the current time step t. Circled vs. boxed variables denote stochastic vs. deterministic variables, and red arrows denote additional inference dependencies. We note that in the representations and behavioural results presented of the model, we do not sample from any variables, i.e. they are all deterministic.

Inference

Place cells: The place cells are inferred as the conjunction between the observed sensory experiences and the structural grid prediction: $p_t = f(g_t \odot x_t)$. This is done within each frequency x_t^f and g_t^f. In order to obtain different frequencies, the sensorium is exponentially smoothed with different decays: $x_t^f = (1 - \alpha^f)x_{t-1}^f + \alpha^f x_t$. Hence we obtain hierarchical temporal scales of the sensory experiences with different smoothing constants α^f. This is similar to the Laplace transform of time observed recently in (Tsao et al., 2018). These different scales provide a unique code for each position even if the same sensory experience occurs at multiple locations in one environment, since the surrounding stimuli, and therefore lower frequency place cells, will likely be different. The grid cells learn different frequencies (/hierarchical representation) to make predictions of the different frequencies in p_t that are in turn due to different frequencies in x_t. As well as facilitating in retrieving the correct memories in the place code, this allows them to efficiently encode space and reuse high frequency information across space. After inferring the place cell code, the memory matrix is updated: $M_t = \lambda M_{t-1} + \eta(p_t - \hat{p_t})(p_t + \hat{p_t})^T$, where $\hat{p_t}$ is the place code generated from inferred grids, λ and η are forgetting and remembering rates, respectively. This memory therefore links the grid prediction of the place code to the inferred place code, allowing the inferred place code to be retrieved from the grid prediction when returning to that location in the environment.

Grid cells: The grid cells are inferred from g_{t-1} and u_t ($f(g_{t-1}, u_t)$ as in the generation) and from x_t via M_{t-1} which allows it to use sensory information that it may have seen in a previous location before.

Training

The agent is optimised end-to-end using backpropagation through time to minimise the following per time step loss function: $L_{total} = \sum_t^T L_{x_t} + L_{g_t} + L_{p_t}$, where L_{x_t} is a cross entropy loss between predicted and observed sensory experiences and L_{g_t} and L_{p_t} are squared error losses between inferred and generated variables. We also add a next time step prediction loss, so that the agent learns to predict the next time step. We note that although we have presented the above as generative and inference networks, and that in Figure 5.7 there are stochastic variables, we treat all variables as deterministic in the results presented. Therefore the architecture used is identical to the generative and inference architectures in Figure 5.7 but treating all variables as deterministic. Best results were obtained this way.

The agent was trained on multiple different environments of differing sizes (squares with possible widths: [8,10,12]) and moved between environments after reaching high prediction accuracy (95%) or 5000 timesteps. There were 12 different possible sensory experiences and 5 different hierarchical frequencies. The agent can move (u_t) up, down, left, right or stay still. When the agent is next to a boundary, the action(s) that would cause it to bump into the boundary(ies) are not available. The agent successfully learns to predict the next sensory experience across multiple 2D graphs of varying sizes and arrangement of sensory experiences. It is important to note that all representations are learned.

5.4.2 Model learns spatial representations resembling those in the brain

Structural and conjunctive units, g_t and p_t, in the artificial neural network learn firing rate maps that resemble those of grid and place cells, respectively. Structural units are grid-like in the sense that they are periodic over the 2D space (Figure 5.8). They form a square lattice, as expected given the statistics of the environment are square (four way connected space and square boundary conditions; as in previous chapter). Periodic grid units of different frequencies are learnt, providing a hierarchical representation of the environment. The grid representations are consistent across environments, even if the environments are of slightly different sizes, suggesting it is a representation of the statistics of the 2D space (rather than memorising a template of a single-sized environment). The grid

representations can therefore be thought of as learning and representing a basis set for the statistics of the environment, rather than storing a specific transition matrix. A basis set representation allows the agent to construct representations of structurally similar environments that differ slightly in their specifics (e.g. size) but share the same underlying statistics (2D space). We additionally observed band cells emerge in the structural units, as are observed in MEC in the brain (Figure 5.8) (Krupic et al., 2012).

We also note that grid cells of a given frequency (same module) have different phases (Figure 5.8)(also as found in the brain (Hafting et al., 2005)). This means that two place cells that respond to the same sensory experience (e.g. banana), will not necessarily both be active if the stimulus is present; each cell will only be active if their corresponding grid cell is also active (which is determined by the grid phase). Therefore, embedded in the place cell code is also information of the location of the sensory experience, and the different grid phases (combined with hierarchy) allow unique coding for the same sensory experience in many different locations.

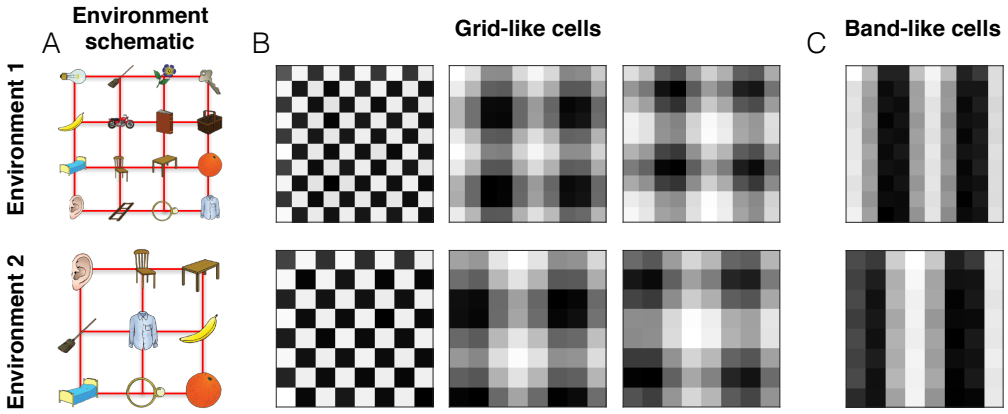


Figure 5.8 **Structural units in the network resemble grid cells**. Top and bottom rows denote two different environments. A) Schematic of two environments – note the shuffled sensory experiences at the vertices and that the environments are different sizes (environments not actual size, as can be seen in the grid rate maps). B) Activity of structural units at each node of the graph reveals they have grid cell-like rate maps. Three grid cells (left-right) in both environments (top-bottom) reveal different frequencies (leftmost vs. other two) and different phases within a frequency (middle vs. right). A given grid cell's representation is similar across environments even though they are of different sizes. C) Some structural units resembled band cells.

Across different environments, a given sensory experience may occur at different locations. Therefore the structure by sensorium conjunction may occur in a different location in the environment, resulting in place cell-like remapping. We observed such an effect in the conjunctive units (Figure 5.9). Furthermore, a given stimulus may appear in a single environment more than once, and if this is at the same corresponding grid phase,

there will be multiple place peaks (Figure 5.9), as observed in the brain (Fenton et al., 2008). Different frequencies of place cell-like conjunctive units are also observed (Figure 5.9).

Hence representations emerge in the network when trained on 2D graphs that resemble those ground truth spatial representations recorded in the brain, suggesting the general principles we have presented in this chapter may account for how information is organised in the hippocampal-entorhinal system.

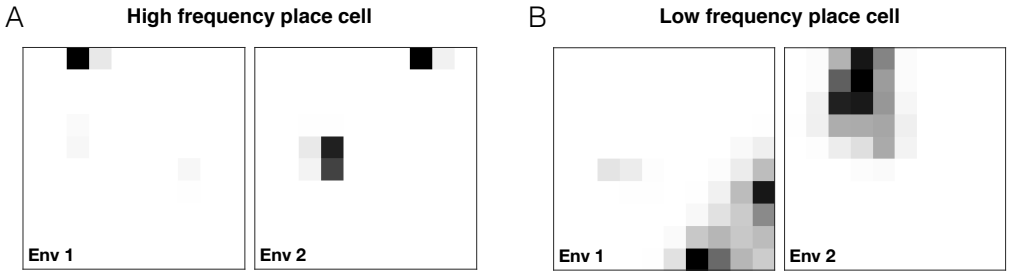


Figure 5.9 **Conjunctive units in the network resemble place cells**. The conjunctive units learn place cell-like representations that exhibit remapping, may have multiple place peaks within an environment and are of both high (A) and low (B) frequencies.

5.4.3 One-shot learning and zero-shot inference

We test whether the fast Hebbian learning allows for one-shot learning of memories. In order to test this, we check whether the agent is able to predict back the sensorium when staying still at that location (i.e. the agent does not transition to a new node at the next time step). This asks whether the correct sensory experience can be retrieved from the grid representation via the fast memory just created in the place cells, which remembers what happened where. We find that after visiting a node only once, the agent is able to correctly predict its associated experience (Figure 5.10).

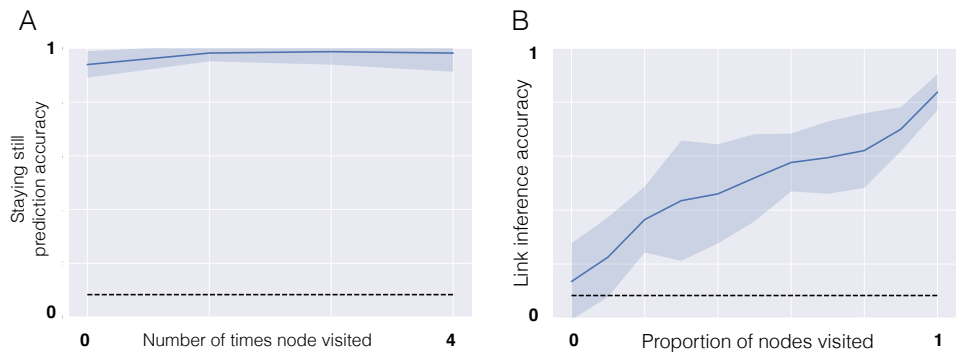


Figure 5.10 **One-shot learning and zero-shot link inference**. A) The agent is able to predict the sensory experience at a given node with high accuracy having only visited the node once. B) The agent is also able to infer untraversed links in the graph and this ability improves with the proportion of nodes in the environment visited. Black line denotes chance (1/number of sensory experiences).

If the agent has learnt a representation of the structure of the environment, it should be able to infer untraversed links in the graph and make next step predictions of a previously visited node when approaching it from a different direction (Figure 5.2). Indeed we found this to be the case, demonstrating the agent is using its structural representation for zero-shot inference (Figure 5.10). This ability was not immediate, however; rather the accuracy of this next-step prediction over untraversed links increased with the proportion of nodes that the agent had visited in the environment. High accuracy was not immediate likely because the agent is not immediately sure of its location in the environment, which improves with time.

5.5 Empirical evidence for a preserved relationship between place and grid cells across environments

5.5.1 Remapping experiment and experimental prediction

Our framework predicts place cells and grid cells retain their relationship across environments, despite place cell remapping, allowing generalisation of structure encoded by grid cells. If this were this case, we would expect the following to be true. 1) When an animal is taken out of one environment and placed in a structurally similar environment but with very different sensory experiences, grid cells should realign and place cells should undergo global remapping, as has been observed (Fyhn et al., 2007). 2) However, within a module, if a place cell's firing field is at a certain phase of an associated grid cell in the first environment, it should remap to a location that is at the same grid phase in the second environment.

We empirically test this prediction in data from a remapping experiment where both place and grid cells were recorded from rats in two different environments (Barry et al., 2012). The environments were of the same dimensions (1m by 1m) but differed in their sensory (texture/visual/olfactory) cues so the animals could distinguish between them. Each of seven rats has recordings from both environments. Recordings on each day consist of five twenty-minute trials in the environments: the first and last trials in one environment and the intervening three trials in a second environment (Figure 5.11).

