Frameworks for generalisation of flexible behaviour: insights from mouse behaviour and neural computation



Veronika Samborska

St Catherine's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

Борітеся — поборете!

Т.Г. Шевченко

Acknowledgements

People get into science for all sorts of reasons. Looking back now, for me, it feels like almost all of it was about the people I got to hang out with.

First and for most, I am grateful to my supervisors who have managed to tolerate me for the past 5 years. It's really by far a much bigger achievement than this thesis. Tim, who did the greatest job of managing the explosions, and thanks to whom it feels like no other job I ever do in my life will ever be enough now. All those days trying to figure out what was happening in the data together felt like we were kids disappearing into another world, and I already regularly miss and crave this feeling. Mark, who has always been there for me when I most needed it, stayed calm, kind, and incredibly supportive even when it seemed impossible that anyone actually could. He also somehow always managed to say just the right thing. Thomas, for taking a chance on me when there was little reason to believe it would pay off, and for teaching me how to transfer soldering skills from mouse boxes to speakers. And David Bannerman, who wasn't my supervisor, but whose sincerity, kindness, and passion for science have impacted every student in Oxford I know who has met him, including myself. If it wasn't for him, I probably would have never started this DPhil.

I got extremely lucky to have been surrounded by many friends who made my DPhil years special. Firstly, thanks to the "lads lads lads" from Mark's lab, otherwise known as *Pussies Pussies Pussies* - Lauren, Marios, TJP, Laura, TB2, and an affiliate member of the lab - Andi. Given the particularities of our group dynamics it's probably not a good idea for me to mention too many specifics of what I am thankful for here. So, I'll save these for when some of you become big-deal PIs. But all the 'holibobs', each treated with utmost scientific approach, all the hard bass and slav squats, too many late nights in Hi Lo, and the nearly daily pub trips, were as integral to my DPhil experience as the actual research I got to do. I wish there was an option to keep doing science with all of you around in the same place forever.

I was also lucky to be around Tim's lab when it was full of absolute superstars with only slightly more reasonable drinking habits (on average) – Tim M, Alon, Fileeeeep, Shirley, Xiao, and even James W, who are all by far the most interesting people to chat about science with, or in fact anything else. Also, Yves who made my final BSB days somewhat

manageable by sharing a disregard for authority and bureaucracy. Jimmy, who would easily win 'the most British person' award, and Arturo, who was my only non-neuroscience friend in Oxford and often suffered the consequences (but who understood more than anyone what it was like to be surrounded by British people). Gioia, whose daughter features in this thesis and who I know will grow up to be just as awesome as her mom. Anna, for providing me with a regular supply of jamón, energy, and sharing the love of pork baps. Thank you all for being your incredible selves!

The hardest part of this DPhil was living through more historical events than I would have liked. During the past 550 days of the war (at time of writing), it would be impossible for me to remain sane without the support of Victoria, Ivan, Oleg, and Sasha. The war turned out to have a lot of indirect effects on many relationships in my life, but I am in no way surprised that it only served to bring me closer to the people above (even if not in distance). Luckily, you are also the best people with whom to celebrate Ukraine's inevitable victory. Thank you for sharing the vision of what real friendship means with me.

To Dayyyyyyviiiiiid. You are okay, I guess...Thank you for never giving up on our little world, soon to be a world of 三人

Finally, my parents and my best friends. My Dad taught me that there are no closed doors in life, and if you find one that's hard to open, you can always just kick it down or try the window. This thesis is largely a result of that lesson. So, this is one is for you, Dad.

List of abbreviations

- **1D, 2D, *n*D** → One-dimensional, two-dimensional, *n*-dimensional
- **AI** → Artificial intelligence
- **ANOVA** → Analysis of variance
- **ANN** → Artificial neural network
- **BCI** → Brain-computer interface
- **BIC** → Bayesian information criterion
- **CA1** → Hippocampal region CA1
- **CA3** → Hippocampal region CA3
- **CNN** → Convolutional neural network
- **CPD** → Coefficient of partial determination
- **dCA1** → Hippocampal region CA1 (dorsal)
- **EC** → Entorhinal cortex
- **EWC** → Elastic weight consolidation
- **fMRI** → Functional magnetic resonance imaging
- **HPC** → Hippocampus
- **ITI** → Inter-trial interval
- **LEC** → Lateral entorhinal cortex
- **MEC** → Medial entorhinal cortex
- **ML** → Machine learning
- **mPFC** → Median prefrontal cortex
- **OFC** → Orbitofrontal cortex
- **PCA** → Principal component analysis

- **PFC** → Prefrontal cortex
- **RDM** → Representational similarity design matrices
- **RL** → Reinforcement learning
- **RNN** → Recurrent neural network
- **RSA** → Representational similarity analysis
- **SR** → Successor representation
- **SSE** → Sum of squares error
- **SVD** → Singular value decomposition
- **TEM** → Tolman-Eichenbaum machine

Notation

- **1D, 2D, *n*D** $\rightarrow$ One-dimensional, two-dimensional, *n*-dimensional
- $x \rightarrow$ Random variable
- $\mathbf{x} \rightarrow$ Vector of variables x_{i}
- $n! \rightarrow$ Factorial of integer *n*
- $\mathbb{R}^2 \rightarrow$ Set of all ordered pairs of real numbers lying on the Cartesian plane
- $\mathbb{E} \rightarrow$ Expectation
- **v[s]** $\rightarrow$ Value function (here expressed as *value of state* **s**)
- $\pi \rightarrow$ Policy
- $r_{(i,j)} \rightarrow$ Elements of RSA matrix
- $RDM_{n(i,j)} \rightarrow$ Elements of the n^{th} RDM matrix
- $s\left(x_{ij}\right) \rightarrow$ Surprise metric for the given neuron on trial *i* in task layout *j*
- $\mu_k \rightarrow$ Baseline mean firing rate of the given neuron in task configuration *k*
- $\sigma_k \rightarrow$ Baseline standard deviation of firing rate of the given neuron in task configuration *k*
- $D \rightarrow$ Data matrix
- U and $V^T \rightarrow$ Singular vectors of D
- $\Sigma \rightarrow$ Diagonal (weight) matrix of D
- $\varphi \rightarrow$ Radial basis function

Equations

- **Chapter 4**
- *Equation 4.1* → Singular value decomposition
- *Equation 4.2* → Matrix representation of task 2 data matrix onto task 1 temporal modes
- *Equation 4.3* → Matrix representation of variance explained in task 2 using cellular modes of task 1
- *Equation 4.4* → Representation of the projection of data set D2 onto cellular and temporal modes of task 1
- **Chapter 7**
- *Equation 7.1* → Linearly combining representational dissimilarity matrices
- *Equation 7.2* → Coefficient of partial determination
- *Equation 7.3* → Calculation of the 'surprise' metric
- *Equation 7.4* → Generalised singular value decomposition (SVD)
- *Equation 7.5* → Eigendecomposition of the covariances between neurons
- *Equation 7.6* → Eigendecomposition of the covariances between timepoints
- *Equation 7.7* → Decomposition of a task into first-half and second-half data matrices
- *Equation 7.8* → Quantifying the generalization of temporal modes across problems
- *Equation 7.9* → Quantifying within-problem variance explained - temporal modes
- *Equation 7.10* → Quantifying cross-problem variance explained - cellular modes
- *Equation 7.11* → Quantifying within-problem variance explained - cellular modes
- *Equation 7.12* → Quantifying cross-problem variance explained - temporal modes
- *Equation 7.13* → Quantifying within-problem variance explained - overall

Abstract

It would be a daunting task to learn everything from scratch each time we are faced with a new problem. Instead, humans and other animals are helped by the fact that few situations in life are truly novel. Thus, we can draw upon previous experiences of similar situations and come up with our 'best guess' of how to deal with a new task in hand. In this thesis, I will attempt to answer the question of how this ability to generalise knowledge between different situations might be achieved in the brain.

To be able to apply previously learnt knowledge to new but related situations, the common structure needs to be abstracted away from the sensorimotor specifics of experience. In my first set of experiments, I showed that mice, like humans, can learn to transfer knowledge between different sensory scenarios, and get better at solving each new sensory example of a problem. A series of behavioural analyses I conducted suggest that animals can abstract strategies away from the specifics on various time scales and could not only learn to generalise the short-term abstract sequences of events, but also a long-term strategy that integrated the recent history of animals' experiences.

In the second set of experiments, I focused on how these abstractions might be achieved in the brain, and how abstract representations might be tied to the sensorimotor specifics of each new situation. I found that neurons in medial prefrontal cortex (mPFC) maintained similar representations across multiple problems, despite their different sensorimotor correlates, whereas hippocampal (dCA1) representations were more strongly influenced by the specifics of each task. These data suggest that mPFC and hippocampus play complementary roles in generalisation of knowledge, with the former abstracting the common structure among related tasks, and the latter mapping this structure onto the specifics of the current situation.

Together, these experiments aim to develop our understanding of a hallmark of biological intelligence – how prior experience is leveraged to solve novel, real-world problems.

Summary of contributions

I declare that this thesis is my own and that attribution has been given where information was derived from other sources. Aside from the exceptions outlined below, I conceptualised the project, designed and assembled the equipment, performed the experiments, analysed the data and constructed all of the figures. While all writing contained herein is my own, the project as a whole has relied on key contributions from others. Throughout, the text uses the pronoun 'we' indicating myself and the team below.

Thomas Akam : Conceptualisation of the project and performed most of the surgeries. **James Butler**: Performed ~10% of the surgeries. **Mark Walton and Tim Behrens**: Conceptualisation of the project and supervision throughout.

The work presented in this thesis has also been published in *Nature Neuroscience*:

Samborska, V., Butler, J.L., Walton, M.E. et al. Complementary task representations in hippocampus and prefrontal cortex for generalizing the structure of problems. Nat Neurosci 25, 1314–1326 (2022). https://doi.org/10.1038/s41593-022-01149-8

Data from all of the experiments in this thesis are freely available to download at https://doi.org/10.6084/m9.figshare.19773334. Code for implementing all of the analyses is available at https://github.com/veronikasamborska1994/notebooks_paper.

Contents

List of Figures **15**

1 General Introduction **17**

1.1 In mice, monkeys, and men 18

"Bau-bau", how a dog barks in Italian 18

Why rats deserve more credit 20

1.2 In the brain 22

Half a brain can be better than one 22

Inspirations from spatial cognition 23

Generalising space 24

Hippocampus and non-spatial reasoning 26

War of Ghosts: schemata in hippocampus and prefrontal cortex 28

1.3 In machines 30

The anatomy of a cognitive map in artificial intelligence agents 31

1.4 Summary 38

2 Abstraction and the underlying behavioural strategies in mice **39**

2.1 Introduction 39

2.2 Behavioural Task 41

2.3 Mice generalise within-trial knowledge between structurally equivalent problems 42

2.4 Mice generalise long-term strategies between structurally equivalent problems 44

2.5 Discussion 46

3 Neural representations underlying generalisation of trial events **49**

3.1 Introduction 49

3.2 Results 51

Recording in mPFC and hippocampus as mice solve multiple problems . 51

Abstract and task-specific representations of trial events by PFC and CA1 units 53

PFC population activity generalises more strongly across tasks than CA1 55

3.3 Discussion 62

4 Generalisation of low dimensional representations **65**

4.1 Introduction 65

4.2 Results 68

Studying generalization of low-dimensional representations using SVD . 68

Temporal structure of activity is invariant across tasks and regions . . . 69

Cell assemblies generalise more strongly in PFC than CA1 72

4.3 Discussion 73

5 Neural representations underlying generalisation of policy **76**

5.1 Introduction 76

5.2 Results 78

5.3 Discussion 81

6 General Discussion **84**

6.1 Summary 84

6.2 Future directions 85

6.3 Outro 88

7 Materials and Methods **91**

7.1 Behavioural apparatus 91

7.2 Subjects 91

7.3 Behavioural training 92

7.4 Behavioural training during recordings 95

7.5 Electrophysiological recordings and spike sorting 95

7.6 Surgery and histology 96
7.7 Data analysis 97
7.8 Time-in-trial alignment 97
7.9 Statistical significance 98
7.10 Representational similarity analysis 99
7.11 Decoding analyses 100
7.12 Surprise Measure 101
7.13 Singular Value Decomposition 102
7.14 Estimating policy 107
7.15 Additional controls for physical movement 107

List of Figures

1 **Chihuahua vs Muffin problem.** 19

2 **Kasparov (white) vs. Deep Blue (black): Game 1, 1997** 31

3 **Transfer learning in mice** 42

4 **Behavioural generalisation of trial structure.** 44

5 **Behavioural generalisation of long-term strategies.** 46

6 **Recording units across multiple problems in a single session and behaviour during recordings.** 52

7 **Example neurons in behavioural task.** 54

8 **Example neurons in physical space and behavioral task.** 56

9 **Port selectivity in CA1 than PFC.** 57

10 **PFC population activity generalises more strongly across tasks than CA1.** 59

11 **Rapid problem-induced 'remapping' in CA1 but not PFC.** 61

12 **Decoding of trial stages across tasks.** 62

13 **Singular Value Decomposition (SVD).** 68

14 **Generalisation of low-dimensional activity patterns of trial events.** . 71

15 **Policy Generalisation in PFC and CA1.** 80

16 **Constructing rich conceptual knowledge from simpler primitives.** . . 87

17 **Counterbalancing and port layouts used throughout the experiments.** 94

18 **Progression of Training, Surgery, and Electrophysiological Recordings.** 95

19 **Trial alignment of activity. Diagram illustrating alignment of spike activity across trials.** 98
20 **Significance of key differences between CA1 and PFC assessed by permuting subjects between regions.** 99
21 **Additional analyses of low-dimensional structure of activity in PFC and CA1.** 106
22 **Fine-grained movement related activity controls.** 110

1

General Introduction

Contents

1.1 In mice, monkeys, and men **18**
"Bau-bau", how a dog barks in Italian 18
Why rats deserve more credit 20
1.2 In the brain **22**
Half a brain can be better than one 22
Inspirations from spatial cognition 23
Generalising space 24
Hippocampus and non-spatial reasoning 26
War of Ghosts: schemata in hippocampus and prefrontal cortex . . . 28
1.3 In machines **30**
The anatomy of a cognitive map in artificial intelligence agents . . . 31
1.4 Summary **38**

As we step into a new restaurant, our minds can effortlessly predict the sequence of events that is about to unfold. We confidently scan the room for an empty table, aware that we will need to be patient and wait for our server. Our brains have already mapped out the progression of the meal, from the appetiser to the dessert, and we instinctively know that the check will arrive when we're finished. This is possible because we have experienced these sequences of events in countless restaurants before, which has equipped us with a general mental map of how all restaurants work.

Similar to how we adeptly handle new dining experiences, we regularly confront and overcome numerous novel challenges in life. Whether it's learning a new language, solving a mathematical problem, or adapting to a new environment, we have a remarkable ability to draw upon our prior knowledge and experiences to make sense of new situations. But how is this ability to generalise knowledge enabled by the brain?

In recent years, scientists have made significant strides in understanding how the brain supports this process. Insights have been drawn from various disciplines – ranging from psychology[1] to neural basis of navigation in space[2], and studies have implicated numerous brain regions[3,4] in generalising knowledge across different situations. More recently, the topic has also been gaining increased attention from computer scientists working on artificial intelligence (AI), as - despite remarkable advances in specific tasks[5,6] - AI typically fails catastrophically when prompted to generalise information across different domains and scenarios. The remainder of this introduction explores these strands of literature in more detail. The main aim here is to provide a rationale highlighting why trying to understand the neural basis for generalising knowledge is important, and to contextualise the empirical work presented in this thesis.

1.1 In mice, monkeys, and men

> The obligation of the theoretical psychologist is to discover general laws of behavior applicable to mice, monkeys and men.
>
> *Harry F. Harlow, 1948*[7]

"Bau-bau", how a dog barks in Italian

My friend's 1-year-old daughter Alice is a big fan of dogs. Every time they pass a dog, she'd point at it and call out "Bau-bau!". Every time – except when they pass a chihuahua. It's hard to blame her for making this mistake (for many reasons). And, in fact, state-of-the art machine learning approaches also struggle with chihuahuas. The key difference, however, is that unlike AI agents, Alice would never think it's a muffin (Figure

2). Even though chihuahuas clearly do not fit her concept of what a dog is, she is much more likely to think it's a cat, or maybe a completely different type of an animal that she hasn't encountered before.

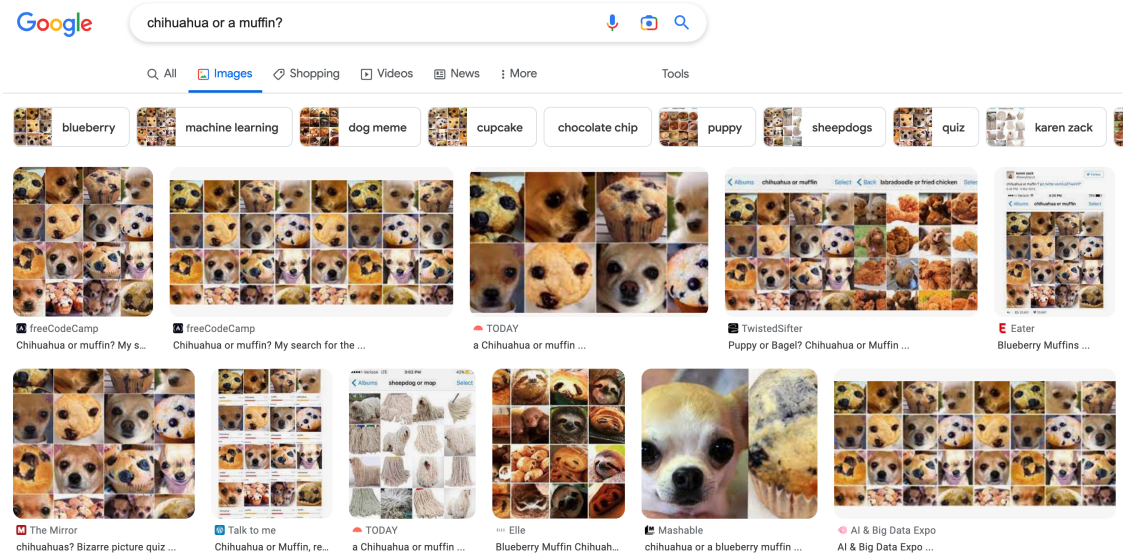


Figure 1: Chihuahua vs Muffin problem. Due to a large overlap in perceptual similarity, deep learning models (convolutional neural networks, to be specific) often misclassify chihuahuas as muffins.

The reason for this is that humans don't perceive objects simply as statistical patterns of sensory qualities. From an early age, we actively engage with the world around us and view objects as entities with spatial and causal relationships[8]. Infants as young as two months old have a basic understanding of physics, anticipating how objects will move and behave[9,10]. Similarly, even before they can speak, babies possess intuitive psychological concepts, distinguishing between animate and inanimate objects[11].

While these very early abilities may rely on recognizing low-level statistical patterns in sensory features (e.g. humans and animals have eyes; inanimate objects usually don't), by 13 months old, children show a remarkable flexibility in their thinking and can easily modify their beliefs to draw new conclusions. For example, if a child is taught about a certain behaviour of a typical bird, they might at first assume that this behaviour applies to visually similar animals like pterodactyls but not to atypical birds like dodos. However, if an adult then labels these creatures (pterodactyl as a 'dinosaur' and dodo as a 'bird'), the child will be more likely to infer behaviours characteristic of the birds to the dodo, labelled as a 'bird', rather than the pterodactyl, labelled as a 'dinosaur'[12].

This capacity to flexibly incorporate new information to draw novel conclusions has undeniable evolutionary motivations for all animals. We don't just need to generalise information but must do so flexibly and appropriately. Getting sick after eating one type of mushroom shouldn't prevent you from eating mushrooms altogether. Instead, you should extract the key features of the poisonous mushroom that might have caused your illness and only generalise and avoid these features in other mushrooms (e.g. foul smell). How animals achieve this high-level generalisation remains to be relatively unknown, but the first insights were gained from controlled psychological experiments.

Why rats deserve more credit

Harry F. Harlow was one of the first psychologists who set out to study how humans and other primates can grasp complex concepts and apply them to new situations in a laboratory environment. He devised a clever experiment where he presented monkeys with pairs of objects, only one of which led to a reward. But here's the twist: the pairs of objects kept changing, while the abstract rule remained the same (win - stay, lose - switch). At first, the monkeys struggled to figure out which object to choose in a new pair. But as they encountered more and more pairs, they began to pick up on the abstract rule - only one object was rewarded. By the end of the training, monkeys had developed a remarkable ability to perform single-shot inference. On the first exposure to a new pair of objects, if the monkey chose a winning shape, it would stick to it on the next trial, and if it failed to get a reward they would switch and choose the other object. Harlow also showed that in scenarios where 'macaroni beads and toys are substituted for food rewards, and the subjects are tested sans iron-barred cages', children aged between one and five years old showed similar improvement in learning across different problems. He termed this ability a 'learning set'[7].

For a while, researchers believed that the ability of animals to perform learning set formation tasks was linked to their general intelligence level[13]. Monkeys were considered superior to cats, while cats were thought to be better than rats and squirrels. However, this view was later challenged for not considering the unique traits and abilities of each animal. One of the first to question this notion were Kay and Oldfield-Box (1963)[14],

who noted that, unlike primates, rats do not preferentially use vision to interpret their environment. They created a situation in which rats could not only see the objects but also touch and move them. Although rats, on average, did not demonstrate a level of mastery of this task that was comparable to single-shot inference that Harlow observed in monkeys, they nevertheless learned to generalise the win - stay, lose - switch strategy significantly faster than the rats in previous experiments that used two-dimensional shapes that animals weren't able to interact with (this idea is also related to the concept of *embodied cognition*[15]).

Another study that further confronted the prevailing belief of rats' learning set formation inferiority was conducted by Slotnick and Katz (1974)[16]. They approached this question by utilising a task in a sensory domain that rats are proficient with and trained them on a series of olfactory discrimination tasks, in which rats were presented with two odours but only one was rewarded. They showed that the animals' performance improved with each new pair odours and their performance was comparable to that achieved by primates in response to visual stimuli. These findings were later replicated not just with odours[17] but also with spatial cues[18]. However, these studies often overlook crucial controls, such as accounting for the effects of simple exposure to the experimental setup, or ensuring that the discrimination is not influenced by environmental features other than the odours (e.g., conducting control tests where the same odour is presented in both rewarded and non-rewarded scenarios). For instance, Reid and Morris (1992) demonstrated that animals exposed to a transfer task, without repeated training on multiple examples of a problem but with prior experience of the apparatus, showed similar performance on a new related problem as those with preceding experience on analogous tasks[19].

More recently, research showed that rats may have the capacity not only to learn abstract rules in one sensory domain but also to transfer that knowledge to other sensory domains. In a study by Murphy et al. 2008[20], rats were trained to associate a particular sequence of light and dark with a reward, and they quickly learned to choose the correct pattern. When presented with a completely different type of sensory stimulus - a sequence of auditory tones with the same structure rats appeared to show a preference for a sequence of

auditory tones that was consistent with the abstract sequence in the previously experienced light-and-dark task.

These findings together suggest that through diverse learning experiences animals, from rats to humans, possess a ubiquitous ability to develop general strategies, and these can influence how well and how fast they learn about the world around them.

1.2 In the brain

Half a brain can be better than one

Let us return to the middle of the 20th century for a moment, when Harry Harlow was conducting his learning set experiments in primates. Harlow was not only interested in the behavioural aspects of how animals generalise knowledge, but also in how the brain supported this process. Lacking the sophisticated technologies of today, he had to rely on good, old-fashioned surgical removal of parts of the brain, along with observations, to come to his conclusions.

In his experiments, Harlow removed half of the brain (a procedure known as a *hemispherectomy*) in monkeys that were well-trained on learning set tasks, and observed their performance in comparison to novice monkeys. He found that despite having half of their brain removed, the trained monkeys were still able to perform better than the untrained monkeys and concluded that "half a brain is better than one if you compare the individuals having appropriate learning sets with the individuals lacking them"[7]. This remarkable ability of animals to perform better than naive animals in new tasks that resemble those that they have encountered before, even after experiencing such severe brain damage, didn't just underscore the tremendous advantage that our ability to generalise knowledge provides for our capacity to make flexible decisions, but also pointed to a possibility that perhaps no brain area was solely responsible for this ability.

These findings were consistent with the contemporary prominent theory of 'mass action' developed by Karl Lashley in the 1920s and 1930s[21], which suggested that certain types of cognitive functions might be distributed throughout the brain and that no distinct

area of the brain is solely responsible for their execution (the principle of 'functional specialization'). Lashley's theory was based on his most famous experiment in which he gradually removed different parts of the cortex of rats that were previously trained to navigate a maze, and found that their performance was not affected by the removal of a particular part of the brain, but instead by the extent of damage to the brain as a whole. Though it's important to note that his ablations were strictly limited to cortical areas and did not involve non-cortical regions, including the hippocampus.

Around the same time as Harlow was studying learning sets, the concept of the 'mass action' theory was about to be scrutinised. Brenda Milner was working with Henry Molaison (or H.M. as he was known when he was alive), who had undergone surgery resulting in a relatively localised damage to his hippocampus[22,23]. Tragically, after the surgery, H.M. suffered from severe anterograde and partial retrograde amnesia[24,25], which lasted until his death in 2008. Although H.M.'s long-term memory and language skills were mostly unaffected during his lifetime[26], he was unable to create new episodic memories. H.M.'s case provided evidence that memory was not as localised as Lashley believed, and demonstrated the importance of the hippocampus in creating new memories.

Inspirations from spatial cognition

Inspired by the captivating case of H.M. and others like it[25], animal researchers pursued electrophysiological studies to examine memory encoding and storage within the hippocampus. O'Keefe and Dostrovsky (1971)[27] conducted experiments on single neurons in the rat hippocampus, discovering that many of these neurons responded specifically to the animal's spatial location within its environment. They coined the term 'place cells' for these neurons and suggested that they could collectively act as an animal's cognitive map - a mental representation of the environment that allows for flexible navigation and decision-making.

The concept of 'cognitive maps' was initially proposed by Edward Tolman in 1948[28] who, after a series of behavioural experiments, concluded that when rats navigate their environment, they learn more than just simple stimulus-response relationships. Instead,

rats displayed a remarkable flexibility in their behaviour and would promptly devise shortcut strategies to reach rewards[29] or rapidly identify new efficient routes if their established paths were blocked[30]. The discovery of place cells provided a potential neural basis for the cognitive maps that Tolman had theorised, thereby bridging the gap between his behavioural findings and the underlying neural mechanisms.

Though Tolman's experiments focused on spatial relationships, he proposed that animals might organise their knowledge more broadly using similar relational cognitive maps. Still, the discovery of place cells has predominantly guided the field towards studying cognitive maps in spatial contexts. When neuroscientists recorded neurons one synapse upstream, they identified medial entorhinal grid cells that fired at locations on a triangular lattice[31]. Other notable yet lesser-known discoveries include cells encoding the animal's current head direction[32], cells encoding vectors from the animal to rewards[33], goals[34], borders[35], objects[36], and even the locations of other animals in social contexts[37,38].

More recently, the same brain structures containing place and grid cells have also been shown to contribute to various neural processes related to a broader view of a cognitive map, including memory, imagination, social cognition, generalisation, and inference[39–42]. Thus, the challenge lies in understanding how these cells might contribute to organising knowledge in the complex, high-dimensional, and non-spatial cognitive map proposed by Tolman.

Generalising space

Spatial settings, like any other aspect of the world, contain repetitions and shared features across different environments. For instance, consider the simple example of walking in a circle around a city block. As you navigate the block, you encounter various landmarks, such as buildings, trees, or signs. As you complete the loop, you find yourself back at your starting point, having traversed a familiar path. This repetitiveness and familiarity allow you to build a mental map of the area that generalises to other similar environments[43]. If you were to walk around a different city block, the general spatial structure would be comparable – you would still be walking in a circle, encountering landmarks, and

ultimately returning to your starting point. Such shared features can therefore simplify the process of navigation and enable animals to capitalise on common aspects when faced with new or altered surroundings.

Although research on how neural representations of space in the brain might facilitate the generalization of spatial information across diverse environments has been limited, some intriguing observations have emerged from studies on place and grid cells. Early research indicated that place cells could generalise their firing patterns to new environments, but the accuracy of the spatial representation diminished with increasing environmental distortion[44,45]. The question remained whether the generalization of place cells stemmed from a geometrically based, metric representation of space or relied on higher-level features such as landmarks or textures[46].

Leutgeb et al. (2005)[47] aimed to address this question by designing an experiment where rats were trained to navigate a square arena with distinct wall patterns. When the patterns were altered, rats displayed a reduction in place cell stability, but only if the pattern change resulted in a change in the metric relationship between the arena walls. This finding suggests that place cells encode a metric representation of space, with their ability to generalise to new environments dependent on this representation. Though other research[48,49] indicates that hippocampal place cells can also be significantly influenced by non geometric contextual factors such as colours and odours, suggesting a more comprehensive and contextually adaptable role for the hippocampus in spatial representation and memory.

Grid cells, known for firing at multiple locations forming a hexagonal grid pattern, have been theorized to offer a metric representation of space for navigation and memory purposes[31]. These cells were found to be less sensitive to environmental changes compared to place cells and exhibited a greater degree of generalization across environmental changes[50,51]. This evidence implies that grid cells may provide a more flexible and resilient representation of space, supporting generalization across various environments.

Hippocampus and non-spatial reasoning

While much of the research on the hippocampus has historically focused on its role in spatial memory processing, a large amount of evidence suggests that it also plays a crucial role in non-spatial cognition. Howard Eichenbaum was one of the leading researchers in this area and viewed Tolman's cognitive map hypothesis as a more general process that allows the brain to establish connections between different aspects of the world and use this information to predict and guide behaviour[52]. Eichenbaum conducted extensive research on the hippocampus and proposed that it is involved in a broader relational memory process[52,53].

