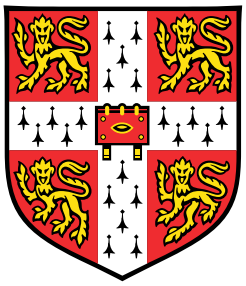
Neural Sympathy

Towards Interfaces Compatible with Neural Plasticity



Charles Gustav Livingston Flyborg Micou

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Magdalene College

July 2025

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Charles Gustav Livingston Flyborg Micou
July 2025

Abstract

Brain-machine interfaces (BMIs) enable a direct mapping from neural activity to interactions with the world. They have clinically validated applications for patients with severe loss of motor function, but also offer valuable scientific tools for probing the neural mechanisms of how actions are learned, planned, and executed. A major challenge for BMIs is the changeable environment in which they operate: as a result of ongoing experience, the brain is constantly reconfiguring its internal neural representations of tasks. To overcome the need for daily BMI recalibrations, we need to understand the interaction between BMI usage and the reorganisation of plastic neural populations.

In this thesis, we describe the design of an optical BMI based on 2-photon imaging of CA1 hippocampus in mice. Populations in CA1 are highly plastic and are critical for spatial navigation, providing an excellent testbed for interpreting changes in neural activity induced by BMI usage. Closed-loop experiments with this BMI reveal multiple representations of the same environment, which are invoked depending on whether BMI or physical locomotion is used to control navigation. We show that BMI usage induces changes in hippocampal representations that persist even when animals stop using their BMI. These changes appear unique to circumstances in which the animal has a causal influence over its travel, and hint at the incorporation of navigational agency into hippocampal maps of space.

We use representational drift, the ongoing changes in representation of familiar environments that take place even after tasks are mastered, to illustrate the eventual degradation of fixed decoder readouts. We show through modelling that, while a variety of drift formulations can degrade fixed decoders at similar rates, changes in representation that are statistically heavy-tailed are far easier for unsupervised adaptive decoders to detect and compensate for. Existing, in vivo observations of drift from both visual cortex and posterior parietal cortex, as well as our own recordings from CA1, exhibit a preference for such heavy-tailed changes in representation. We take this as a subtle hint that downstream readouts within the brain must solve similar design problems to BMI readouts from the brain. This fits into a mindset of 'neural sympathy': interface design stands to benefit from a more fundamental understanding of how and why representations change during BMI usage.

Acknowledgements

This thesis was supervised by Tim O'Leary, who has been a fantastic mentor and colleague, and whose enthusiasm and intellectual outlook are largely responsible for my interest in neuroscience.

The optical BMI experiments in mouse hippocampus described in this thesis were carried out in collaboration with Julija Krupic's lab at the Department of Physiology, Development, and Neuroscience. She has been a second advisor for this work in all but name, and I am extremely grateful for her continued insight. All of these experiments were made possible by the exceptional dedication of Hinze Ho, who has been a wonderful collaborator and a pleasure to work with. These experiments also benefited from initial surgical, imaging, and virtual reality environment groundwork laid by Pauline Kerekes, Marius Bauza, Marino Krstulović, Kou Huang, and Stella Ma.

Ethan Sorrell, a contemporary PhD student in Tim's lab, was invaluable to making quick headway thanks to his experience in optical BMI experiments. His collaborators at Harvard Medical School, Dan Wilson, Helen Yang, and Chris Harvey, were generous with both their time and their expertise.

The work on drift in this thesis was undoubtedly influenced by fruitful discussions with my drift-aligned contemporaries in the Engineering Department: Michael Rule, Dhruva Raman, Adriana Perez Rotondo, and Kristine Heiney.

Over the last three years, Fulvio Forni has humoured my involvement with several short projects with MEng students in his lab. While none of this work features in the thesis, these experiments have been an excellent playground for human interaction with computers through unusual interfaces.

I am grateful to those members of Tim's lab who had no direct involvement in the work in this thesis over the last few years, but nevertheless were always happy to provide feedback and support: Carl Ashworth, Thomas Burger, Thiago Burghi, Rory Byrne, Mónika Jozsa, Prakriti Parthasarathy, Andrea Ramirez Hincapie, Ildi Stark, and Harry Wang. Finally, I am appreciative of the support of George Malliaras, who acted as an impartial advisor for this project.

The work in this thesis was supported by the UKRI EPSRC Doctoral Training Programme studentship with grant reference 2598249 for *Long-term robustness of Brain-Machine Interfaces*.

Table of contents

1 Introduction **1**

1.1 Neural sympathy in BMI Design . . . 1
1.2 Modern approaches to BMI design . . . 3
1.2.1 Clinically viable neuroimaging techniques . . . 3
1.2.2 Optical BMIs . . . 4
1.2.3 Target brain regions . . . 6
1.2.4 Neural decoders . . . 8
1.2.5 Adapting decoders to a variable operating environment . . . 9
1.3 Representational drift . . . 11
1.3.1 The role of plasticity in drift . . . 11
1.3.2 Does drift matter for BMI design? . . . 13
1.4 The hippocampal formation . . . 15
1.4.1 Neural representations of environments and navigation . . . 15
1.4.2 Remapping and modulation of spatial representations . . . 16
1.4.3 Path integration and the cognitive map . . . 19
1.4.4 Hippocampal BMIs . . . 21
1.4.5 A simplified handling of the hippocampus . . . 22
1.5 Thesis outline . . . 23

2 Design of a closed-loop optical hippocampal BMI **25**

2.1 Design overview . . . 26
2.1.1 Design priorities and constraints . . . 26
2.1.2 1D virtual reality navigation task . . . 26
2.1.3 2-photon Calcium imaging . . . 27
2.1.4 Classification of track locations via neural decoding . . . 29
2.2 Linear decision boundaries for classification in CA1 . . . 30
2.2.1 Interpreting decision boundaries for sparse codes . . . 30
2.2.2 Comparison of decoder models for CA1 hippocampus . . . 30
2.2.3 Comparison of decoder models for PPC . . . 33

2.3 Signal Processing . . . 35
2.3.1 Feature-space design overview . . . 35
2.3.2 Downsampling and speckle elimination . . . 35
2.3.3 Motion correction . . . 37
2.3.4 Time-domain filtering . . . 37
2.3.5 Construction of the pixel feature-space . . . 39
2.3.6 Equivalence of pixel feature-space to ROIs . . . 39
2.4 Towards closed-loop BMI . . . 42
2.4.1 Screening of surgery for BMI viability . . . 42
2.4.2 Representations of non-spatial information . . . 42
2.4.3 Remarks on end-to-end BMI latency . . . 45
2.5 Discussion . . . 47
2.5.1 Open-loop evaluation: a green light for closed-loop trials . . . 47
2.5.2 Suitability for cross-day decoding . . . 48
2.5.3 Complexity of the navigation task . . . 48
2.6 Methods . . . 49
2.6.1 Signal processing . . . 49
2.6.2 Classification-based decoding . . . 51
2.6.3 Regression-based decoding . . . 52
2.6.4 Explanatory importance of kinematic variables . . . 53
2.6.5 Neuron ROI contribution to pixel feature space . . . 54
2.6.6 Surgical procedures, animal training, and imaging hardware . . . 54

3 Closed-loop BMI performance during a simple navigation task 57
3.1 Introduction . . . 58
3.1.1 Thought experiment: pairing actions with neural states . . . 58
3.1.2 Disastrous initial closed-loop decoding . . . 60
3.2 Closed-loop BMI experiments and control sessions . . . 62
3.2.1 Session structure and control modality trial types . . . 62
3.2.2 Session descriptions . . . 63
3.2.3 Experimental schedule . . . 68
3.3 Representational incompatibility between locomotion and BMI contexts . . . 70
3.3.1 Retraining of decoders produces good closed-loop performance . . . 70
3.3.2 Over-prediction of the reward location . . . 73
3.4 Factors affecting closed-loop BMI performance . . . 73
3.4.1 Attentiveness . . . 73
3.4.2 Experience and place cell count . . . 76
3.5 Remarks on alternative experimental protocols . . . 81

3.5.1 Altering the target trajectory . . . 81
3.5.2 Boiling the frog: gradually introducing BMI control . . . 82
3.6 Discussion . . . 84
3.6.1 Open-loop and closed-loop BMI are not equivalent . . . 84
3.6.2 Remapping during BMI (or lack thereof) in existing work . . . 85
3.7 Methods . . . 86
3.7.1 Mapping of decoded location to VR velocity . . . 86
3.7.2 Behavioural metrics . . . 87
3.7.3 Closed-loop decoder evaluation . . . 88
3.7.4 Tests of statistical significance and effect size . . . 89

4 Changes in hippocampal activity induced by BMI usage 91
4.1 Introduction . . . 92
4.1.1 Representations sensitive to modes of travel . . . 92
4.1.2 Does BMI usage induce persistent changes in representation? . . . 95
4.1.3 Background: interactions between cues and path integration . . . 97
4.2 Quantifying place cell properties . . . 98
4.2.1 Properties of individual neurons . . . 98
4.2.2 Place cell frequency across various control modalities . . . 100
4.3 Increased importance of visual cues . . . 102
4.4 Superposition of representations following BMI usage . . . 104
4.4.1 Representational similarity viewed at population-level . . . 104
4.4.2 Inspecting individual place cell behaviour . . . 108
4.5 Revisiting path integration with gain modulation . . . 117
4.6 Representations on cross-session timescales . . . 118
4.6.1 Tracking neuron ROIs from session to session . . . 118
4.6.2 Context-specific representations persist between sessions . . . 120
4.7 Discussion . . . 124
4.7.1 Are representations modulated by agency? . . . 124
4.7.2 Leveraging the strengths of optical BMI . . . 126
4.7.3 A remapping that is global, but not total . . . 126
4.8 Methods . . . 128
4.8.1 Quantification of neuron properties . . . 128
4.8.2 Quantification of population-level properties . . . 130
4.8.3 Archetypes of context-specificity . . . 130
4.8.4 Tests of statistical significance and effect size . . . 131
4.8.5 Identification of neuron ROIs between sessions . . . 132

5 Unsupervised adaptive correction of representational drift 135
5.1 Introduction . . . 136
5.1.1 Motivating unsupervised adaptive decoding . . . 136
5.1.2 Existing representational drift datasets . . . 137
5.2 Modelling adaptive decoding . . . 138
5.2.1 Adaptive decoding via continual tuning parameter inference . . . 138
5.2.2 Exploiting statistical regularities in a simulated population model . . . 140
5.2.3 Contrasting statistics of drift at the level of individual neurons . . . 143
5.2.4 Remarks on solving the optimisation . . . 145
5.2.5 Sudden drift is advantageous for adaptive decoding . . . 148
5.3 Is in vivo drift sudden or gradual? . . . 151
5.3.1 Distinguishing between drift types . . . 151
5.3.2 Pre-processing of the in vivo datasets . . . 154
5.3.3 More plastic in vivo populations have more sudden drift . . . 156
5.4 Revisiting cross-session tuning in our BMI experiments . . . 158
5.4.1 Distribution of correlation changes between neuron pairs in CA1 . . . 158
5.4.2 Place field displacement distributions indicate sudden changes . . . 159
5.5 Discussion . . . 161
5.5.1 Drift is gradual at the level of populations, not neurons . . . 161
5.5.2 Causes of sudden jumps in in vivo tuning . . . 162
5.5.3 Implications for adaptive BMI decoders . . . 163

6 Discussion 165
6.1 Summary of contributions . . . 165
6.2 Relation to existing work . . . 166
6.2.1 Representational drift . . . 166
6.2.2 Representations of navigation actions in CA1 . . . 167
6.2.3 Hippocampal BMI design . . . 169
6.2.4 BMI stability in clinical settings . . . 169
6.3 Future work . . . 170
6.3.1 Adaptive correction of in vivo closed-loop decoders . . . 170
6.3.2 Evolution of hippocampal BMI paradigms . . . 171
6.3.3 Neural representations of mixed modes of navigation . . . 172
6.4 Closing remarks . . . 173

References 175

Chapter 1

Introduction

1.1 Neural sympathy in BMI Design

Brain-machine interfaces (BMIs) are devices that use measurements of a subject's brain to control interactions with a computer. For patients with severe loss of motor function, such as those suffering from tetraplegia, clinical BMIs offer a life-changing means of interacting with the outside world. In the decades since the first clinical instances of surgically implanted electrodes for the control of computer actions (Hochberg et al., 2006; Kennedy and Bakay, 1998), BMIs have improved driven by advances in a range of engineering disciplines. We have higher-resolution electrode implants, faster computers to process the larger volumes of neural data, and new machine learning techniques to better unpick the intentions of the patient.

Despite this progress, BMIs still face challenges in maintaining their accuracy over time (Downey et al., 2018; Perge et al., 2013; Sussillo et al., 2016). The brain is a changeable operating environment: it is constantly learning new information and reorganising its internal representations of existing concepts. The apparent patterns of neural activity that correspond to a patient's intentions can change from one day to the next, mandating daily recalibration even for state-of-the-art BMIs. This constant need for recalibration is more than a minor design snag: it will eventually become a limiting factor. As we design BMI hardware capable of achieving neural readouts with ever-higher bandwidths, we need ever greater volumes of training and recalibration data. It is therefore tempting to consider the feasibility of a BMI capable of continuously adapting to changes in patterns of neural activity.

A neural representation describes a relationship between neural activity within the brain and a set of stimuli, tasks, or actions in the outside world. Our goal in this work is not to propose a concrete implementation of a BMI that never needs recalibration, but to take steps towards understanding how changes in neural representations interact with BMI usage. This approach is inspired by a particular design mindset from the world of high-performance

computing called 'mechanical sympathy'[1], which centres on the idea that code will eventually run on a computer constructed from physical components. Algorithms that appear slow on paper but that are implemented with a greater sympathy for the underlying physical machine will often outperform theoretically faster ones that treat the underlying machine as entirely abstract. To eventually design BMI with equivalent sympathy, we require a basic first-principles understanding of the interactions between our components. This thesis presents three main contributions towards 'neural sympathy', with the intent to facilitate the integration between interfaces and their highly plastic operating environment.

First, we describe the design of a novel BMI which uses 2-photon imaging to allow populations from CA1 hippocampus to control a navigation task. This established imaging technique allows us to simultaneously image hundreds of neurons and to scrutinise changes in their tuning over time. Using it as part of a BMI allows us to attribute the contributions of individual neurons to BMI performance. The hippocampus is a particularly exciting target brain region, both because it encodes more abstract representations of the world than traditional targets such as the motor cortex, and because its populations are extremely plastic. It provides a testbed in which learning-driven changes in population-wide activity can take place in seconds rather than in days.

Second, we describe a BMI experiment that follows the traditional paradigm of fitting a neural decoder to training data collected during a task performed via the manual manipulation of a controller. We show that, despite excellent offline decoder accuracy, animals exhibit disastrously poor performance on this task using closed-loop BMI. By considering how CA1 hippocampus synthesises both sensory cues and actions taken during navigation into neural representations, we present evidence for the existence of multiple representations of the same environment that are specific to the means by which that environment is traversed. We subsequently use this insight to demonstrate successful, accurate completion of this navigation task using closed-loop BMI.

Finally, we consider models of neural decoders that can adapt to nonstationary changes in neural representation. These should maintain accurate readouts from populations despite ongoing changes in how those populations encode information. We approach this not as BMI designers, but instead from the perspective of the constraints on learning rules implemented by neural populations in the brain. We demonstrate that not all strategies for changing neural encodings are equal, and that changes that make large localised jumps in neural tuning are preferable to small adjustments by many neurons simultaneously. We show that, fortunately for the design of adaptive decoders for BMIs, the statistics of changes in in vivo populations are favourable for stabilising downstream readouts.

[1]The phrase was originally coined by the Java performance specialist Martin Thompson, though the concept itself dates back to the earliest days of computing.

The following sections in this chapter provide background on three main topics. First, we outline clinical and experimental BMI designs, not because we have any aspirations of implementing a clinically viable product, but to motivate the importance of neural plasticity in the long-term stability of BMI design. Second, we describe the phenomenon of representational drift, treating it as a concrete example of nonstationary changes in neural encoding that may be tractable to compensate for with adaptive neural decoders. Third, we describe the behaviour of neural populations in the hippocampal formation, as this is critical to the implementation of a closed-loop BMI sympathetic to the mechanisms through which the underlying neural population builds up an internal model of the outside world.

1.2 Modern approaches to BMI design

1.2.1 Clinically viable neuroimaging techniques

The operating principle of a BMI is straightforward: measure neural activity, process the signal to identify patterns corresponding to intended actions, and execute the action. We focus first on the measurement component: neuroimaging.

There are a variety of non-invasive techniques that measure second-order effects that are a direct result of neural activity, such as currents measurable at the scalp via electroencephalography (EEG) or localised changes in blood-flow within the brain detectable via functional magnetic resonance imaging (fMRI). These have been used successfully in closed-loop BMI demonstrations (e.g. Wolpaw and McFarland (2004) and Yoo et al. (2004) respectively), but have dramatically limited spatial and temporal resolution when compared to the types of recording made possible with invasive surgeries (Abiri et al., 2019; Ramadan and Vasilakos, 2017; Sorrell et al., 2021). In addition to bandwidth limitations, they are subject to more mundane practical issues for clinical usage. EEG is fundamentally inconsistent over long time periods due to variability in recording conditions such as the dryness of a subject's scalp, which necessitates frequent recalibration and physical adjustment by a third party (Holz et al., 2015); while the physical dimensions, cost, and maintenance burden of MRI machines render them entirely unsuitable as a domestic device. Non-invasive neuroimaging techniques also face competition in clinical settings from other devices that allow patients with tetraplegia or locked-in syndrome to interact with computers, such as eye-tracking software for virtual cursor control, which achieve similar accuracy and responsiveness at lower cost and greater patient convenience (Hamid and Kristensson, 2024).