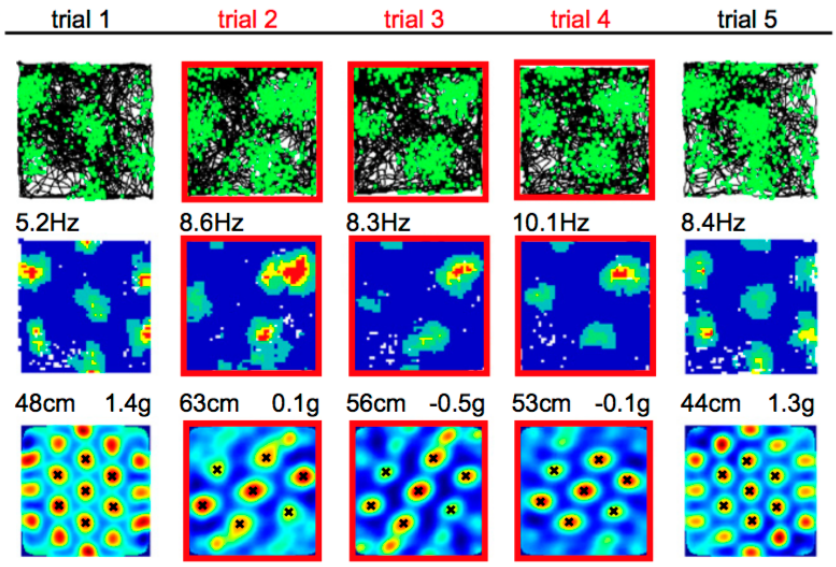


Figure 5.11 **Remapping experiment from Barry et al., 2012**. Rats were exposed to two different environments that were of the same dimensions (1mx1m) but differed in their sensory experiences. On a given day of the experiment, the rat would experience five trials; each trial is twenty minutes long and the rat simply free forages in the environment. Trials 1 and 5 (black text and outline in the figure) are in one environment, which the animals are familiar with. Trials 2-4 (red) are in a second, novel environment the animals have previously not been exposed to. Barry et al., 2012 sought to establish the effect of novelty on grid and place cell properties, finding an increase in grid scale and decrease in grid score, as well as an increase place cell field sizes. An example grid cell is shown in this figure: top; locations of action potentials on the animal's path, middle; firing rate map, bottom; spatial autocorrelogram (the values above the autocorrelogram are the grid scale and score). Both place and grid cells were recorded in both environments, allowing us to test whether their relationships are preserved across environments. Figure adapted from (Barry et al., 2012).

We test the prediction that a given place cell maintains its relationship with a given grid cell across environments using two measures (Figure 5.12). First, whether grid cell activity at the position of the peak place cell activity is correlated across environments (gridAtPlace), and second, whether the minimum distance between the peak place cell activity and a peak of grid cell activity is correlated across environments (minDist; normalised to corresponding grid scale).

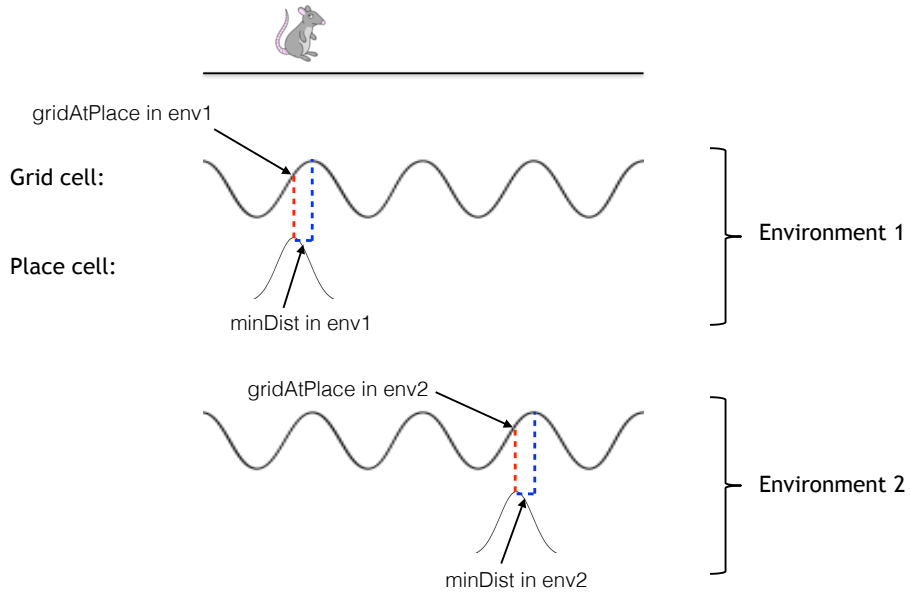


Figure 5.12 **Schematic of preserved place and grid cell relationships across environments on a linear track**. Although we tested our prediction in 2D boxes (Barry et al., 2012), we illustrate it here in 1D for clarity. We can see that although the place cell has remapped across environments, it has remapped to a location that is at the same grid phase in both environments. This can be indexed by testing either whether the grid cell firing rate at the place cell peak (gridAtPlace), or the minimum distance to a grid peak (minDist), are correlated across environments.

To account for potential confounds or biases (e.g. border effects, inaccurate grid peaks), we fit the recorded grid cell rate maps to an idealised grid cell formula (Stemmler et al., 2015), and use this ideal grid rate map to give grid cell firing rates and locations of grid peaks (Figure 5.13). Only grid cell rate maps with high grid scores (> 0.8) were used to ensure good ideal grid fits to the data, and we excluded grid cells with large scales (> 50cm), both computed as in (Barry et al., 2012). Locations of place cell peaks were simply defined as the location of maximum activity in a given cell's rate map. To account for border effects in place cells, we removed place cells that had peaks close to borders (< 10cm from a border).

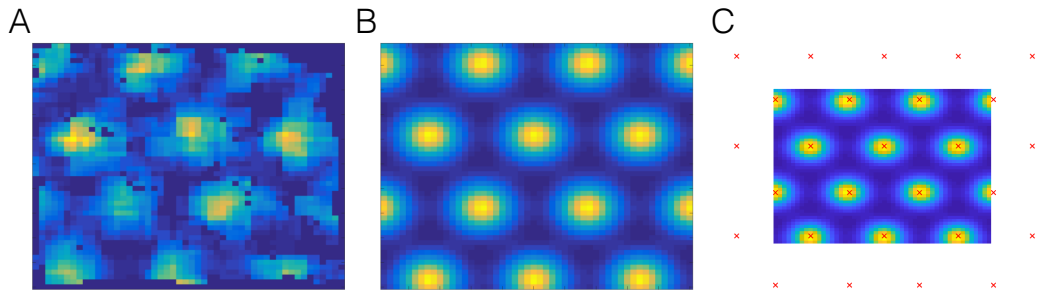


Figure 5.13 **Ideal grid**. We fit an idealised grid rate map using the formula from (Stemmler et al., 2015)(see Appendix 1 for details) to the original grid cell rate maps to remove any possible confounds and to ensure that we obtain accurate grid cell peaks. A) An example original grid cell rate map. B) An idealised rate map of that in A. C) Accurate finding of grid cell peaks (red crosses) on the idealised grid rate map, which also allows peaks that extend outside the box to be used.

5.5.2 Preserved relationship between grid and place cells across environments

Our framework predicts a preserved relationship between place and grid cells of the same spatial scale (module). However, since we do not know the modules of the recorded cells, we can only expect a non-random relationship across the entire population. The general approach for testing this relationship is as follows. We pair each place cell from a given rat with each grid cell from that same rat. Since it was not required the place and grid cells were simultaneously recorded, a place cell may be paired with a grid cell from a different recording session (but the same rat). For each measure (gridAtPlace and minDist), we computed its value for every place cell-grid cell pair on two trials (e.g. trials 1 and 4 to test for a relationship across environments; Figure 5.11). After pooling all the pairs from all rats, a correlation across trials was then performed on these values to obtain a correlation coefficient for that measure over all the place cell-grid cell pairs. To test the significance of this correlation and ensure it is not driven by any bias in the data, we generated a null distribution by randomly shifting the place cell rate maps and recomputing the measures and their correlation across trials. We then examine where the correlation of the non-shuffled data lies relative to the null.

The original remapping experiment sought to establish the effects of environmental novelty on grid and place cell properties (Barry et al., 2012). Trials 1 and 5 were a familiar environment the animal has previously explored. Trials 2-4 were three exposures to a novel environment. They found that exposure to the novel environment caused an increase in grid scale and decrease in grid score, as well as an increase in place cell field size. This effect reduced with exposure to the novel environment, i.e. over the course of trials 2 to 4 such that trial 4 had grid scales and scores most comparable to 1 and 5. We therefore restrict our analyses of the second environment to trial 4.

As a sanity check, we first confirmed these measures were significantly correlated within environments (i.e. across two visits to the same environment - trials 1 and 5), when the cell populations should not have remapped. We see that for both measures there is a significant correlation across the trials (Figure 5.14), indicating that upon returning to the same environment, place cells and grid cells have retained their relationship with each other, as expected.

We then tested across environments (i.e. two different environments - trials 1 and 4), to assess whether our predicted non-random remapping relationship between grid and place cells exists. Here we also find significant correlations for both measures for the 115

place cell-grid cell pairs (Figure 5.14). This result is robust to different parameters, since we note the gridAtPlace result also holds across environments when not fitting an ideal grid and using a wide range of grid score cut-offs (with a grid score cut off of >0 and using original rather than idealised rate maps, we have 255 place cell-grid cell pairs and a significance of $p<0.005$). minDist measure is not calculated without the ideal grid due to inaccurate grid peaks. These are the first analyses demonstrating non-random place cell remapping based on neural activity, and provide evidence for a key prediction of our model: that place cells, despite remapping across environments, retain their relationship with grid cells.

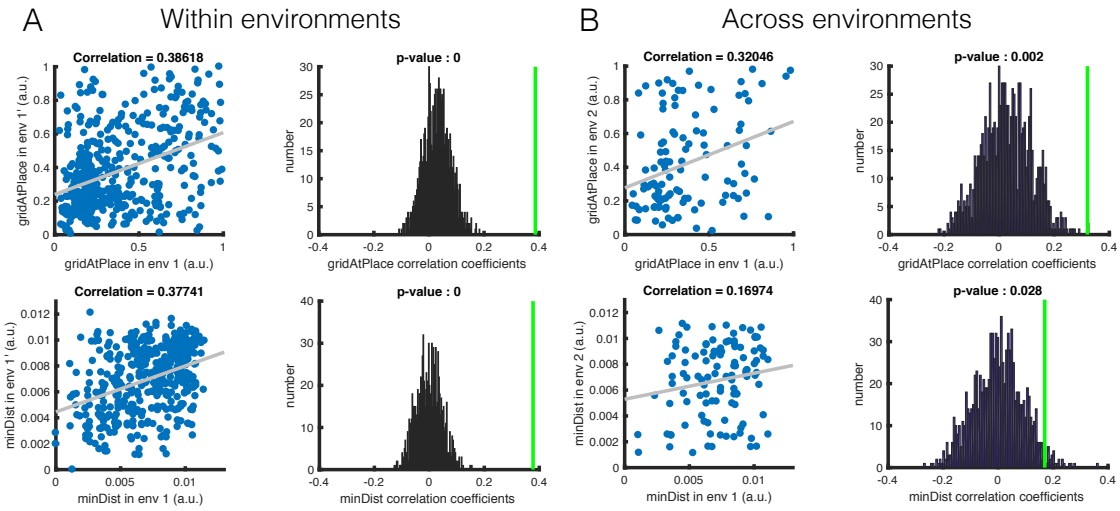


Figure 5.14 **Preserved relationship between place and grid cells across environments**. The scatter plots show the correlation of a given measure across trials, where each point is a place cell-grid cell pair. The histogram plots show where this correlation (green line) lies relative to the null distribution of correlation coefficients. The p value is the proportion of the null distribution that is greater than the unshuffled correlation. A) gridAtPlace (top) and minDist (bottom) measures are strongly significantly correlated over two trials within the same environment, as expected given the same place and grid code should be present. B) These measures are also significantly correlated across the two different environments, providing evidence that place and grid cells retain their relationship across environments.

5.5.3 Grid and place cell remapping

We note that all the grid cells realign across the environments and the place cell population has accordingly undergone global remapping (Figure 5.15); global place cell remapping generally accompanies grid realignment (Fyhn et al., 2007). That grid cells have all realigned means it is non-trivial the place and grid cell relationship is preserved.

Furthermore, in addition to place and grid cells preserving their relationship across environments, to generalise structure, the relationship between grid cells should be preserved across environments. Indeed, although grid cells realign across environments, their correlation structure is preserved (Fyhn et al., 2007). This was true in our data as well, since the correlation structure between the grid cells was itself correlated (i.e. preserved) across environments. This was significant relative to a null distribution generated by permuting grid cell correlations ($p<0.001$). These results are expected if the grid cells encode knowledge that generalises across environments.