Relational reasoning involves binding together different elements of an experience based on shared characteristics, and is crucial for forming new associations between previously unrelated pieces of information. This linking of memories can lay the foundation for arbitrary recombination and generalisation. For example, suppose you are learning how to make a complex recipe that involves several steps and ingredients. You might first learn how to prepare a particular ingredient, such as cooking rice or chopping vegetables. Then, you might learn how to combine these ingredients in a certain way to make a specific dish, such as a stir-fry. Finally, you might be presented with a new recipe that combines some of the ingredients and techniques you have learned, but in a novel way that you have not seen before. By using your relational memory and knowledge of how different ingredients and techniques interact, you can infer how to modify your existing knowledge to successfully make the new dish.

One classic paradigm that has been used to study relational reasoning in a laboratory setting is called *transitive inference*[54–56]. In transitive inference tasks, animals are first trained to prefer one stimulus over another, followed by separate training to prefer the second stimulus over a third. During the test phase, the animal is presented with a choice between the first and third stimuli, which it has not seen before. Animals with intact hippocampi can leverage their relational memories to infer that the first stimulus is preferred over the third, based on the transitive relationships established between the stimuli during training. However, lesions to the hippocampus impair the ability to make

such novel inferences, while leaving intact the ability to make choices based on previously seen stimuli[55,57–59]. Transitive inference tasks involve ordering stimuli along a linear hierarchy and share similarities with other types of linear ordering, such as the ordering of sound frequencies and temporal events. Interestingly, hippocampal cells have also been found to code for time, referred to as time cells[60], suggesting that the hippocampus may be involved in encoding and utilising relational information across various domains.

Learning a general model, or a 'cognitive map' of any given problem requires understanding and remembering relationships between different elements in the environment, whether spatial or non-spatial. Eichenbaum proposed that physical space might therefore be just one instance of relational memory space among many, and when spatial awareness is essential to solve a task, spatial representations would be the dominant mode.Although in the seminal paper by[27], only around 10% of cells were identified as place cells, subsequent research showed that this proportion increases significantly in environments rich in spatial information, reaching up to 50%[61]. Furthermore, in tasks that necessitate spatial information, the proportion of place cells can escalate to as high as almost 90% among the pyramidal complex-spike cells[62]. However, in tasks that require animals to learn odour associations, where spatial information is less crucial, Wood et al. (1999)[63] showed that only 31.5% of the cells in the hippocampus had a spatial correlate, with less than half of those being purely spatial. In another study, where rats were required to learn multiple context-dependent object-reward-location associations, the hippocampal population code indicated a richer representation beyond just spatial location, including a hierarchy of relationships between task dimensions such as context, item identity, and valence[64]. These findings suggest that the hippocampus can use relational memory to form mental models that are tailored to the specific requirements of a given task, whether spatial or non-spatial, to guide behaviour.

Eichenbaum and colleagues observed another essential aspect of hippocampal function, namely what is often referred to as 'conjunctive' coding. In a study by Komorowski et al. (2009)[65], they recorded extracellular spike activity of CA1 and CA3 neurons while rats performed a conditional discrimination task and identified two types of cells:

item-position cells and *position cells*. Item-position cells fired more when rats sampled one of the two items at one or both positions in one of the contexts and predicted accuracy over position-only cells. The findings suggested conjunctive coding, in which individual hippocampal neurons develop responses to specific stimuli in places where they have differential significance, further emphasising the hippocampal role in relational memory processes[52,63]. This conjunctive coding can serve as a neural mechanism for binding information about different aspects of the environment into a coherent representation, allowing for the formation of new associations and inference based on relational memory[53].

Several other types of relational memory tasks have also been shown to depend on the hippocampus, including those involving social value and emotional features[66]. Additionally, hippocampal cells have been shown to encode the history and future of an episode, providing a representation of both place and memory context to facilitate decision-making[67,68]. Together these findings propose a broader role of the hippocampal formation in cognitive mapping, storing relational knowledge across various domains spatial, temporal, and stimulus hierarchy, to guide behaviour.

War of Ghosts: schemata in hippocampus and prefrontal cortex

Abstract and relational reasoning has also been an important topic in human behaviour and learning research, particularly in the context of what are commonly referred to as *schemata*. One of the first insights came from Frederic Bartlett in the 1930s who, inspired by the game of 'Chinese Whispers', asked students to read a Native American story (*The War of the Ghosts*) and recall it after varying time intervals[69]. He noted that students' recall of the story had a lot of differences to the original, and often involved changing unfamiliar parts of the story to familiar ideas in line with their existing knowledge of the world: canoes and paddles became boats and oars; hunting seals became fishing. This led Bartlett to suggest that the participants' pre-existing cultural schemata shaped their understanding of the story and influenced the way they remembered and retold it.

More generally, a schema can be described as a cognitive framework or mental structure that contains knowledge about a concept or a set of concepts and the relations between them and allows individuals to organise and interpret new information based on their existing knowledge and experiences[70]. While this definition appears highly relevant to the themes of generalisation, learning sets, and relational learning discussed earlier, the connections between these lines of work remain surprisingly scarce.

Nevertheless, the hippocampus has not been overlooked by this field either and has also been found to play a crucial role in the representation of schemas. For instance, a landmark study by Tse and colleagues (2007)[71] demonstrated that rats are better at encoding and consolidating new information when it fits within their existing spatial schemas and lesions to the hippocampus can disrupt this ability. In a later study, Tse and colleagues identified another brain region, the prelimbic region of the frontal cortex, to be implicated in this process[72]. Their findings showed that learning schemas were associated with an up-regulation of immediate early genes in this region, suggesting the involvement of prelimbic areas in learning[72]. Finally, these observations are consistent with whole-brain imaging studies in humans, which have shown that hippocampal activity is strongly correlated with frontal cortex activation during tasks that involve processes that bare similarities to schema-based reasoning, such as spatial navigation and inferential reasoning[73–75].

More recently, research has also shed light on more precise neural mechanisms that might underlie the formation and use of schemas in learning and cognition. A study by Zhou and colleagues found that, in a frontal region called the orbitofrontal cortex, the dimensionality of neural ensembles decreases as animals learn a common structure of the task (i.e. *schema*) across multiple sensory instances[76]. Their findings suggested that this evolution of a neural schema might underlie animals' ability to learn new information faster in situations that share the same abstract 'schema'.

The theoretical implications of the roles played by the hippocampus and the frontal cortex in abstract reasoning have been a subject of intrigue. Numerous studies have suggested that both the frontal cortex[77–81] and the hippocampus[53,55,63,82–88] might be involved in

depicting task states and their interrelations. However, our understanding remains limited regarding the unique features of the representations in these two areas and their specific contributions to an animal's capacity to leverage existing 'schemata' to address novel challenges.

1.3 In machines

Computers, like humans, rely heavily on prior experience to solve problems. IBM, the company responsible for developing Deep Blue (the first computer to beat a grandmaster in chess), famously claimed that "Kasparov isn't playing a computer, he's playing the ghosts of grandmasters past"[89]. As Garry Kasparov's eventual defeat demonstrated, with enough prior experience, and especially in deterministic (and Markovian) tasks like chess, where the ability to perform very fast and accurate calculations is critical to success, computers can indeed rival human performance.

However, the 1997 Kasparov-Deep Blue games also revealed a crucial blind spot in computer intelligence. In one game, Kasparov started out by making a very common opening move. Deep Blue responded with a move which, based on its calculations, was associated with a reduction in the opponent's winning chances from 56% to 51%, according to historical databases. However, this computation relied on a crucial assumption that Kasparov would then continue to respond as almost all other players had when faced with this position. Instead, Kasparov not only made a highly unexpected move but achieved a board position that had only occurred once before, out of hundreds of thousands of games stored in Deep Blue's database. This put Deep Blue in an incredibly difficult situation. In this position, it could no longer rely on indexing its deep databases of moves with pre-computed value functions, and instead had to 'decide for itself', and attempt to calculate the strength of each move using its own powers of inference – confronting the vast branching tree of possibilities which characterise large state-spaces[90] *.

*The number of legal positions possible on a chessboard remains an unsolved problem. In 1950, Claude Shannon[91] proposed it to be $\frac{64!}{32!8!^2 2!^6}$ or 10^{43}.

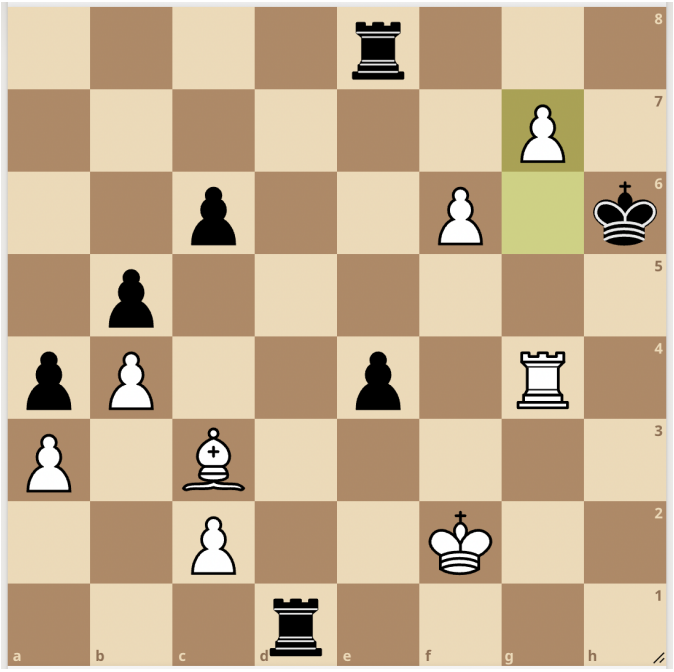


Figure 2: Kasparov (white) vs. Deep Blue (black): Game 1, 1997. In the first game of the 1997 match, Kasparov begins with the moderately rare *King's Indian Attack (Keres variant)* opening, and wins on move 45 by pushing the pawn to g7. Deep Blue resigns upon evaluating (correctly) that this position is a forced mate-in-16 for white.

The anatomy of a cognitive map in artificial intelligence agents

In recent years, much attention has focussed on the bidirectional relationship between neuroscience and machine learning, resulting in a rich body of literature both of neuroscience-inspired machine intelligence[92,93] and empirical neuroscience findings drawing on concepts from AI[94–96].

Hard-coded, brute-force chess engines in the vein of Deep Blue might have proved sufficient to beat the world's leading Grand Master, but few would have claimed that they exhibited *general artificial intelligence*[92,97]. More recently, these models have themselves been surpassed by more general models, capable of attaining super-human performance across a range of domains[98]. Although much of this success is (correctly) attributed to the renaissance of *deep reinforcement learning*[99], the performance of these more general AI systems also highlights the importance of developing a *schema* or *cognitive map* if the agent is to master multiple domains.

The topics central to this thesis – learning the structure of problems and generalising these

to related scenarios – are also major aims of machine learning research, and are critical to the development of an artificial agent capable of replicating biological intelligence. Here, I outline some of the most important of these capacities (*continual learning*, *meta-learning*, *modularity and composition*, and finally *transfer learning*), and draw parallels between findings in neuroscience and machine learning which, though disparate in their respective methods, aim to understand these same capacities.

Continual learning

We take for granted that biological agents can learn a diverse range of tasks throughout life, building on (and flexibly recombining) prior knowledge[1,100]. Interestingly, however, this is not an emergent general property of densely connected neural networks. Various attempts to replicate such continual learning in artificial neural networks have failed, with training on subsequent tasks resulting in an effective over-writing of previous task performance – a phenomenon known as *catastrophic interference*[101]. This raises several questions over the criteria required for continual learning in biological agents. Indeed, many modern approaches from machine learning aiming to alleviate catastrophic forgetting draw inspiration from neuroscience, offering some insight into the neural computations underlying our ability to learn and generalise across time and tasks[3,92,102,103].

Learning algorithms – Deep neural networks are usually trained by stochastic gradient descent and backpropagation[104], but the biologically plausibility of these learning algorithms remains unclear[105]. It seems highly likely that brains have evolved learning algorithms capable of extracting the key statistics of given tasks, enabling them to retain performance on these tasks even as new ones are encountered[3,92], even if these precise learning algoriths are unknown[105]. Nevertheless, many approaches to continual learning in artificial neural networks relate to phenomena known to be involved in learning in biological agents. For instance, stable postsynaptic dendritic spines have been demonstrated to support and maintain memories throughout life[106]. Directly inspired by this, *Elastic Weight Consolidation (EWC)*[107] is an approach to continual learning which aims to reinforce the most important features of previously encountered tasks, mimicking the manner in which biological synaptic consolidation reinforces memories.

The details of this algorithm are beyond the scope of this thesis but, intuitively, it involves identifying the weights of the neural network (which are, in this case, analogous to synaptic connections in the brain) most critical to solving a given task. When a number of tasks are presented sequentially, the loss function of EWC attempts to preserve these most critical weights, thereby leading to accumulated task performance with minimal catastrophic forgetting[107,108].

Training paradigm – Of course, when learning many tasks in sequence, animals seldom do so without re-encountering some of the important features of these tasks, either explicitly or implicitly. One form of implicit re-encountering of tasks is that of *replay*[109,110]. Although replay is theorised to subserve a number of functions in the brain[111], it is perhaps unsurprising that generative replay – whereby samples are drawn from prior experience and presented, interleaved, with current experiences – has been the focus of considerable work in the continual learning literature in machine learning[93,112].

"Deep Generative Replay" was one of the first training paradigms aimed at replicating the potential value of replay in the brain for continual learning in artificial agents[112]. Here, a neural network was trained on multiple tasks in sequence; as training progressed, previously encountered tasks were sampled and presented, interleaved, with novel tasks. While this approach was shown to confer a considerable benefit in performance compared to parameter-matched networks, some limitations exist. The most significant of these is that this replay regime lacks biological plausibility; here, 'replayed' samples were presented in their original form and fidelity, as if they were being encountered anew (rather than remembered). Building on this work, van de Ven et al. (2020)[93] developed a more biologically plausible form of generative replay for deep neural networks, in which compressed representations of previous experiences were stored, and then interleaved with novel experiences using a more sophisticated context-dependent sampling model.

It is worth noting, however, that these synthetic approaches to replay are relatively simplistic in comparison to the rich set of functions known to be subserved by replay in the brain[109,110]. Replay describes the phenomenon by which an assembly of cells active during an awake state are reactivated in a similar (but faster) pattern during sleep or awake

rest. Originally described in the domain of spatial navigation[113,114], replay has long been thought to subserve spatial memory consolidation (emphasised by the coordination of grid and place cell replay during rest[115]. More recently, this account has been expanded to include non-spatial domains, in which human replay re-constructs the experience of abstract tasks[110] and even serves as a mechanism of inference in abstract, compositional tasks[116].

Meta-learning

At an algorithmic level, the mechanisms underlying learning in biological agents are broadly unknown[105]. By contrast, the most common paradigm of 'learning' in artificial neural networks is remarkably simple – consider a network comprised of a series of weights, compute some error metric between actual and desired output, and then update these weights at a number of steps (using backpropagation[117]).

Animals, however, can exhibit *meta-learning* ('learning to learn', which relates to the previously discussed concept of Harlow's *learning sets*). Here, they learn tasks more efficiently with increased exposure and rate of skill acquisition increases across tasks, together pointing towards some optimisation of the learning procedure itself. Beyond the canonical examples of Harlow's learning sets described earlier, this principle has been observed across numerous animal models. In a head-fixed virtual reality paradigm, for instance, mice learned the structure of a reversal learning task more efficiently over time[118]. In a parallel study, freely-moving mice performed a two-alternative forced-choice task, consisting of a number of sub-tasks with different rules[119]. It was found that, upon transitioning to a new sub-task, mice were able to learn the rules governing that sub-task from the very first trials. Meta-learning offers an appealing avenue towards much more efficient learning as more task domains are increased, and insights from machine learning and AI point towards this being a central component of a cognitive map[3,120].

One of the most interesting examples of meta-learning in AI involves meta-reinforcement learning (*meta-RL*), in which a recurrent neural network (RNN) is trained across a range of tasks commonly found in neuroscience literature[121], using reinforcement learning[122]. The critical insight in this study was to set the learning rate of the network (that is, some

metric specifying the frequency of weight updates in the network) sufficiently low that the network could not learn any single task perfectly; instead, it learned to extract the common structure and rules of tasks. Intriguingly, as a result, the behaviours and learning dynamics of the network were strikingly similar to those of animals performing the analogous tasks. This points towards meta-learning being a natural consequence of biological constraints on learning rate in the brain, and potentially crucial mechanism underlying the efficiency of learning in animals.

Modularity in minds and machines

Structurally and functionally, the brain is modular. Detonator synapses in the hippocampal formation for forming rapid memories[123] and skip connections between cortical layers for efficient information transfer over distant regions[124–126] are just some of the examples of this structural modularity influencing function.

Following from this, many of the most influential theoretical accounts of the hippocampal-prefrontal systems emphasise this structural and functional modularity[95,127,128]. These accounts – which often explain a diverse range of behaviours and representations associated with flexible behaviour – build on a rich literature from machine learning and artificial intelligence, in which modular artificial neural networks confer some gains in performance (and biological plausibility) over naive models[117,129,130].

One of the most influential and recent computational models of structure learning in the hippocampal-entorhinal system is the Tolman-Eichenbaum machine[95,128]. In this framework, *structure learning* (Tolman) in the entorhinal formation is cast as one component of generalisation across tasks, complemented by *relational memory* (Eichenbaum) in the hippocampus.

When trained on a number of tasks from the literature of spatial neuroscience, this model learns and reproduces the same famous neural representations discussed earlier (grid, border, and object-vector cells in the entorhinal cortex; place and landmark cells in the hippocampus)[27,31,36,128,131,132]. Critically, however, it is also able to recapitulate the neural representations observed when animals perform non-spatial tasks. For instance,

following an innovative yet simple task design from Sun et al., 2020[82], the artificial TEM agent completes four laps of a track environment to gain a reward. After training, the hippocampal representations of the model resemble units which count across laps, incrementing towards the reward before resetting – precisely the same as the hippocampal representations recorded by Sun et al. These representations offer a parsimonious account of the task itself; the most important component, in terms of structure, is that there are n laps to reward – where features such as n and the physical configuration of the lap environment itself are sensory particularities, while the lap-counting is the underlying (and *generalisable*) structure. Such representations point towards a shared framework for learning and generalising the structure of tasks – regardless of whether the task itself is spatial[82,133,134] or conceptual[82,85,87,135].

Another influential theory of generalisation in the hippocampal-entorhinal system is that of a *predictive map* of the environment and tasks embedded within it[136]. In this framework, the representations of the hippocampus (most notably, of course, place cells) are cast as those most effective for *predicting future state*. Stachenfeld et al.[136] demonstrate this using a reinforcement learning paradigm known as the *successor representation (SR)*[137] – in brief, this is a matrix representation of an environment which describes the expected, discounted future occupancy of any state in terms of a starting point. One of the key insights of the study by Stachenfeld et al. is that columns of this SR matrix look remarkably like hippocampal place cells in a square environment. Furthermore, if the dimensionality of these place cell-like representations is reduced, the resulting matrices very closely resemble the periodicity of grid cells. Together, these findings offer a tantalising account of the hippocampus as encoding a predictive map of task and environment, with the entorhinal cortex encoding a flexible and generalisable low-dimensional decomposition of this. A host of experimental studies have since offered evidence both in favour of[138]and in contention with[139] this theoretical account. Nevertheless, in terms of modularity, this theoretical account of the hippocampus encoding a predictive map of space has offered a number of insights bridging spatial and non-spatial accounts of the hippocampal-entorhinal system[140].

Indeed, another form in which this modularity is expressed in the brain is through a hierarchy of representational abstraction across hippocampal-entorhinal-prefrontal circuits. Some of the most compelling evidence for this in the hippocampal-prefrontal circuit derives from a study[141] in which rats performed a spatial alternation task, resulting in hippocampal representations which fire alternately on left and right turns (these representations, commonly known as 'splitter cells', have been widely described previously[63]). These trajectory-dependent splitter cell representations were observed in both medial prefrontal cortex and the CA1 region of the hippocampus. Notably, however, when the thalamic connection bridging these regions (the *nucleus reuniens*) was lesioned or optogenetically silenced, these trajectory-dependent representations persisted in mPFC, but were ablated in CA1, perhaps pointing towards a pathway by which an abstract representation (such as a splitter cell) might be inherited from prefrontal areas for use in the hippocampus. Capturing such a representational hierarchy across brain regions will surely become a major focus of computational neuroscience in future years[95].

Although these accounts differ in their implementations and precise neural predictions, together they point towards modularity being an important component of a flexible and generalisable cognitive map in the brain.

Transfer learning

The final necessary component of cognitive maps I will discuss here is that of *transfer learning*. Learning the statistics of a given task in a specific environment – whether by backpropagation in a neural network or by (currently unknown) biological learning algorithms – is clearly a useful objective for achieving bespoke behaviours and representations for solving that task. However, it would be vastly computationally expensive (and inefficient) to do so *de novo* when tasks share similar structures and statistics. In both machine learning research[142] and neuroscience[143], the mechanisms by which learned representations and behaviours of one task can be transferred to those of related ones have been extensively studied.

For instance, in the literature of reinforcement learning, learning or approximating a value function is a central problem[122]. After this value function is acquired, however, it can

be flexibly transferred to similar paradigms (that is, tasks and environments), thereby affording more efficient learning in any subsequent tasks. The value function learned by an agent as it (reinforcement) learns to play chess will still confer benefit if transferred to the game of Go[144].

This apparently banal point might have deep implications for neuroscience – many of the well-described neural representations of space are now known to be re-used in *non-spatial* domains. Entorhinal grid cells, which tile spatial environments as animals navigate them, also appear to represent the leg length and neck length of cartoon birds when humans view these two abstract dimensions on a screen[31,85]. Similarly, hippocampal place cells do not fire simply at a given location in physical space, but also in *conceptual space*. When rats must *navigate* a number of tone frequencies and press a lever at the correct tone, hippocampal place cells tile this 'tone space' in the same manner that they would a physical environment[87].

Through this interdisciplinary research from machine learning[98,144] and neuroscience[85,87], we have learned that transfer learning is likely to be a critical component underlying the formation of cognitive maps, whereby the representations developed in one task or environment can also be *transferred* to novel, but related, ones.

1.4 Summary

Despite significant progress in the field, many questions remain unaddressed as to how the brain draws on experience to achieve abstraction, thereby enabling animals to quickly solve novel problems. For example, how does the brain represent knowledge at different levels of abstraction? How does it map abstract knowledge onto new sensory scenarios? The experimental work described in this thesis builds on a growing body of literature discussed above, and seeks to answer these questions and further our understanding of how the brain enables generalisation of knowledge across situations.

2

Abstraction and the underlying behavioural strategies in mice

Contents

2.1 Introduction **39**
2.2 Behavioural Task **41**
2.3 Mice generalise within-trial knowledge between structurally equivalent problems **42**
2.4 Mice generalise long-term strategies between structurally equivalent problems **44**
2.5 Discussion **46**

2.1 Introduction

To understand how the brain produces a particular behaviour, it is useful to first verify that the animal is indeed capable of exhibiting the behaviour in question. As we were interested in how animals generalise prior knowledge to new situations, our first objective was to create an experimental framework in which animals could leverage their knowledge of the abstract problem structure to solve new problems more effectively.

As discussed in the *Introduction*, animals' ability to re-use prior knowledge in new situations has been extensively studies in the context of 'learning set' tasks[7]; most

frequently, in primates and to a lesser extent in rats (see *Why rats deserve more credit*). The general structure of the traditional 'learning set' paradigm involves continuously presenting subjects with novel pairs of objects, only one of which leads to a reward. Despite the changing pairs of objects, the abstract rule remains the same - stay with the same object if you get a reward, and switch if you don't. The behavioural evidence for animals' ability to generalise knowledge of an abstract task structure is demonstrated by their ability to learn faster each time when presented with a new task, a phenomenon known as 'learning to learn' (*meta-learning*).

Borrowing some of these ideas, we developed a novel behavioural paradigm for mice that involved presenting subjects with a series of problems which all shared the same abstract rules, but crucially differed in their physical characteristics and the corresponding sensory and motor demands. The abstract problem we employed is probabilistic reversal learning - a widely used test of cognitive flexibility[145–147]. In classic reversal learning paradigms, subjects are presented with two choices that have anticorrelated reward probabilities (that is, only one choice is good at any given time). After discovering the option that yields the greatest probability of gaining a reward, and consistently selecting it over a number of trials, the rule is suddenly (and without communication to the subject) inverted, and the other option becomes the rewarding one. To efficiently and reliably solve this task, animals must learn to infer when the reversal has occurred, and switch to choosing the alternative option. The key novel feature of our behavioural design was that, instead of presenting animals with just one instance of a reversal learning task, we serially presented mice with multiple instances of this task each encompassing a new sensory-motor set of options.

We hoped that by adopting this design animals would learn to abstract the shared features present in all problems (that is, the reversal learning problem) from the sensory instances unique to each problem (motor and sensory features of various sets of options), enabling them to learn faster each time when presented a new problem (and, thereby, 'learn to learn'). Below, I will describe our behavioural task in more detail, and discuss the

evidence we found to support our hypothesis as well as highlight some possible limitations of our approach.

2.2 Behavioural Task

Subjects serially performed a set of reversal learning problems which shared the same structure but had different physical layouts. In each problem, every trial started with an 'initiation' nose-poke port lighting up. Poking this port illuminated two 'choice' ports, which the subject chose between for a probabilistic reward (Figure 3A). Once the subject consistently chose the high reward probability port (i.e., achieved a 75% success rate over 8 consecutive tries, calculated using the exponential moving average method that places more importance on the most recent attempts), reward contingencies reversed (Figure 3B). Once subjects completed ten reversals on a given port layout (termed a 'problem'), they were moved onto a new problem where the initiation and choice ports were in different physical locations (Figure 3C). All problems therefore shared the same trial structure (initiate in the illuminated poke, then choose between the two illuminated pokes) and a common abstract rule (one port has high and one low reward probability, with occasional reversals), but required different motor actions due to the different port locations. Problem switches occurred between sessions, and subjects completed ten different problems.

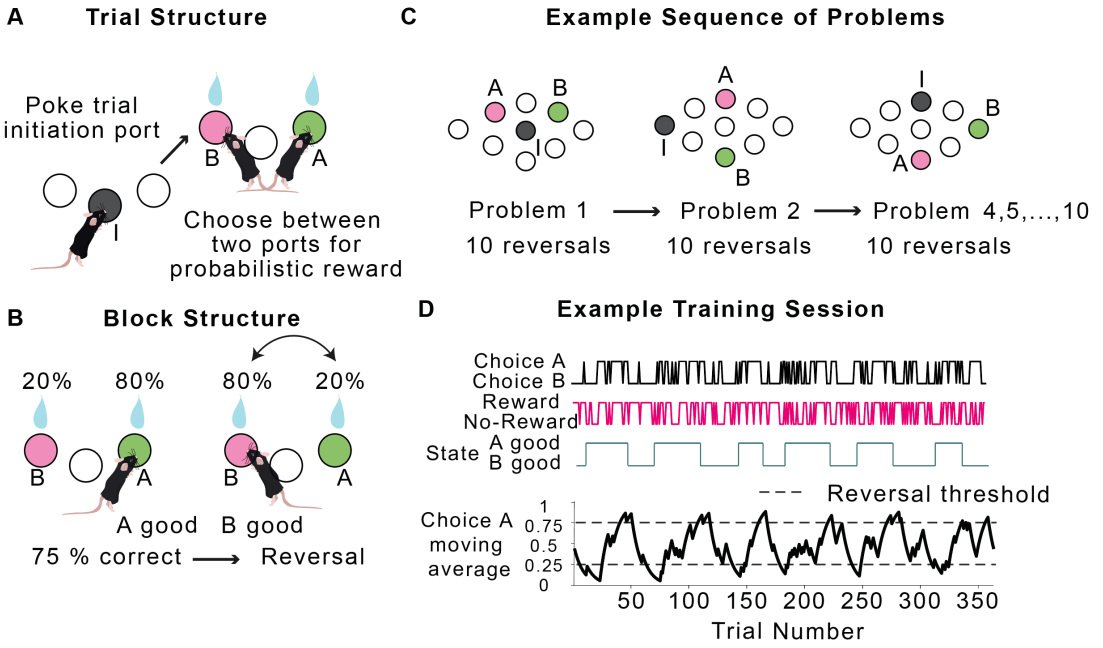


Figure 3: Transfer learning in mice. **(A)** Trial structure of the probabilistic reversal-learning problem. Mice poked in an initiation port (grey), then chose between two choice ports (green and pink) for a probabilistic reward. **(B)** Block structure of the probabilistic reversal-learning problem. Reward contingencies reversed after the animal consistently chose the high reward probability port. **(C)** Example sequence of problems used for training, showing different locations of the initiation **(D)** Example behavioural session late in training in which the animal completed 12 reversals. Top subpanels show animals' choices, outcomes they received, and which side had high reward probability; bottom panel shows exponential moving average of subjects' choices (tau=8 trials).

2.3 Mice generalise within-trial knowledge between structurally equivalent problems

Since the general structure of each trial (initiate then choose) was shared across all problems, we first tested whether subjects could learn to generalise this sequence across problems (Figure 4).

One behavioural proxy of animals' understanding of how a trial works (first initiate, then choose, then initiate again) is how often they make nose pokes that are inconsistent with this sequence (i.e. pokes into the alternative port after already having made a choice, instead of going straight back to initiation). Mice made fewer such out-of-sequence pokes across reversals within each task ($F_{(9,72)} = 17.82, p < .001$; Figure 4A), but importantly also across tasks ($F_{(9,72)} = 18.29, p < .001$; Figure 4B), demonstrating that mice got better at following the trial sequence on each new problem.