We see substantial gains in bandwidth with neuroimaging techniques that employ craniotomies. Early work with electrocortography (ECoG), which places a dense electrode array on the surface of the brain, saw substantially higher accuracy on computer cursor manipulation than EEG-based techniques (Leuthardt et al., 2004; Schalk et al., 2007). More recently, patients

with ECoG implants near brain regions associated with speech have been able to directly produce text output, albeit slowly and with errors, which could subsequently be synthesised into speech (Anumanchipalli et al., 2019; Chartier et al., 2018; Herff et al., 2015). ECoG electrode spacing is typically on the order of millimetres in high-density arrays. While a substantially higher resolution than EEG, this is still a far cry from capturing and identifying spikes from individual neurons, and thereby realising the full bandwidth potential of a BMI.

Increasing the density of ECoG electrodes on the brain surface to sub-millimetre levels provides minimal additional effective information (Slutzky et al., 2010). However, with increasingly invasive surgery, we can implant even denser electrode arrays (spacings on the order of 10-100 μm) fully within brain tissue to provide intracortical recordings. Recently, microelectrode arrays such as the Neuropixel (Jun et al., 2017) have allowed simultaneous recording of hundreds to thousands of individual neurons in humans (Paulk et al., 2022; Steinmetz et al., 2021). Intracortical imaging demonstrates hugely impressive results for BMI applications, including accurate text input at rates comparable to physical handwriting (Willett et al., 2021) and control of prosthetic arms capable of grasping and manipulating objects (Downey et al., 2018; Hochberg et al., 2012). Commercial entities pursuing BMI for clinical applications, such as the widely publicised Neuralink, have dedicated considerable research effort into the development of specialised intracortical electrodes (Hanson et al., 2019). While the surgical procedures sound daunting, thousands of similar operations are performed annually to implant deep-brain stimulation probes for the treatment of tremors in severe instances of Parkinson's disease (Sarica et al., 2023).

The contributions in this thesis are relevant specifically to high-bandwidth BMI designs with the resolution to read out from large populations of neurons simultaneously, as it is at this scale that an understanding of the long-term dynamics of a plastic neural population interacting with an interface become relevant. For clinical applications, this means intracortical electrodes, but in a lab setting we have the option of a somewhat novel tool: 'optical' BMIs that use Calcium imaging.

1.2.2 Optical BMIs

Calcium imaging uses genetically encoded indicators, such as the GCaMP family of indicators (Nakai et al., 2001), that fluoresce in the presence of Ca^{2+} ions released by the firing of a neuron. The main appeal of Calcium imaging techniques is to image populations of hundreds to thousands of neurons while being able to simultaneously identify and track the neurons individually over long periods of time. This allows us to track the long-term dynamics of neural populations and enables the study of representational drift—a phenomenon we elaborate on later in this chapter—in regions such as the hippocampus (Ziv et al., 2013), posterior parietal

cortex (PPC) (Driscoll et al., 2017), and visual cortex (Deitch et al., 2021; Marks and Goard, 2021).

Two different modes of exciting the fluorophore in Calcium indicators constrain the practical experimental details of imaging. In 1-photon imaging, the entire imaged field of view (FoV) is simultaneously excited either by a wide-field or confocal laser and the image is captured by a camera. In 2-photon imaging, where the indicators require simultaneous absorption of two photons, the field of view is raster-swept by a pulsed laser to excite indicators whose emissions are captured by photomultiplier tubes. 2-photon imaging achieves higher signal-to-noise ratios, can penetrate deeper into tissue, and has much lower interference from out-of-plane sources such as neuropil (Grienberger et al., 2022; Kim and Schnitzer, 2022). The main drawback of 2-photon imaging is the requirement to head-fix the subject, typically restricting the scope of performed tasks to those that can be performed when stationary. For the exploration of environments, a topic that is critical to this thesis, this requires that 2-photon imaging experiments take place using virtual reality (VR)—though this is now a well-established methodology (e.g. Dombeck et al., 2010; Harvey et al., 2012).

The nature of Calcium imaging renders it entirely unsuitable for clinical BMIs in human patients. Not only are the indicators introduced via viral vectors, but access to all but the layers near the surface of the brain requires removal and aspiration of cortex. Furthermore, while it boasts excellent spatial resolution, it does not match electrode recordings in temporal resolution. Resonant imaging systems, such as the one we make use of in the following chapters, can typically acquire at around 30 Hz, but the underlying limiting factor comes from the nature of the indicators themselves. The impulse response of the fastest modern indicators to a single action potential have a sharp rise time (roughly 10 ms) but decay more slowly (a decay constant on the order of 150 ms) (Zhang et al., 2020), which means that spike timings must be extracted through deconvolution. These tools exist (Giovannucci et al., 2019), but their accuracy is still imperfect.

Nevertheless, its virtues of simultaneous imaging of large populations of neurons, impressive spatial resolution, and ability to identify and track individual neurons across weeks to months make Calcium imaging a crucial tool for understanding the interactions that take place in a neural population when that population is granted agency over the outside world via a BMI. Early work with optical BMIs typically focused on operant conditioning of individual neurons in paradigms that reinforce firing above gradually ramping thresholds (e.g. Clancy et al., 2014; Hira et al., 2014). More recently, and doubtlessly facilitated by ongoing improvements to the responsiveness and signal-to-noise ratios of Calcium indicators, we have seen entire populations of hundreds of neurons used in closed-loop optical BMI in regions such as mouse PPC (Sorrell et al., 2024). The optical nature of the technique means that it can be readily paired with optogenetic stimulation to allow fine-grained control in conditioning experiments, as performed on small neural ensembles in mouse visual cortex by

Chueh et al. (2025). Other groups have sought to further push the spatial resolution of this imaging technique, with Trautmann et al. (2021) demonstrating a BMI that decodes not from a population of neurons, but from the dendrites that feed into neurons in macaque motor cortex, allowing successful closed-loop BMI control of a standard reaching task.

1.2.3 Target brain regions

With both intracortical recordings and Calcium imaging methods, we record from hundreds to thousands of neurons in a localised population. The brain region targeted for recording determines not only what information is encoded by activity in those populations and the types of patterns used to encode that information, but also the temporal stability of those encodings.

While it is a gross oversimplification to think of the brain as a purely feed-forward system that transforms some sensory input into behaviour, it is nevertheless a convenient metaphor for what may or may not make a brain region a suitable imaging target for BMI. Clinical BMI has typically targeted the motor cortex: its neural activity contains a wealth of information about immediate motions, its neural representations appear to be less volatile than brain regions responsible for more abstract computations (Gallego et al., 2020), and its physical location makes it relatively straightforward to access surgically. While there is ongoing debate as to whether a brain region responsible for motor control has the spare capacity to simultaneously control both the body's limbs and additional degrees of freedom via a BMI (Dominijanni et al., 2021), this is something of a moot point for the rehabilitation of tetraplegic patients.

More recently, BMI design has begun to entertain the possibility of moving the target region 'inward' from motor cortex, aiming to detect patterns of neural activity that correspond not to exact motions but instead to more abstract intentions (Fig. 1.1.a). This has the potential to circumvent issues of unstable moment-to-moment motion typical of motor cortex BMIs (Wolpaw, 2007). BMIs in motor planning areas, such as the PPC, have seen both clinical implementations (Aflalo et al., 2015; Andersen et al., 2019) and purely lab-based BMI investigations (Sorrell et al., 2024). Historically, BMI experiments in sensory regions have been limited to either firing-rate conditioning of single neurons or small ensembles, which we can trace back all the way back to Shinkman et al. (1974), or readouts that are entirely passive and do not control actions (e.g. Shor et al. (2022) for the olfactory bulb). However, newly proposed designs that can implant electrodes at multiple sites simultaneously aim to make use of sensory regions, either to directly compensate for the modulation of motor regions by sensory input (Gallego et al., 2022), or to inject proprioception-like input to sensory areas for fine-grained feedback for BMI-controlled prosthetics (Goueytes et al., 2022).

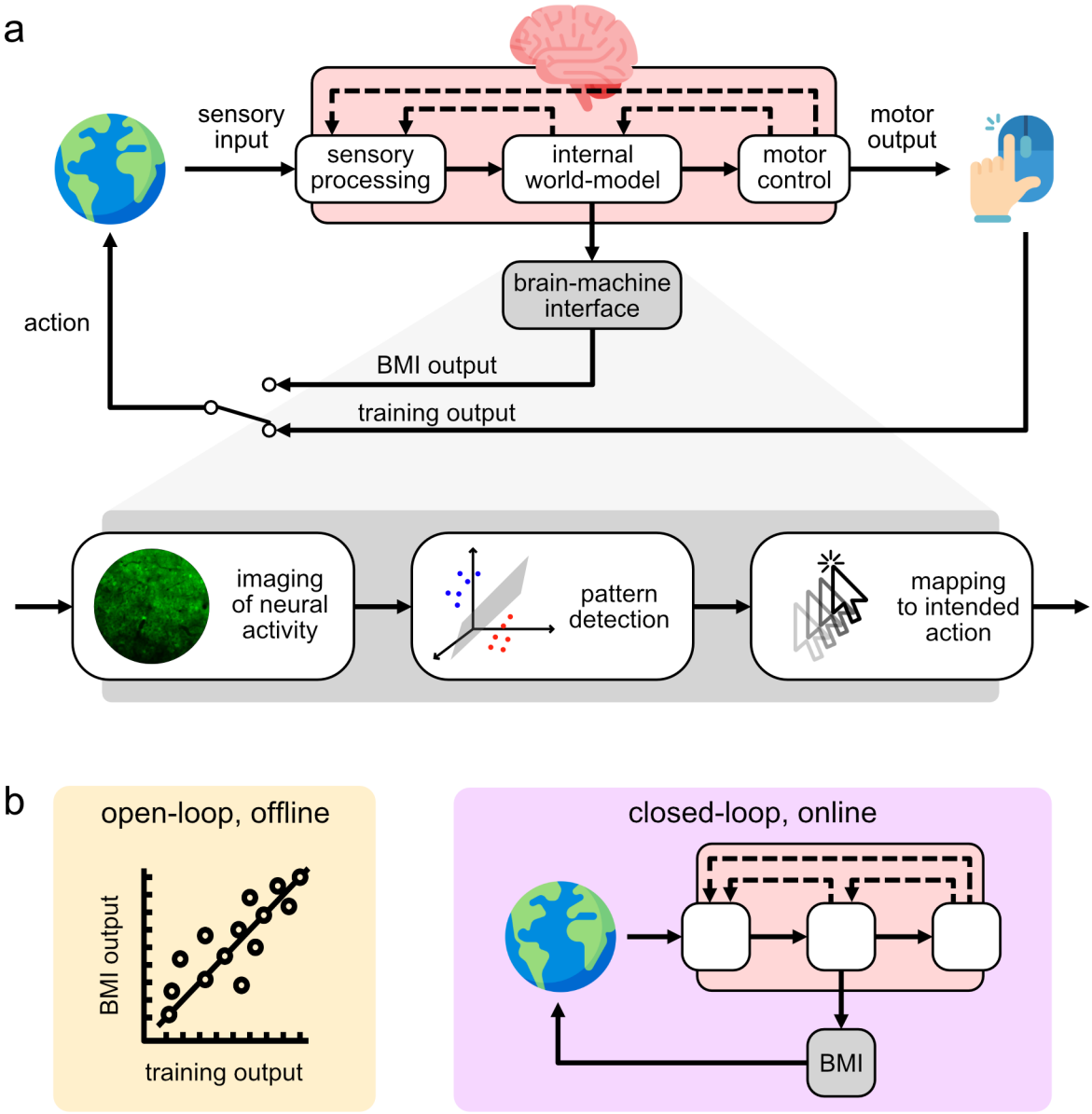


Fig. 1.1 Schematic overview of BMI operating principles. **(a)** Typical closed-loop BMI experiments involve a task that can be performed either through motor output or using a BMI. Motor output is used initially to generate training data for a neural decoder. After the decoder is trained, control of task actions is switched over to BMI control. **(b)** Open-loop neural decoders, which may even operate entirely post facto on offline data, merely compare a predicted output from the BMI to the ground-truth (the action that was actually performed). Closed-loop decoders operate online and form a feedback loop, and participation of the BMI therefore alters the dynamics of the system.

1.2.4 Neural decoders

We turn next to the problem of using imaged neural activity to determine which actions a subject intends to perform. Learning this relationship is the role of a 'decoder', and given the widespread use of machine learning techniques to implement decoders, we often describe decoders as 'trained' by providing some initial data from which to learn the activity-to-intention relation. The choices of model for decoding and the methodologies for their training vary hugely. Different imaging approaches produce different resolutions of data, different regions of the brain may encode the same information through different patterns, and practical considerations in experiment design inform the volume of training data available.

Broadly speaking, there are three methods for generating initial training data: for a subject to perform a task 'manually' through non-BMI means, to show a subject the task being performed, or to ask the subject to imagine performing the task (Gallego et al., 2022; Sorrell et al., 2021). The latter two methods are useful approaches in clinical settings, while the former two methods are useful in BMI experiments with animal subjects.

We draw an important distinction between two types of BMI experiment: open-loop and closed-loop (Fig. 1.1.b). Open-loop BMI is a feasibility assessment: successful experiments demonstrate the accuracy of neural decoding for a given combination of imaging and decoding, but do not involve a participant actually performing a task using BMI. These are typically run offline and allow direct comparisons between decoder architectures by cross-evaluating them on the same data (for example: Herff et al., 2015; Chen et al., 2023; Li et al., 2011; Gallego et al., 2020; Dyer et al., 2017). In closed-loop, which is by necessity an online experiment, the output of a decoder is used to control the outcome of a task in real-time. Closed-loop experiments are fundamentally different from open-loop experiments because the BMI becomes part of the feedback loop through which the brain learns about the world. This is not a bad thing: there is convincing evidence that subjects can improve at BMI usage over time. The clearer, faster, and more frequent the feedback they receive from the task, the better they are able to learn BMI usage (Shanechi et al., 2017; Willett et al., 2019). However, to foreshadow what is to come in Chapters 3 and 4 of this thesis, the active involvement of the brain in the closed-loop condition means that open-loop decoder accuracy is not always predictive of good closed-loop performance (Chase et al., 2009).

Decoders are not necessarily trained from end-to-end by placing neural imaging and observed actions on either side of a machine learning black box. In an excellent example of neural sympathy, Willett et al. (2021) describe a BMI for decoding text from activity in primary motor cortex. This brain region does not typically encode speech or text, and so their design employs an intermediate stage in which the patient imagines drawing letters with a pen. The decoder is trained to decode the x-y coordinates of the pen nib, and the drawn characters are subsequently subjected to a handwriting recognition-like process.

However, constructing interpretable input-sides of these decoders can be challenging. Spikes can be attributed to individual 'units' via spike-sorting algorithms such as KiloSort (Pachitariu et al., 2016), which use a clustering process to determine the identity of spike sources. These units correspond to either a single neuron or multiple neurons in close spatial proximity, as these scenarios can be challenging to distinguish. The spike sorting process becomes difficult as the number of imaged neurons grows (Trautmann et al., 2019) and is more challenging to compute online than in offline analysis, sometimes requiring the design of specialised hardware (as in Lai et al., 2023). In practice, this means that many decoder designs forego attributing activity to neurons (Christie et al., 2014), and instead train decoders either directly from electrode channel voltages (e.g. Gilja et al., 2012; Sussillo et al., 2012), or from a feature space extracted by automatic dimensionality reduction applied to the channel timeseries (e.g. Willett et al., 2021). This strategy is evidently effective at producing accurate decoders, but obfuscates underlying changes in the neural population.

1.2.5 Adapting decoders to a variable operating environment

While they demonstrate impressive accuracies and bandwidths, all the closed-loop demonstrations of BMI we have described so far have a drawback: they require frequent, usually daily, recalibration. Decoders that are trained and then fixed gradually lose their accuracy over time. Some of this variability can be attributed to varying recording conditions within the brain, for example by small movements of electrodes (Chestek et al., 2011; Downey et al., 2018). However, if this were the sole source of variability, we would expect that with a sufficient corpus of training data for an individual, a well-trained decoder would eventually have encountered a full gamut of recording conditions, especially after the initial period of electrode settling.

In practice, the success of an approach using training data that stretches far back in time is mixed: while decoders trained on longer time periods do tend to maintain their accuracy for longer, their accuracies still eventually decay (Sussillo et al., 2012, 2016). In the next section, we discuss how representational drift is one likely culprit, but here we first address the techniques currently used to compensate for apparent changes in the relationship between measured neural activity and intentions.

One family of 'trick' relies on executed behaviours having substantially fewer degrees of freedom than the neural population that generates them. A variety of techniques use an initial one-off identification of a low dimensional manifold. On subsequent days, they then learn a projection of population activity onto that manifold, which requires less training data than training a decoder from scratch (Degenhart et al., 2020; Gallego et al., 2017, 2020; Willett et al., 2021). This does not remove the need for recalibration, but reduces the time burden on the subject and is especially relevant for data-hungry high-bandwidth BMIs.