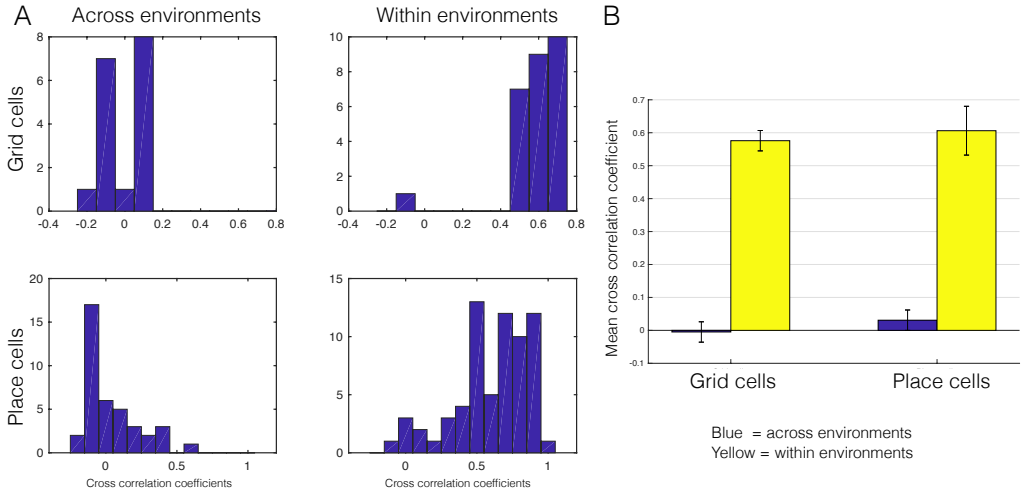


Figure 5.15 **Grid realignment and place cell remapping across environments**. A) Histograms showing the distributions of cross correlations obtained when correlating a given place or grid cell's rate map in one environment with its rate map in a second visit to that same environment (within environments) or its rate map in a different environment (across environments). It can be seen that all the grid cells realign across environments (correlation coefficients are all around 0) and the place cell distributions are very similar to those observed in hippocampal global remapping experiments (Fyhn et al., 2007), suggesting they have undergone global remapping. B) Bar plots showing the mean (±SEM) of these distributions.

5.5.4 Model-data correspondence

Finally, we demonstrate some concordance between these data and the model. Performing the across environment gridAtPlace analysis with the model rate maps, we observed correlations consistent with the data (Figure 5.16). Similarly, the model grid cell correlation matrix was itself highly correlated across environments (Figure 5.16), as expected if structural knowledge is generalised in the grid cells. These results show some qualitative correspondence between the model and the data.

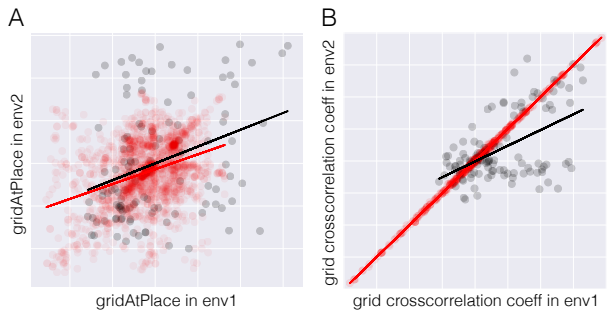


Figure 5.16 **Model-data correspondence**. Model and data are denoted by red and black lines, respectively. A) The gridAtPlace measure across environments. B) The correlation structure of grid cells is itself correlated (i.e. preserved) across environments. The important point here is both model and data are significantly positively correlated, not that the gradients of the lines are matched.

5.6 Discussion

In this chapter we have developed a framework centred on the hippocampal-entorhinal system capable of generalising structural knowledge across environments differing in their sensory particularities. Crucially, the representations of the structure of the world, i.e. how entities in the world relate to each other, are separated from representations of the entities themselves, and this pure structural representation is what grid cells in the brain may be for 2D space. By rethinking the notion of a place cell as the conjunction of this structural representation and sensory experiences, we show how this learnt structural knowledge can be generalised to new situations, as well as account for remapping, and suggest it is non-random. Combining these principles with hierarchical representations and fast Hebbian memory allows learnt structural knowledge to be generalised to perform accurate prediction of the next sensory experience in novel situations. Artificial neural networks endowed with these principles learn and generalise statistics of memories and learn neuronal representations similar to those found in the brain. We demonstrate empirical support for the model prediction that there is a preserved relationship between grid and place cells across environments.

Generalisation as a unifying computational goal of the hippocampal formation?

This framework is able to account for a range of phenomena and functions attributed to the hippocampal-entorhinal system. It offers a unified account of the classic functions of the hippocampus in episodic memory (Squire et al., 2004) and spatial cognition (O'Keefe and Nadel, 1978), as well as structure and concept learning and generalisation (Constantinescu et al., 2016; Kumaran et al., 2009; Mack et al., 2016; McKenzie et al., 2014).

It offers mechanistic explanations for a variety of phenomena in the hippocampal formation and, importantly, how they relate to each other in order to achieve the computational goal of generalisation. These include the emergence of grid cells, place cell responses including remapping, hierarchical representations, temporal representations in LEC, the role of Hebbian learning and recurrent attractors in supporting learning and recall of what happened where, amongst others. An appealing feature of having these different components in one framework is that we may be able to generate explanations for phenomena relying on multiple components. For example, it may explain how, in life, memories are recalled without having to search through a large memory store: constraints on the memory (e.g. where or when it occurred) can be imposed from cortex (grid cells) on to the place cell code allowing recall of only memories consistent with the constraints, i.e. once one knows where the memory is in the structure, one can recall it without searching through memories.

Our framework fits in to a broader view advanced by CLS (Kumaran et al., 2016; McClelland et al., 1995) in which cortex extracts the statistics of memories in hippocampus. However we close the loop by showing how structural representations in cortex can then constrain future experiences and memories in hippocampus, placing them in the context of previous learning. Furthermore, this framework is also able to account for various primary computational goals attributed to the hippocampal formation outlined in the introduction – such as spatial cognition, memory, and representing the relationships between memories. Taken together with the range of empirical data this framework supports, this suggests the computational goal of generalisation could be a candidate unifying goal of the hippocampal formation. However, this suggestion clearly needs further development, such as to account for how conceptual representations may be flexibly adapted in light of new goals by interactions with prefrontal cortex (Mack et al., 2016).

Empirical support for preserved place and grid cell relationships across environments

We found empirical support for the prediction of our model that there is a preserved relationship between place and grid cells across environments in data from a remapping experiment (Barry et al., 2012). Preserved relationships across environments are expected from a conjunctive code that is reused across space and allows for generalisation. This offers a new perspective on remapping, which previous theoretical proposals suggest generates random place cell codes (Monaco and Abbott, 2011). We propose that non-random, precise remapping is necessary for generalising knowledge. Our finding that place

and grid cells retain their relationships across environments, combined with reports of the sensory environment influencing place cells (Komorowski et al., 2009; Tsao et al., 2018; Wood et al., 1999), are supportive of hippocampal conjunctive codes.

Points from the previous chapter developed further

There are additionally a few points in common with the previous chapter that we have developed further here. First, although we focus on 2D graphs in order to compare learnt representations to the well-characterised, 'ground truth' neuronal representations of space, we note that this framework is general to arbitrary graph structures. That representations resembling those observed in the brain for space emerge in our framework suggests, as in the previous chapter, these observed representations may be an instance of general coding mechanisms responsible for organising information across multiple different domains, not just space. In this chapter we have demonstrated such general coding mechanisms may be those used to achieve generalisation.

Second, global codes arise from local learning here as well. Whereas in the previous chapter the agent was given the place cells, i.e. locations, and observed local transitions through the place cells, in the present chapter, we have gone closer to the experience of a real-world agent. Here, the agent only observes sensory experiences and from this must extract a representation of its location in the environment.

Third, this chapter also highlights the utility of having appropriate representations to mitigate expensive computation. In addition to demonstrating, as in the previous chapter, how efficient bases of the structure of the environment may be extracted, the present chapter demonstrates how learning these structural representations such that they can be generalised is highly efficient. By activating the appropriate place cells, i.e. those that are consistent with the grid cells, it is not necessary to perform computationally expensive searches to align the sensory representations with structural representations (the alignment problem); it is already ensured that only those place cells consistent with both will become active. That this is done hierarchically such that we reuse high frequency information across space dramatically reduces the amount of learning that is required, as we are able to reuse weights (between high frequency place and grid cells) across space. Finally, once these structural representations have been learnt over many experiences, they impose constraints on new learning and inference within an experience, such that fewer observations in novel situations are required to make accurate predictions.

Related work

It has recently been shown that grid cell-like representations emerge in artificial neural networks tasked with predicting either location (Cueva and Wei, 2018) or place and head direction cells (Banino et al., 2018) and trained with supervised learning. In these studies, the network is provided with place cells or location – information not directly available to a biological agent. In our framework the network extracts such information from observation of sensory experiences and is only additionally provided action information – all information available to real world agents – and trained with unsupervised learning. We additionally provide a role for place cells, incorporate memory and propose the role of grid cells is to represent and generalise structural knowledge rather than navigation *per se*.

Other networks have tackled a similar computational problem (Fraccaro et al., 2018; Gemici et al., 2017; Wayne et al., 2018). Rather than using biologically plausible memory mechanisms as we do here, these networks use biologically implausible memory similar to RAM, and additionally do not have explicit structural representations. They were also not generally inspired by the brain and did not model the hippocampal-entorhinal system.

Our approach has some similarities to meta-reinforcement learning (Wang et al., 2018) in the sense that over experience the network learns to learn. In our case, the network learns over many experiences a structural code and how to use it to address the correct place cells. This slow learning facilitates fast learning and inference within an experience (new environment). However, there are numerous notable differences. For example, we use unsupervised rather than reinforcement learning, to improve next step prediction rather than optimise reward, and our framework was designed to model the hippocampal-entorhinal system rather than prefrontal cortex. Although meta-RL must learn some form of generalisable structure, again they do not have an explicit structural representation.

We explicitly built the explicit structural representation in to the architecture of our framework. This automatically imposes the bias that relationships between objects are common across sets of objects (Behrens et al., 2018). Imposing such biases has parallels (Behrens et al., 2018) with the relational inductive biases that have been explored recently in novel machine learning algorithms (Battaglia et al., 2018), such as relation networks (Santoro et al., 2017), with the hope this will help them generalise. These approaches allow relationships common across objects to be learnt, but are yet to provide a model of the brain systems that achieve this. It will be interesting to see how these powerful networks relate to our framework.

Some experimental predictions

This framework offers numerous experimental predictions. Some of which, such as designing tasks with certain structure and looking for their predicted neuronal representations in the brain, are common with those of the previous chapter, and so we do not repeat them here.

Predictions from the conjunctive code: First, we mentioned above that hippocampal global remapping occurs in bats when their sensory experiences are switched from echolocation to vision within the same environment (Geva-Sagiv et al., 2016). We predict in this case the remapping is driven solely by sensory experiences since the environment is the same and they have not left it, and therefore the grid (structural) representation would not change. This is in contrast to hippocampal global remapping reported in literature thus far always being accompanied by grid realignment (Fyhn et al., 2007). Second, depending on the exact dynamics, altering activity in either MEC or LEC may give rise to altered activity in different populations of place cells (assuming the other region is still providing normal drive). For example, activating MEC may in theory allow all place cells consistent with the current sensory experience to become active, whereas activating LEC may allow all place cells consistent with the currently inferred location in the structure to become active. More generally, place cell activity should be predicted by both structural and sensory representations in MEC and LEC, respectively (we have provided evidence for the former in this chapter).

Predictions from hierarchical reuse of high frequency information across space: First, we have mentioned a conjunctive code combined with reuse of high frequency statistics across space explains multiple place peaks within an environment. We predict it should be possible, following identification of a place cell's preferred sensory experience, to move/repeat this sensory experience around the environment and observe multiple place cell firing fields that will be at a consistent grid phase and therefore form a triangular lattice. Second (and related to the next paragraph), behavioural errors should be more likely to occur at consistent grid phases. I.e. a high frequency sensory error should be more likely to be of something that shares the high frequency grid phase in a different location of space (e.g. in Figure 5.6 in Environment 1, when performing prediction of what is at the shaded region, an animal should be more likely to erroneously recall a shirt than a key, since the shirt shares the high frequency grid phase with the shaded region).