Our critical question was whether the animals would show improvement and exhibit 'learning to learn' on each new problem and a closer inspection of these data strongly supports this. Consistent with this, in the beginning of the training mice learnt the new poke sequence over many reversals, but by the end of the training they acquired the poke sequence in a single reversal suggesting they learnt how to learn the new poke sequence ($t_{(17)} = 2.81, p = .023$; Figure 4C).

One potential confounding factor in the interpretation of these results is the use of light cues in the study design. The illumination of the initiation port and the choice ports during the trial could have resulted in mice relying solely on these cues instead of learning the abstract trial structure. This possibility is supported by the fact that the most significant reduction in out-of-sequence poked occurred between problem 1 and problem 2, suggesting that instrumental learning of light cues may have played a role. However, it is worth noting that the improvement in the animals' understanding of the trial structure was not solely driven by poor performance on the first problem, as evidenced by the continued improvement throughout training ($F_{(9,64)} = 9.36, p < .001$; Figure 3B).

To further examine the possibility that animal's driven was simply driven by learning to follow port illumination, we also examined behaviour on 'forced choice' trials where only one choice port illuminated, and the other was inactive. Animals did not just follow the light and were equally likely to poke the high reward probability choice port as the choice port that was illuminated, demonstrating that their behaviour was influenced by their belief about reward availability and not just the port illumination (Figure 4D). Still, it remains possible that they used port illumination while acquiring a new problem.

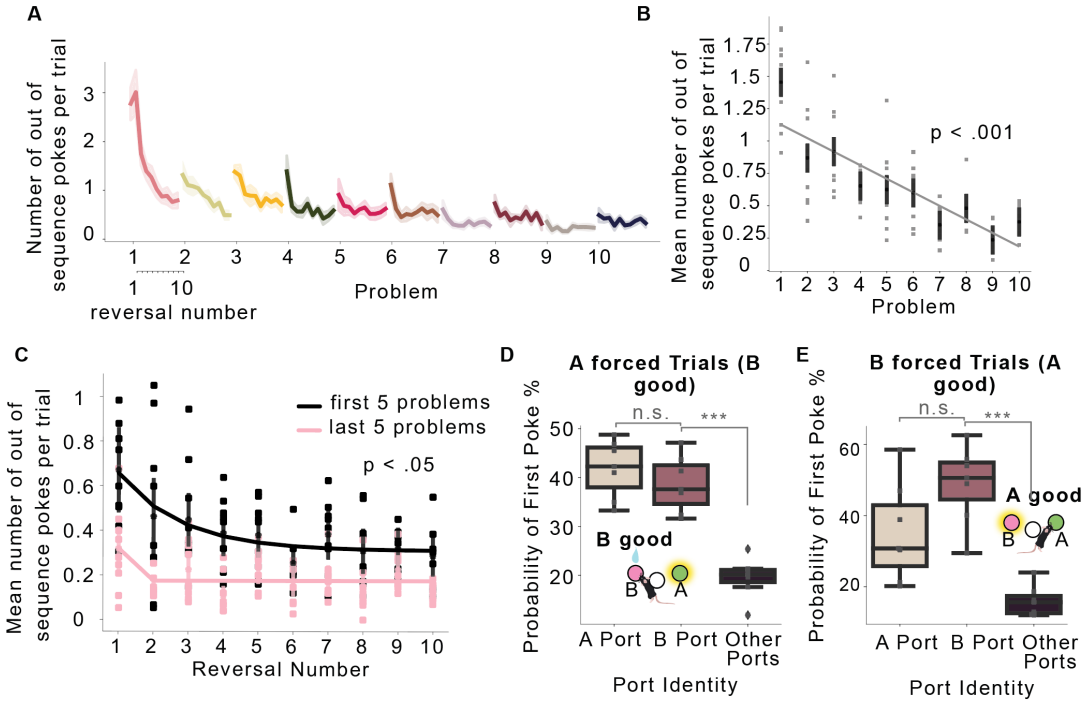


Figure 4: Behavioural generalisation of trial structure. **(A)** Number of pokes per trial to a choice port that was no longer available because the subject had already chosen the other port, as a function of reversal number within each problem and problem number. **(B)** Mean number of pokes per trial to a choice port that was no longer available because the subject had already chosen the other port, as a function of problem number. The *p*-value is from the two-way repeated-measures ANOVA with problem number and reversal number as within-subjects factors.**(C)** Mean number of pokes per trial to a choice port that was no longer available as a function of reversal number on the first 5 problems and the last 5 problems during training. The *p*-value refers to the difference in the log of the time constants from fitted exponential curves in early and late training. **(D)** On forced A trials where A choice was illuminated but B choice was good animals were as likely to first choose the B choice port (not illuminated but good port at the time) as the illuminated A port ($t_{(7)} = 0.76, p = .470$) and were much more likely to choose the good B port than other ports (initiation and ports used in other problems) ($t_{(7)} = 6.25, p < .001$).**(E)** On forced B trials where B choice was illuminated but A choice was good animals were as likely to first choose the A choice port as the B port ($t_{(17)} = 1.45, p = .200$), and were more likely to choose the A choice than the other ports ($t_{(7)} = 2.94, p = .030$). Error bars report the mean ± SEM **(A-C)** or median, inter quartile range and min and max **(D, E)** across mice (n = 9); dots show individual subjects.

2.4 Mice generalise long-term strategies between structurally equivalent problems

Next, we looked at whether subjects showed evidence of learning the abstract task structure (*one port is good at a time*, with reversals) and generalising it to new tasks (Figure 5).

We found that animals adjusted to reversals in reward contingencies faster on each subsequent problem. Animals took fewer trials to reach the 75 % correct threshold for

triggering a reversal within each task ($F_{(9,72)} = 3.52, p = .001$; Figure 5A), and crucially also across tasks ($F_{(9,72)} = 3.91, p < .001$; Figure 5B), consistent with generalising knowledge of this abstract structure. Furthermore, mice exhibited 'learning to learn' from reward, adapting to reversals faster at the end of training than at the beginning ($t_{(17)} = 5.04, p = .001$; Figure 5C).

To efficiently adapt to reversals, one strategy is to integrate past choices and outcomes to determine the optimal stimulus to choose. To investigate whether mice developed this strategy, we employed a logistic regression model that predicted subjects' choices based on their past choices, outcomes, and their interactions. Our findings demonstrated an increase in the influence of both the most recent ($F_{(9,71)} = 5.08, p < .001$; Figure 5E, G) and earlier ($F_{(9,71)} = 5.46, p < .001$; Figure 5E, G) choice-outcome interactions across tasks. Moreover, subjects' choices were increasingly influenced by their previous choices ($F_{(9,71)} = 11.77, p < .001$; Figure 5D, F), indicating a reduction in spontaneous exploration with learning. These results suggest that the subjects' enhanced ability to integrate past experiences led to an improvement in tracking the good port across problems and making more optimal choices.

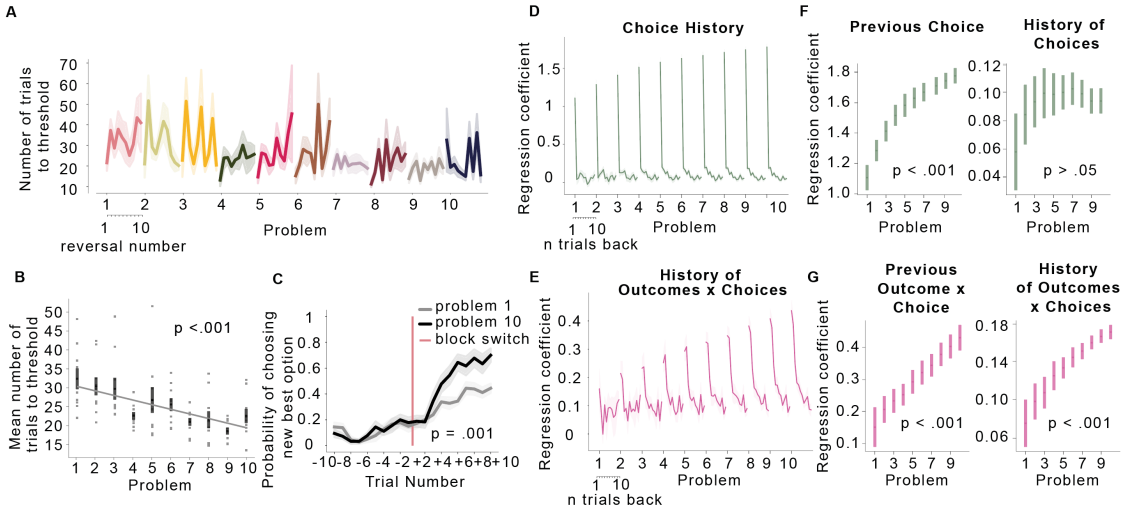


Figure 5: Behavioural generalisation of long-term strategies. **(A)** Number of trials following a reversal taken to reach the threshold to trigger the next reversal, as a function of reversal number within each problem and problem number. **(B)** Mean number of trials following a reversal taken to reach the threshold to trigger the next reversal, as a function of problem number.The *p*-value is from the two-way repeated-measures ANOVA with problem number and reversal number as within-subjects factors. **(C)** Probability of choosing the new best option (the choice that becomes good after the reversal) on the last 10 trials before the reversal and the first 10 trials after the reversal split by the first problem and the last problem. The *p*-value refers to the difference between the slopes after the reversal point in early and late training. **(D,E)** Coefficients from a logistic regression predicting current choices using the history of previous choices **(D)**, outcomes (not shown) and choice - outcome interactions **(E)**. For each problem and predictor, the coefficients at lag 1-11 trials are plotted. **(F,G)** Coefficients for the previous trial (lag 1, left) and average coefficients across lags 2-11 (right), as a function of problem number. Error bars on all plots show mean ± SEM across mice.

2.5 Discussion

In this section, we have explored animals' ability to apply their previous knowledge to more efficiently solve novel problems. To investigate this, we developed a novel behavioural paradigm that provided compelling evidence that mice can transfer their knowledge of the sequence of responses and between-trial policies required to solve a reversal learning task across structurally identical but distinct problems that vary in sensory and motor features. These findings offer additional support for animals' aptitude for 'learning to learn'[7,148].

While some researchers have suggested that reversal learning paradigms primarily assess animals' ability to inhibit a motor response[149], our behavioural paradigm required mice to learn a more general rule beyond motor inhibition (as sensory-motor demands varied across problems) - that only one option is advantageous at a time. We observed that

animals were able to better track the history of their choices and rewards across different problems, indicating that they had learned a general strategy that is relevant for this abstract rule.

However, it is important to note that simply tracking rewards and outcomes is not sufficient for animals to make optimal choices in a reversal learning paradigm. They must also develop a *belief* when the reversals might occur, allowing them to anticipate and adapt to changes in the task structure. While our behavioural analyses did not directly examine this belief formation, recent computational research by Costa and colleagues (2015)[146] suggests that animals might accumulate evidence by tracking their choices and outcomes to generate a likelihood estimate (in Bayesian terms, a simple *prior probability*) that a reversal could occur (that is, *belief*).

To investigate this possibility, I fit this model to our behavioural data, along with other simple computational models. The results of these simulations, however, are omitted from this thesis, as they offered no additional insights into the underlying behavioural processes. The reason for this is that the primary measure for assessing the goodness of fit of models is the BIC score (*Bayesian information criterion*), which is a measure only capable of comparing the *relative* fits of different models to the data[150]. This makes it impossible to determine the absolute fit of a model. As such, demonstrating that any of these models fits our behavioural data well would simply restate the *possibility* that tracking the history of rewards and outcomes is a possible mechanism for belief formation, without providing any further evidence to support or refute this claim. Nevertheless, the theoretical framework advanced by Costa and colleagues (2015)[146] likely offers a more nuanced account of behaviour and learning dynamics in probabilistic reversal learning paradigms than classic RL models. This is because, as presented in Costa et al.[146] in the context of data from monkeys performing a probabilistic reversal learning task, animals typically switch to selecting the alternate option very soon after the current one becomes non-rewarding. Such an immediate reversal indicates some representation of the underlying rule and states of the task, rather than the naïve value function updates that would be performed by

a classic RL model, which would require a greater number of prediction errors to elicit a change in behaviour[122].

Further research is needed to explore this possibility in greater depth, and the strongest tests would involve more direct behavioural manipulations. One possible approach could be to vary the frequency or predictability of reversals in the task and observe how this affects evidence accumulation mechanisms and subsequent belief formation. For example, making reversals completely predictable (e.g. every 10 trials) could help to isolate a change of belief, and study the associated evidence accumulation mechanisms.

It is also worth noting that our study lacked two crucial controls. Firstly, we did not include a control group of mice that were trained on only one sensory-motor instance of a reversal learning task. This control would have enabled us to investigate the extent to which experience with varying examples of structurally similar tasks impacts animals' ability to generalise knowledge.

Secondly, we did not test the mice on a completely novel task that did not share any structural similarities with the original tasks - which might have offered greater insight into the *specificity* of knowledge transfer across tasks. The most intriguing test would be to compare the performance of pre-trained mice on a task with opposite behavioural demands to our reversal-learning task (e.g. win-shift, lose-stay). These limitations highlight the need for further research to fully understand the generalisation abilities of mice and other animals.

3

Neural representations underlying generalisation of trial events

Contents

3.1 Introduction **49**
3.2 Results **51**
Recording in mPFC and hippocampus as mice solve multiple problems 51
Abstract and task-specific representations of trial events by PFC and CA1 units 53
PFC population activity generalises more strongly across tasks than CA1 55
3.3 Discussion **62**

3.1 Introduction

In the preceding chapter, we demonstrated that after experiencing many problems with the same structure but different sensory particularities, mice could leverage their prior experience to solve new tasks more efficiently. Building upon this finding, our next objective was to investigate the neural mechanisms underlying this ability.

Solving a new problem is not simply a matter of understanding the abstract general rules that apply in a given situation; it also requires the capacity to map these rules onto the specifics of the situation at hand. Achieving such abstraction and linking it to specific

sensorimotor details is a complex process, and our understanding of how this is achieved by the brain is remarkably sparse. However, recent data suggest that interactions between frontal cortex and the hippocampal formation play an important role[151].

Neurons[76,152] and fMRI voxels[153,154] in these brain regions can form representations that generalise over different sensorimotor examples of tasks with the same structure, and track different task rules embedded in otherwise similar sensory experience[63,155]. For instance, fMRI voxels located in the prefrontal cortex exhibit generalisation properties across distinct sensory instances of a reinforcement learning problem which share the same task structure[154]. Meanwhile, hippocampal cells can also abstract spatial concepts from the superficial details of an environment and store them as a schema-like representation[152].

The involvement of these regions in abstraction is also of interest from a theoretical perspective. Both frontal cortex[77–80] and hippocampus[53,55,63,82–88] have been hypothesised to represent task states and the relationships between them. It has not been clear what distinguishes the representations in these regions, but some insight might be gained by considering hippocampal representations underlying spatial cognition. In rodent hippocampus, place cells are specific to each particular environment[27,44,47], but firing patterns in neighbouring entorhinal cortex (including grid cells) generalise across different environments – that is, they are *abstracted* from sensorimotor particularities[36,50,156–158]. Similarly, there is evidence that mPFC representations of spatial tasks generalise across different paths[159–161].

One possibility is that, as in space, abstracted or schematic representations of tasks in cortex might be flexibly linked with the sensorimotor characteristics of a particular environment to rapidly construct concrete task representations in hippocampus, thereby affording immediate inferences[3,95,128]. Indeed, hippocampal manipulations appear particularly disruptive when new task rules must be inferred, either at the beginning of training[162] or when task contingencies change[88,163]. Furthermore, several theoretical frameworks have considered the hippocampus as a rapid learning system, and the cortex as a slow learning

system which extracts the statistics of task representations from the hippocampus, offering normative accounts of numerous empirical neural phenomena[127,128,164].

To investigate the cortical and hippocampal contributions to generalisation, we recorded single units in the medial prefrontal cortex (mPFC) and hippocampus (dCA1) in mice pre-trained to solve multiple reversal-learning problems. As discussed earlier, these mice learned to transfer knowledge of both the sequence of responses required to solve a trial and how to track their history of choices and rewards to make better choices (i.e. policy). The two chapters which follow will explore within-trial abstraction and examine how the brain might execute the initiate-choose sequence. Later, in Chapter 5, we will revisit how the brain might facilitate the abstraction of an animal's policy.

3.2 Results

Recording in mPFC and hippocampus as mice solve multiple problems

We recorded single units from dorsal CA1 (345 neurons, n=3 mice, 91 to 162 neurons per mouse) and medial prefrontal cortex (mPFC, 556 neurons, n=4 mice, 117 to 175 neurons per mouse; Figure 6A) using electrophysiology. For recording sessions, we modified the behavioural task such that changes from one task to the next occurred within a session, with the transition to the next task triggered once subjects had completed four reversals on the current task, up to a maximum of three tasks in one session (Figure 6B). Subjects adapted well to this change and in most recording sessions performed at least four reversals in three different task layouts, allowing us to track the activity of individual units across tasks (Figure 6C). Cross-task learning reached asymptote prior to starting recordings, meaning that during recording sessions mice no longer showed improvement across tasks (Figure 6D, F, G) and there were no differences in behavioural performance between CA1 and PFC animals (Figure 6E).

During recording sessions (7 to 16 sessions; 341 to 650 trials in each session per mouse), we used ten different port layouts, but to simplify the analysis they were all reflections of three basic layout types (Figure 6B), each of which occurred once in every session. In

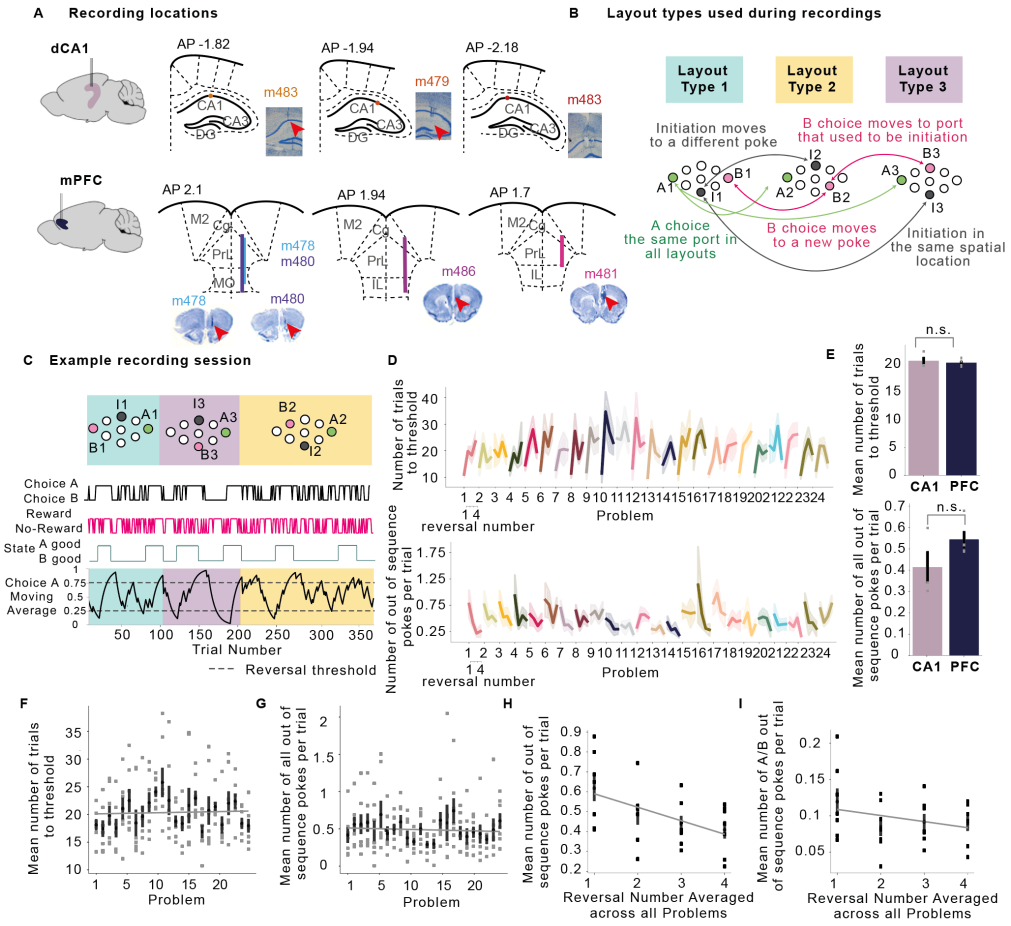


Figure 6: Recording units across multiple problems in a single session and behaviour during recordings. **(A)** Silicon probes targeting hippocampal dorsal CA1 and medial PFC were implanted in separate groups of mice. mPFC probes were advanced on each session whereas CA1 probes were static throughout recordings (for more details see *Electrophysiological Recordings and Spike Sorting Methods*). mPFC recordings were mostly from prelimbic area but also in some cases reached infralimbic area, medial orbital and cingulate.**(B)** Diagram of problem layout types used during recording sessions. **(C)** Example recording session in which a subject completed four reversals in each of three problems. Top panel shows the ports participating in each problem colour-coded by layout type. Bottom panel shows the exponential moving average of choices, with the choices, outcomes and reversal blocks shown above.**(D)** *Top:* Number of trials following a reversal taken to reach the threshold to trigger the next reversal, as a function of reversal number within each problem ($F_{(3,18)} = 24.19, p < .001$) and problem number ($F_{(23,138)} = 1.47, p = .090$) during recordings. *Bottom:* Number of pokes per trial to a choice port that was no longer available because the subject had already chosen the other port, as a function of reversal number within each problem ($F_{(3,18)} = 11.12, p < .001$) and problem number ($F_{(23,138)} = 0.10, p = .474$) during recordings. **(E)** *Top:* No significant difference in the mean number of trials animals took to reach the threshold for a reversal during recordings between PFC and CA1 animals ($t_(7) = 0.42, p = .690$). *Bottom:* No significant difference in the mean number of out of sequence pokes during recordings between PFC and CA1 animals ($t_{(7)} = 1.39, p = .220$). **(F)** Average number of trials following a reversal taken to reach the threshold to trigger the next reversal, as a function of session number during recordings. **(G)** Average number of out of sequence pokes mice made as a function of problem number during recordings. **(H)** All out of sequence pokes plotted as a function reversal block number within problem. **(I)** Out of sequence A/B choice pokes plotted as a function reversal block number within problem. **(H, I)** Mice made more out of sequence pokes per trial during the first reversal block compared to every following reversal (Reversal 1 vs Reversal 2: $t_{(167)} = 2.61, p = .015$, 1 vs 3: $t_(167) = 3.84, p = .001$; 1 vs 4: $t_{(167)} = 4.15, p < .001$).

the first layout type, the initiation port (I1) was the top or bottom port, and the choice ports were the far left and far right ports. One of these choice ports remained in the same location in all three layouts used in a session, and will be referred to as the A choice. This acted as a control for physical location, allowing us to assess how the changing context of the different tasks affected the representation of choosing the same physical port. Both the other choice port (B choice), and the initiation port, moved physical locations between tasks. In the second layout type, both the initiation port (I2) and B choice port (B2) were in locations that were not used in layout type 1. In the third layout type, the initiation port was the same as the initiation port in layout type 1 (I3 = I1), and the B choice port was the same as the initiation port from layout type 2 (B3 = I2). Hence, in every recording session, we had examples of (1) the same port playing the same role across tasks, (2) different ports playing the same role across tasks and (3) the same port playing different roles across tasks (I3 and B2). In addition to these task design controls, we further ruled out the possibility that either potentially invariant aspects of task execution (e.g. velocity and acceleration) or task/trial specific features (animal's precise nose position) might explain differences in representations between different brain areas by combining pose estimation software and general linear methods (for more details see *Controls for Physical Movement Methods*). The order of the layout types was randomised in each recording session.

Abstract and task-specific representations of trial events by PFC and CA1 units

As animals transferred knowledge of the trial structure across problems, we postulated that certain neurons could encode a high-level understanding of the trials, focusing on key stages such as initiation, choice, and outcome, while disregarding the sensorimotor details unique to each task. We discovered that such 'problem-general' neurons were predominantly found in PFC (Figures 7A and 8A) In contrast, neurons in the CA1 region displayed more varied responses. While some did exhibit 'problem-general' activity, their responses were notably more varied across tasks compared to PFC neurons (see Figures 7B and 8B). Intriguingly, certain CA1 neurons maintained consistent firing patterns at the same physical location, irrespective of changes in its functional role within different tasks.

Other CA1 neurons 'remapped' between problems, changing their tuning with respect to both physical location and trial events.

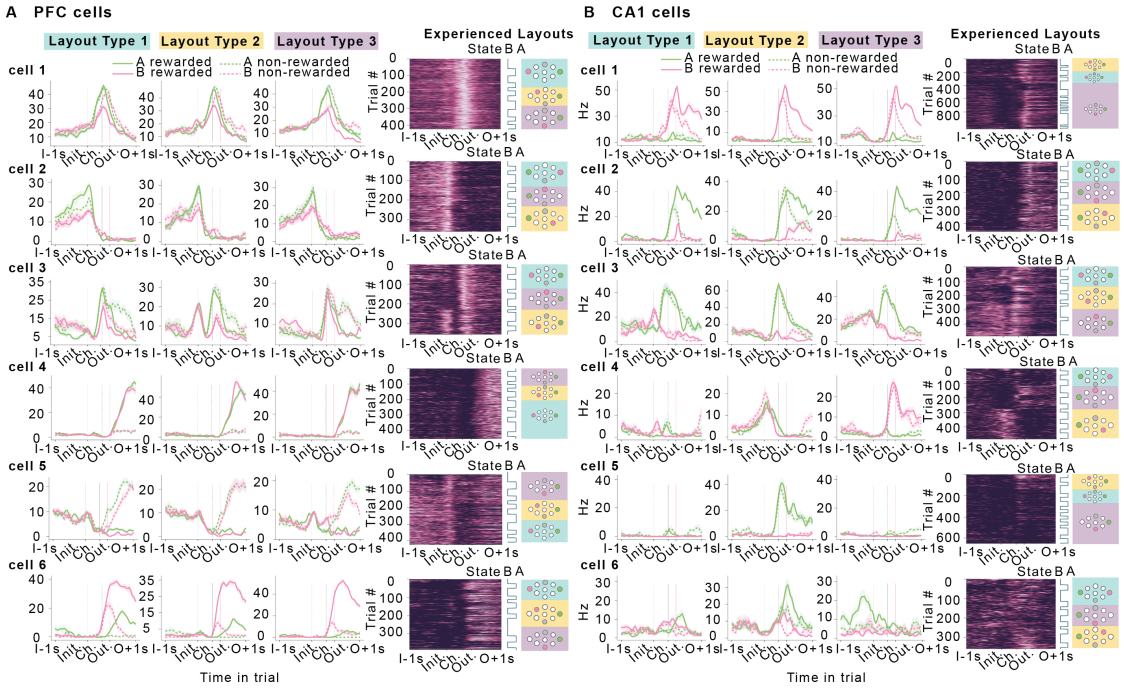


Figure 7: Example neurons in behavioural task. **(A)** Example PFC neurons. Cell 1 in PFC fired selectively to both choice ports (but not initiation) in each problem, even though the physical location of the choice ports was different both within and across problems. Cell 2 fired at the initiation port in every problem, even when its physical location changed. Cell 3 fired at B choice ports in all problems but also gained a firing field when initiation port moved to the previous B choice port (showing that PFC does have some port-specific activity). Cell 4 responded to reward at every choice port in every problem. Cell 5 responded to reward omission and had high firing during the ITI. Cell 6 responded to reward at B choice port (that switched location) in each problem. **(B)** Example CA1 neurons. Some CA1 cells also had problem general firing properties (cells 1 and 2). Cell 1 fired at B choice that switched physical location between problems. Cell 2 responded to the same port in all problems and modulated its firing rate depending on whether it was rewarded or not. Cell 3 fired at the same port in all layouts. Cell 4 switched its firing preference from initiation to B choice that shared physical locations, analogous to 'place cells' firing at a particular physical location. Cells 5 and 6 'remapped', showing interactions between problem and physical port. Cell 5 fired at a given port in one layout but not when the same port was visited in a different layout. Cell 6 fired at choice time at a given port in one layout and changed its preferred firing time to pre-initiation in a different layout. In all plots, average firing rates are arranged by layout types 1, 2 and 3, but the order in which they were experienced is plotted in the 'Experienced layouts' sub-panel. Error bars show firing rates ± s.e.m. across trials.

To dive deeper into the dynamics between problem-general and port-specific neuronal representations, we carried out a more focused analysis. We first categorized neurons according to their times of peak activity in Layout Type 2 (Figure 9 for details). We then scrutinized their activity across Layout Types 2 and 3. It is worth noting that in Layout Type 3, the port designated as 'B Choice' served a different role, specifically

as the 'Initiation' port, in Layout Type 2. To establish a baseline, we also examined 'A Choices', which were consistent in their physical location across both layouts. Our analysis showed that CA1 neurons had a stronger predisposition towards port-specific activity when compared to those in the PFC.

These single unit examples suggest that although task general representations might exist in both regions, PFC activity appears to generalise more across tasks, while CA1 represents physical location more strongly, and additionally exhibits 'remapping' between tasks in which neurons change their tuning to both physical location and task events

PFC population activity generalises more strongly across tasks than CA1

To assess whether our single unit observations hold up at the population level, we sought to characterise how neural activity in each region represented task events, and how these representations generalised across tasks.