For some historical context, in addition to cross-day variability, earlier BMI designs also encountered issues with trial-to-trial variability. This produced a variety of approaches that introduced Kalman filtering stages to decoders (Kim et al., 2008; Wu et al., 2004), which were subsequently extended to techniques for compensating for cross-day variability (Fan et al., 2014; Gilja et al., 2012). These extensions are *not* unsupervised learning techniques: they require knowledge of the user's intent. However, in experimental settings, such intent can be inferred from the task structure (assuming a user wants to move to a correct target), and it is not impossible to imagine examples of how this could be translated to a clinical setting: for cursor input, assuming the user wanted a direct path to their eventual target (Wilson et al., 2023), and for text input, by employing contextual spelling and language models (Jarosiewicz et al., 2015). Decoders that adapt their weightings of neural importance based entirely on unsupervised techniques, such as metrics trying to maintain smoothness of motion, have been historically limited in their effectiveness, even in idealised simulations (Prins et al., 2017). This is gradually changing: recent advances have adapted supervised techniques for tracking latent manifolds between sessions to unsupervised contexts and demonstrated promising results on open-loop validation data (Karpowicz et al., 2025).

Kalman filtering was originally assumed to have led to increased BMI accuracy by extracting cleaner, more informative signals. Retrospective consideration of these results suggests that they may be better explained by how smooth behaviour of the computer-controlled cursors provides better feedback to the BMI user, allowing them to improve their BMI control more easily (Willett et al., 2019). A decade ago, there was substantial optimism that long-term variability could be solved by further improvements to feedback: if we could design BMIs that are easy for the user to learn, the problems of long-term variability might be self-correcting (Dadarlat et al., 2023; DiGiovanna et al., 2009; Shenoy and Carmena, 2014). How exactly to achieve this remains an open question.

Unless decoders can be entirely fixed, which at present seems unlikely, we end up with a system with two learners: a decoder learning to extract a steady readout, and a brain learning to produce stable motions. Such systems require particular care to prevent oscillatory or runaway instabilities (Müller et al., 2017). Efforts to design 'coadaptive' BMIs are still largely works of theory (Ito et al., 2020; Merel et al., 2013), with feasible practical implementations limited by an incomplete understanding of the interactions between interface and population.

In the short term, the populations used for control during BMI may exhibit systematic temporary changes in their underlying dynamics that are 'reversible': when animals return to motor control following BMI usage, the populations return to their original dynamics (Ganguly et al., 2011). Alone, this could be attributed to a difference between the sensory experiences of performing a task manually and with BMI, and may have nothing to do with plasticity. However, a growing body of evidence suggests that BMI control forms specific subspaces of population response that change independently of the learning of new motor

tasks (Jeon et al., 2022; Sadtler et al., 2014; Schroeder et al., 2022). This leads to a critical distinction in how BMI usage affects neural populations: some changes are short-lived and reversible, but others are persistent changes in neural tuning that correspond to learning within the context of BMI usage. This latter category necessarily interacts with ongoing changes in the neural representations of action plans, and is therefore particularly relevant to the long-term stability of BMI output.

1.3 Representational drift

1.3.1 The role of plasticity in drift

Neural representations are not stationary. They change over time, even for tasks that are already mastered and environments that are already familiar (Driscoll et al., 2022; Rule et al., 2019). For a concrete example, we consider a population of neurons in PPC that encode the position of a mouse in a simple T-maze task (Fig. 1.2.a) as imaged by Driscoll et al. (2017). Position is encoded at a similar level of fidelity on any given day, and yet the role individual neurons play in this neural representation changes over time. Despite apparent mastery of the task and no improvement in encoding accuracy, there is no crystallisation into a memory that is static at the level of the neural population. A crucial observation about these changing representations is that they are not mean-reverting: it is not possible to simply average over a long time period to uncover an underlying representation. Instead, they exhibit the nonstationary statistics of a random walk, which results in incremental changes to the population on each day: two representations spaced a single day apart will look more similar to each other than two representations imaged a week apart. This phenomenon, representational drift, has been observed in the visual cortex (Deitch et al., 2021; Marks and Goard, 2021), the hippocampus (Khatib et al., 2023; Ziv et al., 2013), the PPC (Driscoll et al., 2017), the olfactory bulb (Schoonover et al., 2021), and the auditory cortex (Eppler et al., 2025).

In Micou and O'Leary (2023), we have previously argued that a significant component of drift is due to plasticity: a system that is continuously trying to learn will be continuously reorganising itself. We briefly summarise these arguments here, using as a starting point three mutually compatible underlying causes for the observation of drift in recordings (excluding any systematic methodological issues with neuroimaging). First, it is possible that subtle and unobserved changes in animal behaviour[2] between recording sessions result in apparent drift in representation (Sadeh and Clopath, 2022). Second, physiological factors unrelated to task performance or learning, such as neurogenesis or synaptic perturbations, would

[2]This would imply that the behaviour itself demonstrates nonstationary changes orthogonal to task success, corresponding to an exploration of behaviour space. Given that generating this behaviour would require changes in neural activity, the distinction between neural and behavioural plasticity is perhaps obstructive (Waschke et al., 2021).

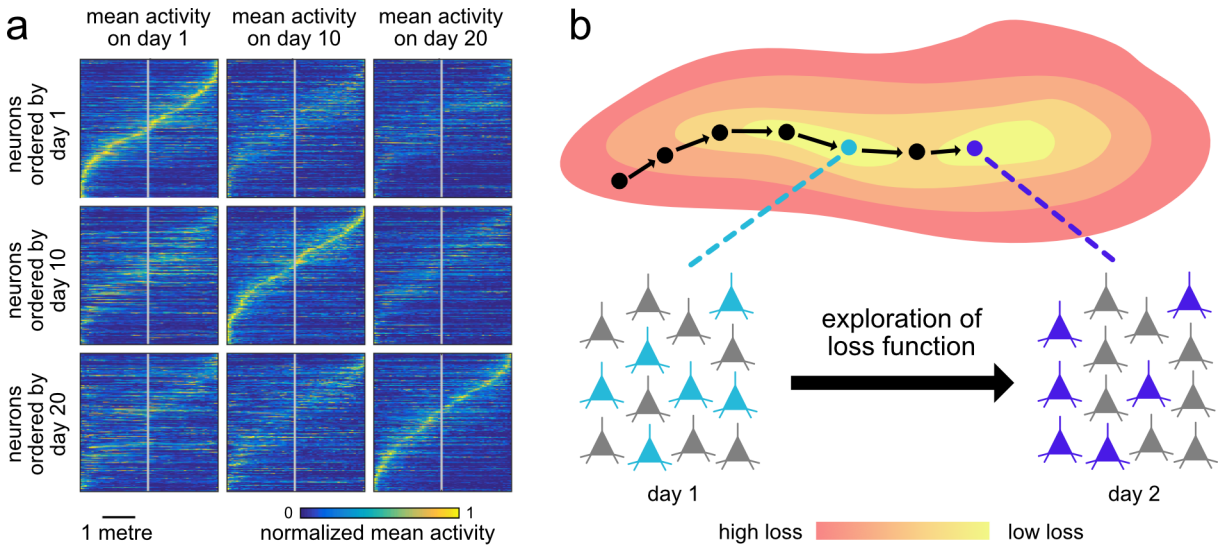


Fig. 1.2 Representational drift as a window into neural plasticity. **(a)** A population of neurons in PPC, with the mean activity of each neuron shown at each location along the main branch a T-shaped maze on three different days. While location remains encoded by the population at a relatively constant level of fidelity, the encoding gradually changes over time. *Adapted from Driscoll et al. (2017).* **(b)** As a neural population applies a learning rule to continually arriving data, it traverses the loss landscape for the quality of a neural encoding. The encoding gradually changes as the population explores roughly equivalent regions of the loss landscape. *Adapted from Micou and O'Leary (2023).*

cause a stochastic turnover in a neural population that would require compensatory plasticity (Raman and O'Leary, 2021; Zenke et al., 2017; Ziv and Brenner, 2018). Third, and most relevant for our purposes, the very learning rules implemented by the brain drive a walk through possible neural configurations that are all 'good' at solving the task. We visualise this as an exploration of a loss landscape (Fig. 1.2.b), in which a neural population that has mastered a task continues to apply a learning rule leading to walk around a high-dimensional basin in the loss landscape.

We justify the relationship between drift and learning both from first principles and from emerging experimental evidence. A system that learns either by continually sampling from an environment, such as the lived experience of an animal, or stochastically, such as in techniques like stochastic gradient descent used to train artificial neural networks, must update their internal parameters incrementally. These update rules do not converge, but instead explore orbit-like limit cycles in the loss landscape that result in distributions of parameter changes that are not mean-reverting on experimentally tractable timescales (Chaudhari and Soatto, 2018; Pashakhanloo and Koulakov, 2023). This directly predicts, and indeed we experimentally observe, that drift takes place more quickly in more plastic neural populations. Drift rates vary across brain regions (Ebrahimi et al., 2022). The hippocampus, which learns representations of environments almost instantaneously, drifts on the order of

hours to days (Khatib et al., 2023; Ziv et al., 2013). For sensory regions, which demonstrate plasticity on longer timescales, drift manifests at the level of weeks in both visual cortex (Deitch et al., 2021; Marks and Goard, 2021) and in the olfactory bulb (Schoonover et al., 2021). Representational drift in motor cortex is also slow, with certain neurons that may be directly coupled to task success even remaining relatively stable (de Laittre and MacLean, 2024; Jensen et al., 2022).

For an extreme example of a low-plasticity region, we consider HVC (proper name) in birdsong cortex. Juvenile zebra finches learn a base song from an adult tutor and subsequently repeat the same base song for life (Leonardo and Konishi, 1999; Nottebohm, 1968). This is a task with a 'before' and 'after' learning, rather than one requiring ongoing learning. HVC stores a lifelong representation of the learned song (Aronov et al., 2008; Mehaffey and Doupe, 2015), and it experiences no drift (Katlowitz et al., 2018). This suggests that any random or unavoidable component of drift due to synaptic turnover is in fact low: if the brain needs an immutable circuit, it can make one.

To complete our point on the fundamental relationship between drift and learning, we highlight the existence of several very recent studies that demonstrate that drift rates are modulated both by exposure to novelty and ongoing experience. We highlight that these results are all relatively recent and not necessarily in agreement with each other on *how* drift is modulated, but maintain that these studies do all report *some* interaction between experience and drift. Khatib et al. (2023) describe faster rates of hippocampal drift in environments that are more frequently explored. In seeming contradiction, Schoonover et al. (2021) describe slower rates of olfactory drift for scents more frequently experienced. Elyasaf et al. (2024) describe hippocampal drift rates that are decreased when animals regularly experience novel contexts, which is in apparent contradiction to models simply describing drift as new memories overwriting the old (Devalle et al., 2024). Finally, and to complicate matters further, different external factors may cause drift in different aspects of the encoding in CA1, with Geva et al. (2023) observing that repeated experience drives drift at which location in an environment a neuron responds to, and elapsed time drives drift in how strongly that neuron responds. This is an ongoing area of research, and the only statement we can make with any confidence is that understanding drift and understanding plasticity must go hand-in-hand.

1.3.2 Does drift matter for BMI design?

If a neural decoder is trained on a single day and used on subsequent days its accuracy will gradually degrade (Fig. 1.3.a). While extending training data to cover historical relationships between neural activity and some stimulus or action may partially improve cross-day performance, nonstationary changes in the underlying representation make the eventual decay of decoder performance to chance levels inevitable (Rule et al., 2020; Sussillo et al., 2016).

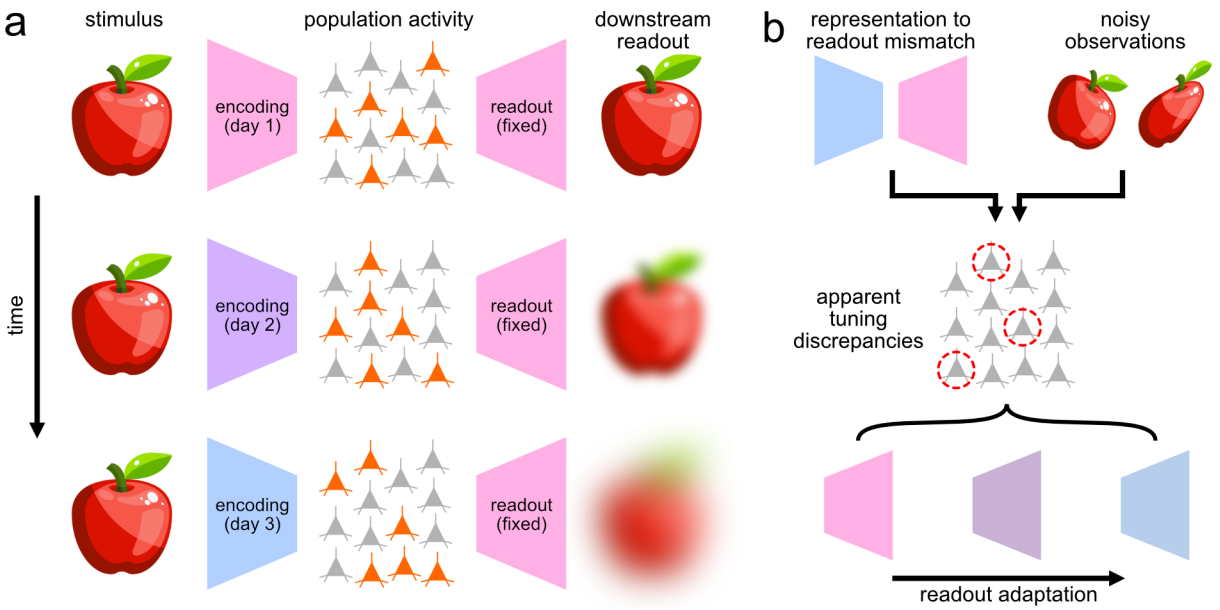


Fig. 1.3 Readout adaptation allows a steady readout from a neural population undergoing nonstationary changes. *Adapted from Micou and O'Leary (2024).* **(a)** A neural representation describes how patterns of population activity can encode a stimulus. A readout correctly matched to the encoding can infer the stimulus from population activity. However, as the representation drifts, the fixed readout becomes increasingly mismatched, and the quality of the downstream readout gradually deteriorates. **(b)** A downstream readout can attempt to correct for drift by identifying discrepancies in activity that suggest changes in tuning, but must take care not to confuse inherently noisy observations of the world for nonstationary changes in representation.

We are not the first to entertain the possibility that closed-loop BMI usage may slow or even eliminate representational drift in a population by constraining which patterns of neural activity produce task success (Driscoll et al., 2022). While there is no current experimental evidence directly comparing drift rates with and without BMI usage[3], we reiterate the widespread practice of retraining or recalibrating BMI on a frequent basis to compensate for what appear to be nonstationary changes in BMI inputs. We infer that drift is not 'frozen' by BMI usage. There is, however, the exciting possibility that an interface designed with an understanding of the underlying mechanisms of plasticity (and therefore drift) in mind would be better positioned to mitigate drift.

In keeping with our 'neural sympathy' tenet, we suggest that instead of seeking to freeze representations in populations that naturally drift, we should be able to adapt to drift—potentially without resorting to supervised learning (and thus without the need for additional

[3] In their very recent preprint, Chueh et al. (2025) describe conditioning the firing rates of small ensembles of neurons in visual cortex using a combination of Calcium imaging and optogenetic stimulation. The current iteration of the work describes qualitative support for a drift model that jumps betweens stable local minima in the loss landscape during BMI usage, and we hope that a future version will include a quantitive comparison of drift rates with and without BMI usage.

retraining data). After all, individual neurons within populations have no direct access to a ground truth: they must instead build up a representation of the external world through statistical regularities in upstream neural activity. Given both that upstream activity is subject to drift and that behaviour remains consistent, we maintain that it must be possible to produce stable readouts from a drifting neural code by identifying irregularities in the activity of upstream neurons and adapting a readout strategy accordingly (Fig. 1.3.b). Indeed, existing work has already described biologically plausible rules that allow for drift compensation by treating it as a homeostasis problem (Rule and O'Leary, 2022). In Chapter 5 of this thesis, we will describe the main results from Micou and O'Leary (2024), which suggest that details in the statistics of how neural populations drift are in fact favourable for unsupervised drift compensation.

Exciting new work motivated specifically by compensating for drift has lead to longer-term stability in clinical BMIs, even when using lower-resolution ECoG imaging (Natraj et al., 2025). There are two key reasons to keep drift in mind in the pursuit of future BMI designs. First, drift rates are faster in more plastic brain regions that might otherwise be desirable BMI targets. Second, interfaces that go beyond controlling a cursor and use their higher effective bandwidths for control in a greater number of dimensions require more training data (and thus more retraining data during calibration), which makes drift a potential limiting factor for scaling. Beyond these two concrete design considerations, we view drift as the steady-state of a learning system, and therefore take interactions between drift and BMI as offering potentially useful insight into leveraging population-level plasticity in BMI design. This is our rationale for implementing and experimenting with a closed-loop BMI in one of the most plastic brain regions: CA1 hippocampus.