Predictions from hierarchical memory recall: First, since the ventral, lower frequency representations are responsible for refining the high frequency code during recall, silencing ventral regions should give rise to increased activity in dorsal regions of hippocampus and increased errors in sensory prediction on the same high frequency grid phases. Effectively this alters low frequency, contextual information. Second, memory specificity will occur at different times during recall depending on the hierarchical level. More concretely, during prediction or recall, representations in hippocampus should start becoming similar to the representation present during the original experience (i.e. storage) sooner in ventral than dorsal hippocampus. I.e. a wave of increased similarity between the recalled and experienced representation should progress from ventral to dorsal hippocampus.

Future directions

There are numerous possible future directions for this framework. We mention a few here. Although we have been able to account for a range of functions and phenomena in the hippocampal formation, we have of course not captured the whole story. The hippocampus is also involved in other functions, such as imagination, causal reasoning and transitive inference, as well as numerous other types of tasks (Buckmaster et al., 2004; Dusek and Eichenbaum, 1997; Hassabis et al., 2007; Van Der Meer et al., 2012). Indeed some of the conjunctive cells found in hippocampus don't only code for the conjunction of location and sensory experiences, but also for abstract task-relevant information such as whether a stimulus is the same now as it was previously (Wood et al., 1999). That our framework is not specific to space but rather general to arbitrary topologies suggests that it may be able to account for some of these other functions and representations of the hippocampal formation, as it may be able to learn and generalise the structure of many other tasks. This is yet to be explored but testing the generality of our framework and its application to tasks is an exciting future direction. When the same model is trained on multiple tasks (underlying structures), it will then become important for the agent to be able to infer which type of structure it is in. Of particular interest would be if the model is able to learn a basis set of structural primitives that it can flexibly recombine to construct novel structural representations of tasks (Behrens et al., 2018; Kemp and Tenenbaum, 2008).

On a related note, an important dimension of episodic memories is missing here: time. We have discussed how 'what' and 'where' may be bound together in hippocampus but not 'when'. Our framework predicts that when time is of importance to solve a task, time will be treated as another structural dimension of the experience and grid-like coding

of time will be required and represented in MEC, as has been experimentally observed (Kraus et al., 2015). Such grid-like coding of time resembles a spectral transform of time resulting in periodicity, analogous to that observed of space explored in the previous and current chapter, and is contrasted with the recently observed ramping activity with time in LEC, as would be expected from a Laplace transform of the recent past (Tsao et al., 2018) that we modelled and predicted to be in LEC. We further predict this grid-like coding of time to be explicit, rather than the predicted inherent encoding of time observed in LEC (Tsao et al., 2018). In addition to adding the temporal dimension to episodic memories, introducing time to the framework may offer insight in to transitive inference (Dusek and Eichenbaum, 1997) and how we can immediately generate plans by inferring which elements should be where in the abstract structure that now includes time, as discussed in (Behrens et al., 2018).

Since the explicit structural representation both forms an efficient basis of the underlying structure and generalises, it is a particularly good representation on which to learn algorithms useful for solving tasks. That it is an efficient basis of the structure means, for example, planning on it can be efficient (as in intuitive planning or (Bush et al., 2015; Stemmler et al., 2015)); that is generalises means that learning how to solve tasks on these representations can be reused in different sensory instantiations of the tasks. These structural representations may therefore be particularly useful world models on which algorithms such as model-based RL could operate.

6 Discussion

6.1 Summary

In this thesis we have investigated the internal models that support flexible behaviour. We have investigated how organisms learn, represent, monitor, update and use internal models to support a variety of flexible behaviours: adapting behaviour in a changing world, flexible planning without search, and generalisation of knowledge to novel situations. Although the different strands of work are fairly different, they broadly speaking share the common theme of examining internal models supporting flexible behaviours of some sort. In this discussion, we first very briefly summarise the work done in this thesis, then, since much of the discussion of the specific projects was done in their respective chapters, here we focus on possible links between the two different parts of this thesis as well as more speculative points.

In the first part of the thesis, we investigated the neural substrates of monitoring the quality of the current strategy and maintaining flexibility in internal models. Consistent with a large body of work implicating the ACC in behavioural flexibility, we found specifically ACC to be engaged when switching from a state of exploitation to exploration as well as when uncertainty in the current belief was increased during stable exploitative behaviour. We found evidence for the long-proposed interaction between ACC and the noradrenergic system, consistent with the possibility that the need for behavioural flexibility computed in ACC may be broadcast to other brain regions via the noradrenergic system. Despite considerable theoretical work proposing how noradrenaline may impact upon internal models as well as suggestive experimental work, evidence demonstrating a relationship between noradrenaline and a directly measured neural model representing beliefs has been lacking. We showed that pupil diameter – an index of NA release – predicted increases in the uncertainty of a neural model measured in mOFC, suggesting NA may play an important role in maintaining flexibility in internal models to facilitate adaptive behaviour.

In the second part of the thesis, we investigated the general principles for organising knowledge in cognitive maps in the MTL that may give rise to its observed representations, how these representations may be used for flexible planning and inference, as well as how we may be able to generalise abstract structural knowledge to novel situations. We took advantage of the well-characterised “ground truth” neuronal representations of space by

attempting to understand under what conditions they emerge. By reconceptualising space as a state space, we showed, as others have previously done, that grid-like firing patterns can be obtained as a PCA of the activity of 2D-distributed place cells, suggesting grid cells may be summarising the statistics of the topology of state spaces. We showed that grid-like eigenvector representations of state spaces contain the necessary information to represent the structure of the environment for flexible planning without search. We then showed that when various desirable information-processing properties of the MTL are put together to solve the computational goal of generalisation, representations mirroring those in the brain emerge. More specifically, we showed that an agent endowed with fast memory and encouraged to learn explicit structural representations to facilitate prediction of sensory experiences across 2D environments learns abstract, generalisable structural knowledge resembling grid cells in 2D. This knowledge representation allows inference of unobserved links and can be combined with novel sensory experiences in a conjunctive code resembling place cells. We then provided empirical support for a prediction of this framework: that place and grid cells maintain their relationships across environments.

These two sets of work have been presented somewhat separately. It is natural to ask how they may interact. Here we briefly examine potential interactions between the questions and brain systems presented in the two different threads of this thesis.

6.2 Discussion

As briefly mentioned in the introduction, it is becoming increasingly clear that both PFC and the hippocampal formation as well as their interactions are important for the various roles ascribed to either set of regions, including spatial cognition, memory, value-based decision making, social cognition, concept learning, and generalisation (e.g. (Barron et al., 2013; Constantinescu et al., 2016; Doeller et al., 2010b; Eichenbaum, 2017c; Kumaran et al., 2009, 2012; Mack et al., 2016; Tavares et al., 2015; Vikbladh et al., 2018; Wikenheiser et al., 2017; Wikenheiser and Schoenbaum, 2016; Wimmer and Shohamy, 2012)). Therefore in general it seems as though to understand these processes it is necessary to consider both systems.

As in chapters 4 and 5, and consistent with others' proposals (Dordek et al., 2016; Stachenfeld et al., 2017), we proposed the MTL is responsible for representing the structure of the environment, and MEC in particular may represent an explicit and generalisable structural code, for which grid cells are an example in 2D spaces. There is structure in many tasks and domains that would benefit from appropriate structural representations.

Therefore, given the importance of both sets of regions and their interactions in their various functions, it is natural to ask whether the involvement of PFC could also be explained in terms of coding for the structure of the environment. Here we discuss evidence PFC may similarly be important for representing the structure of tasks, evidence demonstrating the importance of interactions between PFC and MTL, followed by highlighting some of the possible differences between them. We then speculatively further discuss some of our results in light of these considerations.

Evidence for structural representations in mOFC/vmPFC

As we mentioned in the introduction, patients with lesions to vmPFC/mOFC have difficulty making decisions in a goal-directed fashion (Bechara et al., 1994, 2000). This is a decision making strategy necessitating some form of model of the structure of the environment, suggesting a lack of structural representations could explain their deficits. Indeed, these patients have been characterised as having "difficulty in organising and structuring their problem space" (Goel et al., 1997). More precise evidence comes from animal work. Lesions to OFC in monkeys results in an inability to appropriately assign credit to the choices that caused outcomes (Walton et al., 2010). Instead of learning in a precise fashion, as a structural representation would permit, OFC-lesioned monkeys rather have to rely on more basic, less precise temporal correlations, assigning reward credit to the recent history of choices rather than to the specific part of the task (structure) that caused them (Walton et al., 2010). This is consistent with the observation that only contingent rewards – those requiring an understanding of the task structure – as opposed to randomly delivered rewards engage the OFC (Jocham et al., 2016). Both of these observations can be interpreted as an inability to represent the structure of contingencies between objects in the environment (Behrens et al., 2018) or the correct location in task state space (Wilson et al., 2014). Furthermore, learning sets (Harlow, 1949), which crucially rely on structural representations that generalise across sensory particularities, are thought to rely on the PFC (Wang et al., 2018). While these observations do not necessarily mean the abstract structure *per se* is represented in this region, it is suggestive that it is at least playing a critical part in computing on this structure if represented elsewhere.

Furthermore, we have in this thesis identified grid cells as candidate explicit structural representations, and therefore their representation in PFC would be further evidence of structural codes in these regions. Indeed, in both of the studies previously mentioned that found markers of grid cells in spatial (Doeller et al., 2010b) and non-spatial

(Constantinescu et al., 2016) contexts in MEC, these markers were also found in vmPFC/mOFC. In fact, they were found in a network of brain regions – termed the default mode network (DMN) (Fox and Raichle, 2007) – that is engaged during a variety of tasks including memory, valuation, and social cognition, and engaged during offline periods when these structural representations may be built. Furthermore, direct cellular recordings in humans have found grid cells in some of these regions (Jacobs et al., 2013). Hence grid cells, which we consider structural representations, are represented in PFC. Given these cells are found in a network involved in a variety of tasks, it is possible that, similar to the MTL, a range of the functions ascribed to these regions may be explained by their reliance on general coding mechanisms responsible for organising structural knowledge.

An important feature that we have emphasised of the structural representations in MTL is that they generalise over sensory particularities. This also appears to be true of representations in PFC. Learning sets require the generalisation of a rule to novel sets of stimuli (Harlow, 1949), as do other tasks reliant on PFC (Walton et al., 2010). When humans observe or listen to stories, representations in mOFC (and other regions in the DMN) generalise over the specifics, rather coding for the current location in an abstract schema of the stories (Baldassano et al., 2018). Generalisations have also been observed at the level of singe units (Morrissey et al., 2017; Xie and Padoa-Schioppa, 2016; Yu et al., 2018). For example, when monkeys are choosing between different pairs of juices, OFC neurons code for the same set of decision variables across different pairs of juices (Xie and Padoa-Schioppa, 2016).

As pointed out in (Behrens et al., 2018), representations in mOFC therefore bear similarities to those of grid cells in MEC. Specifically, They encode the current location, or state, in the task state space, the relationship between states, and generalise over sensory particularities. Taken together with the above observations, these results suggest that PFC may contain similar generalisable structural codes to those we have discussed in MTL, and it is possible these may underlie a range of functions ascribed to PFC. It will be an important and interesting challenge for the future to test whether the myriad of tasks reliant upon, or complex and hard to interpret representations found in, PFC can be explained under the general framework of structural representations.

Interactions between PFC and MTL are important for structural representations

Having presented evidence suggestive of structural representations in PFC, we now ask whether these may interact with those in MTL. Impaired communication between these

regions indeed results in impairments in structure-dependent learning set tasks (Browning et al., 2006; M'Harzi et al., 1987). More recent studies offer insight in to the types of representations that their interactions may support.

First of all, there is convincing evidence that state representations in OFC are dependent on hippocampal input. As we discussed in the introduction chapter, a recent study found OFC represents an abstract 'holistic' representation of the current task that cannot be explained by the sum of its specific task parts (Wikenheiser et al., 2017), as expected from this region representing states (Wilson et al., 2014). Importantly, suppressing hippocampal input to OFC abolishes this abstract state coding (Wikenheiser et al., 2017), consistent with hippocampus providing representations that define the task structure.

Since one needs to know their location in state space in order to update the appropriate part of the state space representation (cognitive map) following observations (credit assignment), such state representations in OFC may be important for facilitating updating of hippocampal representations. Indeed, activity in the (lateral) OFC at the time of learning predicts the updating of hippocampal associations between objects in the environment (Boorman et al., 2016). Furthermore, there is evidence that conceptual object representations in the hippocampus are updated by mOFC to reflect new task goals (Mack et al., 2016). Hence interactions between PFC and MTL appear important for organising and updating conceptual or structural task representations.