We first assessed the influence of different task variables in each region using linear regression to predict spiking activity of each neuron, at each time point across the trial, as a function of the choice, outcome, and *outcome* x *choice* interaction on that trial (Figure 10A). Since the progression through trials (initiation $\rightarrow$ choice) was determined by subject's behaviour, we first aligned electrophysiology signals across trials by linearly time-warping the signal (for more details see *Time Warping Methods*). We then quantified how strongly each variable affected population activity as the population coefficient of partial determination (i.e. the fraction of variance uniquely explained by each regressor) at every time point across the trial (Figure 10B). This analysis was run separately for each task in the session and the results were averaged across tasks and sessions. Both regions represented current choice, outcome, and choice x outcome interaction, but there was regional specificity in how strongly each variable was represented. Choice (A vs B) representation was more pronounced in CA1 than PFC (peak variance explained - CA1: 8.4%, PFC: 4.8%, $p < .001$), whereas outcome (reward vs no reward) coding was stronger in PFC (peak variance explained – CA1: 7.1%, PFC: 12.9%, $p < .001$). Furthermore,

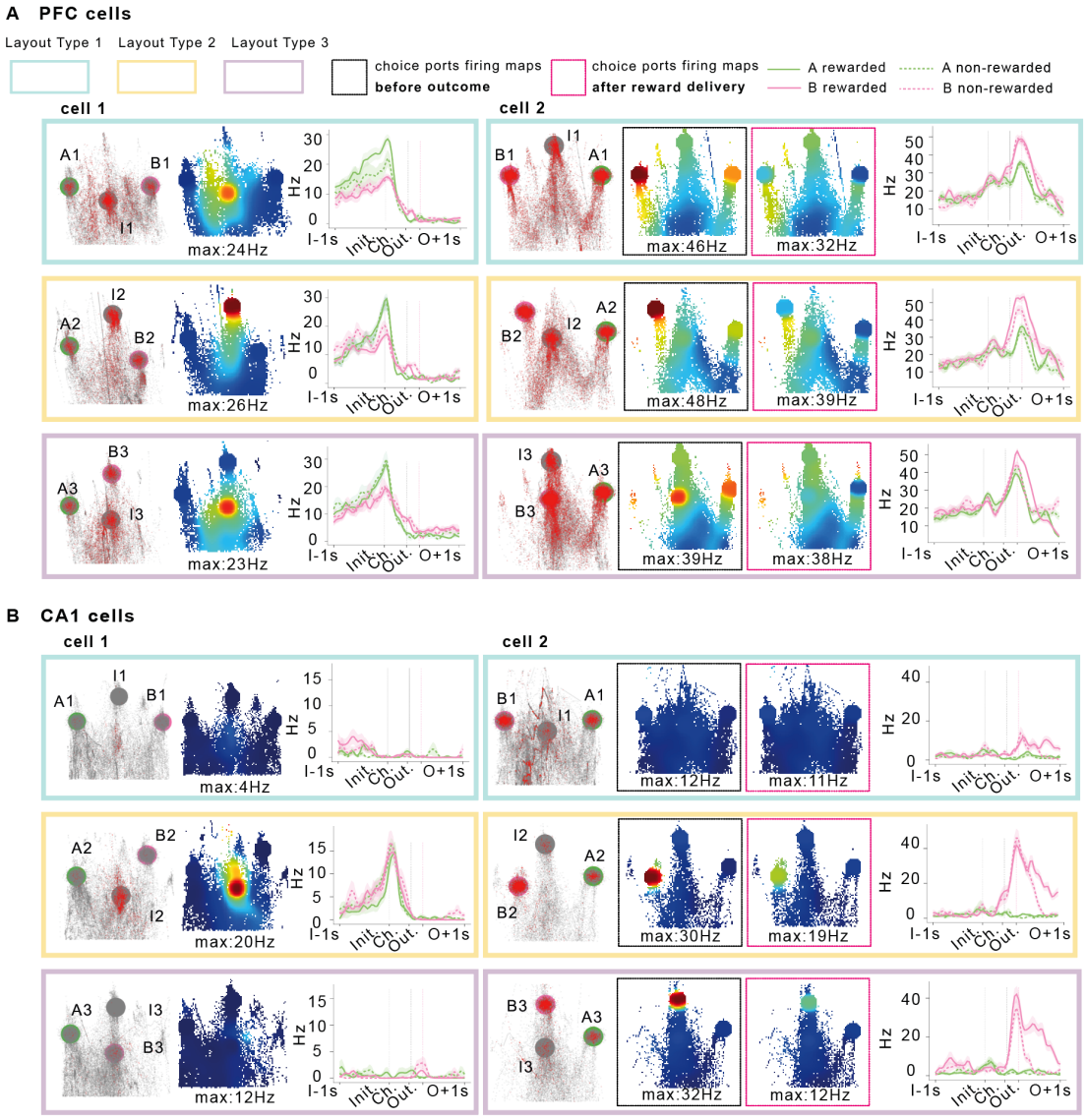


Figure 8: Example neurons in physical space and behavioral task. (A) Example PFC neurons. For each cell, left panels show trajectories of animals' nose (gray) and locations where spikes occurred (red) in a 2D space corresponding to the view of a camera positioned above the box looking at the ports, affine transformed to correct for the oblique view of the ports. Middle panels show firing rate heat maps in this same 2D space. Right panels show average firing rates across the trial for each trial type. Layout types are indicated by the color of boxes (blue, yellow and purple). Cell 1 fired at the initiation port in every problem even when its physical location changed. Cell 2 fired at all choice ports in all problems. For choice port selective cells (PFC and CA1 cell 2), we split the firing rate maps by whether the within-choice port spikes (and occupancies) occurred at times before outcome (left) or during reward consumption (right) to further show that these cells are selective to trial events. **(B)** Example CA1 neurons. Cell 1 fired at the bottom initiation port in layout type 2 but not when this same port acted as a B choice in layout type 3 or when the port was not a part of the current problem but was visited in layout 1. Cell 2 fired at one of the B ports in one layout type 3 and had no selectivity to the same port in layout type 2 when this port was an initiation port and, instead, fired at a different B choice in layout type 2. Error bars show firing rates ± s.e.m. across trials.

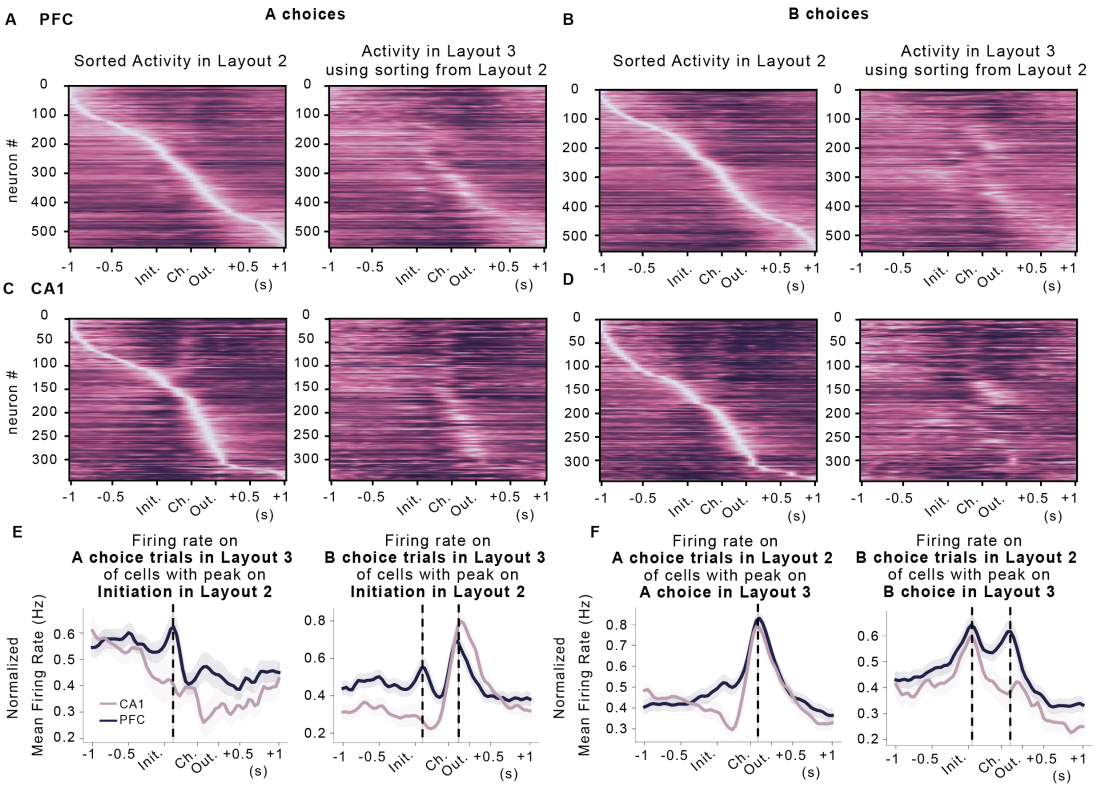


Figure 9: Port selectivity in CA1 than PFC. (A) PFC activity on A choice trials in Layout Type 2 (left) and Layout Type 3 (right). **(B)** PFC activity on B choice trials in Layout Type 2 (left) and Layout Type 3 (right). **(C, D)** Same as A-B but for CA1. **(E)** We identified cells that had their peak firing rate around initiation in Layout Type 2 and plotted their average activity in Layout Type 3 on A (left) and B (right) choice trials. **(F)** We identified cells that had their peak firing rate around A (left) or B (right) choices in Layout Type 3 and plotted their average activity in Layout Type 2 on A (left) and B (right) choice trials. CA1 neurons that fired at initiation in Layout Type 2 fired at B choice in Layout type 3, and vice versa, indicating that they primarily represented physical port. PFC neurons that fired at initiation time in Layout Type 2 generalised to initiation time in layout type 3 but also had a peak at B choice, and vice versa, indicating influence of both task-general and port-specific representations. Error bars on all plots report mean firing rates ± SEM across cells.

choice x outcome interaction explained more variance in CA1 than PFC (peak variance explained – CA1: 3.7%, PFC: 2.4%, $p < .001$).

Though highlighting some differences in population coding between regions, this approach cannot assess the relative contribution of abstract representations that generalise across tasks versus task specific features such as the physical port location. This requires comparing activity both across time points in the trial and across tasks, which we did using representational similarity analysis (RSA[165]). We extracted firing rates around initiation and choice port entries (+/- 20ms around each port entry type) and categorised these windows by which task they came from, whether they were initiation or choice, and

- for choice port entries whether the choice was A or B and whether it was rewarded - yielding a total of 15 categories (Figure 10C). For each session we computed the average activity vector for each category, then quantified the similarity between categories as the correlation between the corresponding activity vectors. We show RSA matrices for this 'choice time' analysis (Figure 10C, left panels), and an 'outcome time' analysis (Figure 10C, right panels) where the windows for choice events were moved 250ms after port entry, holding the time window around trial initiations constant.

To quantify the factors influencing representation similarity, we created representational similarity design matrices (RDMs) which each encapsulated the predicted pattern of similarities under the assumption that activity was influenced by a single task feature (Figure 10D). For example, if the population activity represented only which physical port the animal was at, its correlation matrix would look like Figure 10D, Port. We included design matrices for a set of task-general features; the trial stage ('Initiation vs Choice'), choice (A vs B), trial outcome (both on its own as 'Outcome', and in conjunction with choice 'Outcome at A vs B'). Changes in activity across tasks might occur simply due to neurons being tuned for particular physical locations, which will be captured by the 'Port' RDM. However, it is also possible that the changing context provided by different tasks modifies the representation of choosing the same physical port at the same trial stage. To assess such 'remapping', we included an RDM 'Choice A task specific' which modelled task specific representations of the A choice, which shares the same physical location and meaning across tasks. We modelled the observed pattern of similarities in the data as a linear combination of these RDMs, quantifying the influence of each by its corresponding weight in the linear fit. To be able to examine the temporal evolution of these effects we run a series of regressions onto the data. In each, the data around initiation port entry was the same but the data around the choice port entry progressed serially through time from choice point until after the reward was delivered (Figure 10E).

Consistent with our single unit observations, both PFC and CA1 represented both task specific and task general features to some extent. However, there was a marked regional specificity in how strongly different features were encoded (Figure 10E). PFC had stronger,

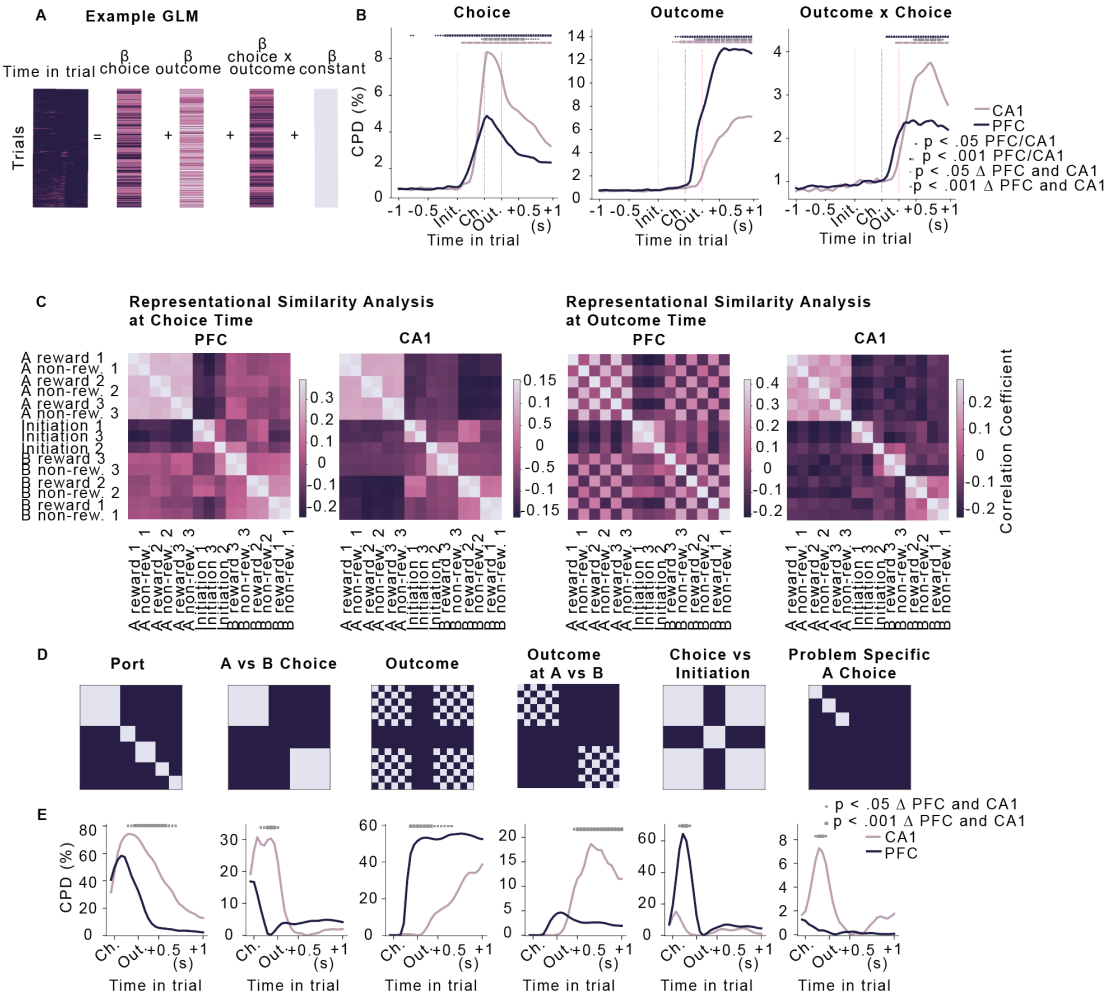


Figure 10: PFC population activity generalises more strongly across tasks than CA1. **(A)** Linear regression predicting activity of each neuron at each timepoint across the trial, as a function of the choice, outcome and outcome × choice interaction. **(B)** CPDs from the linear model shown in a for choice, outcome and outcome × choice regressors in PFC and CA1. Significance levels for within-region effects were based on a two-sided permutation test where firing rates were rolled with respect to trials. Significance levels for differences between regions were based on a two-sided permutation test across sessions. All significance levels were corrected for multiple comparisons over time points. **(C)** Representation similarity at 'choice time' (left) and 'outcome time' (right), quantified as the Pearson correlation between the demeaned neural activity vectors for each pair of conditions. **(D)** RDMs used to model the patterns of representation similarity observed in the data. Each RDM codes the expected pattern of similarities among categories in c under the assumption that the population represents a given variable. The Port RDM models a representation of the physical port poked (for example, far left) irrespective of its meaning in the trial. A vs B choice models a representation of A/B choices irrespective of physical port. The Outcome RDM models representation of reward versus reward omission. The Outcome at A vs B RDM models separate representations of reward versus omission after A and B choices. Choice vs Initiation models representation of the stage in the trial. Problem-specific A choice models separate representation of the A choice in different problems. **(E)** CPDs in a regression analysis modelling the pattern of representation similarities using the RDMs shown in d. The time course is given by sliding the windows associated with choices from being centred on choice port entry to 0.76 seconds after choice port entry while holding time windows centred on trial initiations fixed. Stars indicate timepoints where regression weight for each RDM was significantly different between the two regions ($p < 0.05$ (small stars) and $p < 0.001$ (big stars)), from one-sided permutation tests across sessions corrected for multiple comparison over time points (see *Statistical Significance Methods*).

abstract, sensorimotor-invariant representation of trial stage (initiation vs choice) and trial outcome ($p < .001$). In contrast, CA1 had stronger representation of the physical port the subjects was poking, and whether it was an A vs B choice ($p < .001$). Additionally, CA1 but not PFC showed a task specific representation of A choices ($p < .001$). This is striking because during A choices both the physical port and its meaning are identical across tasks, indicating that the changing task context alone induced some 'remapping' in CA1 but not PFC. Finally, there was a regional difference in the representation of trial outcome. PFC outcome representations were more general (the same neurons responded to reward or reward omission across ports and tasks – $p < .001$). CA1 also maintained an outcome representation, but this was more likely to be conjunctive than in PFC – different neurons would respond to reward on A and B choices ($p < .001$).

To exclude the possibility that task specificity in CA1 might be driven by CA1 representations drifting slowly over time we quantified the change in activity across transitions between problems using a 'surprise' measure indicating how unexpected the population activity was on each trial and time-point given the average activity at that time-point across 'baseline' trials (see *Surprise Measure Methods*). We confirmed that task representation changed abruptly at transitions between tasks (Figure 11). In CA1, when an initiation or choice port moved to a different physical location, the neuronal representation at the corresponding stage of the trial changed immediately, as indicated by an abrupt increase in surprise at the layout transition (Figure 11A-F).

To further show that PFC neurons predominantly represented trial stages (initiation, choice, outcome) in a task general way while CA1 neurons mainly had task specific representations we also conducted a traditional decoding analysis (see *Decoding Analyses Methods*). We trained a simple linear model to decode different phases of the trial (Initiation, A choice, A reward, A no-reward, B choice, B reward and B no-reward) on one of the tasks and tested how well it could decode these trial stages in a different task. PFC predominantly decoded to the correct trial stage (Figure 12). Where PFC made errors, they were predominantly to the other state that would have been valid at the sequential position in the trial (A rather than B choice or outcome). By contrast, CA1 predominantly

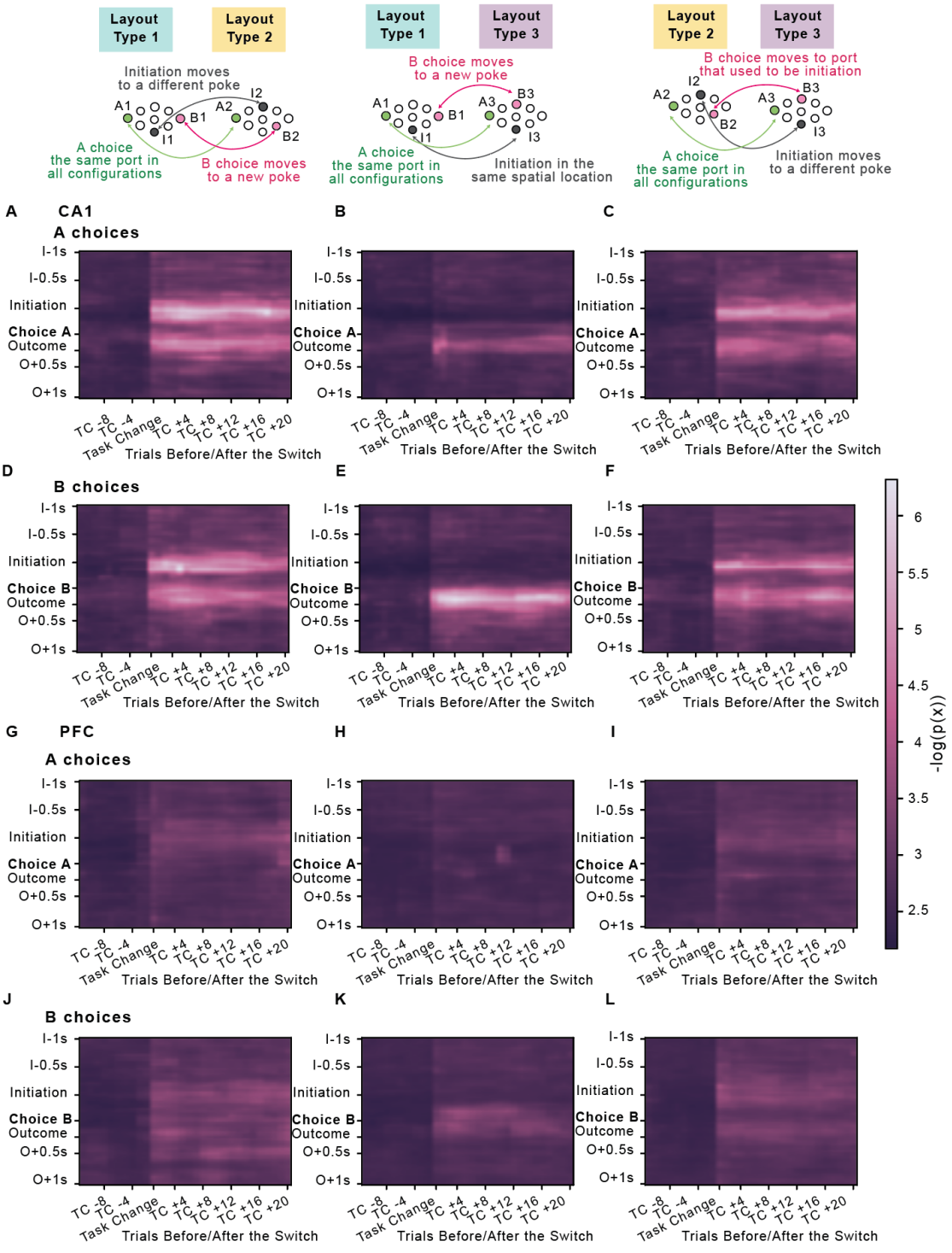


Figure 11: Rapid problem-induced 'remapping' in CA1 but not PFC. Heatmaps showing how surprising the activity at each time point of each trial was around transitions between problems. Activity in CA1 on A and B choice trials is shown in **(A-C)** and **(D-F)** respectively. Activity in PFC on A and B choice trials is shown in **(G-I)** and **(J-L)** respectively.

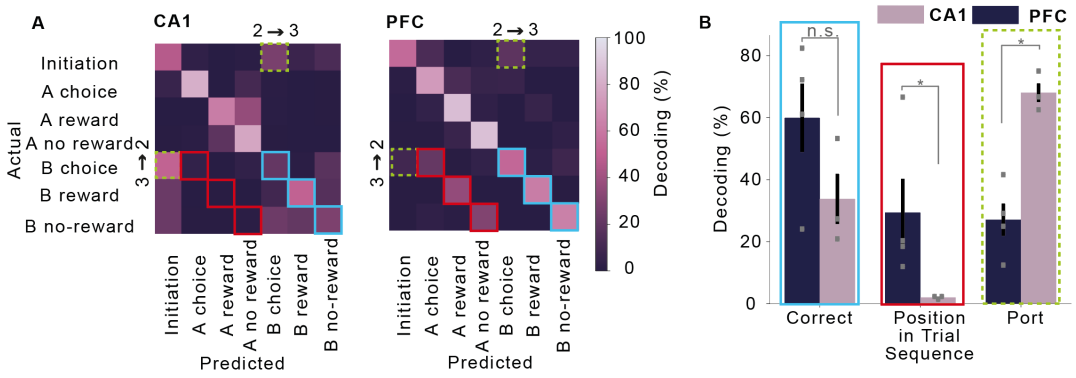


Figure 12: Decoding of trial stages across tasks. **(A)** Confusion matrices from linear decoding of position in trial, using a decoder that was trained on one problem and tested on another, averaged across animals and across all problem pairs. Coloured squares indicate three possible patterns of decoding that indicate different neuronal content. Blue indicates correct cross-task decoding to the same abstract state (for example, B choice decodes to B choice). Red indicates decoding to a different state that could have occurred at the same sequential position in the trial (for example, B choice decodes to A choice). Dashed green corresponds to decoding to the same physical port for those training and test layouts where the Initiation and B choice ports interchanged (for example, B choice decodes to Initiation when the decoder was trained on layout 2 and tested on layout 3). **(B)** Bar plots showing the probability of the cross-task decoder outputting the correct abstract state (blue), the other state that can have the same position in the trial sequence (red) and the state that has the same physical port as the training data (dashed green, computed only from confusion matrices where B choice and initiation ports interchange) computed using the corresponding cells highlighted in A. Error bars report the mean ± s.e.m. across different mice (CA1: n = 3 mice; PFC: n = 4). Significance levels were compared against the null distribution obtained by shuffling animal identities between regions (one-sided permutation tests). NS, not significant.

decoded to the same physical port as the training. Together these population results confirm that PFC had a predominantly generalising representation, and this representation embeds the sequential properties of the task while CA1 encoded task specifics (such as port identity) more strongly.

3.3 Discussion

The primary focus of this chapter was to interrogate how the prefrontal cortex and hippocampal CA1 regions of the brain represent and generalise the sequence of trial events across different reversal-learning problems. We observed that both regions contained neurons that demonstrated 'task-general' representations of abstract trial stages, such as initiation, choice, and outcome. However, PFC neurons generalised more strongly across tasks and encoded abstract meanings of different stages of the task (initiation, choice, and outcome), which might underlie animals' ability to rapidly understand how

to perform a trial on any set of physical ports. In contrast, CA1 neurons were more 'task-specific' and exhibited 'remapping' between tasks, whereby neurons changed their tuning to both physical location and task events[44]. This might allow the hippocampus to keep track of specific contextual and sensory details associated with different tasks and link them with the more abstract meanings of trial phases.

The identification of a common representation of the sequential structure of different states (and associated actions) aligns with theoretical accounts of cognitive maps - in which abstracting task structure is critical for behavioural performance[3,166]. Such theories suggest that these abstractions need not be limited to representing exact sequences but can also abstract the *rules* and *statistical regularities* that constrain possible sequences. It is worth noting here that these ideas were, at least primarily, developed in the context of the entorhinal cortex[3]. While we did not record from entorhinal cortex in the current presented data, recent fMRI evidence in humans in a conceptually similar experiment suggests entorhinal representations will also generalise the structure of reinforcement learning tasks[50,51,154]. It is also notable that abstract representations of trials are present in mPFC in purely spatial contexts[159,160]. It will therefore be intriguing to develop an understanding of how these representations differ across entorhinal and prefrontal cortices, and if, as discussed in the *Introduction*, a hierarchy of representational abstraction exists across these regions.

Our study supports the proposal that *abstracted* or *schematic representations* of tasks in cortex might be linked flexibly with the sensorimotor characteristics of a particular environment to construct concrete task representations in the hippocampus. This might allow for immediate inferences when new task rules must be inferred, either at the beginning of training or when task contingencies change. To investigate this possibility more directly, future studies might consider examining how hippocampal and prefrontal representations of trial events are affected when the trial structure of the task changes slightly - for example, if animals need to initiate, choose, and then go back to initiate to receive a reward. It would be particularly interesting to track the representational change

that occurs in the hippocampus and prefrontal cortex when animals encounter these novel trial structures.

Future studies could also explore the causal relationships between these regions during reversal learning tasks. For example, optogenetic manipulation of PFC or hippocampal activity could be used to assess the respective impact of these regions on the other, and interrogate the effects of these manipulations on behavioural performance throughout learning. Some intriguing (although slightly tangential) evidence has already been presented in this regard, in a study by Ito et al.[141]. Here, rats performed a *spatial* alternation task, resulting in hippocampal representations which fire alternately on left and right turns (these representations, commonly known as 'splitter cells', have been widely described previously[63]). These trajectory-dependent splitter cell representations were observed in both medial prefrontal cortex and the CA1 region of the hippocampus. Notably, however, when the thalamic connection bridging these regions (the nucleus reuniens) was lesioned or optogenetically silenced, these trajectory-dependent representations persisted in mPFC, but were ablated in CA1, perhaps pointing towards a pathway by which an abstract representation (such as a splitter cell) might be inherited from prefrontal areas for use in the hippocampus. Finally, investigating how neural representations change during the learning process, as animals become more familiar with the task structure, could provide further insights into how the brain constructs abstract and concrete representations of tasks[76].

Overall, the findings presented in this chapter have important implications for understanding the neural basis of flexible behaviour and learning. The differences in the level of generalisation between these regions highlight the complementary roles they might play in encoding and integrating different aspects of task-related information.

4

Generalisation of low dimensional representations

Contents

4.1 Introduction **65**
4.2 Results **68**
Studying generalization of low-dimensional representations using SVD 68
Temporal structure of activity is invariant across tasks and regions . . 69
Cell assemblies generalise more strongly in PFC than CA1 72
4.3 Discussion **73**

4.1 Introduction

The goal of neural population analysis is to gain insight into how groups of neurons collectively perform computations which underlie observed behaviours. This is a challenging task, as neural population responses are typically high-dimensional *, with many neurons firing simultaneously and information being distributed throughout structural and functional sub-networks[167]. In order to make such data more tractable (and amenable to intuitive understanding), neuroscientists often turn towards *dimensionality reduction* techniques[167]. Indeed, these techniques are often valuable for intuitively

*Indeed, complete statistical description of a network of n neurons would require n dimensions.

understanding and visualising some preserved features of the high-dimensional data. However, the utility of dimensionality reduction methods is not limited to this; they can also offer profound insights into the fundamental computation being performed by a population of neurons (the elegant oscillatory activity of motor cortex neurons when monkeys reach towards an object is perhaps one of the most famous examples of this[168]).