1.4 The hippocampal formation

1.4.1 Neural representations of environments and navigation

The hippocampal formation plays critical roles in episodic memory and navigation. This last feature is strikingly visible in the patterns of activity generated by individual cells throughout different regions of the hippocampal formation as animals traverse an environment (Fig. 1.4). Place cells (O'Keefe, 1976) elevate their firing rate at single locations within an environment[4]. Grid cells (Hafting et al., 2005) demonstrate activity on a space-tiling hexagonal lattice that allows populations of grid cells to behave as a vector basis for position (Bush et al., 2015; Solstad et al., 2006). Head direction cells (Taube et al., 1990) encode an orientation with

[4]At least, in the typical scale of enclosure in lab settings for *rodentia*. There is considerable evidence that place cells can form fields in several locations, both from open-field recordings of bats traversing great distances (Geva-Sagiv et al., 2015), and from lab experiments that simply use larger enclosures (Park et al., 2011).

respect to a global frame of reference. The anatomical distribution of these cell response types throughout the formation indicates a structured and specialised system for navigating through environments: place cells feature largely in CA1, CA3, and the dentate gyrus; grid cells feature largely in the medial entorhinal cortex (MEC) and CA3; and there are obvious instances of these spatial encodings being combined into 'conjuctive' responses, in which individual neurons will behave as the mixture of cell types, such as grid cell firing becoming modulated by heading direction (Boccara et al., 2010).

In addition to encoding the current position in the world, activity in the hippocampus also anticipates future trajectories at decision points (Johnson and Redish, 2007). It is tempting to speculate that the role of these 'preplay' events is part of a navigation planning process (Buhry et al., 2011). Similarly, when animals are at rest and not actively navigating, we observe 'replay' events, where activity patterns appear to revisit past trajectories, and there is compelling evidence that this is important to the formation of spatial memory (Eschenko et al., 2008; Girardeau et al., 2009; Jadhav et al., 2012). These representations are therefore not merely the output of some straightforward transformation from sensory input to an inferred position, but interact with the need to choose an appropriate navigation action in a given situation.

1.4.2 Remapping and modulation of spatial representations

When animals are actively foraging or are otherwise engaged with a spatial task in an environment, their representations of that environment in CA1 are richer than if they have no specific task: more place cells form and they fire more consistently (Kentros et al., 2004). If animals become bored of a task and disengage, the fidelity at which the environment is encoded drops, as visible by a decrease in the accuracy of decoders trained to predict position (Pettit et al., 2022). When animals are stressed or afraid, not only is long-term learning impaired, but hippocampal representations can collapse entirely (Tomar and McHugh, 2022). Similarly, internal states like hunger or thirst strongly influence these representations (Kennedy and Shapiro, 2009) and can even cause 'remapping'.

Broadly, remapping describes a rapid reconfiguration of an environment's representation (Colgin et al., 2008; Fenton, 2024; Muller and Kubie, 1987). This can take place globally, in which all neurons spontaneously change their firing relationship with the environment (e.g. Fyhn et al., 2007), or partially, in which only some subset of the population is affected (e.g. Anderson and Jeffery, 2003). While remapping is often described as triggered by changes in immediate context or the effects of internal cues, historical context (that is, where an animal has been recently) can also determine which of several valid maps for an environment is used (Plitt and Giocomo, 2021). And, even after removing all of these potential contextual differences, representations of exactly the same environment during the same experimental session have demonstrated switching between multiple maps that stably coexist with one

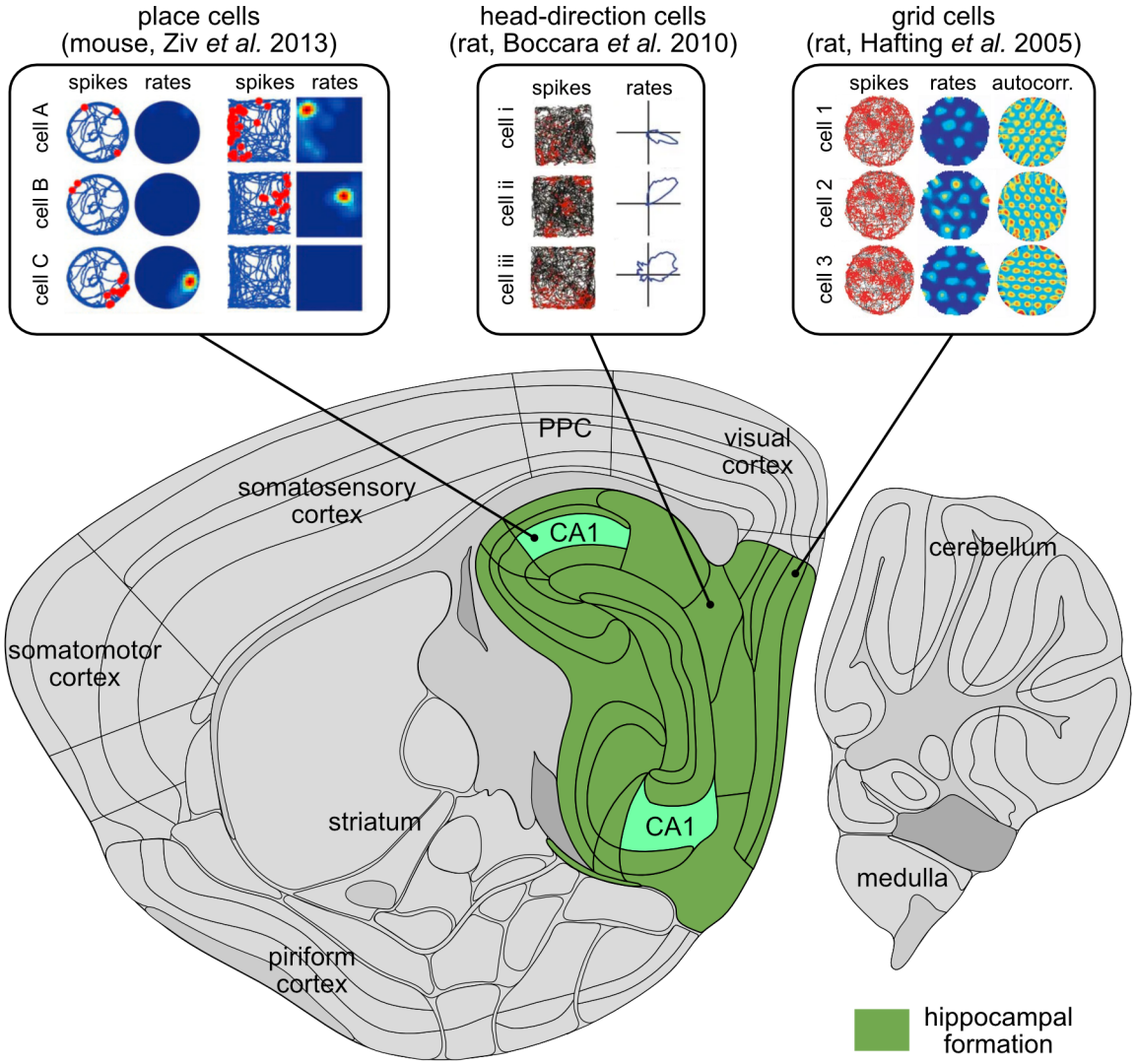


Fig. 1.4 The hippocampal formation is rich in neural encodings essential for navigation. A sagittal section of the mouse brain (*Allen Mouse Brain Atlas, Plate 100883805*) showing the hippocampal formation in relation to other brain areas, highlighting the target of our BMI: CA1 hippocampus. As illustrative aids (indeed, two of the recordings are not even from mice), we include recordings from existing work that illustrate some of the navigation primitives in the hippocampal formation. Left: place cells, firing at distinct locations within two different environments, as recorded from mouse CA1 by Ziv et al. (2013). Centre: head-direction cells, tuned to heading direction and—in this particular example—responsive conditional on grid cell activity, recorded from rat pre-subiculum by Boccara et al. (2010). Right: grid cells, with their distinctive hexagonal tiling of space, recorded from layer 2 of rat medial entorhinal cortex by Hafting et al. (2005).

another (Sheintuch et al., 2020). Exciting emerging evidence, especially within the context of implementing a hippocampal BMI, suggests that the identity of the map that gets used depends on what an animal intends to do within its environment (Zutshi et al., 2025).

In addition to remapping triggered by changes in context, it is also possible to alter neural representations of space by altering the borders of enclosures or landmarks within enclosures. This can locally distort maps, affecting both grid cell-based and place cell-based representations (Krupic et al., 2018). When populations respond to manipulations of an environment, they are sensitive to whether the changes affect proximal cues (their immediate vicinity) or distal cues (landmarks that could be used for orientation) (Knierim and Rao, 2003).

Experiments that gradually alter environments over the course of days through subtle changes have shown both gradual changes in hippocampal representations, indicating the accommodation of new information into an existing representation (Leutgeb et al., 2005), and the seemingly contradictory existence of inflection points at which sharp switches take place between representations (Wills et al., 2005). Capturing both of these kinds of hippocampal reaction simultaneously in models is challenging (Colgin et al., 2010; Jeffery, 2011), and it is perhaps easier to ascribe them each to separate phenomena that take place at two levels of abstraction: the creation of new contexts, and the updating of existing contexts (Heald et al., 2023). There are obvious parallels to the distinction we made earlier in changes in neural populations induced by BMI: some apparent changes in representation can be caused by a sufficiently large change in inputs (such as the emergence of a specific context for BMI usage), while other changes are the result of persistent alterations to tuning (learning that takes place within the BMI context). However, unlike the motor cortex—the typical target of BMI—these changes in tuning can take place very quickly.

Place fields, the regions of space in which place cells elevate their firing, can form even upon the first encounter with an environment. This is a behavioural necessity: animals learn to navigate new spaces very quickly (Rosenberg et al., 2021). Such rapid field formation implies a learning rule that takes place faster than classical Hebbian plasticity (Bittner et al., 2017). An exact formulation remains difficult, particularly because the development of representations of space seems to take place across multiple timescales—from near instantaneously to taking several minutes—may even require multiple mechanisms (Sheffield and Dombeck, 2019). In addition to place field formation being temporally heterogeneous, their distribution in space exhibits certain biases. More salient experiences, especially reward locations, end up more strongly expressed in the hippocampal code. This manifests both as elevated firing rates and as recruitment of a greater number of cells (Hollup et al., 2001; Hölscher et al., 2003; Kaufman et al., 2020; Lee et al., 2020), with strong reward representations even persisting through large remapping events (Gauthier and Tank, 2018).

1.4.3 Path integration and the cognitive map

In order to model place field formation as a learning rule applied to a population, we need some notion of what a 'place' actually is. Physical location in space is an excellent latent variable for explaining observations about the outside world. In this sense, a place could be as simple as a variety of stimuli or cues that all coincide with each other. For example, upon smelling coffee beans, hearing the sound of conversation, and seeing the interior decor, we might infer that we were in our local café (Fig. 1.5.a). The sensory richness of this location means that even in the absence of a single cue—perhaps a bad cold has impaired our sense of smell—we would still have sufficient information to place ourselves.

But associations of stimuli alone are not sufficient to solve complex navigation problems: a representation of an environment, a cognitive map, requires the binding together of places through travel (O'Keefe and Dostrovsky, 1971; Tolman, 1948). The missing piece is a system for path integration: by keeping track of where we are going in addition to where we are, we can form an internally consistent view of space (Etienne and Jeffery, 2004; McNaughton et al., 2006; Moser et al., 2017). At the level of connectivity, projections from MEC (and particularly the grid cells therein) provide signals critical for path integration in CA1 (Gil et al., 2018; Tennant et al., 2018).

Given the importance of path integration to representations in CA1, we choose to describe a view of navigation in terms of a graph structure in which places are nodes, and the actions required to travel between these places are edges (Fig. 1.5.b). This pairing of states with actions to learn space is common to a variety of models that emulate hippocampal function, such as the successor representation (Gershman et al., 2012; Momennejad and Howard, 2018), the Tolman-Eichenbaum Machine (Whittington et al., 2020), and clone-structured cognitive graphs (Dedieu et al., 2019; George et al., 2021). These models are all successful in reproducing, naturally through their learning rules rather than via explicit construction, particular behaviours of in vivo cells in CA1, such as their activity near borders (Lever et al., 2009; Solstad et al., 2008), responses to particular objects (Nagelhus et al., 2023), or the appearance of counting the number of laps of a track (Sun et al., 2020). Despite the success of these models, we do not take the hard line that the hippocampus must mechanistically implement a graph-like structure (as in, for example, Muller et al., 1996). We acknowledge ongoing debate ranging from whether the hippocampus merely learns and generates sequences, resulting in the appearance of graph-like structures (Buzsáki and Tingley, 2018; Raju et al., 2024), to a view that finds graphs incomplete as descriptions, as the modulation of activity within place fields hints at an encoding of space that is continuous rather than discrete (Alexander et al., 2020).

We benefit from outlining our own naïve mental model of CA1, which is sufficient for the treatment of the hippocampus in this thesis. It is deliberately qualitative, but captures ideas common to the more formal work above. We imagine CA1 hippocampus largely as

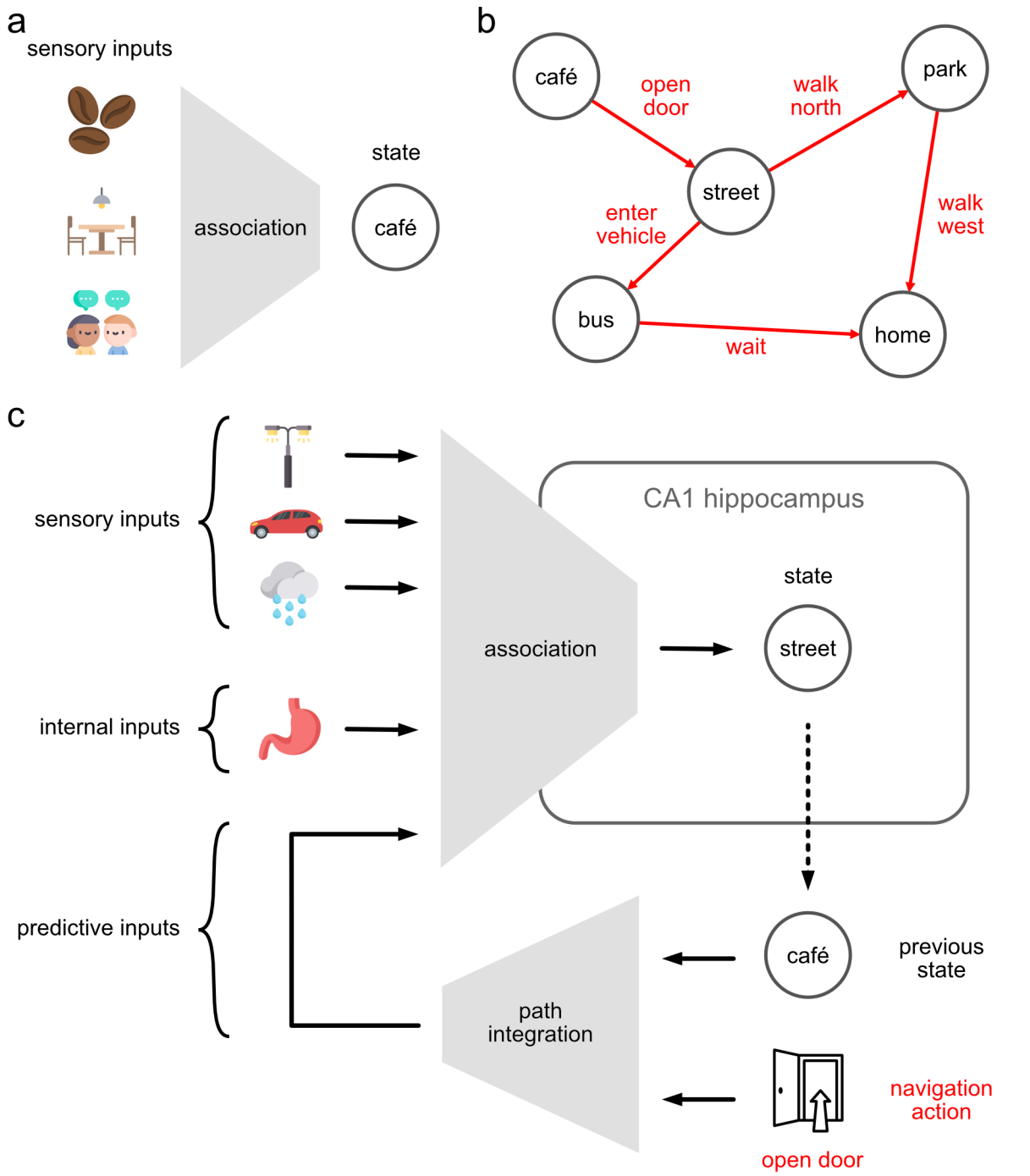


Fig. 1.5 Cognitive maps require a feedback loop to maintain internal consistency. **(a)** Physical locations in space are an excellent latent explanation for why several stimuli occur together. A single place can be constructed purely by associating together the stimuli that occur there. **(b)** Navigation requires knowledge of how places relate to each other. We use a mental model that represents space as a graph in which nodes are places (or states, if we want to generalise beyond purely spatial representations) and directed edges are the actions required to travel from one node to another. **(c)** A conceptual illustration of how stable maps are maintained in the hippocampus. In addition to aggregating sensory and internal input, past states are combined with actions taken via a feedback loop.

an associator that lumps together both coincident sensory stimuli and internal states (such as hunger or boredom) to produce a 'state' or 'place'. In attractor-like fashion, sufficiently familiar associations are recognised as existing ones, while sufficiently novel associations produce new states. Critically, CA1 is part of a path-integration feedback loop, which pairs the current state with a navigation action to produce a prediction of a next state (Fig. 1.5.c). In consistent environments, this feedback loop builds up accurate predictions that form the basis for the graph edges of a cognitive map. Under sufficient perturbation, such as global changes to the sensory cues, this feedback loop generates an entirely new map, resulting in remapping-like changes in the representation.