Differences between MTL and PFC structural representations

Having highlighted similarities between MEC and PFC representations, we now mention some potential differences. Most of these points have been made implicitly already; we just make them explicit here.

A full cognitive map may be represented in MTL. A large piece of the puzzle that we have not explored in much detail is how these abstract structures are used. We touched on planning in Chapter 4, but where did the goals come from? Whereas MTL may represent a full cognitive map, the OFC may represent the current location within this map, i.e. the state in the task state space (Schuck et al., 2016; Wilson et al., 2014). In particular OFC may be important when the current state is unobservable, i.e. for representing latent states (Schuck et al., 2016; Wilson et al., 2014). This state representation may be able to integrate the cognitive map representations in MTL with the agent's goals – consistent with OFC responding to the subjective value of stimuli (for example, OFC neurons will fire to food but only if the animal is hungry (Rolls et al., 1989)). OFC state representations may therefore be

attending to the relevant dimensions of the map in order to flexibly re-represent it to suit the agent's current goals. Evidence for this comes from an elegant study demonstrating that mOFC compresses task representations, so as to filter out irrelevant information, to a degree predicted by participants' ability to attend to relevant information (Mack et al., 2017). Such a role of OFC in attending to relevant dimensions according to current goals may predict that, for example, in intuitive planning presented in chapter 4, the current and goal locations required to construct the plans may be input to MEC from OFC. In a similar vein, it has been proposed that PFC provides a top down context signal to hippocampus for memory retrieval (Eichenbaum, 2017c).

If OFC encodes the current state of the task, then if there is no relevant task at all, we may not expect structured representations of the current environment to be present. There is suggestive evidence this is indeed the case. In a recent study, humans were required to perform a cover task on a sequence of objects (e.g. press left if the current object is upside down). Unbeknownst to them, and irrelevant to the cover task, the order of the sequence of objects was determined by a random walk on a structured graph, with the objects at the vertices. Remarkably, the structure of this graph could be recovered from the MTL using fMRI (Garvert et al., 2017) (offering a candidate neural substrate for Tolman's latent learning experiments (Tolman and Honzik, 1930)). However, there was no evidence for such a representation anywhere else in the brain, including OFC. These results taken together with the above are consistent with the notion that MTL represents the full cognitive map whereas the OFC represents only the currently relevant task state, and may act as a bridge between map-like representations and an agent's goals and actions.

Maintaining flexibility in structural state inferences

With these considerations in mind, it is noteworthy that in our fMRI study presented in the first part of this thesis, the hippocampus and mOFC followed the same activation patterns. In particular, they were both strongly engaged during exploitation, their activity increased with certainty in the current strategy, and there was weak evidence of internal model representations in hippocampus, as well as strong evidence in mOFC. Whilst this study was of course not designed to test for structural representations, it is interesting that the mOFC in this study can be explained as representing one's location in the state space of the current task, and it is possible the hippocampal activation reflects its supporting of this state representation (Wikenheiser et al., 2017).

We provided evidence for a role of the noradrenergic system in maintaining flexibility in this state inference to allow for behavioural adaptation when the current strategy is no longer taking us towards our goals. Although of course highly speculative, NA-induced flexibility in state inferences may add a new piece to the puzzle, offering potential insight in to how animals may be able to flexibly shift or re-represent their cognitive maps in light of changing goals or environment contingencies. Although in a different context, this is consistent with the observation that reducing NA reduces mental search in hippocampus (Amemiya and Redish, 2016). This study was focused on neural simulation of candidate routes, but it is likely similar principles would apply to search of structural representations. NA input may therefore facilitate shifting between different "organisation and structuring of the problem space" (Goel et al., 1997) in vmPFC/mOFC.

Temporary representations in ACC

The ACC had the opposite pattern of activity to this hippocampal-mOFC network. Furthermore, no evidence of entropy-dependent decoding as found in mOFC was found in ACC, nor much, if any, action decoding in the regions of ACC in which we observed the belief certainty and exploitation to exploration transition effects. This raises the question of what type of representation is in the ACC. We discussed a notable study in the introduction that recorded population activity in rats performing a task very similar to ours (Karlsson et al., 2012). They found ACC activity was stable as rats exploited an option, underwent periods of rapid and widespread remodelling as they switched in to exploration, and then became stable again once they resumed exploitation. Intriguingly, when the rats returned to exploiting the same option, the pattern of ACC activity in the majority of cases was different to before. A recent study has more convincingly demonstrated that after switching strategy ACC always establishes a new stable activity state, even upon return to the same strategy (Malagon-Vina et al., 2018). These observations would explain the lack of decoding in our study, as the ACC would never be in the same state in the train and test data partitions. Furthermore, when the optogenetically-induced increase in NA input to ACC that caused stochastic behaviour was ceased, the rats reverted back to their previous strategy (Tervo et al., 2014). This suggests that although the rats disengaged behaviour from their internal model, the internal model remained intact, which may indicate it is contained elsewhere.

If these suggestions are true, then rather than containing consistent task structural representations across time, ACC holds a temporary representation, which may be responsible for monitoring task performance and change (Kolling and O'Reilly, 2018;

Malagon-Vina et al., 2018). In this case it regulates the updating and use of internal models but does not contain them itself. This temporary representation may contain a unique combination of reward history and the current strategy (Kolling and O'Reilly, 2018) that forms a temporary set of expectations about task reward rates, against which true reward rates are compared such that this representation comes to monitor the quality of the current strategy. When the strategy does not meet expectations, this representation becomes volatile until it is discarded along with the behavioural strategy (Karlsson et al., 2012). It is possible that large increases in noradrenergic signalling cause the resetting (Bouret and Sara, 2005) of this temporary representation, that allows the initiation of exploration (Karlsson et al., 2012; Tervo et al., 2014), and puts the network in a state to allow learning of a new temporary representation. This effect of noradrenaline would mediate the alteration of the temperature parameter of the softmax choice model, as opposed to the option values (as discussed in Chapter 3). Importantly, the stable task representations that generalise across time need not be held in ACC in this case. This is of course highly speculative; how such a temporary representation that contains the information necessary to monitor the quality of the current strategy is created is not known. However we discuss this here as it offers a possible explanation for the ACC results we observed in our study, and may be an interesting direction for future research to shed light on the elusive function of the ACC (Ebitz and Hayden, 2016).

6.3 Conclusion

To conclude, using a variety of different techniques and firmly grounded in computational modelling, we have explored how internal models of the environment may allow organisms to exhibit behaviour that is flexible to changes in the environment or internal goals. We have investigated the neural substrates for tracking the uncertainty in beliefs and driving exploration of alternatives, as well as how internal models may be kept flexible to allow for behavioural adaptation. We have shown how the well-known internal model of space may arise under more general principles that represent the topology of the environment to provide a cognitive map for systematically organising knowledge across multiple domains (Tolman, 1948). Although there is of course much work to be done, investigating internal models seems to be a promising direction for understanding the kinds of flexible behaviours that characterise mysterious feats of biological intelligence.

7 Appendix

7.1 Appendix 1 – Grid score and ideal grid rate map

Grid scores:

The grid scores presented in this thesis are computed exactly as in (Barry et al., 2012), which is described in their paper as the following. Gridness is a measure of spatial periodicity and is calculated by defining a mask over the spatial autocorrelation centred on but excluding the central peak and bounded by a circle passing around the outside edge of the outermost of the six central peaks. This area of the spatial autocorrelation is then rotated in 30° increments up to 150°, and for each rotation, the Pearson product moment correlation coefficient was calculated against the unrotated mask. The grid score is then expressed as the lowest correlation obtained for rotations at 60° and 120° minus the highest correlation obtain at rotations of 30°, 90°, or 150°. Intuitively, it can be seen that this will be high if there is 6-fold symmetry as the spatial autocorrelation is rotated, as there will be in a triangular lattice.

Ideal grid rate map:

In order to obtain idealised grid cell spatial firing rate maps, we used the grid cell rate map equation in (Stemmler et al., 2015), which is described in their paper as the following. To obtain the firing rate map $\Omega(\vec{x})$ of a grid cell, which describes the cell's expected spike count when the animal is at location $\vec{x} = (x, y)$, superimpose three plane waves with wave vectors $\omega\overrightarrow{k_l}$, where $\left|\overrightarrow{k_l}\right| = 1$ for $l \in \{1,2,3\}$, and exponentiate the result:

$$\Omega(\vec{x}) = n_{max} \cdot \exp\left[\varkappa/3 \sum_{l=1}^{3}\{cos\ (\omega\overrightarrow{k_l} \cdot \vec{x}) - 1\}\right]$$

This is a 2D von Mises function. For a hexagonal grid that is aligned to the x axis, the three vectors are $\overrightarrow{k_l} = (\cos(\phi_l), \sin(\phi_l))$ with $\phi_l = -\pi/6 + l\pi/3$. The term ω is given by $\omega = 2\pi/(\sin(\pi/3)\lambda)$, where λ is the grid scale, $1/\varkappa$ measures the cell's relative tuning width and n_{max} is the maximal expected spike count. To capture the spatial phase $\overrightarrow{c_j}$ of cell j's firing pattern, the argument $\vec{x}$ is replaced by $\vec{x} - \overrightarrow{c_j}$. We obtain these parameters (scale, phase, orientation (which can be added to ϕ_l)) from the original grid rate maps. We can also

fit all parameters by minimising the sum of squares between the idealised and original grid rate maps.

7.2 Appendix 2 – NeuroRisk

We include here something else done during my DPhil that I am proud to have been a part of; that is, with Alon Baram, co-organising two lab retreats, a two day intensive computational neuroscience 'training course' ("pre-retreat"), and creating a fun game to learn neuroanatomy to play at the retreat: NeuroRisk, which we describe here:

Introduction: We should know more neuroanatomy and games are fun.

Methods: We developed NeuroRisk, inspired by the board game Risk, a strategy board game wherein a world map is divided in to territories that players compete for by using their armies to capture territories, with the outcome of an attack determined by dice rolls. We adapted Risk to NeuroRisk, where instead of the world, the map is the human cortex. People are split up in to multiple teams (of ~5 people) that are each given a mission, such as "conquer the frontal lobe". Each team will start the game with their soldiers randomly distributed throughout the brain. The success of an attack on a brain region is, instead of by chance, determined by which team can present some pitch about that brain region that wins a popular vote. So in order to attack some brain region, the attacking team pitches, for example, an interesting or funny fact about that brain region. The defending team then does the same. The rest of the players (not on either of those teams) then votes on who deserves to win the battle based on their pitch. The rules are below.

Importantly, attacks can only be made to brain regions that are functionally connected. To determine functional connectivity, we downloaded and used the parcellation by (Glasser et al., 2016) along with their HCP Workbench viewer. We clustered brain regions together to reduce the number of parcels from ~200 to ~50, resulting in our own map of human cortex, i.e. the board of the game (Figure 7.1). Using the HCP Workbench, we could then online – during the game – determine functional connectivity, i.e. between which brain regions (or territories) attacks could be made.

Results: The game proved to be a lot of fun, with frantic searches for facts about brain regions (Figure 7.2) and innovative pitches or attacks. As with normal Risk, however, the game took a very long time and did not reach a conclusion. We all learnt a lot about various brain regions.

Conclusion: Learning neuroanatomy can be fun.

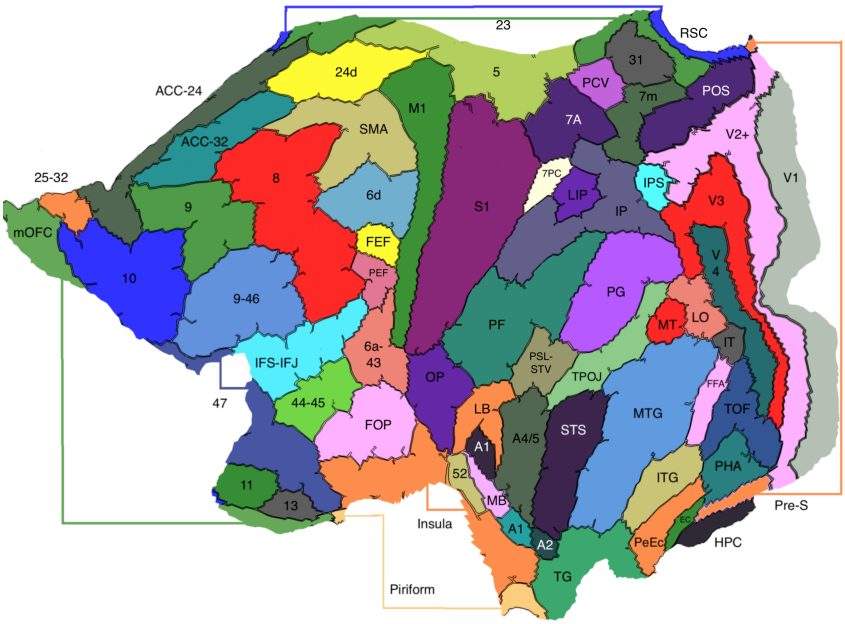


Figure 7.1 **Map of the human cortex, i.e. the board in NeuroRisk**.