One common approach to dimensionality reduction is to project the data onto a lower-dimensional space while aiming to preserve its relevant features. This technique has been used to reveal structure in neural population activity, providing insights into the neural code and neural computations[167]. For instance, in the auditory cortex, neural population activity rotates in a low-dimensional space determined by a stimulus-dependent initial condition[169]. In the frontal cortex, the initial state of the neural population depends on a subject's prior expectations, allowing the dynamics of the network to combine these expectations with sensory evidence in an interpretable Bayesian manner[170]. The value of dimensionality reduction in these and other cases is that, while neural activity cannot be fully explained by two or three dimensions, these dimensions often explain a lot of the *variance* of the neural population activity and relate to behaviours in an interpretable manner[168]. By identifying patterns and structure in the data, we can begin to understand how neural populations work together to perform flexible computations, which are often hidden by the 'curse of dimensionality' of the data[171].

However, a critical question arises once a low-dimensional representation has been identified: how well does it generalise to novel stimuli or tasks? In other words, can the same low-dimensional structure be used to explain population activity across different situations, or is it specific to the original set of stimuli or task statistics from which it derives?

To address this question, we investigated whether the low-dimensional structure of neuronal activity is invariant across different reversal-learning problems that all shared the same abstract structure, but were defined by differing sensory-motor correlates. To do this, we used the mathematical technique of *singular value decomposition (SVD)* to break

down the activity of different groups of neurons into cellular and temporal *modes* and investigate how these generalise between different tasks. Cellular modes can be thought of as groups of neurons that work together in a coordinated manner, while temporal modes can be interpreted as patterns of neural activity that vary over time.

To help illustrate this concept, consider a football team. Just as a team is composed of different players in different positions who all work together to achieve a common goal, a neural population is made up of different cells that work together to perform computations. If we think of each player as a neuron, then a cell assembly (*cellular mode*) is analogous to a group of players with the same position, such as the offensive line or the wide receivers. Just as the players in each position group work together to perform specific tasks, the neurons in a cell assembly work together to perform specific computations. And just as the players on a football team need to adapt to different opponents and game situations, the neural population might adapt to different stimuli or tasks at hand. Similarly, the temporal modes of neuronal activity can be thought of as different sequences of play that the team uses to achieve different goals. For example, a team might use quick passing play to move the ball down the field, or a long-bomb play to score a touchdown. In the same way, neurons might use different patterns of activity to perform different computations over time.

In this chapter, I will outline the use of SVD to analyse neural population activity during different reversal-learning tasks, and how these results can provide us with some insight into the 'generalisability' of low-dimensional representations across different tasks that share the same structure but have different sensory-motor correlates. In the discussion section, I will discuss how our results relate to existing findings in the literature, as well as highlight some of the limitations of this approach and the potential for future research.

4.2 Results

Studying generalization of low-dimensional representations using SVD

Using singular value decomposition (SVD), we decomposed activity in each task into a set of cellular and temporal 'modes' (Figure 13A). Cellular modes correspond to sets of neurons whose activity covaries over time, and hence can be thought of as cell assemblies. Each cellular mode is specified by a vector with a weight for each cell, indicating how strongly the cell participates in the mode. Cellular and temporal modes come in pairs, such that each cellular mode has a corresponding temporal mode, which is a vector of weights across time-points indicating how the activity of the cellular mode varies over time.

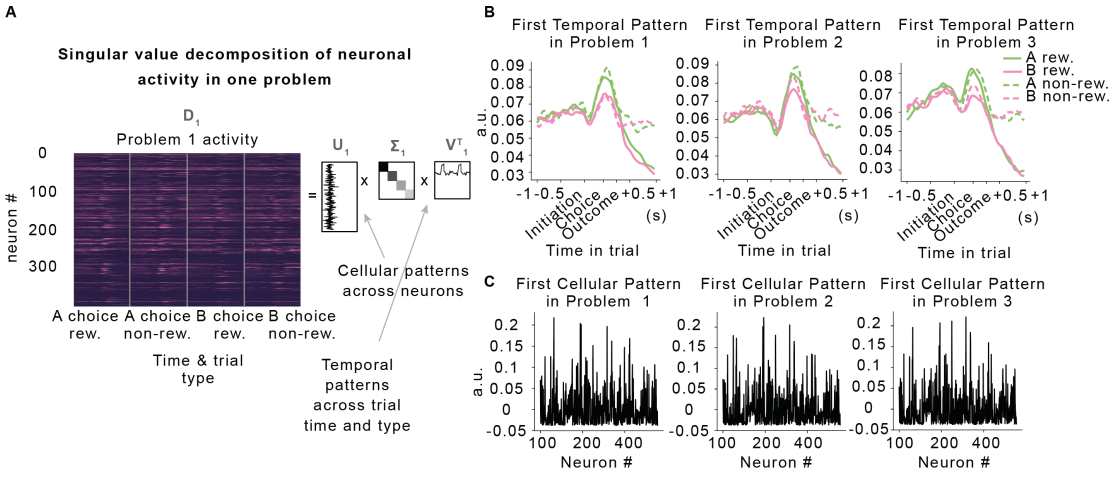


Figure 13: Singular Value Decomposition (SVD). (A) Illustration of SVD analysis. A data matrix comprising the average activity of each neuron across timepoints and trial types was decomposed into the product of three matrices, where diagonal matrix Σ linked a set of temporal patterns across trial type and time (rows of V_T) to a set of cellular patterns across cells (columns of U). **(B)** First temporal mode in V_T from SVD decomposition of data matrix from PFC plotted in each problem separately for clarity and separated by (green) and B (pink) rewarded (solid) and non-rewarded (dashed) choices. **(C)** First cellular mode from SVD decomposition of data matrix from PFC in each problem showing that similar patterns of cells participate in all problems.

To evaluate the cellular and temporal modes for a given task, we computed for each cell the average firing rate at each time point across the trial, for four types of trials – rewarded A choices, non-rewarded A, rewarded B, and non-rewarded B (non-rewarded trials included both correct trials (20% chance of no reward) and incorrect trials). For each cell, we

concatenated these four time series to create a single time series containing the average activity of the cell across each time point of the four trial types. The temporal modes span this same set of time points and hence capture variation across both time-in-trial and trial-type. We then stacked these single cell activity time series for all neurons, to create an activity matrix D where each row contained the activity of one neuron (Figure 13A). Using SVD, we decomposed this activity matrix into cellular and temporal modes U and V, linked by a diagonal weight matrix Σ.

$$D = U\Sigma V^T \tag{4.1}$$

The cellular modes are the columns of U and the temporal modes are the rows of V_T. Both modes are unit vectors, so the contribution of each pair to the total data variance is determined by the corresponding element of the diagonal matrix Σ. The modes are sorted in order of explained variance, such that the first cellular-temporal mode pair explains the most variance. The first cellular and temporal mode of PFC activity in three different tasks is shown in (Figure 13B, C). It is high throughout the ITI and trial with a peak at choice time, but strongly suppressed following reward (similar to cell 5 in Figure 7A).

We reasoned that: (i) if the same events were represented across tasks (for instance, $\mathrm{initiation}$, $\mathrm{A/B\ choice}$, and reward), then the temporal modes would be exchangeable between tasks, no matter whether these representations were found in the same cells; (ii) if the same cell assemblies were used across tasks, then the cellular modes would be exchangeable across tasks, no matter whether the cell assemblies played the same role in each task; and (iii) if the same cell assemblies performed the same roles in each task, then pairs of cellular and temporal modes would be exchangeable across tasks.

Temporal structure of activity is invariant across tasks and regions

To see whether the same representations existed in each task, we first asked how well the temporal modes from one task could be used to explain activity from other tasks. Since the set of temporal modes V is an orthonormal basis, any data of the same rank or less can be perfectly explained when using all the temporal modes. However, population activity

in each task is low dimensional so a small number of modes explain a great majority of the variance. Modes that explain a lot of variance in one task will only explain a lot of variance in the other task if the structure captured by the mode is prominent in both tasks. The question is therefore how quickly variance is explained in task 2's data, when ordering the modes according to variance explained in task 1. To assess this, we projected the data matrix D_2 from task 2 onto the temporal modes V_1 from task 1, giving a matrix M_V whose elements indicate how strongly each temporal mode contributes to the task 2 activity of each neuron.

$$M_V = D_2 V_1 \tag{4.2}$$

The total variance explained by each temporal mode is given by squaring the elements of M_V and summing over neurons. We plot the cumulative variance explained as a function of the number of temporal modes used (Figure 14A). To control for drift in neuronal representations across time, we computed the data matrices separately for the first and second halves of each task. We compared the amount of variance explained using modes from the first half of one task to model activity in the second half of the same task, with the variance explained using modes from the second half of one task to model activity from the first half of the next task.

In both PFC and CA1, the cumulative variance explained as a function of the number of temporal modes used, did not depend on whether the two data sets were from the same task (solid) or different tasks (dashed) (Figure 14E, $p > .05$). This indicates that the temporal patterns of activity, and therefore the trial events represented, did not differ across tasks in either brain area. However, as this analysis used only the temporal modes, it says nothing about whether the same or different neurons represented a given event across tasks. In fact, we can even explain activity in one brain region using temporal modes from another region and mouse (Figure 14B).

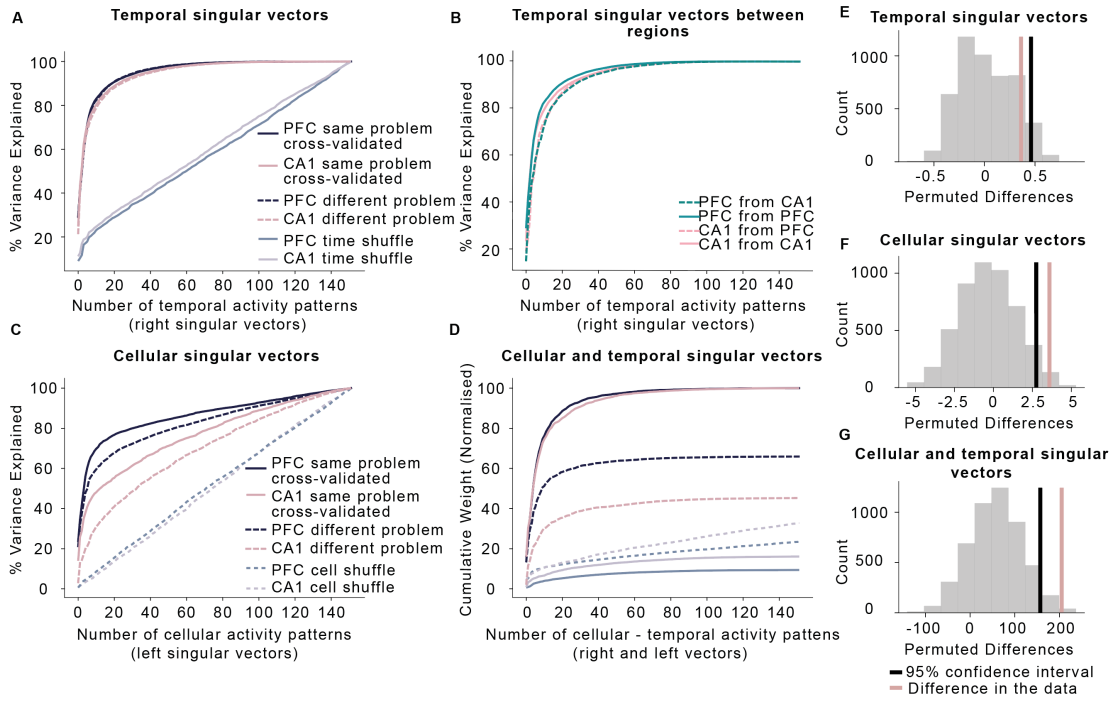


Figure 14: Generalisation of low-dimensional activity patterns of trial events. (A) Variance explained when using temporal activity patterns V_1^T from one problem to predict either held-out activity from the same problem (solid lines) or activity from a different problem (dashed lines). Light purple and lilac lines indicate variance explained when shuffling timepoints in the firing rates matrices. **(B)** Variance explained when using temporal activity patterns V_1^T to predict either activity from the same problem and brain region (solid lines) or a different brain region (and, therefore, different animal) and the same problem (dashed lines) D_2. **(C)** Variance explained when using cellular activity patterns U_1 from one problem to predict either held-out activity from the same problem (solid lines) or activity from a different problem (dashed lines). Dashed light purple and lilac lines indicate variance explained when shuffling cells in the firing rates matrices. **(D)** Cumulative weights along the diagonal Σ using pairs of temporal V_1^T and cellular U_1 activity patterns from one problem to predict either held-out activity from the same problem (solid lines) or activity from a different problem (dashed lines). Weights were normalized by peak cross-validated cumulative weight computed on the activity from the same problem. **(E)** To assess whether the temporal singular vectors generalised significantly better between problems in PFC than CA1, we evaluated the area between the dashed and solid lines in d for CA1 and for PFC separately, giving a measure for each region of how well the singular vectors generalised. We computed the difference in this measure between CA1 and PFC (pink line in **A**) and compared this difference to the null distribution obtained by permuting sessions between brain regions (gray histogram; black line shows the 95^{th} percentile of distribution). Temporal singular vectors generalised equally well between problems in the two regions. **(F)** Cellular singular vectors generalised significantly better between problems in PFC than CA1. Computed as in **(E)** but using the solid and dashed lines from **(C)**. **(G)** Pairs of cellular and temporal singular vectors generalised significantly better between problems in PFC than CA1. Computed as in **(E, F)** but using the solid and dashed lines from **(D)** a.u., arbitrary units.

Cell assemblies generalise more strongly in PFC than CA1

The pattern was very different when we used cellular modes (i.e. assemblies of co-activating neurons) from one task to explain activity in another. We quantified variance explained in task 2 using cellular modes from task 1, by projecting the task 2 data matrix D_2 onto task 1 cellular modes U_1, giving a matrix M_U whose elements indicate how strongly each cellular mode contributes to task 2 the activity at each timepoint.

$$M_U = U_1^T D_2 \tag{4.3}$$

The total variance explained by each temporal mode is given by squaring the elements of M_U and summing over timepoints. In both PFC and CA1, cellular modes in U that explained a lot of variance in one task, explained more variance in the other half of the same task than they did in an adjacent task (Figure 14C - differences between solid and dashed lines). However, the within task vs cross task difference was larger in CA1 than PFC (Figure 14F, $p < .05$). This indicates that PFC neurons whose activity covaried in one task were more likely to also covary in another task, when compared to CA1 neurons. As this analysis considered only the cellular modes it does not indicate whether a given cell assembly carried the same task information across tasks.

To assess how well the cellular-temporal activity patterns from task 1 explained activity in task 2, we projected data set D2 onto the cellular and temporal mode pairs of task 1 (U_1^T, V_1).

$$\Sigma_2 = U_1^T D_2 V_1 \tag{4.4}$$

If the same cell assemblies perform the same roles in two different tasks, the temporal and cellular modes will align, and Σ_2 will have high weights on the diagonal. We therefore plotted the cumulative weight of the diagonal elements of Σ within and between tasks (Figure 14D). In both PFC and CA1 cellular and temporal modes aligned better in different data sets from the same task (solid lines), than for different tasks (dashed lines). However, this difference was substantially larger for CA1 than PFC (Figure 14G, $p < .05$).

These data show that although the temporal structure of activity in both regions generalises perfectly across tasks, brain regions and subjects – a consequence of the same set of trial events being represented in each, the cell assemblies used to represent them generalised more strongly in PFC than CA1.

4.3 Discussion

Here we aimed to understand whether the low-dimensional representations in PFC and CA1 would be similar when the same task (reversal-learning) was performed in different sensory-motor situations. We found that, while both regions had similar patterns of brain activity that described the task (called 'temporal modes'), different neuronal assemblies participated in the representation of each task in CA1, but less so in PFC.

Recent data have highlighted the low-dimensional structure of task representations in rodent OFC[76]. In this chapter, we show that these low-dimensional temporal modes are also consistent across tasks in both mPFC and CA1. Furthermore, we confirm that they are consistent between animals and further demonstrate they are consistent between different brain areas (mPFC and CA1), suggesting this low-dimensional structure does not reflect the unique representational properties of a particular brain area.

Our analysis presented here makes further unique contributions. Because we record across the same neurons across different tasks, we are able to ask not only whether the temporal dimensions are preserved across tasks, but whether these temporal modes align to the same neurons in each task - that is, whether the same neurons represent the same trial events across tasks. *They do so significantly more in PFC than CA1.*

Another recent line of work has examined related ideas in primate PFC and hippocampus[152,172]. Where data are available from both structures in the same task, representations are found to be geometrically arranged in line with task coordinate space, and no clear differences are observed between structures. While it is tempting to postulate a species difference, careful examination reveals another possibility. Because these data are acquired in tasks that share no sensory elements (monkeys obersving visual stimuli, with no overlapping stimuli between tasks), and because motor coordinates are aligned to task

coordinates, it is not possible in these data to discern whether there is also a sensorimotor component to the hippocampal representation, as we observe in our data.

While this work provides a number of novel insights into the low-dimensional structure of neuronal activity and its invariance across different reversal-learning problems, there are several limitations to our approach. Firstly, we only investigated whether low-dimensional representations were shared between tasks with the same abstract structure. It remains to be seen how our findings extend to other types of abstract structures. Further research, such as performing recordings while animals complete two different tasks that do not share the same abstract structure could explore this topic further. Similarly, conducting an experiment to examine the low-dimensional representations as animals execute identical motor commands -integrated within *different abstract structures* - would present another interesting avenue for understanding how the brain incorporates both sensory-specific and abstract information to enable complex behaviour.

Additionally, while we utilised dimensionality reduction techniques to identify low-dimensional structure in the data, the relationship between this structure and actual behavioural performance remains to be fully explored. Our findings suggest that the low-dimensional structure of neural activity is consistent across different reversal-learning problems, but it is unclear whether this consistency is related to consistent behavioural performance across these problems. To further explore the mechanisms underlying low-dimensional representations, future studies could incorporate perturbation techniques to probe the causal relevance of these representations.

For example, an innovative study by Sadtler et al. (2014)[173] hypothesised that the relative difficulty of learning different tasks might be a consequence of the *relative difficulty of generating the neural population activity necessary to perform these tasks*. Here, Rhesus macaques learned to perform a series of tasks which involved moving a cursor on a screen, and were implanted with a brain-computer interface (BCI) which could be used to perturb the neural population activity in primary motor cortex. Firstly, it was observed that this activity in primary motor cortex had a low-dimensional 'intuitive manifold' of activity. When the BCI was used to perturb neural population activity within

this intuitive manifold, the macaques were able to learn new tasks with relative ease. When perturbations of a same magnitude were applied *outwith* the intuitive manifold, learning new tasks proved much more difficult. This indicates that such low-dimensional representations (or manifolds) might influence learning - potentially, a very valuable insight into the causality of neural representations and the behaviours they subserve. Incorporating similar perturbation studies into our research could provide further insight into the mechanisms underlying the low-dimensional representations observed in our study. It would be particularly interesting to study cognitive processes such as combinatorial creativity, where animals would be required to fuse seemingly unrelated behavioural motifs in novel ways[1]. This process might involve the generation of new, low-dimensional patterns of neural activity, while still residing within the intrinsic manifolds of the two pre-existing distinct motifs [†].

Despite these limitations, the findings presented in this chapter highlight the potential for using dimensionality reduction techniques to gain insights into the complex relationship between neural activity and the behavioural correlates of knowledge generalisation, and tame the 'curse of dimensionality' to discover interpretable representations in neural data.

[†]As mathematicians working in the field of topology have demonstrated, a manifold can itself be composed of *sub-manifolds*[174]. Intuitively, for instance, a *circle* S^1 satisfies the criteria to qualify as a 1-dimensional manifold, but is itself a sub-manifold of the Cartesian plane $\mathbb{R}^2$.

5

Neural representations underlying generalisation of policy

Contents

5.1 Introduction **76**
5.2 Results **78**
5.3 Discussion **81**

5.1 Introduction

So far, we have focused on the neural representations that might underpin animals' grasp of the trial sequence and their ability to adapt to new sensory instances of the same trial structure. However, in reversal learning tasks, it is also essential for animals to make well-informed decisions about which option to choose and to anticipate when reversals might occur. This process requires the transfer of information across trials, leading us to hypothesise the existence of an accompanying neural representation that would aid animals in determining when a change in reward contingencies is likely to take place.

In seeking to understand the neural representations that enable generalisation of *across*-trial structure of reversal-learning problems, it is crucial to consider the strategies animals

employ when navigating the task. In reversal-learning tasks, the shifting of reward between the two options necessitates the animals suppressing previously rewarded responses, and instead shift to making actions that were previously deemed unproductive. Hence, a prevalent idea for some time was that reversal learning paradigms were primarily a way to gauge the inhibitory control over responses[149]. This concept has been alternatively described as a measure of 'cognitive flexibility', 'behavioural flexibility', 'cognitive control', 'inhibitory control', 'impulse control', 'response inhibition', and 'behavioural inhibition'[175].

More recently, empirical work elucidating the learning process during reversal tasks has been increasingly influenced by the mathematical framework of *reinforcement learning (RL)*[122]. Classic reinforcement learning (RL) models support the process of behavioural adaptation by strengthening advantageous actions and reducing the significance of incorrect or unrewarded actions. This is accomplished by calculating prediction errors, which capture the disparity between the expected reward and the actual reward received[176]. A reversal in reward contingencies, where the previously rewarded action no longer leads to a reward, triggers a negative prediction error for the RL model for the formerly rewarded action, as the actual reward falls short of the expected reward. This error decreases the value estimate for that action, while positive feedback for an alternative action increases its value estimate. Over several trials, the model adapts to the new reward contingencies, effectively 'unlearning' the previous action-reward association and 'learning' the new one. In essence, RL models dynamically update action-reward associations to adapt to changing contingencies in the environment, providing a cogent simulation of the reversal learning process. These models have been successfully applied to reversal learning tasks, elucidating behavioural data across diverse species like rodents, macaques, and humans[94,146,177–179].

However, the primary limitation of classic RL models in the context of reversal learning is that they operate without utilising any representation of the task structure, such as rules that describe the relationship between actions, rewards, and external variables like time[122,180]. In reality, reward contingencies in reversal learning tasks change in predictable

ways, and understanding these abstract rules can enable animals to perform well at any new task. In line with this, previous studies have demonstrated that humans[181] and other animals[146,182,183] do not simply adapt to a change in reward contingencies through trial and error, but instead learn the high-order structure of the task (only one option is good at a time) and use this knowledge to enhance their decision-making. For example, animals appear to identify two temporal contexts through repetition, anticipate reversals, and adjust their behaviour accordingly[146,183].

In our paradigm, where the same reversal learning problem is introduced in varying sensory and motor contexts, learning the abstract structure of the task becomes particularly important. In line with this, we observed in Chapter 2 that animals 'learned to learn' to track their history of choices and rewards to make better choices (i.e. policy) as they encountered each new problem, implicating cognitive processes beyond simple inhibitory control. In the following chapter, we explore whether there is a corresponding neural representation that tracks this policy across trials, enabling animals to determine the currently optimal choice. Furthermore, we investigate whether this representation exhibits a level of abstraction that might enable it to be applied in novel situations, even when there are differences in sensory experiences across tasks.

5.2 Results

To obtain an estimate of subjects' beliefs about which option was best, we used a logistic regression predicting current choices as a function of the history of previous rewards, choices and their interactions (Figure 15A). This allowed us to compute, trial-by-trial, the probability that the animal would choose A vs B – that is, computing the animal's policy. When we used this policy as a predictor of neural activity, it explained variance that was not captured by within-trial regressors such as choice, reward and *choice* x *reward* interaction. Specifically, the subjects' policy interacted with the current choice explained variance (Figure 15B, $p < .001$). Notably, this signal became prominent around the time of trial initiation, when it would be particularly useful for guiding the decision.

To examine whether policy representations generalised across tasks, we evaluated the correlation across tasks between the policy weights in the neural regression. Because the A port was the same on each task, but the B port varied between tasks, we computed policy regression weights at each time point separately for A and B choices (controlling for reward). We then computed the average across-task correlation of these weights between every pair of timepoints (Figure 15C). The diagonal elements of these matrices show the average correlation across tasks at the same time point in each task. Visual inspection (Figure 15C), and permutations tests of differences between sums of the diagonals of Policy on A and B choices correlation matrices ($p < .05$), revealed that these correlations were larger in PFC than CA1 (Figure 15D). On average, therefore, cellular representations of policy generalised across tasks better in PFC than CA1 on both A and B choices.

One possible explanation is that PFC simply represented action values in a task-general way. A more interesting possibility is that current policy shapes the representation of each trial stage differently, but in CA1 these representations are more tied to the sensorimotor specifics of the current task. To test this, we examined time-slices through the correlation matrices at initiation, choice, and outcome times (Figure 15E). In PFC, all three correlation profiles on both A and B trials peaked at the correct time point (the equivalent to the diagonal elements of the matrix) – i.e. the policy representations generalised across problems, but were specific to the different parts of the trial ($\mathrm{initiate}$, choose, $\mathrm{outcome}$). A similar pattern was present in CA1, but only on A choices (which are the same physical port across tasks). No CA1 correlation was significantly above zero on B choices. Indeed, whilst PFC policy correlations were greater than CA1 correlations for all representations (all $p < .05$) on both A and B choices, CA1 correlations showed a greater difference between A and B trials at outcome time (Figure 15E, all $p < .05$).

Overall, therefore, both PFC and CA1 maintained representations of the subject's current policy that were not simple value representations – as they differed depending on the trial stage. These representations were abstracted across tasks in PFC, but tied to the sensorimotor specifics in CA1. A portion, but not all, of this task specificity in CA1 was accounted for by the port identity.

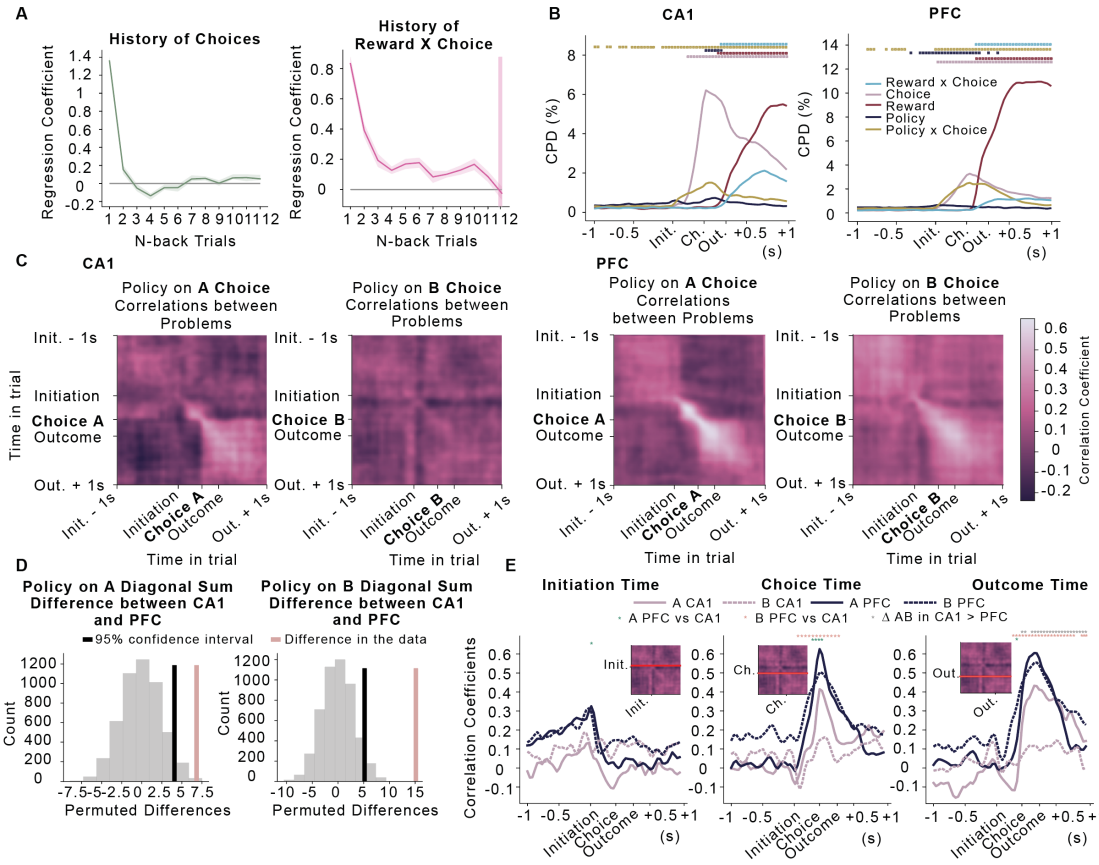


Figure 15: Policy Generalisation in PFC and CA1. **(A)** Weights from logistic regression predicting choices in recording sessions using choices, rewards and choice × reward interactions over the previous 12 trials as predictors. The effect of choice × outcome interaction history was significantly above zero on up to 11 trials back (one-sided t-test, $p < .05$) except for the 7th trial $t_{(6)} = 1.99, p = .094$. Error bars report the mean ± s.e.m. across mice. **(B)** CPDs from regression models predicting neural activity using current trial events, subjects' policy (estimated using the behavioral regression in a) and policy interacted with current choice. Stars denote the timepoints at which each regressor explained significantly more variance than expected by chance (permutation test based on rolling firing rates with respect to trials, $p < 0.001$, corrected for multiple comparisons; for more details on permutation tests, see the 'Statistical significance' section). **(C)** Correlations across problems between policy weights in regressions predicting neural activity. Regressions were run separately for A (left panels) and B (right panels) choices in each problem and at each timepoint across the trial. Correlations of policy representations between all problem pairs were evaluated for each pair of timepoints; values on the diagonal show how correlated policy representation was at the same timepoint in both problems. Positive correlation indicates that the same neurons coded policy with the same sign in both problems. **(D)** To quantify whether policy generalized more strongly between problems in PFC than CA1, we computed the between-region difference in the sum along the diagonal of the correlation matrices in c, separately for A and B choices, and compared it against the null distribution obtained by permuting sessions between brain regions. Policy representation on both A and B choices generalized more strongly in PFC than CA1. **(E)** Slices through the correlation matrices at initiation (left), choice (center) and outcome (right) times for A (solid) and B (dashed line) choices. Significant differences between conditions are indicated by stars as shown in the legend.