We will discuss at greater length in Chapter 4 how path integration is effectively a mixture of actively taken navigation actions, such as the decision to move, and signals that are available even when the navigator is entirely passive. For example, both optical flow (Madhav et al., 2024) and elapsed time (Kraus et al., 2013) feature prominently in the hippocampal formation. One exciting aspect of BMI-based navigation is the ability for the animal to retain 'navigational agency', causal control over its travel, without the sensory input associated with the physical manipulation of a controller being coupled to that travel.

1.4.4 Hippocampal BMIs

The hippocampus is by no means a typical target for BMI. It is more invasive to access surgically than the cortex and its representations are relatively volatile.

To our knowledge and to-date, there have been no successful closed-loop demonstrations of optical BMIs based on CA1 hippocampus. However, a variety of feasibility assessments based on offline, open-loop decoding have laid out initial groundwork (Chang et al., 2021; Chen et al., 2023; Sun et al., 2022). These three studies all use data from 1-photon imaging of animals navigating in simple environments and demonstrate that decoding an animal's position in real-time from CA1 populations is tractable. We will make more detailed comparisons of their signal-processing and machine learning methodologies in Chapter 2, where we discuss the implementation of our own real-time decoder.

Individual neurons in CA1 have successfully been trained to fire by operant conditioning in patch-clamp single cell recordings, but this is much more difficult in depression model mice or mice that are acutely stressed (Ishikawa et al., 2014)—we emphasise again the importance of motivation, attention, and internal states for modulation of hippocampal activity. The only existing instance of a population-wide closed-loop hippocampal BMI is the hugely impressive work in Lai et al. (2023). This describes training decoders from 40-minute 128-channel electrode recordings of CA1 hippocampus in Long-Evans rats during a head-fixed, spherical treadmill-based exploration of a 2D virtual environment. In closed-loop experiments, this BMI averaged decoded locations over a 3-second window and either moved the animal along

the decoded trajectory or, in an experimental variant, moved an object along that trajectory while the animal remained stationary. Animals succeeded at this task and demonstrated the ability to produce patterns of hippocampal activity on-demand, even at the level of large populations, to obtain a reward. This work was presented largely in demonstrating volitional control in the context of applications for clinical BMI. Its success reassures us of the technical feasibility of what we hope to illustrate in this thesis: interactions between interface and population.

1.4.5 A simplified handling of the hippocampus

In this work, there are necessarily features of the hippocampus that we either ignore or aggressively simplify. This is part pragmatism, and part focus on a subset of problems relevant to neural representations of navigation. Before we outline the structure of this thesis and proceed to a description of our BMI, we justify the treatment of several notable items as 'out of scope'.

The first of these is the theta rhythm. This is an oscillatory pattern of neural activity common to mammalian brains with a frequency of 3-12 Hz (see Watrous et al. (2013) for a review across species). Neural function in the hippocampus is strongly influenced by the cycle of this rhythm (Buzsáki and Moser, 2013). At the level of the entire hippocampal formation, the rhythm seems to coordinate a variety of behaviours, such as the timescale of remapping (Jezek et al., 2011) and whether information is being stored or retrieved (Eckmann et al., 2024; Lengyel et al., 2005). On a more fine-grained scale, the phase at which individual neurons fire within the cycle shifts as animals traverse place fields (O'Keefe and Recce, 1993), which makes for compelling evidence that information is conveyed by the timing of spikes within the hippocampus—not just by rates of firing. This suggests that it is possible to improve decoder accuracy by consideration of precise spike timing. However, the temporal resolution of Calcium imaging is too poor to accurately place spikes relative to the phase of the theta rhythm, and the actions performed via BMI take place on a longer behavioural timescale.

The second item concerns the fidelity at which we reason about the anatomy of the hippocampal formation. The imaging we perform for BMI is exclusively of dorsal CA1 in stratum pyramidale. This contains primarily excitatory cells that serve as the 'output' of CA1, part of which is subsequently fed back into the entorhinal cortex via the subiculum. CA1, in turn, receives path integration signals from the entorhinal cortex, thereby completing the feedback loop we qualitatively described in Fig. 1.5. However, CA1 is not homogeneous: deep pyramidal cells fire more strongly than superficial ones (Mizuseki et al., 2011), and the size of the place fields formed varies along the proximodistal axis (Henriksen et al., 2010). Similar observations hold true for grid cells in the entorhinal cortex, with the scale of the grid increasing from dorsal to ventral subregions (Brun et al., 2008). We image at a consistent

location within CA1, and choose to refrain from arguments relying on more fine-grained anatomy within the hippocampal formation (see Valero and de la Prida (2018) for an example of an argument that hinges on higher anatomical fidelity and specific cell types).

The final item pertains to evidence for hippocampal encoding of abstract structures that are not strictly spatial in nature (Garvert et al., 2017), such as social hierarchies (Kumaran et al., 2012). Learning rules that can generically capture all forms of relation and structure necessarily perform well at encoding representations of environments. While of great potential interest for eventual BMI designs, the BMI experiments we describe in this work concern rodents performing a strictly spatial navigation task.

1.5 Thesis outline

This thesis incrementally builds from the details of the implementation of an optical BMI for CA1 hippocampus through to a theoretical consideration of how different forms of drift affect the ability of downstream decoders to adapt. As we span a variety of topics, the results chapters are loosely broken up by underlying discipline:

- Chapter 2 concerns engineering. It describes the design of our novel 2-photon imaging-based hippocampal BMI, highlighting practical realities and trade-offs for an interface that is viable in a context of scientific inquiry.

- Chapter 3 concerns animal behaviour during BMI usage. It discusses factors affecting the closed-loop decoding accuracy of the CA1 BMI, lays out experimental structure and techniques, and illustrates the possibility of distinct representations for BMI and locomotion performance of the same task.

- Chapter 4 concerns systems neurophysiology. It describes changes in neural activity that take place in CA1 populations both during BMI usage and as a longer-lasting consequence of that BMI usage. It relates these observations back to how CA1 hippocampus synthesises information from a variety of internal and external sources to produce coherent cognitive maps.

- Chapter 5 concerns computational neuroscience. It models how downstream readouts can compensate for drift, with an emphasis on how this is impacted by differences in the statistics of neuron-level retuning. It predicts that certain forms of retuning are advantageous, and finds a preference for these forms in existing in vivo datasets.

With the exception of Chapter 5, which walks through an extensive description of modelling choices alongside its results, detailed methodologies are placed at the end of each chapter.

Chapter 2

Design of a closed-loop optical hippocampal BMI

Acknowledgement

The hippocampal imaging dataset in this chapter used as offline validation for the BMI design was collected in experiments performed in collaboration with Hinze Ho and Stella Ma, who collectively performed all animal handling, training, and surgeries, within Julija Krupic's lab at the Department of Physiology, Development, and Neuroscience at the University of Cambridge.

Chapter overview

This chapter outlines the design of a closed-loop BMI based on decoding the activity of neurons in CA1 hippocampus as imaged by 2-photon Calcium imaging. We demonstrate that our decoder:

- Processes hundreds of neurons from CA1 and decodes from them in real-time with negligible latency
- Achieves high decoding accuracy in offline, open-loop validation
- Can be interpreted to determine which neurons are relevant for decoding which states
- Does not inadvertently rely on factors relating to physical manipulation of a controller, such as blood flow or motion artefacts

Effectively, this chapter serves as methodological background for the following two chapters. Those will eventually describe how the use of this BMI in closed-loop results in changes in the hippocampal encoding of an environment. We focus here on explaining design decisions and practical trade-offs.

2.1 Design overview

2.1.1 Design priorities and constraints

Our experimental paradigm is to image CA1 in an animal while it performs a mastered navigation task in a familiar environment. We use this to train a decoder that can classify distinct locations within that environment, after which point we can 'close the loop' and allow those decoded states to causally control progress through that environment. The 'closed-loop' part of the title in this chapter is false advertising: we will, in fact, be leaving the exact details for how decoded states generate movement actions for the next chapter. Instead, we will describe relevant details of `ReTiCaDe` (the REal TIme CAlcium imaging DEcoder), a modular software library for composing signal processing and decoding that can interface with proprietary 2-photon microscopes (Micou, 2022), and assess the open-loop decoding characteristics of various BMI design choices.

One of the major constraints on our design is that we are operating in a low training data regime. Unlike existing work, which uses upwards of half an hour of data (Chang et al., 2021; Lai et al., 2023), we train decoders from a mere 10 minutes. Our maximum session length is limited by the risk of indicator damage through photobleaching. We use a short initial training period both to allow the animal to spend longer using the BMI, and to allow time for the animal to return to non-BMI control at the end of each experiment for us to search for permanent changes in tuning activity induced by BMI usage.

We prioritise the interpretability of our BMI system over marginal gains in decoder accuracy. This means favouring both signal processing techniques that are already widespread in the neuroscience literature and decoder models that are accessible to a broad audience without an extensive background in machine learning. We aim to facilitate reproducibility by removing any steps requiring human intervention during decoder training, such as the manual feature selection in Lai et al. (2023). All decoding and signal processing described in this chapter is achieved without dedicated hardware, with all real-time processing taking place on a relatively old (circa 2011) consumer-grade desktop CPU[1].

2.1.2 1D virtual reality navigation task

We use a virtual reality (VR) setup in which a head-fixed mouse initially controls its navigation along a cue-rich corridor-like track by running on a wheel (Fig. 2.1.a, Fig. 2.2). The animal learns a navigation task which requires it to stop at a specific fixed location on the VR track (250 cm from the track start). The corridor environment is rich in visual cues, with both proximal cues in the form of varying wall patterns and distal cues in the form of tower-like

[1] During the closed-loop experiments, we perform some offline training on more recent hardware to minimise downtime between trials, but this is still entirely single CPU-based.

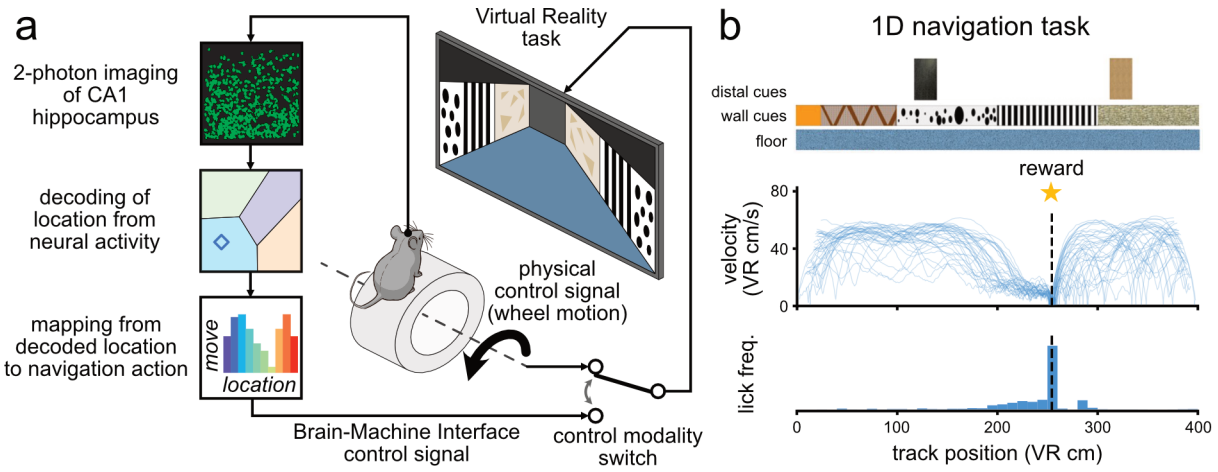


Fig. 2.1 A schematic overview of the BMI rig and corresponding navigation task. (a) A mouse views a VR environment while head-fixed on a wheel. Control of navigation in this environment can be switched between a physical control signal, allowing the mouse's locomotion on the wheel to directly drive progress along the track, and a BMI control signal, which first decodes a neural state based on 2-photon Calcium imaging and subsequently maps this state onto a navigation action. **(b)** The VR environment consists of a virtual 4 metre-long track rich with visual cues in the form of varied local wall patterns and distal tower-like objects. When running on the wheel, animals learn to stop at a fixed reward location which is not explicitly marked by cues. Each trace on the velocity profile plot denotes a unique lap from an example 10-minute trial. The lick frequency histogram shows the distribution of licking events over the course of that same example trial.

objects visible beyond the walls. The VR nature of the task means that there are no useful olfactory, vestibular, tactile, or auditory cues (excepting a *click* noise on delivery of reward, which is sweetened soy milk) to supplement the animal's navigation. The reward location is positioned in the middle of a wall cue patch and is therefore not directly indicated: its location must instead be inferred by a combination of distance to distal cues and path integration. Reaching the end of the track will place the animal in darkness for 2 seconds and then return it to the start of the track, allowing the animal to run multiple laps over the course of a 10-minute trial. Expert animals develop a stereotyped 'parabolic' velocity profile, in which they accelerate at the start of the track and slow down ahead of the reward site, and will lick the spout in anticipation of receiving their reward (Fig. 2.1.b).

2.1.3 2-photon Calcium imaging

We image dorsal CA1 using 2-photon Calcium imaging using GCaMP8f as an indicator (see Methods 2.6.6), using a resonant galvanometer mirror scanner that allows us to sample a complete field of view (FoV) at 30 Hz at a resolution of 512x512 pixels. For retrospective analysis of activity we identify neuron regions of interest (ROIs) within the FoV and extract

fluorescence traces with the widely-used `suite2p` software package (Pachitariu et al., 2017). Individual recording sessions contain roughly between 250 and 1,000 ROIs, dependent on the success of the surgical procedure and indicator expression, of which approximately two thirds are typically identified as neurons with high confidence by `suite2p` based on footprint shape (Fig. 2.3, see Fig. 4.5 in Chapter 4 for a more complete animal-by-animal quantification). In the following section, we perform offline evaluation of decoders that use as input the fluorescence of each of these neuron ROIs on each imaging frame, with the intent to inform the design of an eventual BMI pipeline that operates on data ingested in real-time.

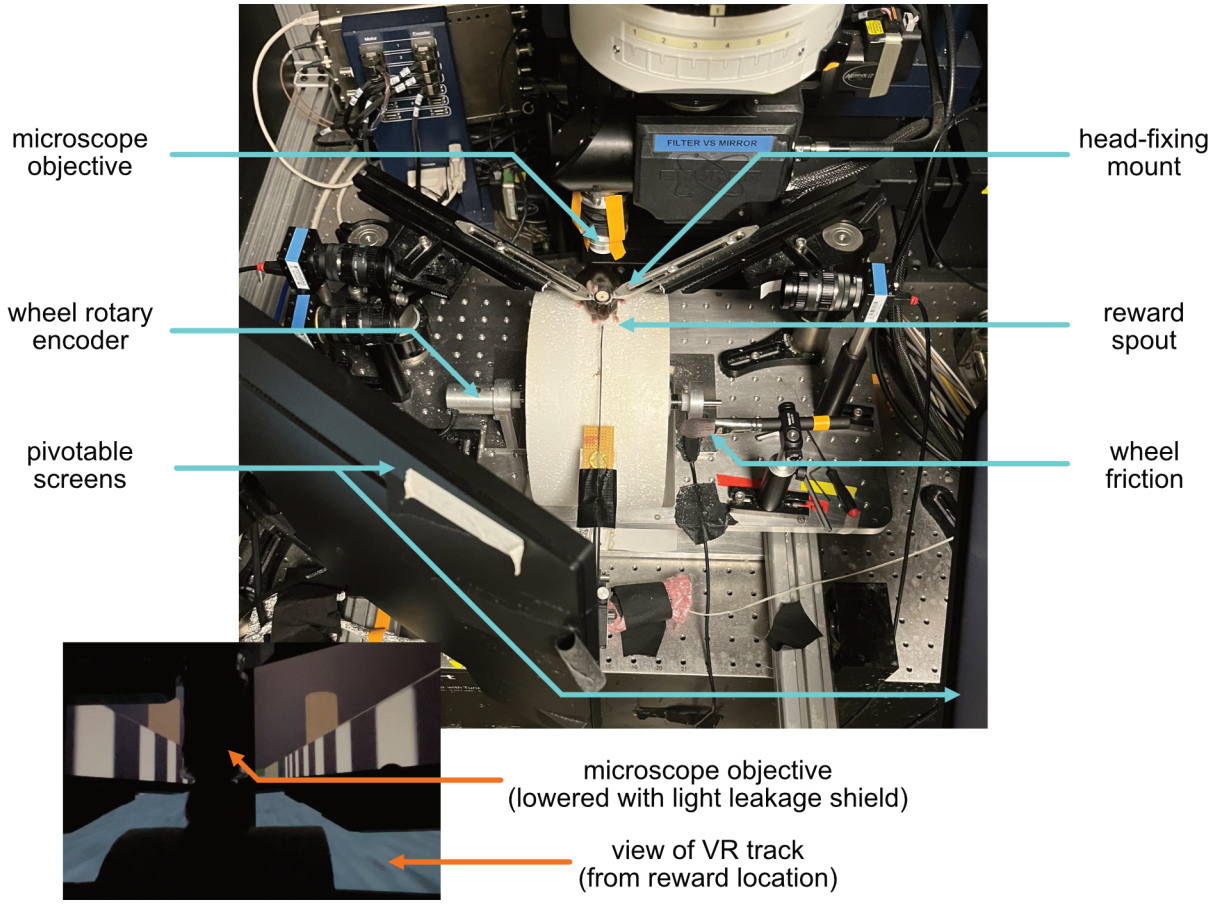


Fig. 2.2 Physical realisation of the BMI rig. A picture of the BMI rig showing the main components. The VR environment is displayed across two screens that can be angled to occupy the entirety of the animal's field of view. A rotary encoder tracks the wheel's orientation for navigation of the environment via locomotion. A brush can be pushed up against the wheel to provide additional friction to control the amount of effort required to complete a lap. The reward spout is instrumented to record licking events. **(inset)** A view from behind the animal during a trial showing the VR environment at the location where the reward is dispensed.