Figure 7.2 **Game in progress**. Below: can see the board and soldiers (skittles) on it. Above: can see people searching for facts on their laptops. Age of Empires battle music is playing in the background.

NeuroRisk Rules:

The world:

- Human cortex.

Teams & soldiers:

- Five teams.
- Each team will start with a soldier on ~ 1/5 of the territories.
- Plus more soldiers to distribute as you please, according to your mission.
- At the end of each go, you can redistribute your soldiers.
- You can only eat skittles of your team colour (this is not recommended).

Your turn and attacking:

- At the start of your turn, you will receive more new soldiers; the number will depend on how many regions you conquered on the last go & number of current territories; According to: ceil(current_territories/3) + randi(conquered_territories)
- You can keep attacking regions for as long as you want when it's your turn.
- Challenge a brain region you wish to attack, and state from which region you are attacking (that you currently occupy).
- You must be functionally connected to the region you are attacking from your current region. Being adjacent territories is NOT enough to attack.
- Attacker states with how many soldiers they are attacking. They must have at least this number +1 on the region from which they are attacking.
- To attack, the attacker pitches something interesting or cool about that region.
- Defender responds with their own interesting or cool fact about that region.
- Everyone else that is not in either of those two teams, votes on which pitch wins.
- The loser loses the number of soldiers that the attack was made with.
- If the defender now has no soldiers left on that region, the attacker takes that region.
- The attacker cannot leave their original territory empty.
- At the end of your turn, you can redistribute your soldiers (but must keep at least one on each of your occupied regions).

- Optional rule – To get around the problem of regions no one knows, there can be dice attacks:
- Attacker can choose to attack with dice, but with bad odds.
- If attacker chooses to attack with facts, defender can choose dice, but they will have bad odds.
- Whether it is a facts or dice attack is decided before any facts are pitched.
- Therefore the game will not halt at these regions, but there is still incentive to pitch facts.

8 References

Akam, T., Costa, R., and Dayan, P. (2015). Simple Plans or Sophisticated Habits? State, Transition and Learning Interactions in the Two-Step Task. PLoS Comput. Biol. *11*, 1–25.

Akam, T., Rodrigues-vaz, I., Zhang, X., Pereira, M., and Oliveira, R. (2017). Single-Trial Inhibition of Anterior Cingulate Disrupts Model-based Reinforcement Learning in a Two-step Decision Task . Abstract : Introduction : 1–49.

Amemiya, S., and Redish, A.D. (2016). Manipulating Decisiveness in Decision Making: Effects of Clonidine on Hippocampal Search Strategies. J. Neurosci. *36*, 814–827.

Aprahamian, M., Higham, D.J., and Higham, N.J. (2016). Matching exponential-based and resolvent-based centrality measures. J. Complex Networks *4*, 157–176.

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

Aston-Jones, G., and Cohen, J.D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. Annu. Rev. Neurosci. *28*, 403–450.

Baldassano, C., Hasson, U., and Norman, K.A. (2018). Representation of real-world event schemas during narrative perception. BioRxiv 252718.

Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action : contingency and incentive learning and their cortical substrates. Neuropharmacology *37*, 407–419.

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., Pritzel, A., Chadwick, M.J., Degris, T., Modayil, J., et al. (2018). Vector-based navigation using grid-like representations in artificial agents. Nature *557*, 429–433.

Bär, K.J., De la Cruz, F., Schumann, A., Koehler, S., Sauer, H., Critchley, H., and Wagner, G. (2016). Functional connectivity and network analysis of midbrain and brainstem nuclei. Neuroimage *134*, 53–63.

Barrash, J., Tranel, D., and Anderson, S.W. (2000). Acquired Personality Disturbances Associated With Bilateral Damage to the Ventromedial Prefrontal Region. Dev. Neuropsychol.

Barron, H.C., Dolan, R.J., and Behrens, T.E.J. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. Nat. Neurosci.

Barry, C., Ginzberg, L.L., O'Keefe, J., and Burgess, N. (2012). Grid cell firing patterns signal environmental novelty by expansion. Proc. Natl. Acad. Sci. *109*, 17687–17692.

Bartra, O., Mcguire, J.T., and Kable, J.W. (2013). The valuation system : A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage *76*, 412–427.

Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al. (2018). Relational inductive biases , deep learning , and graph networks. ArXiv 1–38.

Bechara, A., Damasio, A.R., Damasio, H., and Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition *50*, 7–15.

Bechara, A., Damasio, H., and Damasio, A.R. (2000). Emotion , Decision Making and the Orbitofrontal Cortex. 295–307.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., and Rushworth, M.F.S. (2008). Associative learning of social value. *456*, 245–250.

Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A., Stacenhfeld, K.L., and Kurth-Nelson, Z. (2018). What is a cognitive map? Organising knowledge for flexible behaviour. BioRxiv.

Benzi, M., and Klymko, C. (2013). Total communicability as a centrality measure. 1–27.

Boorman, E.D., Behrens, T.E.J., Woolrich, M.W., and Rushworth, M.F.S. (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. Neuron *62*, 733–743.

Boorman, E.D., Rajendran, V.G., O'Reilly, J.X., and Behrens, T.E. (2016). Two Anatomically and Computationally Distinct Learning Signals Predict Changes to Stimulus-Outcome Associations in Hippocampus. Neuron *89*, 1343–1354.

Bostock, E., Muller, R.U., and Kubie, J.L. (1991). Experience-dependent modifications of hippocampal place cell firing. Hippocampus *1*, 193–205.

Botvinick, M.M. (2012). Hierarchical reinforcement learning and decision making. Curr. Opin. Neurobiol. *22*, 956–962.

Bouret, S., and Sara, S.J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. Trends Neurosci. *28*, 574–582.

Browning, P.G.F., Easton, A., and Gaffan, D. (2006). Frontal-Temporal Disconnection Abolishes Object Discrimination Learning Set in Macaque Monkeys. Cereb. Cortex *17*, 859–864.

Buckmaster, C.A., Eichenbaum, H., Amaral, D., Suzuki, W., and Rapp, P. (2004). Entorhinal

Cortex Lesions Disrupt the Relational Organization of Memory in Monkeys. J. Neurosci. *24*, 9811–9825.

Bunsey, M., and Eichenbaum, H.B. (1996). Conservation of hippocampal memory function in rats and humans. Nature *379*, 255–257.

Burwell, R.D., Witter, M.P., and Amaral, D.G. (1995). Commentary: A perirhinal and postrhinal cortices of the rat: A review of neuroanatomical literature and comparison with findings from the monkey brain. Hippocampus *5*, 390–408.

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

Butter, C.M. (1969). Perseveration in Extinction and in Discrimination Reversal Tasks Following Selective Frontal Ablations in Macaca Mulatta ’. 163–171.

Chan, S.C.Y., Niv, Y., and Norman, K.A. (2016). A Probability Distribution over Latent Causes, in the Orbitofrontal Cortex. J. Neurosci. *36*, 7817–7828.

Chen, G., King, J.A., Lu, Y., Cacucci, F., and Burgess, N. (2018). Spatial cell firing during virtual navigation of open arenas by head-restrained mice. Elife *7*, e34789.

Cohen, N., and Squire, L. (1980). Preserved learning and retention of pattern analysing skill in amnesia: dossiciation of knowing how and knowing that. Science (80-.). *210*, 207–210.

Cohen, N.J., and Eichenbaum, H. (1993). Memory, Amnesia, and the Hippocampal System (Cambridge: MIT Press).

Colgin, L.L., Moser, E.I., and Moser, M.B. (2008). Understanding memory through hippocampal remapping. Trends Neurosci. *31*, 469–477.

Constantinescu, A.O., O’Reilly, J.X., and Behrens, T.E.J. (2016). Organizing conceptual knowledge in humans with a gridlike code. Science (80-.). *352*, 1464–1468.

Coull, J.T., Middleton, H.C., Robbins, T.W., and Sahakian, B.J. (1995). Contrasting effects of clonidine and diazepam on tests of working memory and planning. Psychopharmacology (Berl). 311–321.

Cueva, C.J., and Wei, X.-X. (2018). Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. 1–19.

Danjo, T., Toyoizumi, T., and Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. Science *359*, 213–218.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711.

Daw, N.D., O’Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. Nature *441*, 876–879.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron *69*, 1204–1215.

Dayan, P. (1993). Improving Generalisation for Temporal Difference Learning: The Successor Representation. Neural Comput. *5*, 613–624.

Derdikman, D., Whitlock, J.R., Tsao, A., Fyhn, M., Hafting, T., Moser, M.B., and Moser, E.I. (2009). Fragmentation of grid cell maps in a multicompartment environment. Nat. Neurosci. *12*, 1325–1332.

Deshmukh, S.S., and Knierim, J.J. (2013). Influence of local objects on hippocampal representations: Landmark vectors and memory. Hippocampus *23*, 253–267.

Dickinson, A., and Balleine, B. (1994). Motivational control of goal-directed action. Anim. Learn. Behav. *22*, 1–18.

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

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

Dolan, R.J., and Dayan, P. (2013). Goals and habits in the brain. Neuron *80*, 312–325.

Dordek, Y., Soudry, D., Meir, R., and Derdikman, D. (2016). Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. Elife *5*, 1–36.

Duncan, K., Doll, B.B., Daw, N.D., Shohamy, D., Duncan, K., Doll, B.B., Daw, N.D., and Shohamy, D. (2018). More Than the Sum of Its Parts : A Role for the Hippocampus in Configural Reinforcement Learning Article More Than the Sum of Its Parts : A Role for the Hippocampus in Configural Reinforcement Learning. Neuron 1–13.

Dusek, J.A., and Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. Proc. Natl. Acad. Sci. *94*, 7109–7114.

Ebitz, R.B., and Hayden, B.Y. (2016). Dorsal anterior cingulate : a Rorschach test for cognitive neuroscience. Nat. Publ. Gr. *19*, 1278–1279.

Ebitz, R.B., and Platt, M.L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. Neuron *85*, 628–640.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. Neuron *44*, 109–120.

Eichenbaum, H. (2017a). Memory : Organization and Control.

Eichenbaum, H. (2017b). The role of the hippocampus in navigation is memory. J. Neurophysiol. *117*, 1785–1796.

Eichenbaum, H. (2017c). Prefrontal–hippocampal interactions in episodic memory. Nat. Rev. Neurosci. *18*, 547–558.

Eichenbaum, H., and Cohen, N.J. (2014). Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? Neuron *83*, 764–770.

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., and Tanila, H. (1999). The Hippocampus, Memory, Review and Place Cells: Is It Spatial Memory or a Memory Space? Neuron *23*, 209–226.

Eslami, S.M.A., Rezende, D.J., Besse, F., Viola, F., Morcos, A.S., Garnelo, M., Ruderman, A., Rusu, A.A., Danihelka, I., Gregor, K., et al. (2018). Neural scene representation and rendering. *1210*, 1204–1210.

Estrada, E., and Higham, D.J. (2010). Network properties revealed through Matrix Functions. *52*, 696–714.

Estrada, E., Higham, D.J., and Hatano, N. (2009). Communicability betweenness in complex networks. Phys. A Stat. Mech. Its Appl. *388*, 764–774.

Fenton, A., Kao, H., Neymotin, S.A., Olypher, A., Vayntrub, Y., Lytton, W.W., and Ludvig, N. (2008). Unmasking the CA1 Ensemble Place Code by Exposures to Small and Large Environments : More Place Cells and Multiple , Irregularly Arranged , and Expanded Place Fields in. *28*, 11250–11262.

Fox, M.D., and Raichle, M.E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *8*.