5.3 Discussion

Here we showed that prefrontal cortex and hippocampus both contain representations that can help guide animals' choices - that is, representations which are useful for *policy*. These policy representations were tied to the ports in the hippocampus (generalising on A but not B), but were abstracted from the sensory-motor features in mPFC (albeit with a small port component).

The existence of policy signals in mPFC is consistent with ideas about the role of this region in flexible decision-making. This line of investigation was first sparked by neuropsychological observations pertaining to patients with damage to the dorsal-lateral prefrontal cortex, who displayed enduring difficulties in deducing and adapting to evolving rules on card-sorting tests[184]. Later studies have further shown that lesions to medial frontal cortex resulted in deficits in rule-shifting behaviour and shifting of attention between perceptual categories in analogous to card-sorting tasks in rats[185–187]. As in reversal learning problems, these paradigms involve the requirement that animals change their responses based on which option they believe to be rewarding at a given time. In contrast to reversal learning problems, however, these tasks are characterised by shifting of response based on changing rules of perceptual categories, rather than the accumulation of evidence through rewards.

The effects of lesions of medial frontal cortex have been widely attributed to inflexibility or perseveration, but could also be interpreted in terms of an animal's beliefs about task structure[183]. In line with this, previous work has shown that lesions to medial prefrontal cortex caused reversal learning deficits only if the stimuli were difficult to discriminate[188], or where there were several options to select from[189]. Hence, lesions to medial prefrontal cortex do not result in a problem with perseverative responding or inflexibility in a general sense. Rather, perseveration is restricted to situations where animals are required to build a task structure and determine a behavioural policy based on this structure. Indeed, medial prefrontal activity has been shown to be crucial for remembering which of two familiar tasks was most recently successful[190], and numerous studies have found evidence for the role of mPFC in representing task structure in terms of clusters of behaviorally

related elements[160,161,191,192]. Here we expand on these ideas and provide evidence for representations of behavioural policy in mPFC, and show that in situations where there is an abstract task structure that the animals must learn, the *associated policy representations in mPFC are also abstract*.

Notably, we found that policy coding was not unique to prefrontal cortex, as hippocampus also contained policy representations, corroborating existing findings for the existence of signals relevant for decision-making in hippocampal formation[122,193]. In spatial alternation tasks, 'splitter cells' in hippocampus fire differentially in the same spatial locations depending on what the animal is about to do next and maybe be encoding the current policy (concretely, turn left next or turn right next)[63,194–196]. Similarly, motivational state affects hippocampal firing, and hippocampal units differentiate between the same location depending on whether they are food or water deprived accompanied by the preference in their respective choices[197]. More broadly, existing evidence suggests that the hippocampus supports rapid encoding of multiple items, sequences, and events in relation to the temporal context[198] by creating a flexible representation of the environment[199–201]. This representation can later be used to guide performance[202–204], and thus, the hippocampus may be essential for maintaining representations of the contingency structure of a task within a specific context. These representations can then be adaptively applied to anticipate potential changes in the environment[205]. We expand on these observations to provide further evidence that hippocampal activity might represent sensorimotor specifics of an animal's decision-making processes in the context of broader task structures.

Our observations, however, do not point to a precise mechanism on how the decision-making process takes place on a neural level when animals solve multiple reversal-learning tasks with the same structure but different sensory particularities. Our main barrier for this was that, in our paradigm, the reversals were not entirely predictable which made it hard to identify precisely when the decision to switch takes place, and to dissociate various correlates of evidence accumulation processes. To tackle this question, future work could adopt a design with reversals that are more deterministic. Additionally, varying the number of trials until the deterministic switch would enable the study of how the

associated neural processes adapt to help decision-making, and provide an opportunity for establishing a more precise link between the brain and behaviour.

Nevertheless, the present findings have important implications for the neural basis of value-based decision-making. The notion that policy representations may be abstracted in situations where tasks share the same structure aligns directly with recent ideas from computer science such as meta-RL[127]. Indeed, as discussed in the *Introduction*, when meta-RL agents are trained on a number of canonical tasks from neuroscience literature, their behaviour and internal representations resemble those seen in the frontal cortex in these tasks[127,206,207]. Our findings show that these abstracted signals can exist on markedly different behavioural timescales. In our experiment, temporally extended policy signals that help guide choices across-trials are embedded in the representation of the short-term structure of the trial sequence ($\mathrm{initiate} \rightarrow \mathrm{choose}$). This suggests that neither the prefrontal cortex or hippocampus needs to factorise and construct an independent representation for the two components of the task, and can maintain information about both of these aspects within the same representation, abstract in prefrontal cortex and sensory-specific in hippocampus. Whether this observation holds in tasks that require animals to generalise but have other behavioural demands remains to be tested.

6
General Discussion

6.1 Summary

Humans and other animals effortlessly generalise prior experience to novel situations that are only partially related. To do this, we must condense experiences to abstractions - features which are shared across different situations. Critically, we must also bind these abstractions to the specifics of the current situation.

This thesis makes three main contributions to understanding how, and when, this process happens. *Firstly*, it demonstrates that this focus on abstraction - well-established in studies of spatial reasoning and memory[71,72] - is also important in standard reinforcement learning paradigms, such as reversal learning. Whereas the dominant focus in these paradigms has been on variables such as value and prediction error[166,208,209] (which are, of course, important for learning actions *tabula rasa*), here I show that the neural representation in medial prefrontal cortex reflects the temporal structure of the problem itself, allowing actions to be generalised from similar previous experiences. One intriguing possibility is that such representations are formed during the shaping process that precedes most operant experiments.

Secondly, I have shown that mPFC and CA1 contain different representations which

imply different functional roles. Population responses in mPFC were dominated by problem-invariant representations that might form the abstraction. By contrast, the CA1 responses contained major sources of variance that were either invariant to the sensorimotor particularities (port-selective) or, intriguingly, the interaction of these with the problem structure (demonstrating 'remapping' between problems or reflecting the interaction of task policy and individual port). Representations such as these are required to bind task-general abstractions to the current sensory problem with which the animal is presented.

Thirdly, I found that task abstractions in mPFC simultaneously represent behaviour over markedly different temporal scales. Part of the mPFC representation pertained to the immediate next action in the sequence (for example, 'go to the initiation port'), but part of the representation pertained to the integrated history of rewards and actions over many trials that allowed the animal to make profitable choices. Notably, both parts of the representation were largely maintained in an abstract form that generalised over problems with different sensory particularities.

More broadly, the experimental work done as a part of this thesis demonstrated that mice can acquire sophisticated models of tasks they frequently experience in their environment, and apply this knowledge to solve new problems faster. I further demonstrate that prefrontal cortex contains representations of what can be thought of as a 'learning-set', or 'schema' of abstract relationships and variables needed to solve new related problems, while the hippocampus combines sensorimotor and abstract information to represent an interaction between the two - which might be crucial for both interpreting our ongoing experiences as well as encoding and recalling our memories.

6.2 Future directions

Much of this thesis has focussed on leveraging prior knowledge to solve problems sharing a similar structure. However, when we interact with the world, we do not simply re-use our most appropriate prior knowledge – we also flexibly *construct* new knowledge. Indeed, this is somewhat intuitive – only some of our experiences are going to be sufficiently

structurally similar to what we have already experienced. For example, when learning a new language, only some parts of another language will be relevant. When a chef creates a new recipe, they draw upon their past culinary experiences and knowledge to combine ingredients in novel ways to create a unique dish. This ability to combine and recombine past experiences is a key aspect of our creativity and problem-solving skills.

Together with a growing body of research, the work presented int his thesis shows that the brain can develop abstract neural representations that are divorced from the sensory specifics of any given situation – a principle which might serve as the substrate for the ability of animals to flexibly apply previously learned knowledge to solve new, but structurally related, problems[76,81]. There is therefore an opportunity to expand this work to more complex scenarios, where entirely new structural knowledge needs to be built to quickly solve novel problems. How does the brain support decision-making in complex tasks that can be broken down into simpler primitives? What sort of *primitives* are most efficient to represent?

Whether mathematical or culinary, when attempting to solve a novel, complex problem, it is often productive to decompose it into a number of smaller and more tractable problems. Many situations encountered in life can also be decomposed into combinations of common computational elements (these can be thought of as common *motifs* or *primitives* of thought[1]).

Cyclical structures, for example, abound in the life experiences of animals. For humans, a cycle of the Gregorian calendar year comprises four seasons, but also fifty-two weeks (which are, themselves, cycles of seven $\mathrm{day} \rightarrow \mathrm{night} \rightarrow \mathrm{day} \rightarrow ...$ sub-cycles) (Figure 16A). Perhaps, given their abundance in nature, cycles have some privileged representation in the brains not just of humans, but many animals (hibernation in bears[210], seasonal migration in birds[211], and molting in crustaceans[212] are just some examples of cycles). Other stereotypical *motifs* or *primitives* – such as lines, grids, and simple hierarchies – might also be privileged representations.

Future research could focus on designing tasks for mice that approximate common motifs which dominate the environments of laboratory animals. The basic behaviour can be

simple: one or more nose ports lights up, and the animal must select an illuminated port, then another one or more ports lights up, and so on (Figure 16B). Each port can be thought of as a state in a task motif and the transitions between ports could follow simple relational structures such as loops of unidirectionally connected states or lines of bidirectionally connected states with rewards always given at particular states (for instance, ABCDA → reward on a loop; ABCDC → reward on a line; Figure 16C).

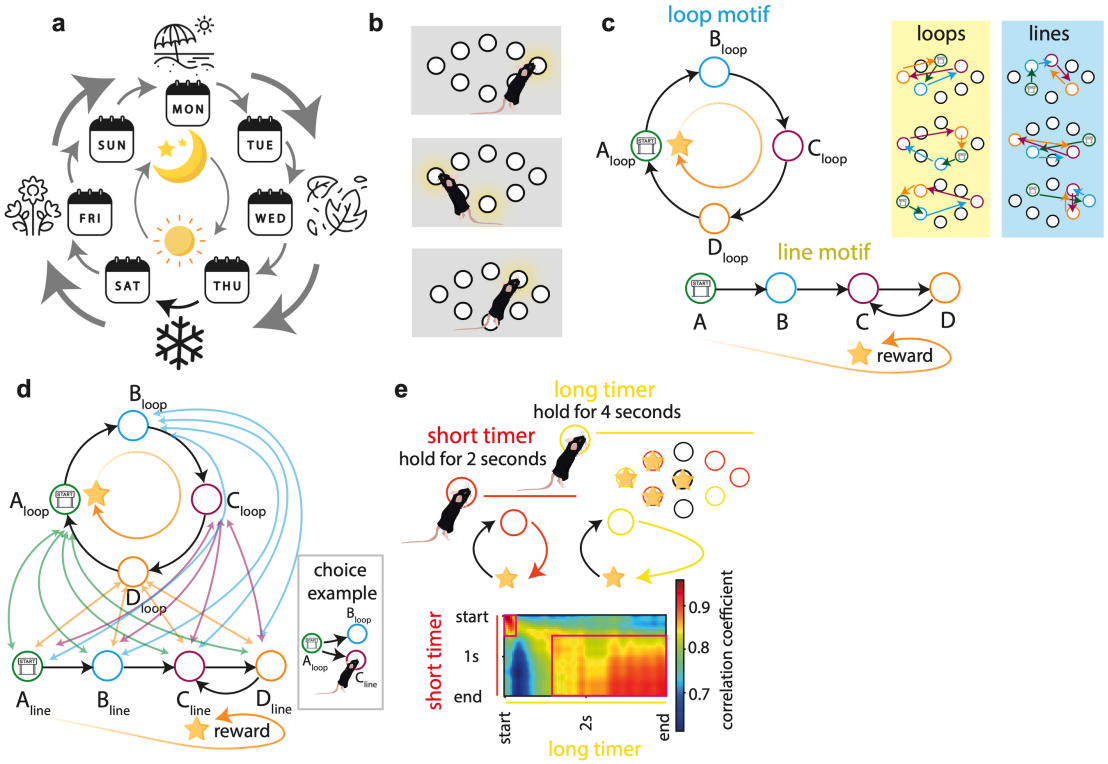


Figure 16: Constructing rich conceptual knowledge from simpler primitives. **(A)** Seasonal, weekly and daily patterns all have stereotypical cyclicity. **(B)** Basic mouse behaviour. **(C)** Examples of how line and loop motifs can be defined (right) and examples of various sensory examples of these motifs that can be presented to animals to encourage them to generalise these motifs (left). **(D)** On compositional problems where a line and a loop are combined, an animal will sometimes get a choice to transition to the other motif (i.e. line ↔ loop). Animals should choose depending on which state will get them to a reward faster. **(E)** Top: Mice were trained to hold in different nose ports for 2 or 4 seconds before receiving a reward; this task can be thought of as a simple 'hold-reward-hold' circular motif of different sizes (i.e. analogous to day-night vs weekly cycles). Bottom: Correlation matrix of all mPFC neurons between responses on a short timer versus a long timer. mPFC neurons represent the starts and ends of the timers similarly across different timer durations and port locations (highlighted by red squares).

One could train mice on various sensory instances of these simple motifs, ensuring that they can generalise abstract elementary structures to any new sensory situation. Next, animals could be presented with a new task that would be composed of previously

independently presented elementary motifs (in this case, a loop and a line) and determine if they can integrate them to solve a new 'compositional' problem which involves combining pre-learned primitives (Figure 16D). To check their understanding of the new task structure, mice will be presented with choices where they can either stay on the same task motif or transition to the other one. Depending on where they are in the sequence when the choice is presented, one of the options will take them to a reward faster (e.g. if presented with a choice to continue to B_{loop} or move to C_{line}, mice should choose C_{line} as it takes them to the rewarding state in fewer steps; Figure 16D).

The results presented earlier in this thesis suggest that mPFC should represent abstract task motifs in a sensory-invariant way. But what will happen when these motifs are combined to form a structurally new task?

Some preliminary data I have collected suggests that mPFC might be learning to represent common primitives within a larger task structure (following from the previous analogy, representing *days within weeks*), even in scenarios where there are no explicit boundaries with which to decompose it (Figure 16E). This raises the possibility that mPFC will continue to represent simpler sub-components independently, ensuring that any new composition of motifs can be solved without interference of prior compositional experience. An alternative hypothesis is that this experiment might ellucidate novel mechanisms through which the brain *composes* these elements into a seamless, unified cognitive map framework.

Overall, the exploration of how mice navigate and integrate complex task motifs offers a tantalising avenue towards understanding the neural underpinnings of complex cognition. This sets the stage for further investigations, challenging us to decipher the remarkable capabilities of the brain in the face of complex, multifaceted challenges.

6.3 Outro

Imagine if we perceived every object or event as completely unique - it would be impossible to even develop a language to express our thoughts, as the number of words or concepts required would grow (at least linearly) with the thousands of new objects, environments,

and experiences we encountered every day. Generalisation of experiences enables us to group objects, concepts, and experiences according to some principled taxonomy of similarity which aids understanding and facilitates behaviours[1]. In additional to augmenting learning, this capacity of generalisation enables us to make predictions upon entering new scenarios.

Beyond the vast benefits apparently conferred by this are some subtle grounds for caution. Indeed, upon encountering a new scenario, we might make predictions or seek some normative explanations for observed phenomena; in doing so, however, we also generalise from our own experiences. When we decide to study cognitive processes in mice and other animals, we are not exempt from this caution either. Designing experiments or analyses, we might (inadvertently) assume that their experiences are going to be similar to ours, or that their behaviour means something similar to what it would in the context of humans. This is where it becomes important to recognise the limitations of our own perspectives. The question being considered here is not one of whether animals have identical experiences to us, but whether our experimental design and analyses are appropriate for the behaviour that we aim to target in the chosen animal model[213]. Perhaps this accidental hubris is why the general intelligence of rodents (and, thus, experimental paradigms used to interrogate it) was underestimated for so long (see *Why rats deserve more credit* in the *Introduction*).

Throughout the work presented in this thesis, we took perhaps the most commonly accepted perspective in modern philosophy of science – that of *falsifiability*, whereby hypotheses were formulated in such a manner that we could determine if they were wrong[214,215]. Nevertheless, in the design of the experiments and formulation of novel analyses, our focus (perhaps inevitably) drifted towards finding suporting evidence, rather than that which contradicts our theoretical frameworks. As a result over the course of the project, it became increasingly evident how vital it was to maintain an open-minded perspective and actively seek out evidence that challenged our preconceived notions. We had to adopt alternative theoretical frameworks and consider dozens of alternative explanations for the observed behaviour and neural responses. By doing so, despite the

story ending up quite different to what we anticipated (and perhaps less complete), we did manage to learn a few new things about the brain and behaviour. And maybe that's what makes science so much fun - never knowing what you'll end up finding.

7 Materials and Methods

7.1 Behavioural apparatus

Experiments were performed in custom-made operant boxes, controlled using pyControl[216]. The boxes used in the training phase of the experiment had six nose-poke ports on the back wall, each with infrared beam, stimulus LED and solenoid valve for dispensing liquid rewards and a speaker for auditory stimuli. For recording experiments, mice were transferred to operant boxes with nine nose-poke ports located in electrically shielded sound-attenuating chambers. The operant box design is detailed at https://github.com/pyControl/hardware.

7.2 Subjects

Nine male C57BL/6J mice were used in the study, aged 6 weeks at the start of the experiment. Animals were group-housed before surgery and individually housed after surgery on a 12-hour light/dark cycle. All nine animals were implanted with silicon probes, but we obtained data from only seven, due to one probe being damaged during surgery and having to cull one animal before recordings. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[76,160,161]. Animals were pseudo-randomly assigned to the CA1 and PFC

groups. Data collection and analysis were not performed blinded to the conditions of the experiments. Experiments were carried out in accordance with Oxford University animal use guidelines and performed under UK Home Office Project Licence P6F11BC25.

7.3 Behavioural training

Mice were placed on water restriction 48 hours before starting behavioral training, with 1 hour of water access provided 24 hours before the first session. Mice were trained 6 days per week, and, on the day off, they received 1 hour ad libitum water access in their home cage. On training days, mice typically received all their water in the task but were given additional water if required to maintain their body weight above 85% of their pre-restriction baseline weight.

Mice were trained on a sequence of reversal learning problems, each with the same structure but a different physical port layout. Each reversal learning problem used three nose-poke ports, out of the six or nine ports available in the operant box. One port was used for trial initiation; the other two were choice ports where reward could be obtained. During the initial training phase, ports not used in the current problem were covered. During recording sessions, ports used in all three problems presented in the session were exposed throughout, and unused ports were covered.

Each trial started with the initiation port lighting up, until the subject poked it, after which two choice ports both lit up. Mice chose one of the choice ports, which triggered a sound cue (250ms long), indicating the trial outcome, with a pure tone (5kHz) indicating that they will get a reward and white noise indicating reward omission. Reward was delivered at the termination of the auditory cue. A 2-second ITI started once the animal left the port after reward consumption or a non-rewarded choice. One in four randomly selected trials was a forced-choice trial, where a single randomly selected choice port lit up that the animals had to select. At any given point in time, one choice port had a high reward probability, and the other one had a low probability. Reward probability reversals were triggered 5–15 trials after the subject crossed a threshold of 75% correct choices (exponential moving average, τ=8 trials). This means that the 75% threshold is not based

on any specific set of 8 consecutive trials, but rather on a weighted average where more recent trials have a greater impact on the calculation than older trials.

In the initial training stage of experiment (*Abstraction and the underlying behavioural strategies in mice*), mice encountered a single problem (that is, a specific port layout) per session and moved to the next problem the session after they had completed ten reversals on the current problem. In each problem, the first three reversals had reward probabilities of 0.9 and 0.1 at the good/bad choice ports. The fourth and fifth reversals had reward probabilities of 0.85 and 0.15, and the remaining reversals had reward probabilities of 0.8 and 0.2. In this phase, each session was 30 minutes long, and animals performed two sessions per day. The reward sizes during this stage were incrementally decreased from 15μl in the beginning of the training to 4μl, based on the animals performance. Each session started with a free reward given from each of the two choice ports. Mice were divided into three groups, with each group starting on a different layout. Sequentially presented layouts were chosen to be as different as possible, and the sequence of problem layouts was counterbalanced across animals. The data from this phase of the experiment is contained in Chapter 1.

Once mice had completed ten problems, we started presenting multiple problems in each session to prepare them for recording sessions where we sought to record neurons across multiple problems. Initially, mice were trained on two problems in a session in the nine port operant boxes subsequently used for recordings. Mice completed 12 different problems in this stage, with the port layout used in each chosen to be as different from the previous one as possible. The reward probabilities in this phase were always 0.8 and 0.2, and the reward size was 4μl. After mice completed two reversal blocks on one layout, choice ports that were going to be a part of the new problem layout both lit up. Mice received a free reward from each of the new choice ports. Next, the new initiation port lit up, signaling mice where they could initiate a trial. All port layouts and counterbalancing used in the experiment are presented in (Figure 17).



Figure 17: Counterbalancing and port layouts used throughout the experiments. Port layouts used for initial training **(A)**, intermediate training **(B)** and recording sessions **(C)**, showing the locations of the initiation (grey), A choice (green) and B choice (pink ports) on the wall of the box. Ports that were not used in any of the problems presented in a session were covered up and are not shown on the diagrams. Ports that were exposed but not used in a given problem are shown in white. In B and C the set of problems used in a single session are arranged vertically. During initial and intermediate training stages, the presentation order of different configurations was counterbalanced across animals by randomly assigning mice to 3 in groups each of 3 animals. The starting configurations for the different animals are indicated under the layouts, with an arrow indicating the direction the group subsequently progressed through the layouts across training. The layouts shown for the recordings are labelled 1,2 and 3 for consistency with how they are described in the results text, but the actual 4 presentation order within each session was randomised.

7.4 Behavioural training during recordings

During recordings (Chapter 3, 4, 5), subjects completed four reversal blocks in each of three different problem layouts in every session. Task parameters were the same as during the two-layout-per-session training stage, with the exception that now subjects needed to complete four blocks on each problem before they were moved onto a new one. As before, the problem change was signaled by the two new choice ports lighting up until the subject collected a reward from each, followed by the new initiation port lighting up. Port layouts used during recording sessions were designed to allow us to ask specific questions of the neural activity and were all reflections of three basic layout types, each of which was presented once per session in a randomized order. The timeline of behavioural training is further illustrated in (Figure 18).

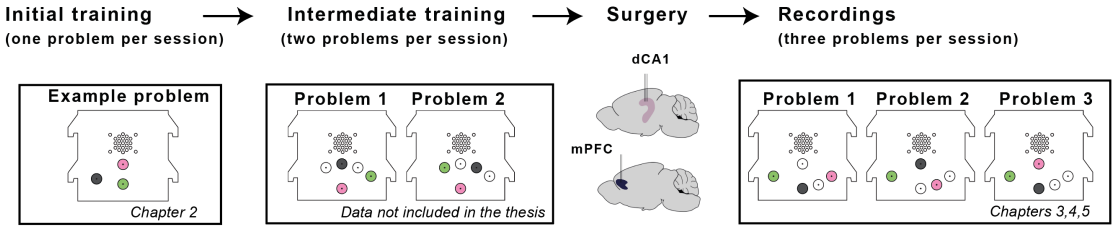


Figure 18: Progression of Training, Surgery, and Electrophysiological Recordings., intermediate training Mice were first trained on one problem per session (data included in Chapter 2), followed by intermediate training on two problems per session (data not included in the thesis). The subjects then underwent the surgery targeting the dorsal CA1 (dCA1) and medial prefrontal cortex (mPFC) regions of the brain. Post-operative testing involved electrophysiological recordings during sessions with three problems each (findings discussed in Chapters 3, 4, and 5).

7.5 Electrophysiological recordings and spike sorting

Cambridge NeuroTech 32 silicon channel probes were used for all recordings, with F series probes used for hippocampus and P series for mPFC. For hippocampal recordings, probes were implanted above the CA1 cell layer and lowered after surgery until they were in the layer, as assessed by local field potential and spike activity. For mPFC recordings, we lowered the probe $\sim$ 100 µm on every recording day. Neural activity was acquired at 30 kHz with a 32-channel Intan RHD 2132 amplifier board (Intan Technologies) connected to an OpenEphys acquisition board. Behavioral, video and ephys data were synchronized

using sync pulses output from the pyControl system. Recordings were spike sorted using Kilosort[217] and manually curated using phy. Clusters were classified as single units and retained for further analysis if they had a characteristic waveform shape, showed a clear refractory period in their autocorrelation and were stable over time.

In our analyses, we work with the assumption that the term cells denotes cell x day, suggesting that independent cells are recorded on each experimental day. However, it is crucial to acknowledge that we have not explicitly validated the independence of cells across different experimental days. Consequently, any statistical tests that rely on the count of cells as *n* could potentially be flawed, as we are unable to confirm their true independence. However, our analyses survived a group test (see *Statistical significance*) for more details), by shuffling animal IDs between the hippocampal and prefrontal groups, which means the effect was robust in each animal.

7.6 Surgery and histology

Subjects were taken off water restriction 48 hours prior to surgery, then anaesthetised with isoflurane (3% induction, 0.5–1% maintenance), treated with buprenorphine (0.1 mg/kg) and meloxicam (5 mg/kg), and placed in a stereotactic frame. A silicon probe mounted on a Microdrive (Ronal Tools) was implanted into either mPFC (AP:1.95, ML:0.4 , DV:-0.8), or dCA1 (AP:-2 , ML:1.7 , DV:-0.7), and a ground screw was implanted above the cerebellum. Both of the DV coordinates are relative to the brain surface. Mice were given additional doses of meloxicam each day for 3 days after surgery, and were monitored carefully for 7 days post-surgery, then placed back on water restriction 24 hours before restarting task behaviour. At the end of the experiment, electrolytic lesions were made under terminal pentobarbital anaesthesia to mark the probe location, animals were perfused, and the brains fixed in formal saline for subsequent histology to identify lesion locations.

7.7 Data analysis

All analyses were carried out using custom Python code (found here:
https://github.com/veronikasamborska1994/notebooks_paper).
Only sessions where animals completed three problems and four reversals in each problem were used for neural analyses.

7.8 Time-in-trial alignment

Activity was aligned across trials by warping the time interval between trial initiation and choice to match the median interval across all recorded trials. Activity before trial initiation or after choice was not warped. Spike times that occurred between initiation and choice were converted into the aligned reference frame by linear interpolation between initiation and choice time. The firing rate of each neuron was calculated in the aligned reference frame at timepoints evenly spaced every 40 ms, from 1 second before trial initiation to 1 second after trial outcome, using a Gaussian kernel with 40-ms standard deviation. To compensate for the change in spike density due to time warping, spikes in the warped interval between initiation and choice were weighted by the stretch factor applied before evaluating the firing rate (Figure 19).

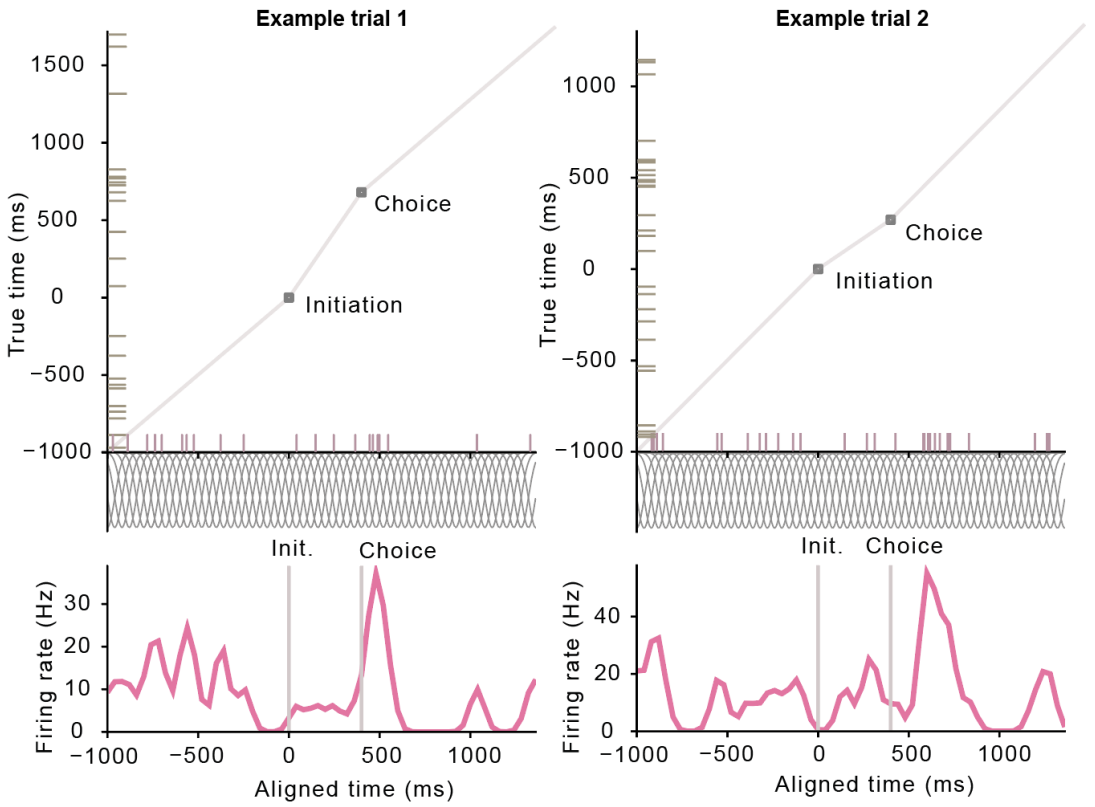


Figure 19: Trial alignment of activity. Diagram illustrating alignment of spike activity across trials. Trials were aligned using the times of initiation and choice port entry, by warping the interval between these two events to match the median interval. Top panels show spike times in the true (blue ticks) and aligned (orange ticks) time reference frames. Spike times were transformed by linear interpolation between the reference points. The output firing rate (bottom panels) was calculated at points spaced every 40ms in the aligned reference frame using Gaussian smoothing (40ms standard deviation) of the spike train. To compensate for the change in spike density due to the time warping, spikes were weighed by the stretch factor between the true and aligned reference frames (weighting is indicated by height of the orange ticks) prior to Gaussian smoothing.