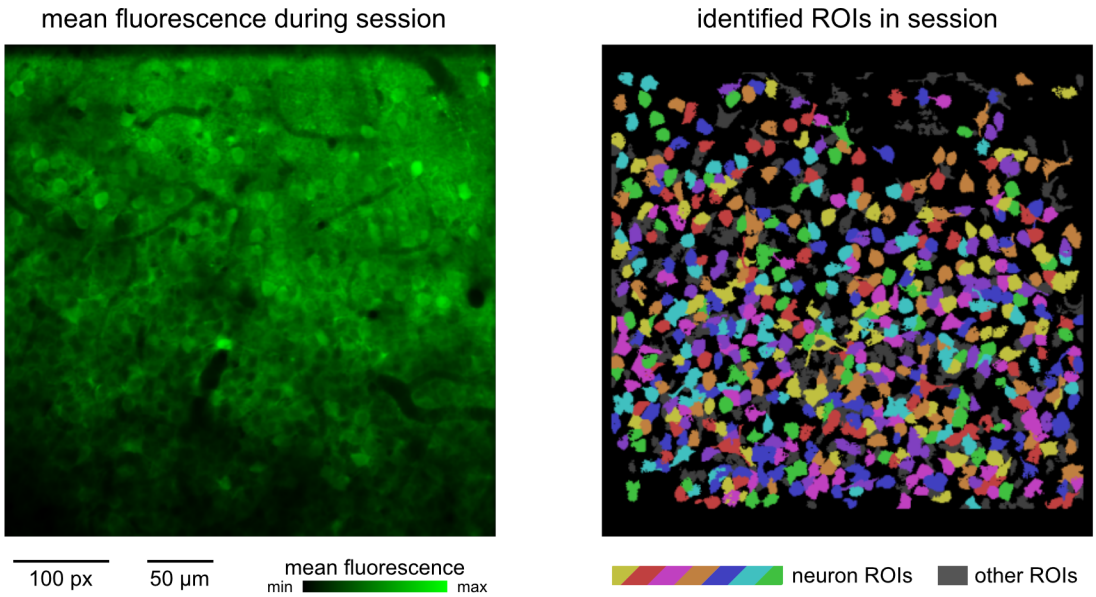


Fig. 2.3 A typical field imaging field of view. (left) Mean fluorescence levels in the image over the course of a 10 minute-long recording during performance of the navigation task. **(right)** ROIs identified in the same recording as corresponding to neurons with high confidence (various colours), in addition to other ROIs (grey), as extracted by the `suite2p` software package.

2.1.4 Classification of track locations via neural decoding

We decode the location of the animal on the track based on its neural activity. Rather than decoding this location as a continuous position (i.e. a regression for a coordinate along the dimension of travel), we instead discretise the track into spatial bins and classify locations (Fig. 2.4.a), similarly to Chen et al. (2023). This discretisation is motivated by our mental model of the cognitive map: by treating places as discrete, we can reason about defining concrete navigation actions to travel between them.

More pragmatically, the discretisation also allows us to skirt around some thorny decoder training issues arising from attempting to decode a continuous position with minimal training data. The teleportation upon lap completion which returns the animal to the start of the track means that our environment is topologically circular: the end of the track is 'near' to the start. While it is possible to construct loss functions for a circular geometry, this limits our ability to use widely-understood off-the-shelf machine learning models to map from neural activity to position. In the preliminary stages of design, we observed that training regressors that minimised a mean squared-error (MSE) cost function would end up biased towards predicting positions at the centre of the track, as mispredictions near the extrema are more severely punished. By discretising the track into bins and treating this as a classification problem, giving the decoder training no knowledge of spatial bin adjacency, and by balancing training samples by bin occupancy, we can sidestep this bias. We use 10 bins, each 40 cm

in length. While the exact choice of bin size is somewhat arbitrary, this division of the track has two main desirable properties: the reward location is well-contained in a bin, and only a minority of transitions between wall cues line up with bin boundaries (allowing us to evaluate whether local decoder accuracy is contingent on proximal cues).

2.2 Linear decision boundaries for classification in CA1

2.2.1 Interpreting decision boundaries for sparse codes

Place cells in CA1 hippocampus result in an encoding of space that is sparse: neurons fire at specific locations, and are silent (or fire at a far lower rate) elsewhere (Fig. 2.4.a). This encoding strategy makes it straightforward to describe how a population of neural activity is used by a classifier. Consider a plot of the samples of our training data in a space where each dimension corresponds to the activity of each neuron for each sample, and then marking each sample as either 'in' or 'not in' some spatial bin y. For a large, idealised population where all place cells behave like switches that are either 'on' or 'off', it is possible to entirely separate those samples in each bin y from those outside it using a dividing plane (a linear decision boundary). To decode whether some future sample with an unknown bin is in bin y, we simply determine which side of the decision boundary the sample falls in (Fig. 2.4.b). The orientation of the decision boundary, which we will refer to collectively as 'the weights', tells us which neurons are and are not important for determining whether a given sample is in bin y. By constructing such a decision boundary for each bin, we end up with 10 one-versus-rest (OvR) classifiers, each with their own set of weights informative about which neurons matter for decoding their own bin, that can be composited together to classify any given sample into its decoded bin (Fig. 2.4.c).

Obviously, place cells do not behave like idealised switches. While some cells do consistently fire at a given location on every visit (Fig. 2.4.d), it is more common for cells to only fire on some fraction of visits (Fig. 2.4.e). A given place cell not firing is therefore not a good indication that an animal is outside of that cell's place field. Decoders need a population sufficiently large to have redundancy in decoding each bin, as well as enough training data to learn which cells are reliable predictors of bin occupancy. The question is whether place cells behave sufficiently closely to idealised switches in our own experiments that linear decision boundaries are appropriate, or whether 'heavier duty' models are required.

2.2.2 Comparison of decoder models for CA1 hippocampus

The decoder we use in closed-loop experiments is a classifier based on a support vector machine (SVM, sometimes abbreviated SVC for support vector classifier) with a linear kernel

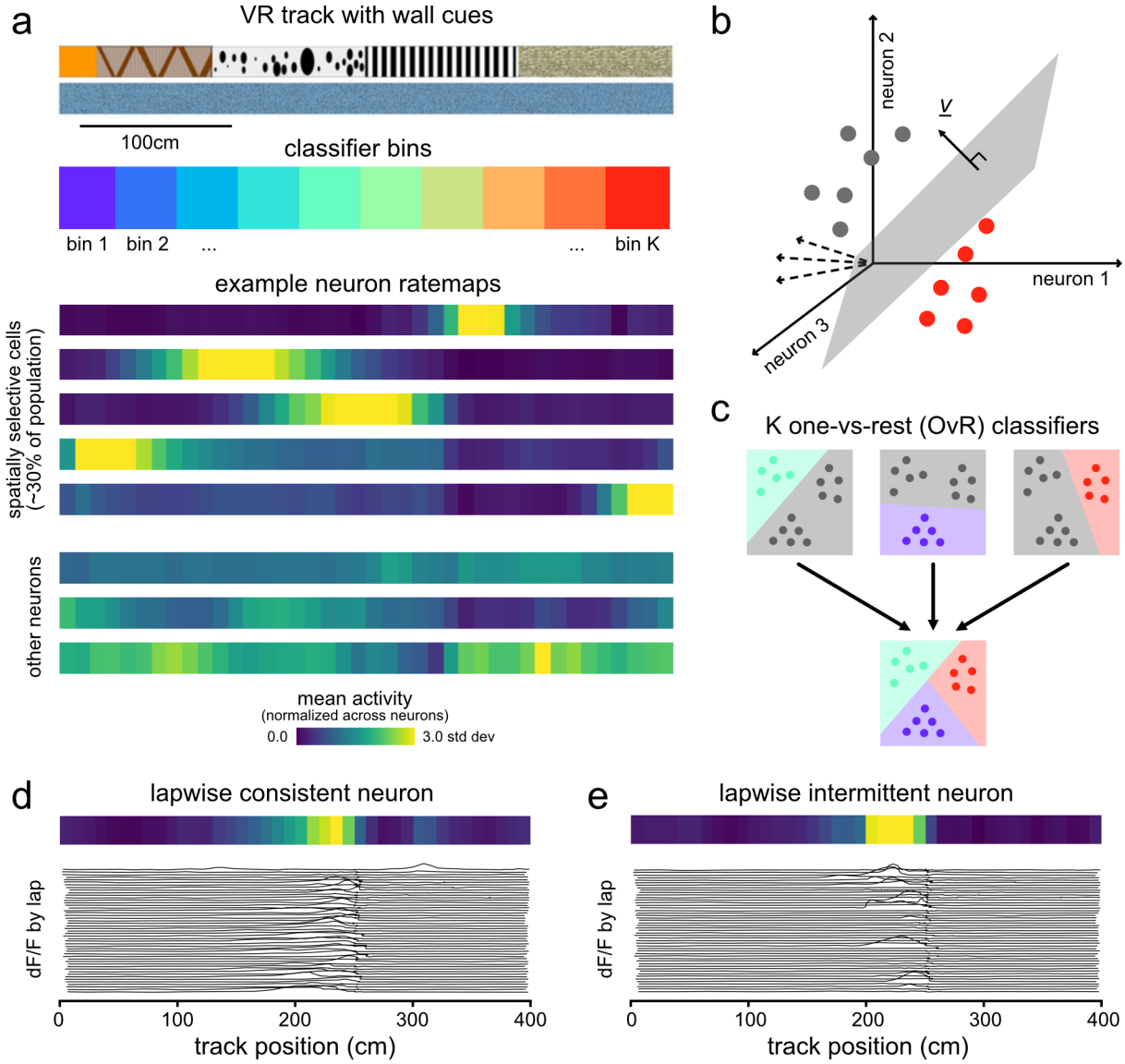


Fig. 2.4 Representations of position in CA1 lend themselves to simple linear classifiers. **(a)** We train models to predict the location of an animal as one of K discrete bins on the 1D track based on neural activity. In populations of CA1 neurons, neurons carrying information about location use a sparse encoding, typically firing in a single region on the track. **(b)** This sparse encoding means that distinguishing samples at one location (red) from samples from all other locations (grey) is tractable with a simple linear decision boundary. Furthermore, the orientation of this decision boundary $\underline{v}$ describes which dimensions of the feature space (individual neurons) are relevant for decoding this location. **(c)** By producing a one-vs-rest classifier for each of the K bins and combining their decision boundaries, we are able to classify new samples into a discretised location on the track. **(d)** The ratemap of an individual neuron and the underlying filtered Calcium traces for each lap of the track during a 10-minute recording session. **(e)** As in (d), showing a neuron with a ratemap that is convincing on the surface, but reveals inconsistent firing behaviour when considered lapwise.

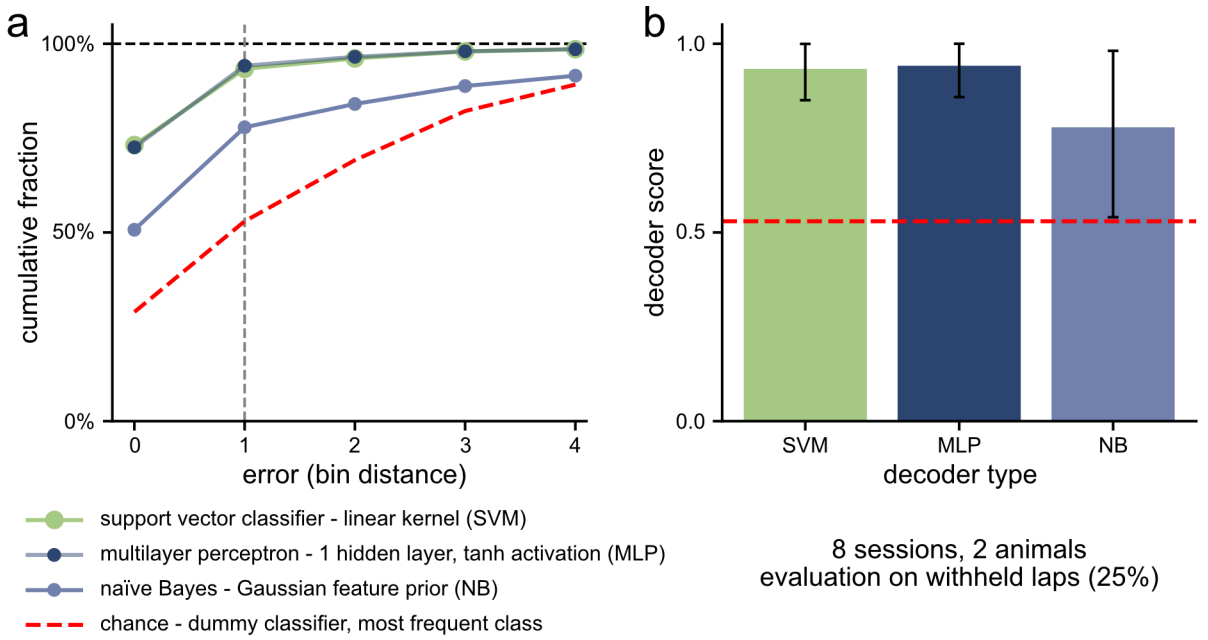


Fig. 2.5 Linear kernel SVMs classify track positions with high accuracy. **(a)** A cumulative error plot comparing SVM, MLP, and Gaussian Naïve Bayes classifiers. The x-axis tracks the error of a classified sample (a sample classified in the correct bin has an error of 0, a sample classified in a bin immediately adjacent to the correct bin has an error of 1, and so on). The y-axis indicates the cumulative fraction of samples classified at a given error level, averaged over 8 sessions from 2 different mice. **(b)** A simple decoder score corresponding to the fraction of the withheld testing dataset correctly classified to within 1 neighbouring bin of error. Error-bars indicate the minimum and maximum scores over all 8 sessions over 10 shuffles of which track laps are used as training data and which are used as testing data.

(Cortes and Vapnik, 1995). This explicitly constructs a linear decision boundary for each class (spatial bin). It trains and decodes quickly, and is a reasonable fit for our volume of training data and feature input size. We are far from the first to apply SVMs to hippocampal data (it is difficult to pinpoint the first application, but for some early examples see Takahashi and Sakurai (2009), which is based on small ensembles, or Hassabis et al. (2009), which is based on fMRI data). Nevertheless, we find it important to demonstrate its suitability for our specific experimental setup.

In Fig. 2.5, we present an open-loop decoding comparison between three types of classifier (see Methods 2.6.2). The first is the SVM. The second is a multi-layer perceptron (MLP) with a single hidden layer, capable of constructing a non-linear decision boundary and therefore in theory heavier duty than the SVM. The third is a Bayesian approach, specifically a Gaussian Naïve Bayes classifier (NB), which became a popular decoding approach in the hippocampal literature starting with Zhang et al. (1998). For each of eight 10-minute sessions (2 animals, 4 sessions each) of the VR task controlled by the animal running on the wheel, we perform a 75%-25% split of the laps in that session into training and testing data and evaluate decoder

performance on the unseen test laps. We repeat this 10 times per session, shuffling which laps are allocated to training and testing, and visualise mean decoding performance in Fig. 2.5.a as a cumulative error plot. This visualisation allows us to account for some misclassifications being worse than others based on spatial proximity of the bins. For example, in the case of NB, ~50% of samples are correctly classified into exactly the correct bin, and ~75% of samples are classified in either the correct bin or one of the spatially adjacent bins, and so on. We find that the SVM and the MLP perform almost identically, while NB performs relatively poorly.

We extract a simple decoder score from the cumulative error evaluation by taking the fraction of samples correctly classified to within one neighbouring bin of accuracy (Fig. 2.5.b). This is a somewhat generous scoring method, but it is in keeping with the scheme used in Chen et al. (2023), and has the virtue of leniently handling situations where the animal is near the boundary between two spatial bins. Our result, that SVMs perform just as well as more complex models, is in agreement with a similar assessment in Sun et al. (2022), who also report no improvement for even heavier-duty decoders that use models that are recurrent in addition to non-linear. While more sophisticated models are likely to eventually achieve higher accuracy with larger volumes of training data, we find SVMs entirely satisfactory for our experimental needs.