Fraccaro, M., Jimenez, D., Yori, R., Alexander, Z., Ali, P.S.M., and Fabio, E. (2018). Generative Temporal Models with Spatial Memory for Partially Observed Environments.

Franklin, N.T., and Frank, M.J. (2015). A cholinergic feedback circuit to regulate striatal population uncertainty and optimize reinforcement learning. Elife *4*, 1–29.

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

Garvert, M.M., Dolan, R.J., and Behrens, T.E.J. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. Elife *6*, 1–20.

Gauthier, J.L., and Tank, D.W. (2018). A Dedicated Population for Reward Coding in the Hippocampus. Neuron *0*, 1–15.

Gemici, M., Hung, C., Santoro, A., Wayne, G., Mohamed, S., Rezende, D.J., Amos, D., and Lillicrap, T. (2017). GENERATIVE TEMPORAL MODELS WITH MEMORY. 1–25.

Gershman, S.J. (2015). A Unifying Probabilistic View of Associative Learning. 1–20.

Geva-Sagiv, M., Romani, S., Las, L., and Ulanovsky, N. (2016). Hippocampal global remapping

for different sensory modalities in flying bats. Nat. Neurosci. *19*, 952–958.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature *536*, 171–178.

Goel, V., Grafman, J., Tajik, J., Gana, S., and Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. 1805–1822.

Grieves, R.M., and Jeffery, K.J. (2017). The representation of space in the brain. Behav. Processes *135*, 113–131.

Grieves, R.M., Jenkins, B.W., Harland, B.C., Wood, E.R., and Dudchenko, P.A. (2016). Place field repetition and spatial learning in a multicompartment environment. Hippocampus *26*, 118–134.

Gustafson, N.J., and Daw, N.D. (2011). Grid cells, place cells, and geodesic generalization for spatial reinforcement learning. PLoS Comput. Biol. *7*.

Hafting, T., Fyhn, M., Molden, S., Moser, M.B., and Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature *436*, 801–806.

Harlow, H.F. (1949). The formation of learning sets. Psychol. Rev. *56*, 51–65.

Harlow, J. (1868). Recovery from the passage of an iron bar through the head. Publ. Massachusetts Med. Soc.

Hartley, T., Lever, C., Burgess, N., Keefe, J.O., and Hartley, T. (2013). Space in the brain : how the hippocampal formation supports spatial cognition.

Hassabis, D., Kumaran, D., Vann, S.D., and Maguire, E.A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. Proc. Natl. Acad. Sci. *104*, 1726–1731.

Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. Neuron *95*, 245–258.

Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. Nat. Neurosci. *14*, 933–939.

Høydal, Ø.A., Skytøen, E.R., Moser, M.-B., and Moser, E.I. (2018). Object-vector coding in the medial entorhinal cortex. BioRxiv.

Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. *148*, 574–591.

Huys, Q.J.M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S.J., Dayan, P., and Roiser, J.P. (2015). Interplay of approximate planning strategies. Proc. Natl. Acad. Sci. *112*, 3098–3103.

Izquierdo, A., Suda, R.K., and Murray, E.A. (2004). Bilateral Orbital Prefrontal Cortex Lesions

in Rhesus Monkeys Disrupt Choices Guided by Both Reward Value and Reward Contingency. *24*, 7540–7548.

Jacobs, J., Weidemann, C.T., Miller, J.F., Solway, A., Burke, J.F., Wei, X.-X., Suthana, N., Sperling, M.R., Sharan, A.D., Fried, I., et al. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. Nat. Neurosci. *16*, 1188–1190.

James, A., Oertzen, T.J. Von, Norbury, R., Huppertz, H., and Brandt, K.R. (2018). Left entorhinal cortex and object recognition. Neuroreport 363–367.

Jepma, M., and Nieuwenhuis, S. (2011). Pupil Diameter Predicts Changes in the Exploration–Exploitation Trade-off: Evidence for the Adaptive Gain Theory. J. Cogn. Neurosci. *23*, 1587–1596.

Jocham, G., Brodersen, K.H.H., Constantinescu, A.O.O., Kahn, M.C.C., Ianni, A.M., Walton, M.E.E., Rushworth, M.F.F.S., and Behrens, T.E.E.J. (2016). Reward-Guided Learning with and without Causal Attribution. Neuron *90*, 177–190.

Jodo, E., Chiang, C., and Aston-Jones, G. (1998). Potent excitatory influence of prefrontal cortex activity on noradrenergic locus coeruleus neurons. Neuroscience *83*, 63–79.

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

Jones, J.L., Esber, G.R., McDannald, M.A., Gruber, A.J., Hernandez, A., Mirenzi, A., and Schoenbaum, G. (2012). Orbitofrontal Cortex Supports Behavior and Learning Using Inferred But Not Cached Values. Science (80-.). *338*, 953–956.

Joshi, S., Li, Y., Kalwani, R.M., and Gold, J.I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. Neuron *89*, 221–234.

Kamin, L. (1969). Selective association and conditioning. In: Fundamental issues in instrumental learning.

Kane, G.A., Vazey, E.M., Wilson, R.C., Shenhav, A., Daw, N.D., Aston-jones, G., and Cohen, J.D. (2017). Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. 1073–1083.

Karlsson, M.P., Tervo, D.G.R., and Karpova, A.Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science (80-.). *338*, 135–139.

Kell, A.J.E., Yamins, D.L.K., Shook, E.N., Norman-haignere, S. V, Mcdermott, J.H., Kell, A.J.E., Yamins, D.L.K., Shook, E.N., and Norman-haignere, S. V (2018). Article A Task-Optimized Neural Network Replicates Human Auditory Behavior , Predicts Brain Responses , and Reveals a Cortical Processing Hierarchy Article A Task-Optimized Neural Network Replicates

Human Auditory Behavior , Predicts Brain Responses , and Reveals a Cortical Processing Hierarchy. Neuron *98*, 630–644.e16.

Kemp, C., and Tenenbaum, J.B. (2008). The discovery of structural form. Proc. Natl. Acad. Sci. *105*, 10687–10692.

Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., and Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulate cortex. Nat. Neurosci. *9*.

Khaligh-Razavi, S.-M., and Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Comput. Biol. *10*.

Kim, M., and Maguire, E.A. (2018). Hippocampus, Retrosplenial and Parahippocampal Cortices Encode Multicompartment 3D Space in a Hierarchical Manner. Cereb. Cortex 1898–1909.

Kjelstrup, K.B., Solstad, T., Brun, V.H., and Hafting, T. (2008). Finite Scale of Spatial Representation in the hippcampus. Science (80-.). 2–5.

Knox, W.B., Otto, A.R., Stone, P., and Love, B.C. (2012). The nature of belief-directed exploratory choice in human decision-making. Front. Psychol. *2*, 1–12.

Kolling, N., and O'Reilly, J.X. (2018). State-change decisions and dorsomedial prefrontal cortex : the importance of time. COBEHA *22*, 152–160.

Kolling, N., Behrens, T.E.J., Mars, R.B., and Rushworth, M.F.S. (2012). Neural mechanisms of foraging. Science *336*, 95–98.

Kolling, N., Wittmann, M.K., Behrens, T.E.J., Boorman, E.D., Mars, R.B., and Rushworth, M.F.S. (2016). Value , search , persistence and model updating in anterior cingulate cortex.

Komorowski, R.W., Manns, J.R., and Eichenbaum, H. (2009). Robust Conjunctive Item-Place Coding by Hippocampal Neurons Parallels Learning What Happens Where. J. Neurosci. *29*, 9918–9929.

Kraus, B.J., Robinson, R.J., White, J.A., Eichenbaum, H., and Hasselmo, M.E. (2013). Hippocampal "Time Cells": Time versus Path Integration. Neuron *78*, 1090–1101.

Kraus, B.J., Brandon, M.P., Robinson, R.J., Connerney, M.A., Hasselmo, M.E., and Eichenbaum, H. (2015). During Running in Place, Grid Cells Integrate Elapsed Time and Distance Run. Neuron *88*, 578–589.

Kropff, E., Carmichael, J.E., Moser, M., and Moser, E.I. (2015). Speed cells in the medial entorhinal cortex.

Krupic, J., Burgess, N., and O'Keefe, J. (2012). Neural Representations of Location Composed of Spatially Periodic Bands. Science (80-.). *337*, 853–857.

Kumaran, D., and McClelland, J.L. (2012). Generalization through the recurrent interaction of

episodic memories: A model of the hippocampal system. Psychol. Rev. *119*, 573–616.
Kumaran, D., Summerfield, J.J., Hassabis, D., and Maguire, E.A. (2009). Tracking the Emergence of Conceptual Knowledge during Human Decision Making. Neuron *63*, 889–901.
Kumaran, D., Melo, H.L., and Duzel, E. (2012). The Emergence and Representation of Knowledge about Social and Nonsocial Hierarchies. Neuron *76*, 653–666.
Kumaran, D., Hassabis, D., and Mcclelland, J.L. (2016). What Learning Systems do Intelligent Agents Need ? Complementary Learning Systems Theory Updated. Trends Cogn. Sci. *20*, 512–534.
Leutgeb, S., Leutgeb, J., Barnes, C., Moser, E., McNaughton, B., and Moser, M.-B. (2005). Independent Codes for Spatial and Episodic Memory in Hippocampal Neuronal Ensembles. Science (80-.). *309*, 619–623.
M'Harzi, M., Palacios, A., Monmaur, P., Willig, F., Houcine, O., and Delacour, J. (1987). Effects of selective lesions of Fimbria-Fornix on learning set in the rat. Physiol. Behav. *40*, 181–188.
Mack, M.L., Love, B.C., and Preston, A.R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *113*.
Mack, M.L., Preston, A.R., and Love, B.C. (2017). Medial prefrontal cortex compresses concept representations through learning. BioRxiv 1–22.
Malagon-Vina, H., Ciocchi, S., Passecker, J., Dorffner, G., and Klausberger, T. (2018). Fluid network dynamics in the prefrontal cortex during multiple strategy switching. Nat. Commun. *9*, 309.
Manns, J., and Eichenbaum, H. (2006). Evolution of Declarative Memory. Hippocampus *16*, 795–808.
Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.
Marr, D. (1971). Simple memory: a theory for archicortex. Philos. Trans. R. Soc. London *120*, 187–200.
Marr, D., and Poggio, T. (1976). From understanding computation to understanding neural circuitry.
Martins, A.R.O., and Froemke, R.C. (2015). Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. Nat. Neurosci. *18*, 1483–1492.
Mathis, A., Stemmier, M.B., and Herz, A.V.M. (2015). Probable nature of higher-dimensional symmetries underlying mammalian grid-cell activity patterns. Elife *2015*, 1–29.
McClelland, J.L. (2013). Incorporating rapid neocortical learning of new schema-consistent

information into complementary learning systems theory. J. Exp. Psychol. Gen. *142*, 1190–1210.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. *102*, 419–457.

McKenzie, S., Frank, A.J., Kinsky, N.R., Porter, B., Rivière, P.D., and Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. Neuron *83*, 202–215.

McNaughton, B.L., Barnes, C.A., and O'Keefe, J. (1983). The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. Exp. Brain Res. *52*.

McNaughton, B.L., Battaglia, F.P., Jensen, O., Moser, E.I., and Moser, M.B. (2006). Path integration and the neural basis of the "cognitive map." Nat. Rev. Neurosci. *7*, 663–678.

Van Der Meer, M., Kurth-Nelson, Z., and Redish, A.D. (2012). Information processing in decision-making systems. Neuroscientist *18*, 342–359.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. Nature *518*, 529–533.

Momennejad, I., Russek, E.M., Cheong, J.H., Botvinick, M.M., Daw, N.D., and Gershman, S.J. (2017). The successor representation in human reinforcement learning. Nat. Hum. Behav. *1*, 680–692.

Monaco, J.D., and Abbott, L.F. (2011). Modular Realignment of Entorhinal Grid Cell Activity as a Basis for Hippocampal Remapping. *31*, 9414–9425.

Morris, R.G.M., Garrud, P., Rawlins, J.N.P., and O'Keefe, J. (1982). Place Navigation Impaired in Rats with Hippocampal Lesions. Nature.

Morrissey, M.D., Insel, N., and Takehara-Nishiuchi, K. (2017). Generalizable knowledge outweighs incidental details in prefrontal ensemble code over time. Elife *6*, 1–20.

Muller, R.U., and Kubie, J.L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J. Neurosci. *7*, 1951–1968.