7.9 Statistical significance

Significance of differences between brain areas in analyses reported throughout the paper were computed by shuffling the sessions of CA1 and PFC animals to obtain null distributions. All comparisons successfully passed a group test (Figure 20), which involved systematically shuffling animal identities between regions and generating 35 permutations of the original groupings. For each reported effect, we established a null distribution by calculating the differences between regions based on this shuffling and then compared the observed actual difference to the 95^{th} and 99^{th} percentiles within our null distribution. To address multiple comparisons across time points, we consistently selected

the peak difference between CA1 and PFC across time points for each permutation and compared each individual time point to this maximum possible difference.This approach is a commonly used method for family-wise error correction for permutation tests[218].

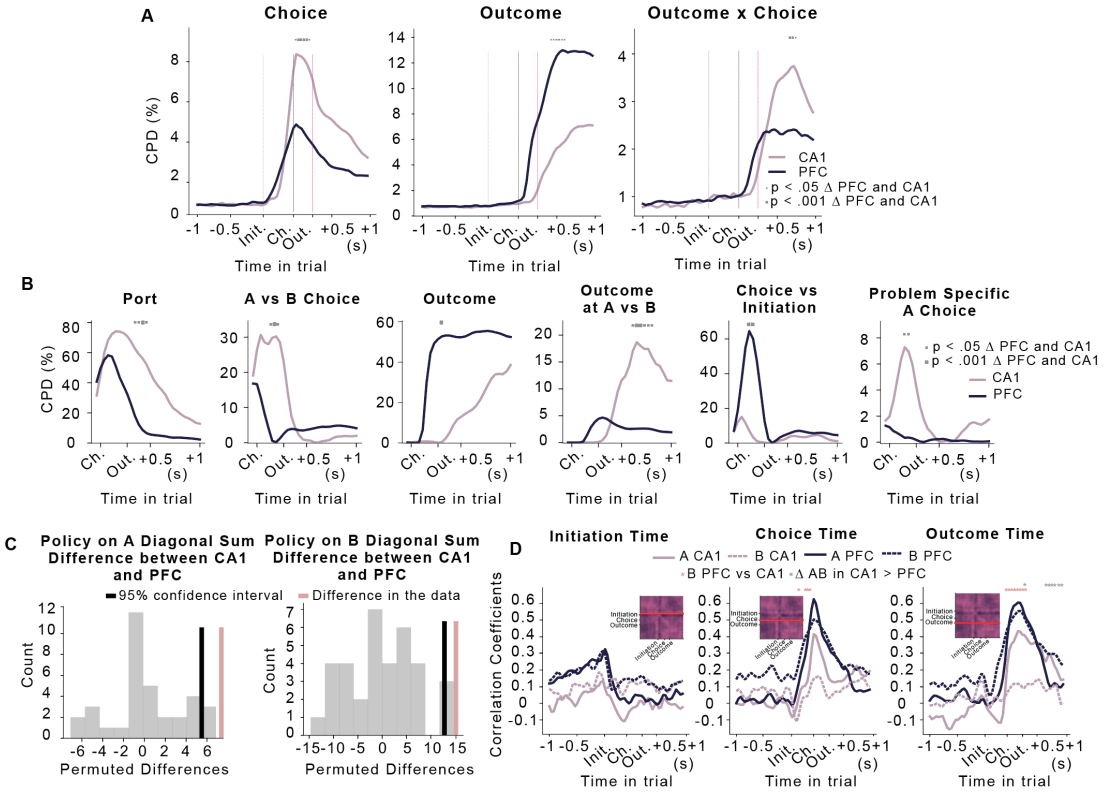


Figure 20: Significance of key differences between CA1 and PFC assessed by permuting subjects between regions. **(A)** Coefficients of partial determination from the linear model shown in (Figure 10A) for choice, outcome, and outcome x choice regressors in PFC and CA1. **(B)** Coefficients of partial determination in a regression analysis modelling the pattern of representation similarities using the RDMs shown in (Figure 10). **(C)** Sums along the diagonal of the correlation matrices shown in (Figure 15) separately for A and B choices. **(D)** Slices through the correlation matrices at initiation (left), choice (centre) and outcome (right) times for A (solid) and B (dash line) choices. For animal shuffles in singular value decomposition analyses see (Figure 21).

7.10 Representational similarity analysis

We created representational similarity matrices that consisted of the Pearson correlation coefficients of neurons in 15 different conditions, defined by the trial stage, choice, outcome and problem number. Because neurons were not simultaneously recorded, we collapsed data across recording sessions for each brain region into a single matrix (*cells* × *trial* events) and then calculated the correlation matrix across cells between different

trial events (that is, representational similarity). We used a linear regression to model the patterns of representation similarity in the data as a linear combination of RDMs:

$$r_{i,j} = \beta_0 + \sum_{n=1}^{9} \beta_n RDM_{n(i,j)} + \epsilon_{i,j} \tag{7.1}$$

Where $r_{(i,j)}$ are elements of the RSA matrix and $RDM_{n(i,j)}$ are elements of the n$^{\text{th}}$ RDM. Before regressing the correlation matrices onto the RDMs the diagonal elements from both were deleted and a constant matrix of ones was added to the design matrix to account for any condition independent correlation between neurons. We plotted the coefficients of partial determination (CPDs) from the regression model described above. The CPD was defined as:

$$CPD\left(RDM_i\right) = \left(SSE_{\sim i} - SSE_{\text{full model}}\right) / SSE_{\sim i} \tag{7.2}$$

Where $SSE_{\sim i}$ refers to sum of squares from a regression model excluding the RDM_i of interest and $SSE_{\text{full model}}$ is the sum of squares from a regression model including all the RDMs. CPDs describe how much unique variance does each RDM account for in the RSA matrix calculated from firing rates.

7.11 Decoding analyses

We trained a support vector classifier (implemented using $\mathrm{sklearn.svm.SVC}$) to classify stages of the trial (Initiation, A choice, B choice, A reward, B reward, A no-reward and B no-reward) from neural activity on one problem and tested how it performed on a different problem. This was computed for all problem pairs, and the mean decoding accuracy for each trial stage was shown in a form of a confusion matrix (Figure 11A).

We then analyzed these confusion matrices to look for patterns of decoding associated with a representation of (1) physical port, (2) trial stage (initiation, choice and type of outcome) and (3) abstract choice. In one of our problem layout pairs, initiation port became a B choice (layout 2 to layout 3), and, in another, initiation became a B choice

(layout 3 to layout 2), so mistakes made by the decoder between B choice and Initiation in these pairs indicate a prominent representation of port location. Decoding errors between A choices and B choices, A rewards and B rewards and A no-rewards and B no-rewards indicate a representation of trial stage. Lastly, representation of an abstract choice (A vs. B) independent of port location was computed by summing cells corresponding to the same abstract choice but in a different physical location across problems. Statistical significance of differences between PFC and CA1 in decoding patterns was established by permuting animal identities between regions and comparing the real differences against the 95% confidence interval of the shuffle.

7.12 Surprise Measure

To investigate the time course of how quickly the firing rates of neurons change in response to layout changes (Supplementary Figure 4), we used the 'surprise' measure from Information Theory:

$$s\left(x_{ij}\right) = \left(x_{ij} - \mu_k\right)^2 / \sigma_k^2 \tag{7.3}$$

where x_{ij} is the firing rate of one neuron on a given trial i and task layout j, μ_k and σ_k are the baseline mean and the standard deviation of the firing rate of that neuron on a particular task layout. If $j = k$, then the $s(x_{ij})$ on each trial i is calculated based on the mean firing rate μ and standard deviation σ of the withheld trials from the same task. More precisely, to calculate how much the firings rates change during the same task layout $s(x_{ij})$ was calculated on the 10 trials before the task layout switch ('test' within task), where μ_k and σ_k were calculated on the 10 trials before those 'test' trials ('train' within task). If $j \neq k$, then the $s(x_{ij})$ on each trial i was calculated based on the mean firing rate μ and standard deviation σ of the withheld trials from a different task. So, to estimate how much the firings rates change after the task layout switch $s(x_{ij})$ was calculated on the 20 trials after the task layout switch ('test' between tasks), where μ_k and σ_k were calculated from the 'train' trials from a different task layout. This measure was calculated for each neuron separately and then averaged across all neurons for each brain region.

7.13 Singular Value Decomposition

SVD was performed using the $\mathrm{numpy\ linalg.svd}$ function in Python[219]. SVD is a principal component analysis technique that decomposes any *n* × *m* matrix into a product of three matrices:

$$D = U\Sigma V^T \tag{7.4}$$

where D comprises the data matrix to be decomposed; U and V^T are sets of singular vectors capturing patterns of covariation in the data; and Σ is a diagonal weight matrix [*].

In our SVD analyses, each row of *D* was the demeaned, trial-aligned activity of one neuron across each timepoint of four concatenated trial types: rewarded A choices, non-rewarded A, rewarded B and non-rewarded B. So the shape of *D* was [$n_{neurons}$, 4 × $n_{timepoints}$]. The columns of *U* are vectors that we term cellular modes because each is a set of weights over neurons, representing groups of neurons whose activity covaries. Each cellular mode has a corresponding row in V^T that we term a temporal mode, as it is a set of weights over timepoints, representing the time course of the cellular mode's activity. Each temporal mode spans the same set of timepoints as the data matrix and, hence, captures variation both over time-in-trial and trial-type. As both modes are unit vectors, their contribution to the total data variance is determined by the corresponding element of the diagonal matrix Σ.

The cellular modes are given by eigendecomposition of the covariances between neurons, as can be seen from the following:

$$\begin{aligned} DD^T &= \left(U\Sigma V^T\right)\left(U\Sigma V^T\right)^T \\ DD^T &= \left(U\Sigma V^T\right)\left(V\Sigma U^T\right) \\ DD^T &= U\Sigma^2 U^T \end{aligned} \tag{7.5}$$

[*]Note that SVD can be considered a generalisation of eigendecomposition to non-square matrices, and is thus more flexible.

As DD^T is the non-normalized covariance between neurons across timepoints, Σ^2 is a diagonal matrix of eigenvalues, U are the corresponding eigenvectors and $U^T = U^1$ because U is an orthonormal basis.

Similarly, the temporal modes are given by eigendecomposition of the covariances between timepoints:

$$\begin{aligned} D^T D &= \left(U\Sigma V^T\right)^T \left(U\Sigma V^T\right) \\ D^T D &= \left(V \sum U^T\right) \left(U\Sigma V^T\right) \\ D^T D &= V\Sigma^2 V^T \end{aligned} \tag{7.6}$$

As $D^T D$ is the non-normalized covariance between timepoints across neurons, Σ^2 is a diagonal matrix of eigenvalues, V are the corresponding eigenvectors and $V^T = V^1$ because V is an orthonormal basis.

Our goal was to test whether cellular and temporal patterns generalize across different problems by quantifying how well cellular and/or temporal modes from one problem explained variance in another. As a control for drift in representations over time, we compared generalization between problems with generalization to held-out data from the same problem. To do this, we constructed separate data matrices for the first and second half of each problem:

$$D_{i,h} = U_{i,h}\Sigma_{i,h}V_{i,h}^T \tag{7.7}$$

where i is the problem number $i = 1, 2, 3$, and h is the half of the problem that the data are taken from $h = f, s$. We can then compare generalization between the second half of one problem with the first half of the next with generalization between first and second half of the same problem, to ensure that any drift is matched between within-problem and cross-problem comparisons.

We quantified three different ways in which activity patterns might generalize between problems. (1) Generalization of temporal modes irrespective of whether they recruited the same neurons. This corresponds to the same trial events being represented but not

necessarily by the same neurons. (2) Generalization of cellular modes irrespective of whether they have the same time course. This corresponds to the same cell assemblies co-activating but not necessarily representing the same trial events. (3) Generalization of cellular and temporal mode pairs. This corresponds to the same cell assemblies representing the same trial events across problems.

To quantify how well temporal modes generalized across problems, we projected the data matrix from half of one problem on the temporal modes from an adjacent half of a different problem:

$$M_V^{cross} = D_{2,f} V_{1,s} \tag{7.8}$$

The total variance explained by each temporal mode for this problem pair is given by squaring the elements of M_V^{cross} and summing over neurons. We average across all adjacent problem pairs and plot the cumulative variance explained as a function of the number of temporal modes used.

The corresponding within-problem variance explained is given by projecting the data matrix from half of one problem onto the temporal modes from the other half of the same problem:

$$M_V^{same} = D_{1,f} V_{1,s} \tag{7.9}$$

Similarly for the cellular modes, the cross-problem generalization was given by projecting the data matrix from half of one task on the cellular modes from an adjacent half of a different problem:

$$M_U^{cross} = U_{1,s}^T D_{2,f} \tag{7.10}$$

The total variance explained by each cellular mode for this problem pair is given by squaring the elements of M_U^{cross} and summing over timepoints. Again, we average across

all adjacent problem pairs and plot the cumulative variance explained as a function of the number of cellular modes used.

The corresponding within-problem variance explained is given by projecting the data matrix from half of one problem onto the cellular modes from the other half of the same problem:

$$M_U^{same} = U_{1,s}^T D_{1,f} \tag{7.11}$$

To quantify how well pairs of neural and temporal patterns generalized between problems, we projected the data matrix from half of one problem on the cellular and temporal modes from an adjacent half of a different problem:

$$M_U^{same} = U_{1,s}^T D_{1,f} \tag{7.12}$$

Σ^{cross} is not diagonal; however, if the same cell assemblies perform the same roles in two problems, the temporal and cellular modes will align, and Σ^{cross} will have high weights on the diagonal. We, therefore, plotted the cumulative sum of the squared weights of the diagonal elements. Because we had different numbers of neurons in each brain region, Σ^{cross} was normalized by the number of neurons recorded from the respective brain region.

The corresponding within-problem variance explained is given by projecting the data matrix from half of one problem onto the cellular and temporal modes from the other half of the same problem:

$$\Sigma^{same} = U_{1,s}^T D_{1,f} V_{1,s} \tag{7.13}$$

To determine the significance of the differences between two regions, we compared differences in the data between PFC and CA1 against a null distribution of differences between areas under the curve by shuffling the sessions between CA1 and PFC animals.

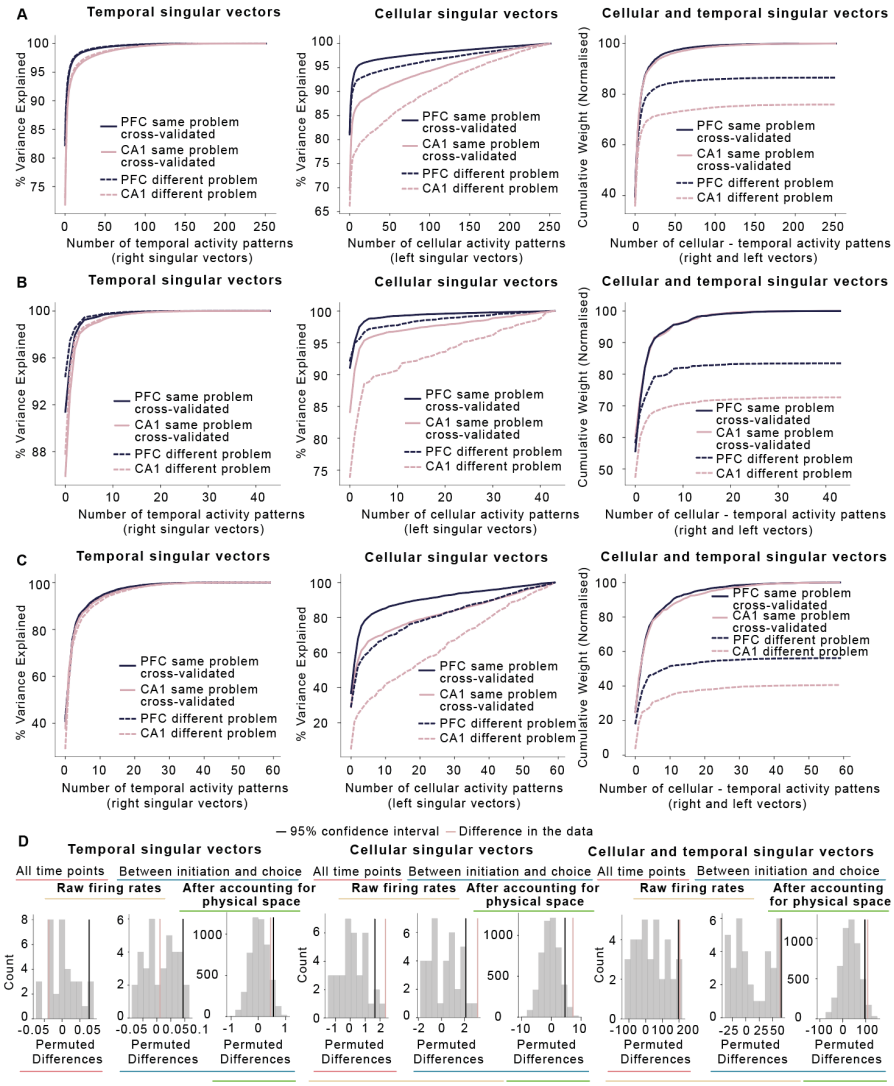


Figure 21: Cellular vectors, as well as combinations of cellular and temporal vector assemblies, exhibit a more robust generalisation pattern in the prefrontal cortex (PFC) compared to CA1. This holds true irrespective of whether we consider **(A)** all time points in the trial (using raw firing rates), **(B)** only time points between initiation and choice (using raw firing rates), or **(C)** solely time points between initiation and choice, while incorporating residual firing rates after accounting for physical factors. **(A-C)** Generalization of low-dimensional structure of activity across problems when considering; **(A)** all time-points in the trial (raw firing rates), **(B)** only time-points between initiation and choice (raw firing rates), **(C)** only time-points between initiation and choice using residual firing rates after accounting for physical movement. Left: Variance explained when using temporal activity patterns V_1^T from one problem to predict either held out activity from the same problem (solid lines) or activity from a different problem (dash lines). Middle: Variance explained when using cellular activity patterns U_1 from one problem to predict either held out activity from the same problem (solid lines) or activity from a different problem (dash lines). Right: Cumulative weights along the diagonal Σ using pairs of temporal V_1^T and cellular U_1 activity patterns from one problem to predict either held out activity from the same problem (solid lines) or activity from a different problem (dash lines). **(D)** Permutation tests for significance of differences between CA1 and PFC in generalization of temporal (left), cellular (middle) and cellular and temporal (right) singular vector, based on the null distribution obtained by shuffling animals across groups (raw firing rates) or sessions (residual firing rates after accounting for physical space). We could not permute animals in the analyses of residual firing rates because we were not set up for recording video data for our first implanted animal.

7.14 Estimating policy

We obtained a trial-by-trial estimate of subjects behavioral policy using a logistic regression predicting current trial choice, using the history of choices, rewards and *choice* x *reward* interactions (Figure 15A). This gave an estimate on each trial of the probability that the animal would choose A, which we term the animals *policy*. We used this policy and its interaction with current choice (*policy* x *choice*), together with current trial events (choice, outcome and *outcome* x *choice* interaction), to predict neural activity in a linear regression, quantifying the variance explained by each predictor at each timepoint as the coefficient of partial determination (CPD) (Figure 15B).

To understand whether policy representations generalized between problems, we conducted this linear regression separately for A and B choices (dropping the current trial choice predictor from the regression), obtaining one vector of coefficients for A choices and one for B choices, indicating how policy affected the activity of each neuron, which we term policy representations. As the policy representation may change across the trial, we did this for a set of time windows across the trial, obtaining a policy representation for each timepoint for A and B choices. We quantified how similar policy representations were between problems and timepoints as the Pearson correlation, to obtain the matrices shown in (Figure 15C). Finally, to understand whether these across trial policy signals might also be tied to representations of unique trial stages, we examined time slices through the correlation matrices at initiation, choice and outcome times. The differences between these signals at each timepoint were then compared against null distributions described in the section titled *Statistical significance*.

7.15 Additional controls for physical movement

To provide additional controls for movement-related activity, we sought to eliminate the effect of an animal's position, velocity and acceleration on firing rates before performing all our subsequent analyses. To do this, we used DeepLabCut[220] pose estimation software to extract the animal's nose position in each session. Because the cameras in the operant boxes were located above the animal, there were some artifacts in the tracking caused

by the occlusion of the nose by the ports and the head cap, which caused the estimated position to jump to incorrect locations. To correct for this, we first removed all samples where the likelihood of correct estimation output by DeepLabCut was below 90%. We then removed samples adjacent to jumps in position larger than ten times the standard deviation of displacements between frames, estimated using the 16^{th} and 84^{th} percentiles of the displacement distribution. We then removed samples that were not in contiguous groups of at least five. After this artifact removal step, we interpolated the missing data, taking advantage of the fact that the movements of the ears and nose are highly correlated, such that the trajectories of the ears provide information about movements of the nose when the nose is occluded. The interpolation was implemented by minimizing a cost function with two terms: (1) the sum of squared derivatives of the nose position, which promotes linear interpolation of missing data, and (2) the sum of squared differences between the derivatives of the ear and nose positions, which promotes the interpolated trajectory of the nose tracking those of the ears.

Next, because we had the ground truth of our port locations in physical space, we performed a linear registration and transformed the 2D coordinates extracted from the video from the oblique camera view to a more informative horizontal view of the wall of the ports. Finally, we used our behavioral data to find when the animals were inside the ports and corrected for any inaccuracy in our DeepLabCut data by placing these coordinates inside the ports.

As we did not expect that the 2D coordinates of animal's nose position would be linearly related to neural firing rates (for example, due to previously reported existence of 'place cells' in CA1), we first needed to create vectorised 'occupancy maps' (Figure 22A). Specifically, we defined a set of Gaussian 'radial basis functions' with the centers randomly selected from an animal's 2D coordinates in each session and a standard deviation of 1cm. Next, for each timepoint, we calculated the activity of each basis function (Gaussian in distance from center of this field), resulting in a matrix of shape [*time*, $n_{basisfunctions}$].

To account for cross-correlations in this matrix, we next performed a principal component analysis (PCA) to extract the first ten orthogonal occupancy components across time accounting for >95% of variance. To confirm that our key results could not be explained by movement-related parameters, we repeated our main analyses using the residual firing rates from a linear model predicting the firing of each neuron using these occupancy components, as well as the acceleration and velocity of the animal at each timepoint (Figure 22A, B). Because we did not have video data for all our animals due to technical limitations at the time of experiments, the significance between brain areas in these analyses was only computed by shuffling the sessions of CA1 and PFC animals to obtain null distributions and correcting for multiple comparisons as before (see *Statistical significance*).

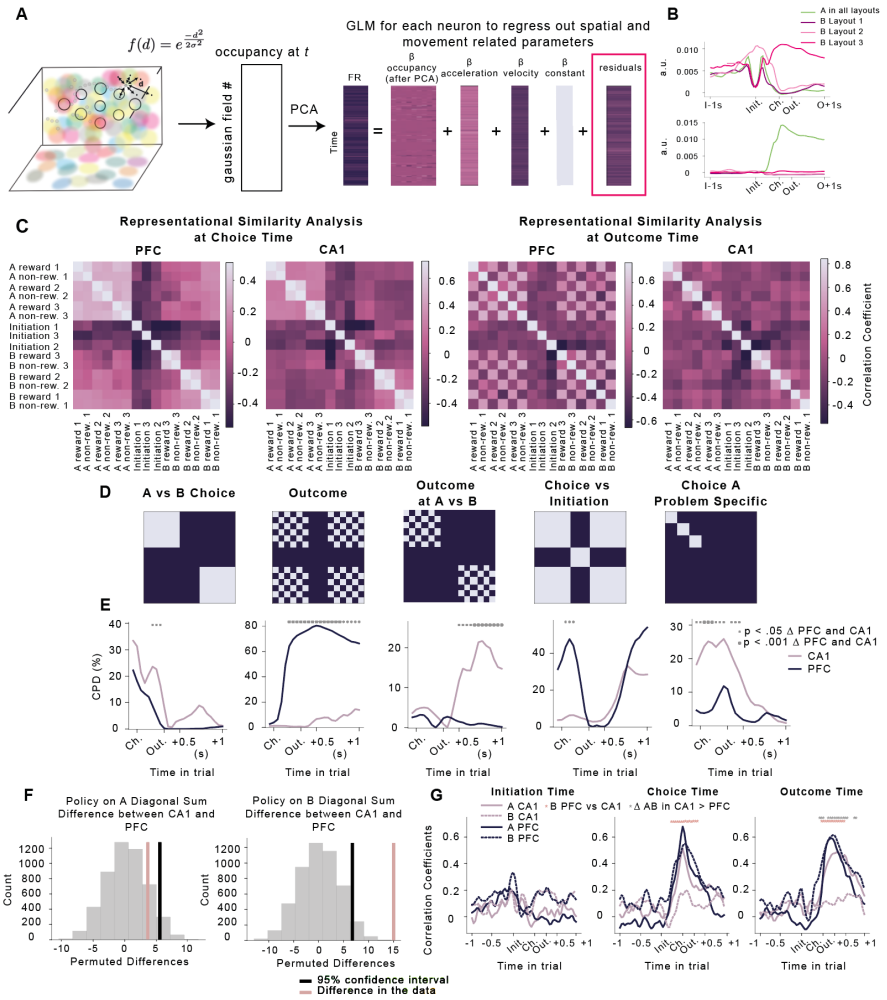


Figure 22: Fine-grained movement related activity controls. (A) We do not expect 2D nose coordinates to be linearly related to firing rates, so to account for place cell-like coding of nose position in the firing rates of neurons we defined a set of Gaussian 'radial basis functions' with the centres randomly selected from an animal's 2D coordinates in each session (different coloured circles). Next, for each time point we calculated the activity of each basis function (Gaussian in distance from centre of this field) resulting in a *time* x $n_{BasisFunctions}$ matrix (left). To account for cross-correlations in this matrix, we next performed principal component analysis to extract the first ten orthogonal occupancy components across time (middle). Next, we fit a linear regression model predicting firing rates of neurons with occupancy as well as nose velocity, and acceleration predictors, resulting in residual firing rates that do not contain variability related to these movement and position parameters (right). **(B)** Top two principal components from the PCA analysis of occupancies in A from an example session. The first component differentiates initiation in Layout 1 and port B in Layout 2 (same physical location) from other ports. The second component differentiates port A (same physical location) from all other ports. **(C)** Representation similarity at 'choice time' (left) and 'outcome time' (right), quantified as the Pearson correlation between the residual neural activity (after accounting for movement related parameters) vectors for each pair of task conditions. **(D)** RDMs used to model the patterns of representation similarity observed in the data. Port RDM was not included in this analysis as the PCs we regress out to account for movement-related activity correlate strongly with the port position in the task. Using an RDM that is so highly correlated with a parameter that has already been regressed out can lead to false correlations (similar to Berkson's paradox). **(E)** CPDs in a regression analysis modelling the pattern of representation similarities in residual firing rates using the RDMs in D. **(F-G)** Policy analyses conducted on residual firing rates after accounting for movement-related activity. Controls of precise physical movements in singular value decomposition analyses are presented in Figure 21.

Bibliography

1. Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. How to grow a mind: Statistics, structure, and abstraction. *science* **331,** 1279–1285 (2011).
2. Hafting, T., Fyhn, M., Molden, S., Moser, M.-b. B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436,** 801–806 (2005).
3. Behrens, T. E., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A., Stacenhfeld, K. L. & Kurth-Nelson, Z. What is a cognitive map? Organising knowledge for flexible behaviour. *BioRxiv* **0** (2018).
4. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81,** 267–278 (2014).
5. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žıdek, A., Potapenko, A., *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).
6. Brown, N. & Sandholm, T. Superhuman AI for multiplayer poker. *Science* **365,** 885–890 (2019).
7. Harlow, H. F. The formation of learning sets. *Psychological review* **56,** 51 (1949).
8. Carey, S. The origin of concepts. *Journal of Cognition and Development* **1,** 37–41 (2000).
9. Spelke, E. S. Principles of object perception. *Cognitive science* **14,** 29–56 (1990).
10. Spelke, E. S., Phillips, A. & Woodward, A. L. Infants' knowledge of object motion and human action. (1995).
11. Johnson, S., Slaughter, V. & Carey, S. Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Developmental Science* **1,** 233–238 (1998).
12. Gelman, S. A. & Coley, J. D. The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental psychology* **26,** 796 (1990).
13. Koronakos, C. & Arnold, W. J. The formation of learning sets in rats. *Journal of Comparative and Physiological Psychology* **50,** 11 (1957).
14. Kay, H. & Oldfield-Box, H. Use of 3-dimensional shapes for investigation of learning sets in the rat. *Nature* **199,** 1018–1018 (1963).
15. Wilson, M. Six views of embodied cognition. *Psychonomic bulletin & review* **9,** 625–636 (2002).
16. Slotnick, B. M. & Katz, H. M. Olfactory learning-set formation in rats. *Science* **185,** 796–798 (1974).
17. Eichenbaum, H., Fagan, A. & Cohen, N. Normal olfactory discrimination learning set and facilitation of reversal learning after medial-temporal damage in rats: Implications for an account of preserved learning abilities in amnesia. *Journal of Neuroscience* **6,** 1876–1884 (1986).
18. Fagan, A. M. & Olton, D. S. Spatial learning set acquisition in rats. *The Quarterly Journal of Experimental Psychology* **39,** 65–75 (1987).