2.2.3 Comparison of decoder models for PPC

The parity of a linear method with more complex models is a particular and convenient feature of the sparse encodings in CA1. This is facilitated by an imaging technique well-suited to isolating the signals of individual neurons from each other. As a short detour from the design of our CA1-based BMI, we illustrate this point by performing a comparison between decoding techniques applied to populations in the PPC, which simultaneously encodes multiple kinematic variables (such as heading direction, position, and velocity) with a denser code.

We consider sessions of an animal performing a VR T-maze experiment, in which it manipulates a running ball to traverse a long corridor and then makes a choice by turning left or right (Fig. 2.6.a). On each attempt at the maze within a session, the correct turn direction is indicated by wall cue patterns at the start of the corridor. These wall cues disappear during the latter half of the corridor, and the animal must make its turn choice by remembering the correct turn (or by adopting some behavioural strategy such as hugging the correct wall). Using a largely similar methodology[2], including the same spatial and temporal resolution of imaging, the same Calcium indicator, and the same software (`suite2p`) to extract neuron ROI traces, this experiment produces a dataset comparable to our own hippocampal setup.

[2]We do not provide a full methodology in this thesis. Consult Sorrell et al. (2024), Driscoll et al. (2017), and Harvey et al. (2012) for a complete description.

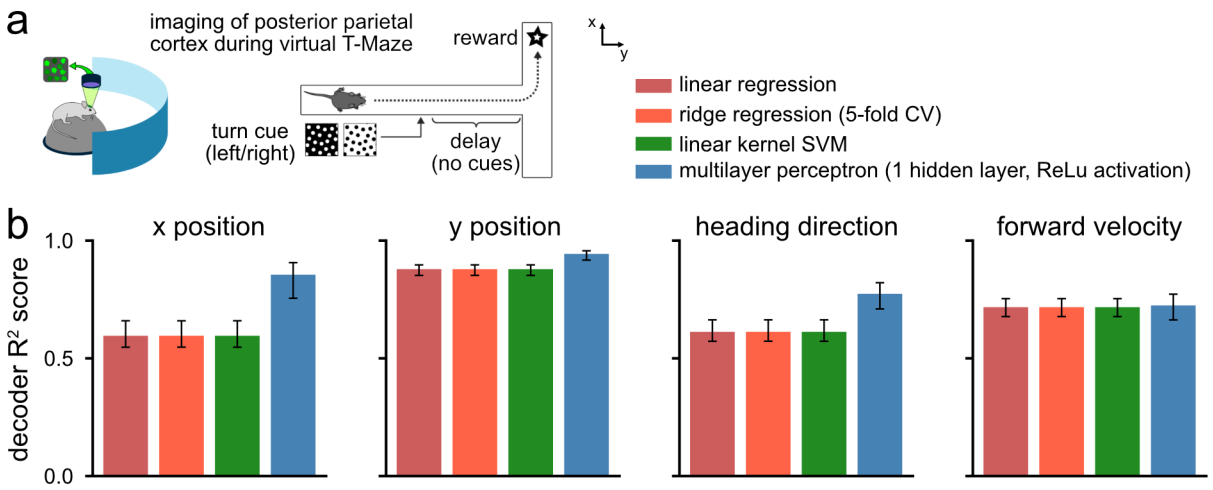


Fig. 2.6 Certain kinematic variables in posterior parietal cortex are decoded more accurately with non-linear decoders. (a) A two-dimensional VR environment in which an animal performs a T-maze task, selecting a left or right turn at a decision point based on a wall cue. *Schematic adapted from Rule and O'Leary (2022).* **(b)** Decoding accuracy for of a variety of kinematic variables from 2-photon Calcium imaging of the PPC during T-maze trials. Three linear decoders and one non-linear decoder are shown. Evaluated on a single session for each of 5 animals, scored on withheld maze attempts of a 75%-25% train-test split for 10 shuffles of split. Error bars show minimum and maximum values over shuffles and sessions. Number of maze attempts per session varies (lowest: 131, highest: 319). *Unpublished data, courtesy of Helen Yang, Dan Wilson, Ethan Sorrell, and Chris Harvey.*

We evaluate the accuracy of several models trained to decode continuous kinematic variables: position in the maze, heading direction, and forward velocity (Fig. 2.6.b, see Methods 2.6.3). Three of these models are linear (simple linear regression, ridge regression, and an SVM using a linear kernel), while one is non-linear (MLP with a single hidden layer). With the exception of forward velocity, which is equally well decoded by all models, the MLP outperforms the linear methods.

The takeaway message relevant to our hippocampal BMI is that the combination of Calcium imaging with CA1's encoding strategy is rather uniquely positioned to construct accurate BMI decoders with easy-to-interpret weights. While it is certainly possible to measure the impact of an individual neuron's activity on the output of a more sophisticated decoder, the complexity of that decoder necessarily dictates the complexity of the mapping from neuron to outcome. This is not just a stumbling block for analysis: if it is difficult to determine how a neuron's activity relates to BMI control, then it also becomes difficult for the neural population to improve BMI performance in closed-loop. Decoder design for other brain regions must compromise by trading off potential decoding accuracy against a clean mapping between activity and outcome to facilitate learning via feedback, but we are in the fortunate position of needing to make no such concessions in our CA1 setup.

2.3 Signal Processing

2.3.1 Feature-space design overview

The decoders trained in the previous section processed offline data and used neuron ROI fluorescences as attributed to ROIs by `suite2p`'s registration process. It would be plausible to do this in real-time by first extracting ROI footprints from the training data used to initialise our BMI decoder and then attributing subsequent activity to ROIs from an image via masking, as in Chen et al. (2023). However, there is an issue of bias towards active cells, exacerbated by our short 10-minute recordings of training data: automatic registration that detects ROIs from fluorescence is more confident in cells that are more active, as they fluoresce more often and have better-defined borders. If we want to adapt or update our decoder to account for a significant remapping in activity, we would either have to re-register cells or need to set our confidence intervals so low that 'neuron ROI' might lose any interpretability benefits.

We do not have a large enough training corpus with a mere 10 minutes of training data to ingest an entirely raw image and let a convolutional neural network take care of signal processing for us, as in Chang et al. (2021), and this approach would also render interpretation of the signal difficult. Instead, we manually construct a signal processing pipeline that transforms a raw image input into a feature space in which pixels correspond closely to neurons (Fig. 2.7). Readers will notice a high-level resemblance in this pipeline to the image processing in Sorrell et al. (2024), which solves what is essentially the same problem, though with different choices in the minutiae of motion correction algorithm, filtering, order of operations, and the deliberate choice to scale the image resolution to force a close correspondence between pixel features and neurons.

The signal processing pipeline needs to execute in under 33 ms to keep pace with incoming image frames. An added benefit of using a feature space of fixed dimensions is that the decoder will execute in a consistent amount of time: we do not run the risk of an unusually large number of ROIs altering the performance characteristics of the pipeline (or indeed taking longer to train, causing a larger gap between training and closed-loop evaluation of the BMI). We outline the main steps of this pipeline here, and then describe how the features in the pixel space can be related back to neuron ROIs to determine the influence each neuron has on decoding outcome.

2.3.2 Downsampling and speckle elimination

With an imaging resolution of 2 pixels/μm, we have an abundance of pixels per neuron (which are roughly 10-15 μm in diameter as imaged). However, resonant imaging at 30 Hz produces a speckled image rich in high-frequency noise in the spatial domain. This speckling interferes with motion correction, generating spurious corrections, and so we apply a low-pass spatial

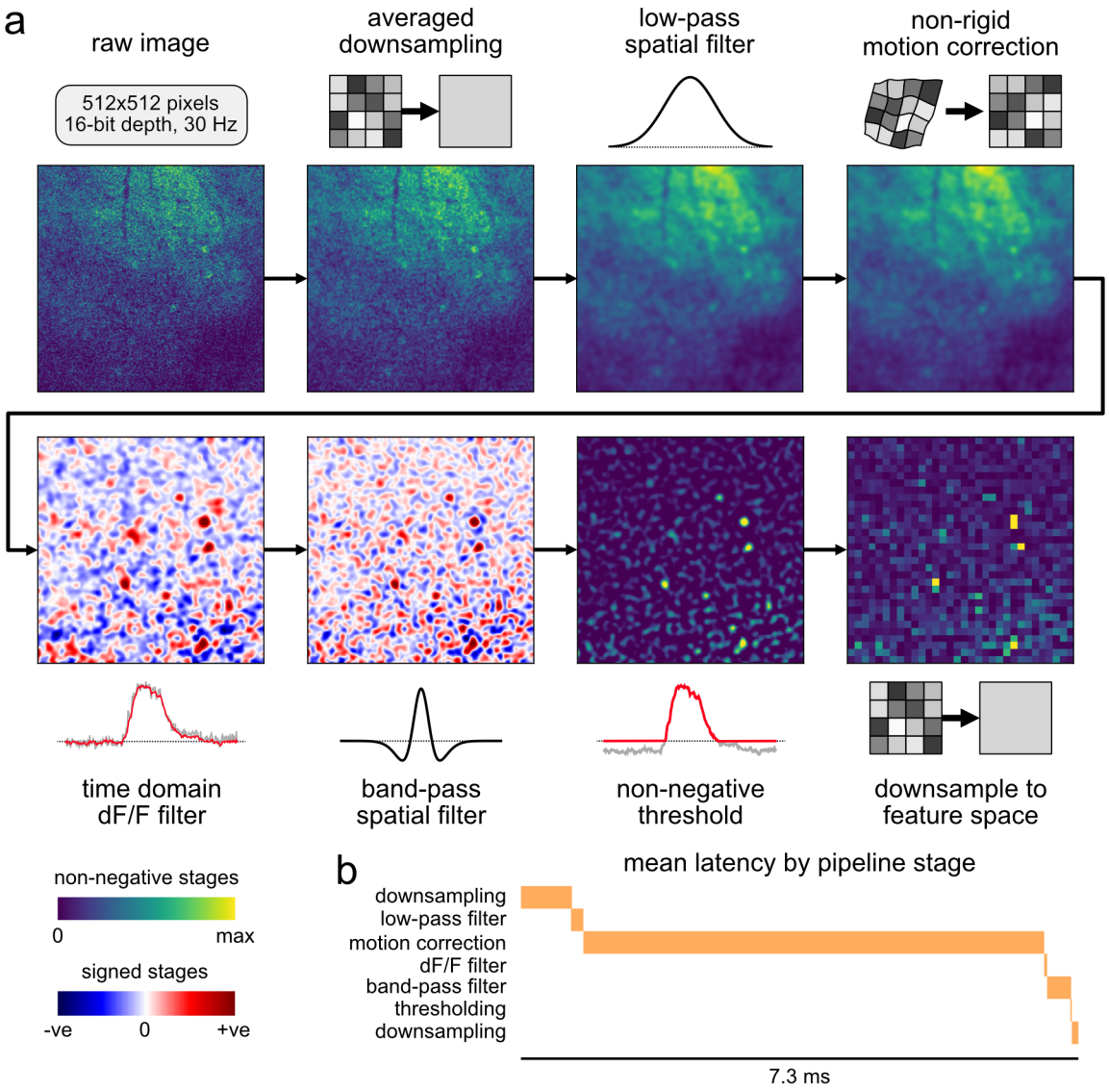


Fig. 2.7 An overview of the signal processing pipeline for extracting a feature space from a feed of images of the field of view. **(a)** The sequence of pipeline operations with their effects illustrated by an example frame. Prior to application of the dF/F filter and band-pass spatial filter, and following the thresholding stage, values in the frame are non-negative. **(b)** A waterfall latency diagram attributing computation time to each stage of the pipeline, averaged over 17,994 frames of a 10-minute trial and measured on the same computer hardware used during the closed-loop BMI trials described in Chapters 3 and 4.

filter in the form of a Gaussian averaging over pixels. In this stage, we also shrink the resolution of the image from 512x512 pixels to 128x128 pixels to make subsequent motion correction less computationally costly. These are displayed as separate stages in Fig. 2.7 for visual clarity, but are implemented as a single step in practice.

2.3.3 Motion correction

Despite head-fixing, the imaged tissue is subject to small motions. In addition to the noise this adds to the neuron features, which can impair decoder accuracy, there is a risk that information about the animal's running velocity could 'leak' via motion artefacts during periods of fast running. It is important for us to correct for these artefacts to ensure that we are decoding from neural activity, and this is not only a matter of rigour: if motion artefacts are learned by the decoder, the animal will not be able to decouple physical performance of the task from maintaining decoder accuracy. While existing optical BMIs (e.g. Sorrell et al., 2024; Chen et al., 2023) use rigid motion correction based on phase correlation, which allows for corrections that translate and rotate of the entire image field of view, we found that our imaging contained small deformations within the FoV: tissue in the image moved relative to other tissue. We partially attribute this in-image movement to the cavity introduced by our surgical procedure (see Methods 2.6.6), which is not part of the process to optically image the PPC. The extent of within-image deformation varied from animal to animal, but in several cases attempts to correct frames using rigid transformations resulted in catastrophic misalignment. We use a non-rigid motion correction algorithm (see Methods 2.6.1), despite its substantial additional computational cost (which accounts for the vast majority of the signal processing pipeline's frame budget usage, see Fig. 2.7), and demonstrate in Section 2.4.2 that there is little evidence to suggest 'leaked information' from motion distortions that could be used to decode variables such as velocity.

2.3.4 Time-domain filtering

While Calcium traces can be deconvolved to recover underlying spikes, the process is imperfect and challenging to perform without acausal processing, which renders it inappropriate for a real-time application. Fortunately, this is not necessary for our setup: elevations in underlying mean firing rate are informative about location, and exact spike timing is not relevant on a behavioural timescale. As a back of the envelope calculation, consider that the average place field in our data is roughly 50 cm in length—even an animal travelling at 100 cm/s, well in excess of what we observe, would take half a second to traverse it.

We decode from Calcium traces as a proxy metric for an underlying mean firing rate, but perform some temporal filtering of the signal to correct for DC drift (changes in the baseline

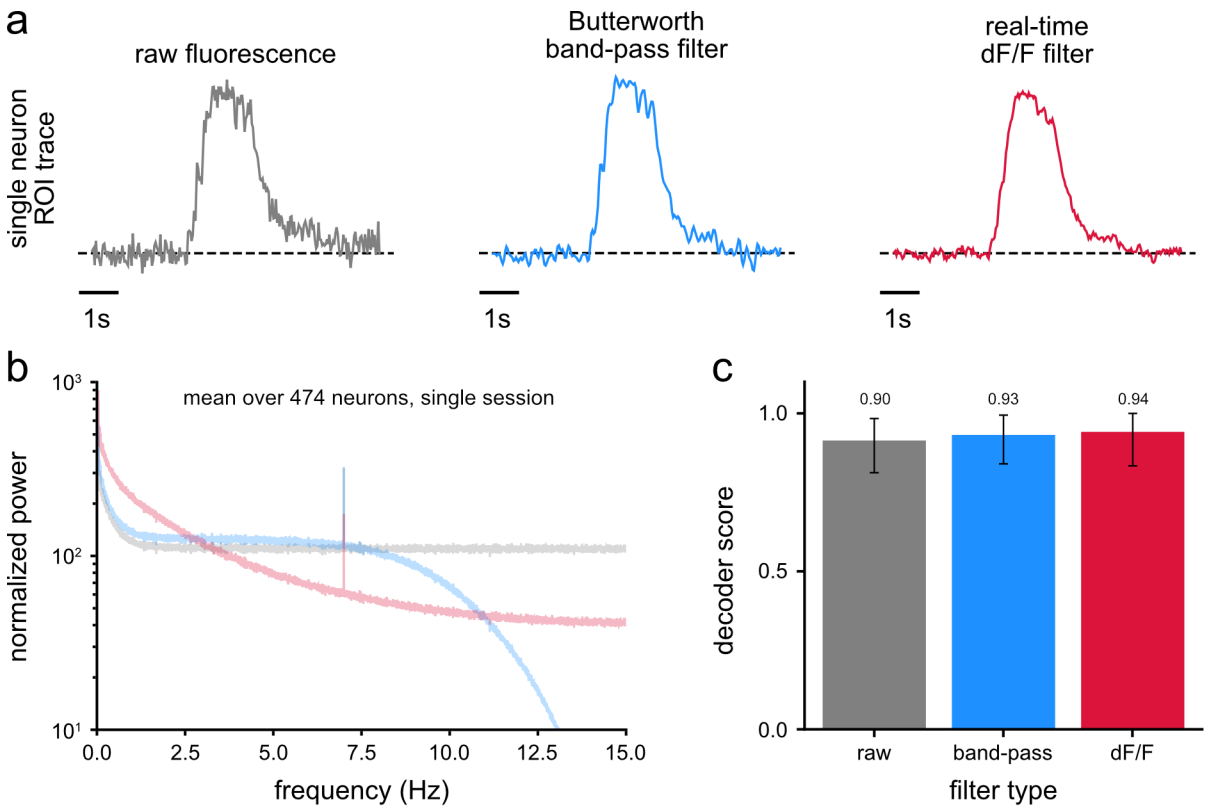


Fig. 2.8 Time-domain filtering of neuron ROI signals. **(a)** An example trace of the fluorescence of a neuron ROI as the animal passes through the neuron's place field, shown both as a raw signal and after filtering with either a Butterworth or dF/F filter with equivalent cut-off frequencies. **(b)** Mean spectral content over 474 neurons from the same 10-minute recording session under the different filtering schemes. The hippocampal theta rhythm is clearly visible as a spike around 7 Hz. Note that the filtering effect of the lower cut-off frequencies (approx. 1/30 Hz) are not visible at this scale of plot. **(c)** Decoder scores (on unseen test laps, using a SVM classifier) with and without the application of time-domain filtering to the feature-space over 8 sessions across two animals. Error bars denote minimum and maximum scores achieved over all sessions and over 10 random splits of which laps are used as training or testing data.

fluorescence over the course of an imaging session) and to reduce susceptibility to high-frequency noise. While this filtering is applied to image pixels in practice, we demonstrate the filtering on a sample neuron ROI in Fig. 2.8.a as an illustrative example.