Muller, R.U., Bostock, E., Taube, J.S., and Kubie, J.L. (1994). On the directional firing properties of hippocampal place cells. J. Neurosci. *14*, 7235–7251.

Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., and Cai, X. (2014). A hierarchy of intrinsic timescales across primate cortex. Nat. Publ. Gr. *17*, 1661–1663.

Nakazawa, K., Quirk, M.C., Chitwood, R.A., Watanabe, M., Yeckel, M.F., Sun, L.D., Kato, A.,

Carr, C.A., Johnston, D., Wilson, M.A., et al. (2002). Requirement for Hippocampal CA3 NMDA Receptors in Associative Memory Recall. *297*, 211–219.

Nakazawa, K., Sun, L.D., Quirk, M.C., Rondi-reig, L., Wilson, M.A., and Tonegawa, S. (2003). Hippocampal CA3 NMDA Receptors Are Crucial for Memory Acquisition of One-Time Experience. *38*, 305–315.

Nassar, M.R., Rumsey, K.M., Wilson, R.C., Parikh, K., Heasly, B., and Gold, J.I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. Nat. Neurosci. *15*, 1040–1046.

Neves, G., Cooke, S.F., and Bliss, T. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. Nat. Rev. Neurosci. *9*, 1030–1034.

Nicolle, A., Klein-Flügge, M.C., Hunt, L.T., Vlaev, I., Dolan, R.J., and Behrens, T.E.J. (2012). An Agent Independent Axis for Executed and Modeled Choice in Medial Prefrontal Cortex. Neuron *75*, 1114–1121.

Noonan, M.P., Walton, M.E., Behrens, T.E.J., Sallet, J., Buckley, M.J., and Rushworth, M.F.S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proc. Natl. Acad. Sci. *107*, 20547–20552.

O'Keefe, J., and 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., and Nadel, L. (1978). The hippocampus as a cognitive map.

O'Reilly, Jbabdi, S., Rushworth, M.F.S., and Behrens, T.E.J. (2013a). Brain Systems for Probabilistic and Dynamic Prediction : Computational Specificity and Integration. *11*.

O'Reilly, J.X., Schuffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., and Rushworth, M.F.S. (2013b). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc. Natl. Acad. Sci. *110*, E3660–E3669.

Ólafsdóttir, F.H., Barry, C., Saleem, A.B., Hassabis, D., and Spiers, H.J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. Elife *4*, 1–17.

Ólafsdóttir, H.F., Carpenter, F., and Barry, C. (2016). Coordinated grid and place cell replay during rest. Nat. Neurosci. *19*, 1–6.

Omer, D.B., Maimon, S.R., Las, L., and Ulanovsky, N. (2018). Social place-cells in the bat hippocampus. Science (80-.). *359*, 218–224.

Packard, M.G., and Knowlton, B.J. (2002). Learning and Memory Functions of the Basal Ganglia. Annu. Rev. Neurosci. *25*, 563–593.

Packard, M.G., and McGaugh, J.L. (1996). Inactivation of Hippocampus or Caudate Nucleus with Lidocaine Differentially Affects Expression of Place and Response Learning. *72*, 65–72.

Pavlov, I.P. (1927). Conditional reflexes: an investigation of the physiological activity of the cerebral cortex.

Pickens, C.L., Saddoris, M.P., Setlow, B., Gallagher, M., Holland, P.C., and Schoenbaum, G. (2003). Different Roles for Orbitofrontal Cortex and Basolateral Amygdala in a Reinforcer Devaluation Task. *23*, 11078–11084.

Preuschoff, K., 't Hart, B.M., and Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. Front. Neurosci. *5*, 1–12.

Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.

Rescorla, R., and Wagner, A. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement.

Rolls, E.T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. Front. Syst. Neurosci. *7*, 1–21.

Rolls, E.T., Sienkiewicz, Z.J., and Yaxley, S. (1989). Hunger Modulates the Responses to Gustatory Stimuli of Single Neurons in the Caudolateral Orbitofrontal Cortex of the Macaque Monkey. *I*.

Santoro, A., Raposo, D., Barrett, D.G.T., Malinowski, M., Pascanu, R., Battaglia, P., and Lillicrap, T. (2017). A simple neural network module for relational reasoning.

Saxe, A.M., McClelland, J.L., and Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. 1–22.

Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B., and Botvinick, M.M. (2013). Neural representations of events arise from temporal community structure. Nat. Neurosci. *16*, 486–492.

Schuck, N.W., Gaschler, R., Wenke, D., Heinzle, J., Frensch, P.A., Haynes, J.D., and Reverberi, C. (2015). Medial prefrontal cortex predicts internally driven strategy shifts. Neuron *86*, 331–340.

Schuck, N.W., Cai, M.B., Wilson, R.C., and Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. Neuron *91*, 1402–1412.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A Neural Substrate of Prediction and Reward. Science (80-.). *275*, 1593–1599.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatry *87*, 230.

Shenhav, A., Botvinick, M.M., and Cohen, J.D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron *79*, 217–240.

Shenhav, A., Straccia, M.A., Cohen, J.D., and Botvinick, M.M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. Nat. Neurosci. *17*, 1249–1254.

Shenhav, A., Cohen, J.D., and Botvinick, M. (2016). Dorsal anterior cingulate cortex and the value of control. Nat. Neurosci. *19*, 1280–1285.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature *529*, 484–489.

Singer, Y., Teramoto, Y., Willmore, B.D.B., Schnupp, J.W.H., King, A.J., and Harper, N.S. (2018). Sensory cortex is optimized for prediction of future input. Elife 1–31.

Solstad, T., Boccara, C.N., Kropff, E., Moser, M.-B., and Moser, E.I. (2008). Representation of geometric borders in the entorhinal cortex. Science *322*, 1865–1868.

Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A.G., Niv, Y., and Botvinick, M.M. (2014). Optimal Behavioral Hierarchy. PLoS Comput. Biol. *10*.

Song, A.H., Kucyi, A., Napadow, V., Brown, E.N., Loggia, M.L., and Akeju, O. (2017). Pharmacological Modulation of Noradrenergic Arousal Circuitry Disrupts Functional Connectivity of the Locus Ceruleus in Humans. J. Neurosci. *37*, 6938–6945.

Spiers, H.J., Hayman, R.M.A., Jovalekic, A., Marozzi, E., and Jeffery, K.J. (2015). Place field repetition and purely local remapping in a multicompartment environment. Cereb. Cortex *25*, 10–25.

Squire, L.R., Kosslyn, S., Zola-Morgan, S., Haist, F., and Musen, G. (1992). Memory and the Hippocampus: A Synthesis From Findings With Rats, Monkeys, and Humans. Psychol. Rev. *99*, 195–231.

Squire, L.R., Stark, C.E.L., and Clark, R.E. (2004). THE MEDIAL TEMPORAL LOBE. Annu. Rev. Neurosci.

Stachenfeld, K.L., Botvinick, M.M., and Gershman, S.J. (2017). The hippocampus as a predictive map. Nat. Neurosci. *20*, 1643–1653.

Stemmler, M., Mathis, A., and Herz, A.V.M. (2015). 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., and Moser, E.I. (2012). The entorhinal grid map is discretized. Nature *492*, 72–78.

Strange, B.A., Duggins, A., Penny, W., Dolan, R.J., and Friston, K.J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? Neural Networks *18*, 225–230.

Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. Nat. Neurosci. *18*, 1025–1033.

Sutton, R., and Barto, A. (1998). REINFORCEMENT LEARNING : an introduction.

Taube, J.S., Muller, R.U., and Ranck, J.B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J. Neurosci. *10*, 420–435.

Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., and Schiller, D. (2015). A Map for Social Navigation in the Human Brain. Neuron *87*, 231–243.

Tervo, D.G.R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., and Karpova, A.Y. (2014). Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. Cell *159*, 21–32.

Thorndike, E. (1911). Animal Intelligence: Experimental Studies.

Tolman, E.C. (1948). Cognitive maps in rats and men. Psychol. Rev. *55*, 189–208.

Tolman, E.C. (1949). There is more than one kind of learning. Psychol. Rev.

Tolman, E., and Honzik, C. (1930). Introduction and removal of reward, and maze performance in rats. Univ. Calif. Publ. Psychol. *4*, 257–275.

Tolman, E.C., Ritchie, B.F., and Kalish, D. (1946). Studies in spatial learning. I. Orientation and the short-cut. J. Exp. Psychol. *36*, 13–24.

Treves, A., and Rolls, E.T. (1991). What determines the capacity of autoassociative memories in the brain? Netw. Comput. Neural Syst. *2*, 371–397.

Treves, A., and Rolls, E.T. (1994). Computational analysis of the role of the hippocampus in memory. Hippocampus *4*, 374–391.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J.J., Moser, M., and Moser, E.I. (2018). Integrating time from experience in the lateral entorhinal cortex.

Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P., and Morris, R.G.M. (2007). Schemas and Memory Consolidation TL - 316. Science (80-.). *316 VN-*, 76–82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., and Morris, R.G.M. (2011). Schema-Dependent Gene Activation. Science (80-.). *333*, 891–895.

Usher, M., and McClelland, J.L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. Psychol. Rev. *108*, 550–592.

Vikbladh, O., Michael, M., John, K., Blackmon, K., Orrin, D., Daphna, S., Burgess, N., and Daw, N.D. (2018). Two Sides of the Same Coin: The Hippocampus as a Common Neural Substrate

for Model-Based Planning and Spatial Memory. Submitted.
Walton, M.E., Devlin, J.T., and Rushworth, M.F.S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. Nat. Neurosci. *7*, 1259–1265.
Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rudebeck, P.H., and Rushworth, M.F.S. (2010). Separable Learning Systems in the Macaque Brain and the Role of Orbitofrontal Cortex in Contingent Learning. Neuron *65*, 927–939.
Wang, J.X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J.Z., Hassabis, D., and Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. Nat. Neurosci. *21*, 860–868.
Wayne, G., Hung, C., Amos, D., Mirza, M., Ahuja, A., Grabska-barwi, A., Rae, J., Mirowski, P., Leibo, J.Z., Santoro, A., et al. (2018). Unsupervised Predictive Memory in a Goal-Directed Agent. ArXiv *1803* . *107*.
Wei, X., Prentice, J., and Balasubramanian, V. (2015). A principle of economy predicts the functional architecture of grid cells. 1–29.
Whittington, J.C.R., Muller, T.H., Barry, C., and Behrens, T.E.J. (2018). Generalisation of structural knowledge in the Hippocampal-Entorhinal system.
Wikenheiser, A.M., and Schoenbaum, G. (2016). Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex. Nat. Rev. Neurosci. *17*, 513–523.
Wikenheiser, A.M., Marrero-Garcia, Y., and Schoenbaum, G. (2017). Suppression of Ventral Hippocampal Output Impairs Integrated Orbitofrontal Encoding of Task Structure. Neuron *95*, 1197–1207.e3.
Willshaw, D.J., Dayan, P., and Morris, R.G.M. (2015). Memory, modelling and Marr: a commentary on Marr (1971) 'Simple memory: a theory of archicortex.' Mem. Model. Marr a Comment. Marr *370*, 20140383.
Wilson, M., and McNaughton, B.L. (1993a). Reactivation of Hippocampal Ensemble Memories During Sleep. Science (80-.). *5*, 14–17.
Wilson, M.A., and McNaughton, B.L. (1993b). Dynamics of the hippocampal ensemble code for space. Science (80-.). *261*, 1055–1058.
Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. Neuron *81*, 267–278.
Wimmer, G.E., and Shohamy, D. (2012). Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. Science (80-.). *338*, 270–273.
Witter, M.P., and Amaral, D.G. (2004). Hippocampal Formation. In The Rat Nervous System,

(Elsevier), pp. 635–704.

Wood, E.R., Dudchenko, P.A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. Nature *397*, 613–616.

Xie, J., and Padoa-Schioppa, C. (2016). Neuronal remapping and circuit persistence in economic decisions. Nat. Neurosci. *19*, 855–861.

Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. *19*, 356–365.

Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. *7*, 464–476.

Yu, A.J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681–692.

Yu, J.Y., Liu, D.F., Loback, A., Grossrubatscher, I., and Frank, L.M. (2018). Specific hippocampal representations are linked to generalized cortical representations in memory. Nat. Commun. 1–11.

Zeithamova, D., and Preston, A.R. (2010). Flexible Memories: Differential Roles for Medial Temporal Lobe and Prefrontal Cortex in Cross-Episode Binding. J. Neurosci. *30*, 14676–14684.