19. Reid, I. C., Morris, R. & Graham, M. Smells are no surer: rapid improvement in olfactory discrimination is not due to the acquisition of a learning set. *Proc. R. Soc. Lond. B* **247,** 137–143 (1992).
20. Murphy, R. A., Mondragón, E. & Murphy, V. A. Rule learning by rats. *Science* **319,** 1849–1851 (2008).
21. Lashley, K. S. Mass action in cerebral function. *Science* **73,** 245–254 (1931).
22. Corkin, S., Amaral, D. G., González, R. G., Johnson, K. A. & Hyman, B. T. HM's medial temporal lobe lesion: findings from magnetic resonance imaging. *Journal of Neuroscience* **17,** 3964–3979 (1997).
23. Penfield, W. & Milner, B. Memory deficit produced by bilateral lesions in the hippocampal zone. *AMA archives of Neurology & Psychiatry* **79,** 475–497 (1958).
24. Gabrieli, J. D., Cohen, N. J. & Corkin, S. The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and cognition* **7,** 157–177 (1988).
25. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry* **20,** 11 (1957).
26. Kensinger, E. A., Ullman, M. T. & Corkin, S. Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: Evidence from amnesic patient HM. *Hippocampus* **11,** 347–360 (2001).
27. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research* **34,** 171–175 (1971).
28. Tolman, E. C. Cognitive maps in rats and men. *Psychological review* **55,** 189 (1948).
29. Tolman, E. C. & Honzik, C. H. Introduction and removal of reward, and maze performance in rats. *University of California publications in psychology* (1930).
30. Tolman, E. C., Ritchie, B. F. & Kalish, D. Studies in spatial learning. I. Orientation and the short-cut. *Journal of experimental psychology* **36,** 13 (1946).
31. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436,** 801–806 (2005).
32. Taube, J. S., Muller, R. U. & Ranck, J. B. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience* **10,** 420–435 (1990).
33. Gauthier, J. L. & Tank, D. W. A dedicated population for reward coding in the hippocampus. *Neuron* **99,** 179–193 (2018).
34. Sarel, A., Finkelstein, A., Las, L. & Ulanovsky, N. Vectorial representation of spatial goals in the hippocampus of bats. *Science* **355,** 176–180 (2017).
35. Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B. & Moser, E. I. The entorhinal grid map is discretized. *Nature* **492,** 72–78 (2012).
36. Høydal, Ø. A., Skytøen, E. R., Andersson, S. O., Moser, M.-B. & Moser, E. I. Object-vector coding in the medial entorhinal cortex. *Nature* **568,** 400–404 (2019).
37. Danjo, T., Toyoizumi, T. & Fujisawa, S. Spatial representations of self and other in the hippocampus. *Science* **359,** 213–218 (2018).
38. Omer, D. B., Maimon, S. R., Las, L. & Ulanovsky, N. Social place-cells in the bat hippocampus. *Science* **359,** 218–224 (2018).
39. Hassabis, D., Kumaran, D., Vann, S. D. & Maguire, E. A. Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences* **104,** 1726–1731 (2007).
40. Van der Meer, M., Kurth-Nelson, Z. & Redish, A. D. Information Processing in Decision-Making Systems. *The Neuroscientist* **18,** 342–359 (2012).

41. Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D. & Spiers, H. J. Hippocampal place cells construct reward related sequences through unexplored space. *Elife* **4,** e06063 (2015).
42. Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y. & Schiller, D. A map for social navigation in the human brain. *Neuron* **87,** 231–243 (2015).
43. Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R. C., Wiener, J. M., Hölscher, C., Hornberger, M., *et al.* Entropy of city street networks linked to future spatial navigation ability. *Nature* **604,** 104–110 (2022).
44. Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience* **7,** 1951–1968 (1987).
45. Lever, C., Wills, T., Cacucci, F., Burgess, N. & O'Keefe, J. Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature* **416,** 90–94 (2002).
46. Knierim, J. J. & Rao, G. Distal landmarks and hippocampal place cells: effects of relative translation versus rotation. *Hippocampus* **13,** 604–617 (2003).
47. Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L. & Moser, M.-B. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science* **309,** 619–623 (2005).
48. Anderson, M. & Jeffery, K. Heterogeneous modulation of place cell firing by changes in context. *J Neurosci* **23,** 8827–8835 (2003).
49. Jeffery, K. & Anderson, M. Dissociation of the geometric and contextual influences on place cells. *Hippocampus* **13,** 868–872 (2003).
50. Fyhn, M., Hafting, T., Treves, A., Moser, M.-B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446,** 190–194 (2007).
51. Barry, C., Hayman, R., Burgess, N. & Jeffery, K. J. Experience-dependent rescaling of entorhinal grids. *Nature neuroscience* **10,** 682–684 (2007).
52. Eichenbaum, H. Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* **44,** 109–120 (2004).
53. Eichenbaum, H. On the integration of space, time, and memory. *Neuron* **95,** 1007–1018 (2017).
54. Burt, C. *et al.* Experimental tests of higher mental processes and their relation to general intelligence (1911).
55. Dusek, J. A. & Eichenbaum, H. The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences* **94,** 7109–7114 (1997).
56. McGonigle, B. O. & Chalmers, M. Are monkeys logical? *Nature* **267,** 694–696 (1977).
57. Gilboa, A., Sekeres, M., Moscovitch, M. & Winocur, G. Higher-order conditioning is impaired by hippocampal lesions. *Current Biology* **24,** 2202–2207 (2014).
58. Wikenheiser, A. M. & Schoenbaum, G. Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience* **17,** 513–523 (2016).
59. Buckmaster, C. A., Eichenbaum, H., Amaral, D. G., Suzuki, W. A. & Rapp, P. R. Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. *Journal of Neuroscience* **24,** 9811–9825 (2004).
60. MacDonald, C. J., Lepage, K. Q., Eden, U. T. & Eichenbaum, H. Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron* **71,** 737–749 (2011).
61. O'Keefe, J. Place units in the hippocampus of the freely moving rat. *Exp Neurol* **51,** 78–109 (1976).
62. Hok, V., Lenck-Santini, P.-P., Roux, S., Save, E., Muller, R. U. & Poucet, B. Goal-Related Activity in Hippocampal Place Cells. *Journal of Neuroscience* **27,** 472–482 (2007).

63. Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. The global record of memory in hippocampal neuronal activity. *Nature* **397,** 613–616 (1999).
64. McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D. & Eichenbaum, H. Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron* **83,** 202–215 (2014).
65. Komorowski, R. W., Manns, J. R. & Eichenbaum, H. Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. *Journal of Neuroscience* **29,** 9918–9929 (2009).
66. Kumaran, D., Melo, H. L. & Duzel, E. The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron* **76,** 653–666 (2012).
67. Ferbinteanu, J. & Shapiro, M. L. Prospective and retrospective memory coding in the hippocampus. *Neuron* **40,** 1227–1239 (2003).
68. Frank, L. M., Brown, E. N. & Wilson, M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27,** 169–178 (2000).
69. Bartlett, F. C. Some experiments on the reproduction of folk-stories. *Folklore* **31,** 30–47 (1920).
70. Gilboa, A. & Marlatte, H. Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences* **21,** 618–631 (2017).
71. Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P. & Morris, R. G. Schemas and memory consolidation. *Science* **316,** 76–82 (2007).
72. Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H. & Morris, R. G. Schema-dependent gene activation and memory encoding in neocortex. *Science* **333,** 891–895 (2011).
73. Barron, H. C., Dolan, R. J. & Behrens, T. E. Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience* **16,** 1492–1498 (2013).
74. Bowman, C. R. & Zeithamova, D. Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *Journal of Neuroscience* **38,** 2605–2614 (2018).
75. Zeithamova, D., Dominick, A. L. & Preston, A. R. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* **75,** 168–179 (2012).
76. Zhou, J., Jia, C., Montesinos-cartagena, M. & Gardner, M. P. H. Evolving schema representations in orbitofrontal ensembles during learning. *Nature* (2020).
77. Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H. & Rushworth, M. F. Separable Learning Systems in the Macaque Brain and the Role of Orbitofrontal Cortex in Contingent Learning. *Neuron* **65,** 927–939 (2010).
78. Takahashi, Y. K., Roesch, M. R., Wilson, R. C., Toreson, K., O'donnell, P., Niv, Y. & Schoenbaum, G. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature neuroscience* **14,** 1590–1597 (2011).
79. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81,** 267–279 (2014).
80. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human orbitofrontal cortex represents a cognitive map of state space. *Neuron* **91,** 1402–1412 (2016).
81. Samborska, V., Butler, J. L., Walton, M. E., Behrens, T. E. & Akam, T. Complementary task representations in hippocampus and prefrontal cortex for generalizing the structure of problems. *Nature Neuroscience* **25,** 1314–1326 (2022).

82. Sun, C., Yang, W., Martin, J. & Tonegawa, S. Hippocampal neurons represent events as transferable units of experience. *Nature Neuroscience* **23,** 651–663 (2020).
83. Manns, J. R. & Eichenbaum, H. Evolution of declarative memory. *Hippocampus* **16,** 795–808 (2006).
84. Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-browne, N. B. & Botvinick, M. M. Neural representations of events arise from temporal community structure. *Nature Neuroscience* **16,** 486–492 (2013).
85. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352,** 1464–1468 (2016).
86. Garvert, M. M., Dolan, R. J. & Behrens, T. E. E. E. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife* **6,** 1–20 (2017).
87. Aronov, D., Nevers, R. & Tank, D. W. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature* **543,** 719–722 (2017).
88. Knudsen, E. B. & Wallis, J. D. Closed-Loop Theta Stimulation in the Orbitofrontal Cortex Prevents Reward-Based Learning. *Neuron* **106,** 537–547 (2020).
89. Stevenson, D. & Wagoner, N. J. Bargaining in the shadow of big data. *Fla. L. Rev.* **67,** 1337 (2015).
90. Bellman, R. *Markovian decision processes* 1957.
91. Shannon, C. E. XXII. Programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **41,** 256–275 (1950).
92. Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. Neuroscience-inspired artificial intelligence. *Neuron* **95,** 245–258 (2017).
93. Van de Ven, G. M., Siegelmann, H. T. & Tolias, A. S. Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications* **11,** 4069 (2020).
94. Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C. K., Hassabis, D., Munos, R. & Botvinick, M. A distributional code for value in dopamine-based reinforcement learning. *Nature* **577,** 671–675 (2020).
95. Whittington, J. C., McCaffary, D., Bakermans, J. J. & Behrens, T. E. How to build a cognitive map. *Nature neuroscience* **25,** 1257–1272 (2022).
96. Whittington, J. C., McCaffary, D., Bakermans, J. J. & Behrens, T. E. How to build a cognitive map: insights from models of the hippocampal formation. *arXiv preprint arXiv:2202.01682* (2022).
97. OpenAI. *GPT-4 Technical Report* 2023.
98. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., *et al.* A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362,** 1140–1144 (2018).
99. 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.* Human-level control through deep reinforcement learning. *nature* **518,** 529–533 (2015).
100. Hinton, G. E. *et al. Learning distributed representations of concepts* in *Proceedings of the eighth annual conference of the cognitive science society* **1** (1986), 12.
101. French, R. M. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences* **3,** 128–135 (1999).
102. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C. & Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Networks* **113,** 54–71 (2019).
103. McCaffary, D. Towards continual task learning in artificial neural networks: current approaches and insights from neuroscience. *arXiv preprint arXiv:2112.14146* (2021).
104. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521,** 436–444 (2015).

105. Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J. & Hinton, G. Backpropagation and the brain. *Nature Reviews Neuroscience* **21,** 335–346 (2020).
106. Yang, G., Pan, F. & Gan, W.-B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462,** 920–924 (2009).
107. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., *et al.* Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences* **114,** 3521–3526 (2017).
108. Kemker, R., McClure, M., Abitino, A., Hayes, T. & Kanan, C. *Measuring catastrophic forgetting in neural networks* in *Proceedings of the AAAI Conference on Artificial Intelligence* **32** (2018).
109. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440,** 680–683 (2006).
110. Liu, Y., Dolan, R. J., Kurth-Nelson, Z. & Behrens, T. E. Human replay spontaneously reorganizes experience. *Cell* **178,** 640–652 (2019).
111. Ólafsdóttir, H. F., Bush, D. & Barry, C. The role of hippocampal replay in memory and planning. *Current Biology* **28,** R37–R50 (2018).
112. Shin, H., Lee, J. K., Kim, J. & Kim, J. Continual learning with deep generative replay. *arXiv preprint arXiv:1705.08690* (2017).
113. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265,** 676–679 (1994).
114. Skaggs, W. E. & McNaughton, B. L. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* **271,** 1870–1873 (1996).
115. Ólafsdóttir, H. F., Carpenter, F. & Barry, C. Coordinated grid and place cell replay during rest. *Nature neuroscience* **19,** 792–794 (2016).
116. Schwartenbeck, P., Baram, A., Liu, Y., Mark, S., Muller, T., Dolan, R., Botvinick, M., Kurth-Nelson, Z. & Behrens, T. Generative replay underlies compositional inference in the hippocampal-prefrontal circuit. *Cell* **186,** 4885–4897 (2023).
117. Rumelhart, D. E., Hinton, G. E. & McClelland, J. L. A General Framework for Parallel Distributed Processing. *Parallel Distributed Processing: Explorations in the microstructure of cognition. Volume I,* 45–76 (1986).
118. Zhao, X., Gattoni, R., Kozlosky, A., Jacobs, A., Morrow, C., Lindo, S. & Spruston, N. Meta-learning in head fixed mice navigating in virtual reality: A Behavioral Analysis. *bioRxiv,* 2023–05 (2023).
119. Levi, A., Aviv, N. & Stark, E. Learning to learn: Single session acquisition of new rules by freely-moving mice. *bioRxiv,* 2023–03 (2023).
120. Hochreiter, S., Younger, A. S. & Conwell, P. R. Learning to learn using gradient descent. *Proc. of the Int. Conf. of Artificial Neural Networks (ICANN 2001),* 87–94 (2001).
121. Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D. & Botvinick, M. Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience* **21,** 860–868 (2018).
122. Sutton, R. S. & Barto, A. G. Reinforcement learning. *Learning* **3,** 322 (2012).
123. Vyleta, N. P., Borges-Merjane, C. & Jonas, P. Plasticity-dependent, full detonation at hippocampal mossy fiber–CA3 pyramidal neuron synapses. *Elife* **5,** e17977 (2016).
124. Winding, M., Pedigo, B. D., Barnes, C. L., Patsolic, H. G., Park, Y., Kazimiers, T., Fushiki, A., Andrade, I. V., Khandelwal, A., Valdes-Aleman, J., *et al.* The connectome of an insect brain. *Science* **379,** eadd9330 (2023).

125. Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., *et al.* A mesoscale connectome of the mouse brain. *Nature* **508,** 207–214 (2014).
126. He, K., Zhang, X., Ren, S. & Sun, J. *Deep residual learning for image recognition* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), 770–778.
127. Wang, J. X., Kurth-nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D. & Botvinick, M. Learning System. *Nature Neuroscience* (2018).
128. Whittington, J. C. R., Muller, T. H., Mark, S., Barry, C., Burgess, N., Behrens, T. E. E., Chen, G., Barry, C., Burgess, N. & Behrens, T. E. E. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* **183,** 1249–1263 (2020).
129. Graves, A., Wayne, G. & Danihelka, I. Neural Turing Machines. *Arxiv,* 1–26 (2014).
130. Sabour, S., Frosst, N. & Hinton, G. E. Dynamic routing between capsules. *Advances in neural information processing systems* **30** (2017).
131. Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B. & Moser, E. I. Representation of Geometric Borders in the Entorhinal Cortex. *Science* **322,** 1865–1868 (2008).
132. Deshmukh, S. S. & Knierim, J. J. Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus* **23,** 253–267 (2013).
133. Carpenter, F., Manson, D., Jeffery, K., Burgess, N. & Barry, C. Grid cells form a global representation of connected environments. *Current Biology* **25,** 1176–1182 (2015).
134. Derdikman, D., Whitlock, J. R., Tsao, A., Fyhn, M., Hafting, T., Moser, M.-B. & Moser, E. I. Fragmentation of grid cell maps in a multicompartment environment. *Nature neuroscience* **12,** 1325–1332 (2009).
135. Bao, X., Gjorgieva, E., Shanahan, L. K., Howard, J. D., Kahnt, T. & Gottfried, J. A. Grid-like Neural Representations Support Olfactory Navigation of a Two-Dimensional Odor Space. *Neuron* **102,** 1066–1075 (2019).
136. Stachenfeld, K. L. K. L. K. L. K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nature Neuroscience* **20,** 1643–1653 (2017).
137. Dayan, P. Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation* **5,** 613–624 (1993).
138. Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D. & Gershman, S. J. The successor representation in human reinforcement learning. *Nature Human Behaviour* **1,** 680–692 (2017).
139. Duvelle, É., Grieves, R. M., Liu, A., Jedidi-Ayoub, S., Holeniewska, J., Harris, A., Nyberg, N., Donnarumma, F., Lefort, J. M., Jeffery, K. J., Summerfield, C., Pezzulo, G. & Spiers, H. J. Hippocampal place cells encode global location but not connectivity in a complex space. *Current Biology* **31,** 1221–1233 (2021).
140. Gershman, S. J. The successor representation: its computational logic and neural substrates. *Journal of Neuroscience* **38,** 7193–7200 (2018).
141. Ito, H. T., Zhang, S. J., Witter, M. P., Moser, E. I. & Moser, M. B. A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. *Nature* **522,** 50–55 (2015).
142. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. & He, Q. A comprehensive survey on transfer learning. *Proceedings of the IEEE* **109,** 43–76 (2020).
143. Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L. & Kurth-nelson, Z. What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100,** 490–509 (2018).

144. Banerjee, B. & Stone, P. *General Game Learning Using Knowledge Transfer.* in *IJCAI* (2007), 672–677.
145. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature neuroscience* **10,** 1214–21 (2007).
146. Costa, V. D., Tran, V. L., Turchi, J. & Averbeck, B. B. Reversal learning and dopamine: a bayesian perspective. *Journal of Neuroscience* **35,** 2407–2416 (2015).
147. Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H. & Holmes, A. The neural basis of reversal learning: an updated perspective. *Neuroscience* **345,** 12–26 (2017).
148. Xia, L. & Collins, A. G. Temporal and state abstractions for efficient learning, transfer, and composition in humans. *Psychological review* **128,** 643 (2021).
149. Jones, B. & Mishkin, M. Limbic lesions and the problem of stimulus—reinforcement associations. *Experimental neurology* **36,** 362–377 (1972).
150. Raftery, A. E. Bayesian model selection in social research. *Sociological methodology,* 111–163 (1995).
151. Xu, W. & Südhof, T. C. A neural circuit for memory specificity and generalization. *Science* **339,** 1290–1295 (2013).
152. Baraduc, P., Duhamel, J. R. & Wirth, S. Schema cells in the macaque hippocampus. *Science* **363,** 635–639 (2019).
153. Baldassano, C., Hasson, U. & Norman, K. A. Representation of real-world event schemas during narrative perception. *Journal of Neuroscience* **38,** 9689–9699 (2018).
154. Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M. & Behrens, T. E. J. Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron* **109,** 713–723 (2021).
155. Guise, K. G. & Shapiro, M. L. Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron* **94,** 183–192 (2017).
156. Lever, C., Wills, T., Cacucci, F., Burgess, N. & O'Keefe, J. Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature* **416,** 90–94 (2002).
157. Barry, C., Ginzberg, L. L., O'Keefe, J. & Burgess, N. Grid cell firing patterns signal environmental novelty by expansion. *Proceedings of the National Academy of Sciences* **109,** 17687–17692 (2012).
158. Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N. & Fiete, I. R. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience* **16,** 1077–1084 (2013).
159. Morrissey, M. D., Insel, N. & Takehara-Nishiuchi, K. Generalizable knowledge outweighs incidental details in prefrontal ensemble code over time. *eLife* **6,** 1–20 (2017).
160. Yu, J. Y., Liu, D. F., Loback, A., Grossrubatscher, I. & Frank, L. M. Specific hippocampal representations are linked to generalized cortical representations in memory. *Nature Communications* **9,** 1–11 (2018).
161. Kaefer, K., Nardin, M., Blahna, K. & Csicsvari, J. Replay of Behavioral Sequences in the Medial Prefrontal Cortex during Rule Switching. *Neuron* **106,** 154–165 (2020).
162. Bradfield, L. A., Leung, B. K., Boldt, S., Liang, S. & Balleine, B. W. Goal-directed actions transiently depend on dorsal hippocampus. *Nature neuroscience* **23,** 1194–1197 (2020).
163. Park, A. J., Harris, A. Z., Martyniuk, K. M., Chang, C.-y., Abbas, A. I., Lowes, D. C., Kellendonk, C., Gogos, J. A. & Gordon, J. A. Reset of hippocampal-prefrontal circuitry facilitates learning. *Nature* **591,** 615–619 (2021).
164. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and

failures of connectionist models of learning and memory. *Psychological review* **102,** 419 (1995).
165. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience,* 4 (2008).
166. Sul, J. H., Kim, H., Huh, N., Lee, D. & Jung, M. W. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* **66,** 449–460 (2010).
167. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience* **17,** 1500–1509 (2014).
168. Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I. & Shenoy, K. V. Neural population dynamics during reaching. *Nature* **487,** 51–56 (2012).
169. Bondanelli, G., Deneux, T., Bathellier, B. & Ostojic, S. Network dynamics underlying OFF responses in the auditory cortex. *Elife* **10,** e53151 (2021).
170. Sohn, H., Narain, D., Meirhaeghe, N. & Jazayeri, M. Bayesian computation through cortical latent dynamics. *Neuron* **103,** 934–947 (2019).
171. Bellman, R. Dynamic programming. *New Jersey Google Scholar,* 24–73 (1957).
172. Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S. & Salzman, C. D. The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell* **183,** 954–967 (2020).
173. Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., Yu, B. M. & Batista, A. P. Neural constraints on learning. *Nature* **512,** 423–426 (2014).
174. Lee, J. M. & Lee, J. M. *Smooth manifolds* (Springer, 2012).
175. Izquierdo, A. & Jentsch, J. D. Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology* **219,** 607–620 (2012).
176. Watkins, C. J. & Dayan, P. Technical Note: Q-Learning. *Machine Learning* **8,** 279–292 (1992).
177. Harris, C., Aguirre, C., Kolli, S., Das, K., Izquierdo, A. & Soltani, A. Unique features of stimulus-based probabilistic reversal learning. *Behavioral neuroscience* **135,** 550 (2021).
178. Metha, J. A., Brian, M. L., Oberrauch, S., Barnes, S. A., Featherby, T. J., Bossaerts, P., Murawski, C., Hoyer, D. & Jacobson, L. H. Separating probability and reversal learning in a novel probabilistic reversal learning task for mice. *Frontiers in behavioral neuroscience* **13,** 270 (2020).
179. Kanen, J. W., Ersche, K. D., Fineberg, N. A., Robbins, T. W. & Cardinal, R. N. Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. *Psychopharmacology* **236,** 2337–2358 (2019).
180. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69,** 1204–1215 (2011).
181. Hampton, A. N., Bossaerts, P. & O'doherty, J. P. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience* **26,** 8360–8367 (2006).
182. Huh, N., Jo, S., Kim, H., Sul, J. H. & Jung, M. W. Model-based reinforcement learning under concurrent schedules of reinforcement in rodents. *Learning & Memory* **16,** 315–323 (2009).

183. Jang, A. I., Costa, V. D., Rudebeck, P. H., Chudasama, Y., Murray, E. A. & Averbeck, B. B. The role of frontal cortical and medial-temporal lobe brain areas in learning a Bayesian prior belief on reversals. *Journal of Neuroscience* **35,** 11751–11760 (2015).
184. Milner, B. Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of neurology* **9,** 90–100 (1963).
185. Joel, D., Weiner, I. & Feldon, J. Electrolytic lesions of the medial prefrontal cortex in rats disrupt performance on an analog of the Wisconsin Card Sorting Test, but do not disrupt latent inhibition: implications for animal models of schizophrenia. *Behavioural brain research* **85,** 187–201 (1997).
186. Ragozzino, M. E., Detrick, S. & Kesner, R. P. Involvement of the prelimbic–infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. *Journal of Neuroscience* **19,** 4585–4594 (1999).
187. Birrell, J. M. & Brown, V. J. Medial frontal cortex mediates perceptual attentional set shifting in the rat. *Journal of Neuroscience* **20,** 4320–4324 (2000).
188. Bussey, T. J., Everitt, B. J. & Robbins, T. W. Dissociable effects of cingulate and medial frontal cortex lesions on stimulus-reward learning using a novel Pavlovian autoshaping procedure for the rat: implications for the neurobiology of emotion. *Behavioral neuroscience* **111,** 908 (1997).
189. Bussey, T. J., Muir, J. L., Everitt, B. J. & Robbins, T. W. Triple dissociation of anterior cingulate, posterior cingulate, and medial frontal cortices on visual discrimination tasks using a touchscreen testing procedure for the rat. *Behavioral neuroscience* **111,** 920 (1997).
190. Rich, E. L. & Shapiro, M. L. Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *Journal of Neuroscience* **27,** 4747–4755 (2007).
191. Böhm, C. & Lee, A. K. Canonical goal-selective representations are absent from prefrontal cortex in a spatial working memory task requiring behavioral flexibility. *Elife* **9,** e63035 (2020).
192. Jung, M. W., Qin, Y., McNaughton, B. L. & Barnes, C. A. Firing characteristics of deep layer neurons in prefrontal cortex in rats performing spatial working memory tasks. *Cerebral cortex (New York, NY: 1991)* **8,** 437–450 (1998).
193. Frans, K., Ho, J., Chen, X., Abbeel, P. & Schulman, J. Meta learning shared hierarchies. *arXiv preprint arXiv:1710.09767* (2017).
194. Dudchenko, P. A. & Wood, E. R. Place fields and the cognitive map. *Hippocampus* **25,** 709–712 (2015).
195. Frank, L. M., Brown, E. N. & Wilson, M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27,** 169–178 (2000).
196. Hasselmo, M. E. What is the function of hippocampal theta rhythm?—Linking behavioral data to phasic properties of field potential and unit recording data. *Hippocampus* **15,** 936–949 (2005).
197. Kennedy, P. J. & Shapiro, M. L. Motivational states activate distinct hippocampal representations to guide goal-directed behaviors. *Proceedings of the National Academy of Sciences* **106,** 10805–10810 (2009).
198. Allen, T. A., Salz, D. M., McKenzie, S. & Fortin, N. J. Nonspatial sequence coding in CA1 neurons. *Journal of Neuroscience* **36,** 1547–1563 (2016).
199. Howard, M. W. & Kahana, M. J. A distributed representation of temporal context. *Journal of mathematical psychology* **46,** 269–299 (2002).
200. Konkel, A. & Cohen, N. J. Relational memory and the hippocampus: representations and methods. *Frontiers in neuroscience* **3,** 679 (2009).

201. Hsieh, L.-T., Gruber, M. J., Jenkins, L. J. & Ranganath, C. Hippocampal activity patterns carry information about objects in temporal context. *Neuron* **81,** 1165–1178 (2014).
202. Eichenbaum, H. A cortical–hippocampal system for declarative memory. *Nature reviews neuroscience* **1,** 41–50 (2000).
203. Squire, L. R., Stark, C. E. & Clark, R. E. The medial temporal lobe. *Annu. Rev. Neurosci.* **27,** 279–306 (2004).
204. Aggleton, J. P., Sanderson, D. J. & Pearce, J. M. Structural learning and the hippocampus. *Hippocampus* **17,** 723–734 (2007).
205. Bornstein, A. M. & Daw, N. D. Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLoS computational biology* **9,** e1003387 (2013).
206. Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., Kennedy, R. T., Aragona, B. J. & Berke, J. D. Mesolimbic dopamine signals the value of work. *Nature neuroscience* **19,** 117–126 (2016).
207. Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S. & Salzman, C. D. The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* **183,** 954–967 (2020).
208. Bari, B. A., Grossman, C. D., Lubin, E. E., Rajagopalan, A. E., Cressy, J. I. & Cohen, J. Y. Stable representations of decision variables for flexible behavior. *Neuron* **103,** 922–933 (2019).
209. Hamid, A. A., Frank, M. J. & Moore, C. I. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* **184,** 2733–2749 (2021).
210. Hellgren, E. C. Physiology of hibernation in bears. *Ursus,* 467–477 (1998).
211. Richardson, W. J. Timing and amount of bird migration in relation to weather: a review. *Oikos,* 224–272 (1978).
212. Chang, E. S. & Mykles, D. L. Regulation of crustacean molting: a review and our perspectives. *General and comparative endocrinology* **172,** 323–330 (2011).
213. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93,** 480–490 (2017).
214. Popper, K. *The logic of scientific discovery* (Routledge, 2005).
215. Hutchison, T. W. & Popper, K. R. Conjectures and Refutations: The Growth of Scientific Knowledge. *Economica* **31,** 217 (1964).
216. Akam, T., Lustig, A., Rowland, J. M., Kapanaiah, S. K., Esteve-Agraz, J., Panniello, M., Márquez, C., Kohl, M. M., Kätzel, D., Costa, R. M., *et al.* Open-source, Python-based, hardware and software for controlling behavioural neuroscience experiments. *ELife* **11,** e67846 (2022).
217. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Kenneth D, H. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *BioRxiv,* 061481 (2016).
218. Nichols, T. E. Multiple testing corrections, nonparametric methods, and random field theory. *Neuroimage* **62,** 811–815 (2012).
219. Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., *et al.* Array programming with NumPy. *Nature* **585,** 357–362 (2020).
220. Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W. & Bethge, M. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience* **21,** 1281–1289 (2018).