The delta-F-over-F filter (dF/F) is widely used for cleaning up Calcium imaging traces, removing both DC drift and performing magnitude normalisation. We use an adaptation of this filter that allows it to run causally and introduces a small amount of low-pass filtering to attenuate high-frequency noise (see Methods 2.6.1). When compared to a band-pass filter with comparable cut-off frequencies (we use a Butterworth filter as an example of a more traditional signal processing technique), with these frequencies selected to keep features

within the passband on time scales ranging from behavioural changes all the way through to the theta rhythm, the real-time dF/F filter does not have a sharp cut-off or uniform gain across its passband (Fig. 2.8.b). The normalisation component of the dF/F filter means that its positioning within the signal processing pipeline requires caution: it is ill-conditioned when ingesting a zero-mean signal (fluorescence is a non-negative quantity), and so we place it before the spatial band-pass filter in our pipeline.

Despite these minor drawbacks, we opt to use this dF/F filter both for its familiarity to a neuroscientific audience and for consistency with how we eventually process ROIs in post hoc analysis. Compared to a raw fluorescence signal, the dF/F filter improves to the decoder accuracy by a similar amount as the Butterworth band-pass filter (Fig. 2.8.c)[3].

2.3.5 Construction of the pixel feature-space

After temporal filtering, we apply a spatial band-pass filter that selects for neuron-sized objects in the image. We subsequently apply a threshold that zeroes-out any pixels with negative values[4]. This stage helps generalisation to unseen test data by avoiding decoders assigning weights to small fluctuations in firing rate taking place outside of place fields. This also helps with the interpretability of weights in our decoder: it forces the decoder to predominantly allocate positive weights to neurons firing within each spatial bin. As the last stage of the pipeline, the image is downsampled by local averaging to a resolution of 32x32 pixels, which gives us a fixed input size to our decoder of 1,024 features.

2.3.6 Equivalence of pixel feature-space to ROIs

SVM-based decoders trained on the feature space and decoders trained on neuron ROIs achieve almost identical performance when evaluated on the same data (Fig. 2.9.a, Fig. 2.9.b). We can further verify the faithfulness of the pixel feature space by confirming that the pixels allocated weights correspond to underlying neurons. The vast majority of pixels with non-zero weights overlap with the footprint of a neuron ROI (Fig. 2.9.d), and, in those instances in which non-zero weights are allocated to locations not within the neuron ROIs identified by `suite2p`, the weights are smaller than those allocated within ROIs (Fig. 2.9.c)[5].

[3] The dF/F filter actually performs marginally better than the Butterworth filter. We attribute this to its relatively higher attenuation of higher frequency components, which are not necessarily relevant on the behavioural timescales of traversing the spatial bins.

[4] A reader with a penchant for machine learning might interpret this as a ReLu function, but the objective is to clean up the signal rather than to deliberately introduce a non-linearity. Pixels in the feature space should correspond closely to neurons, and these obviously cannot achieve negative firing rates.

[5] The confidence intervals in Fig. 2.9.c are largely a result of the FoVs across different animals or sessions containing different densities of neurons, and therefore having a lesser or greater number of zero-weight pixel features. This distribution is sharper for individual sessions.

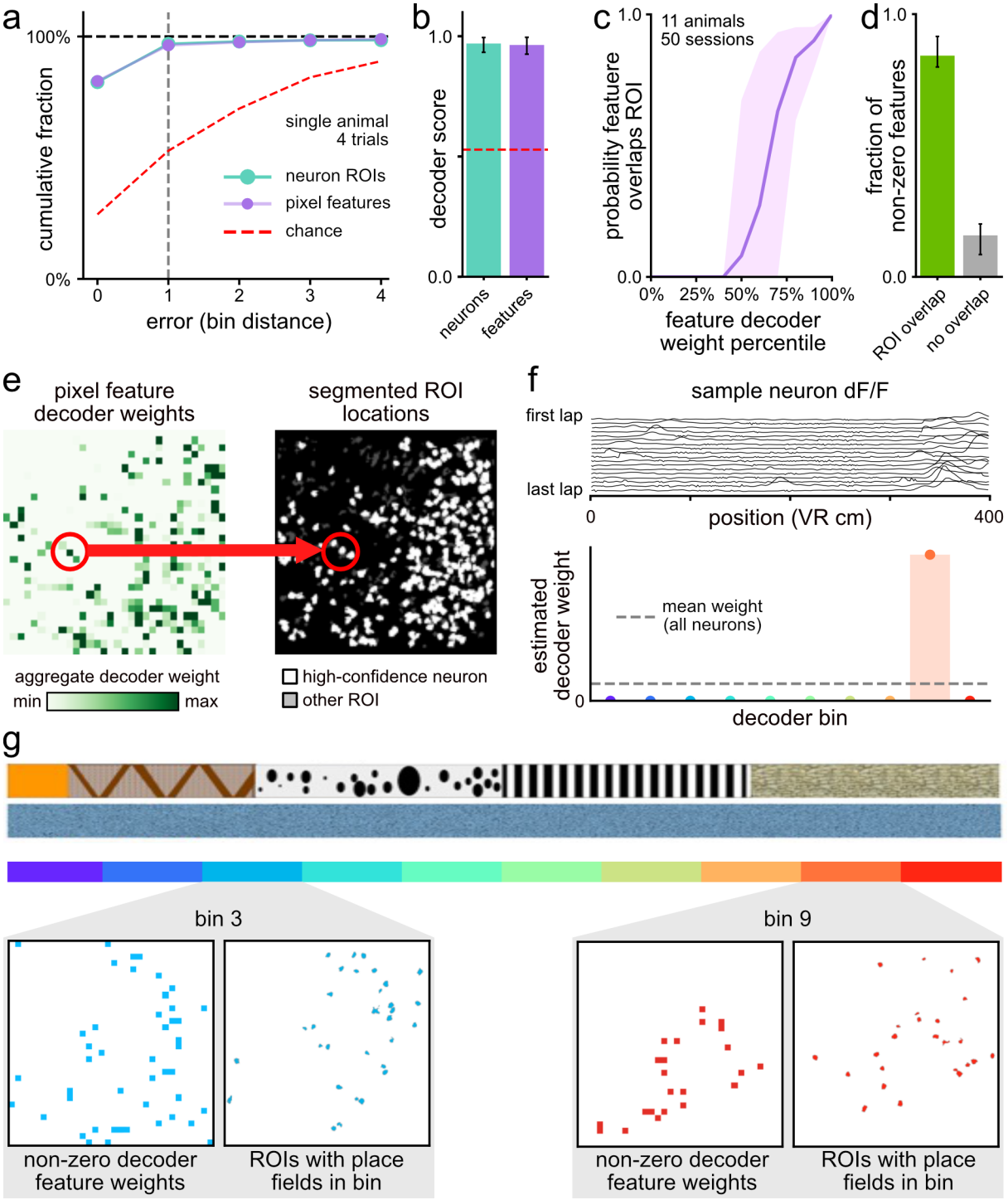


Fig. 2.9 A comparison between decoders based on ROIs and the pixel feature space. **(a)** Decoder accuracy and **(b)** decoder scores, as previously visualised in Fig. 2.5, for decoders using neuron ROIs and pixel features derived from the same trials. **(c)** The probability that a pixel feature overlaps in the field of view with a ROI as a function of the decoder weight percentile allocated to that feature. Confidence interval shows the 10^{th} and 90^{th} percentiles over the decoders trained in 50 BMI sessions detailed in Chapter 3. [*Caption continues...*]

Fig. 2.9 [*Continued*] **(d)** The fraction of pixel features with non-zero weights that overlap a ROI. Error bars correspond to the confidence interval in (c). **(e)** A view of the aggregate weight (a sum of the absolute value of the weight for each bin) for each pixel of the feature space for an example decoder, shown against the ROI locations identified by `suite2p` for the same session. Mapping between the pixel feature locations and ROI locations permits an estimation of the decoder weights attributed to individual neuron ROIs. **(f)** Top: an example neuron and its lapwise dF/F traces during a trial. Bottom: that same neuron's estimated contribution to decoding each spatial bin on the track as inferred from decoder weights. The dashed grey line indicates the mean contribution over all neurons to each decoding bin. **(g)** Each bin location on the track has associated decoder weights. This shows at two example locations the non-zero decoder weights in the feature space, and also shows the ROI locations of the place cells with fields that are partially or completely in the bin.

We can further reassure ourselves that underlying place cell activity is responsible for decoding by determining which neurons are contributing to which decoder features. By mapping the feature weights to underlying neurons (Fig. 2.9.e, see Methods 2.6.5), we can determine each neuron's relevance for decoding each bin: Fig. 2.9.f shows an example neuron with a place field near the end of the track and decoder weights that are zero for all bins except one. This method is *not* a perfect one-to-one mapping: depending on alignment to the pixel grid, it is possible both for an underlying neuron to straddle more than one feature space pixel, and for multiple underlying neurons to simultaneously contribute to the same feature space pixel. As most neurons in the population are not place cells, the majority of such overlaps are of little importance to decoder accuracy (as evidenced by the near-identical accuracy between ROI and pixel features), but this does result in neurons without a place field appearing to have high weights by association—for this reason, we treat analyses of decoder weights stochastically (see Section 4.4.2 in Chapter 4).

Finally, for any given spatial bin on the track classified by the decoder, we can compare which pixel features are used to decode this location with which neurons have place fields in that location. On visual inspection, these are in broad agreement (Fig. 2.9.g). However, discrepancies exist in both directions. Decoder training does not require a feature to strictly conform to the definition we use to automatically identify place fields (see Methods 4.8.1), which enforce minimum and maximum field sizes and impose requirements on relative firing rates; they merely require the feature to be spatially informative. Conversely, a neuron meeting the criteria for being a place cell is not necessarily incorporated by the decoder: if it displays sufficient irregularity in its firing and there is sufficient redundancy in the population, the decoder is better off without including it.

2.4 Towards closed-loop BMI

2.4.1 Screening of surgery for BMI viability

The surgical procedures for preparing an animal for imaging have some inherent variance. In addition to verifying that the imaged layer contains place cells and is indeed stratum pyramidale in CA1, we screened for two other key factors that limit our potential decoder accuracy. The first is the nature of tissue motion during imaging, which appears surgery specific. While we can correct for in-plane motion, we are limited when dealing with motion in the z-axis. This is visible as rippling, wave-like distortions between image frames. The second factor is indicator expression, which can be inhomogeneous. We required sufficient expression in an imaged region for at least half of the FoV to be dense in neuron ROIs.

We did not set a hard target on minimum open-loop BMI decoder accuracy before proceeding to closed-loop experiments[6]. However, after sufficient time for the indicator to become well-expressed, and after the animal had mastered the VR task, we performed a screening session consisting of several 10-minute trials, from which we were able to evaluate open-loop decoding accuracy.

We show the results of an example screening session for animal MH188 in Fig. 2.10, which consists of two trials performed in darkness (i.e. the VR screens are dark, but the experiment is otherwise unchanged) and four trials performed with the screens displaying the cue rich environment used for the VR navigation task. In darkness, the animal is unsurprisingly unable to perform the task successfully. It is a reassuring sanity check that a decoder trained and evaluated on these darkness trials performs no better than chance—if the decoder were able to recover location despite the animal being entirely unaware of its position on an unseen track, this would indicate a serious methodological flaw.

We take the accurate offline decoded laps and the high open-loop decoding scores as grounds to proceed confidently to closed-loop BMI without concern that our signal processing methodology presents issues for decoder accuracy.

2.4.2 Representations of non-spatial information

CA1 hippocampus encodes position, but to what extent does it encode other variables that we could potentially decode for use in BMI? Attempting to decode velocity during traversals of the VR track produces an R^2 score that suggests this is plausible, even with a simple linear decoder (see Methods 2.6.3). However, when we attempt (and subsequently fail) to perform the same decoding for the trials performed in darkness, it becomes apparent that neural activity in our imaged population does not encode velocity in a way that is straightforward for

[6] Largely because surgical outcomes were binary: decoding was either good or entirely impossible, accountable to the two factors above.

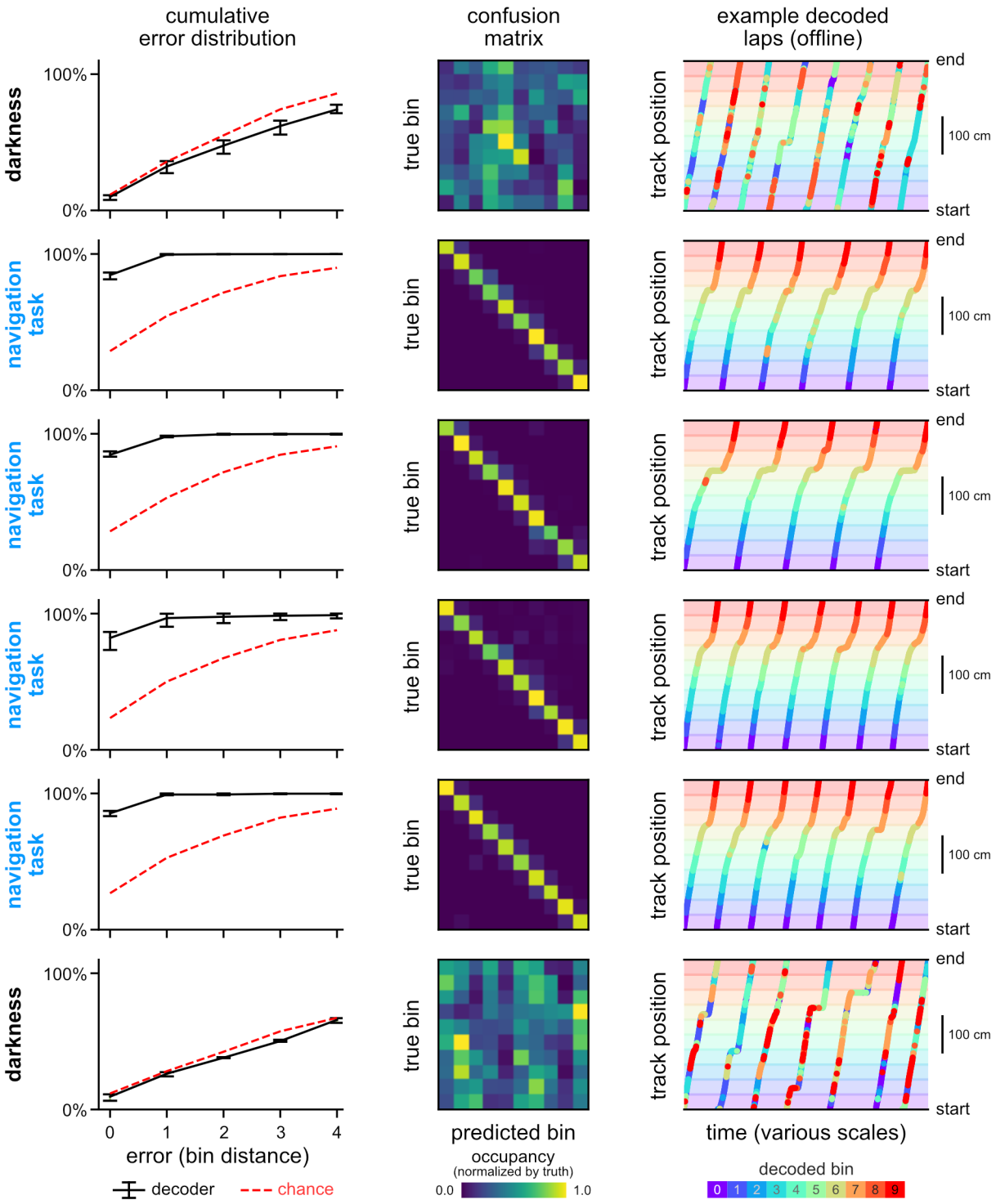


Fig. 2.10 Validation of open-loop decoder across several 10-minute trials. Each row represents a 10-minute trial performed either in darkness or in the cue-rich virtual environment of the navigation task. Decoders are trained on 75% of the laps performed during the trial and evaluated on the remaining 25%. The cumulative error distribution (left column) illustrates with error-bars the minimum and maximum decoder performance over 10 shuffles of which laps constitute training and testing data. The confusion matrices (central column) are averaged over these 10 shuffles. Traces of sample testing laps (right column) display the animal's true position, with the colour of the trace indicating the decoded bin at each timestep.

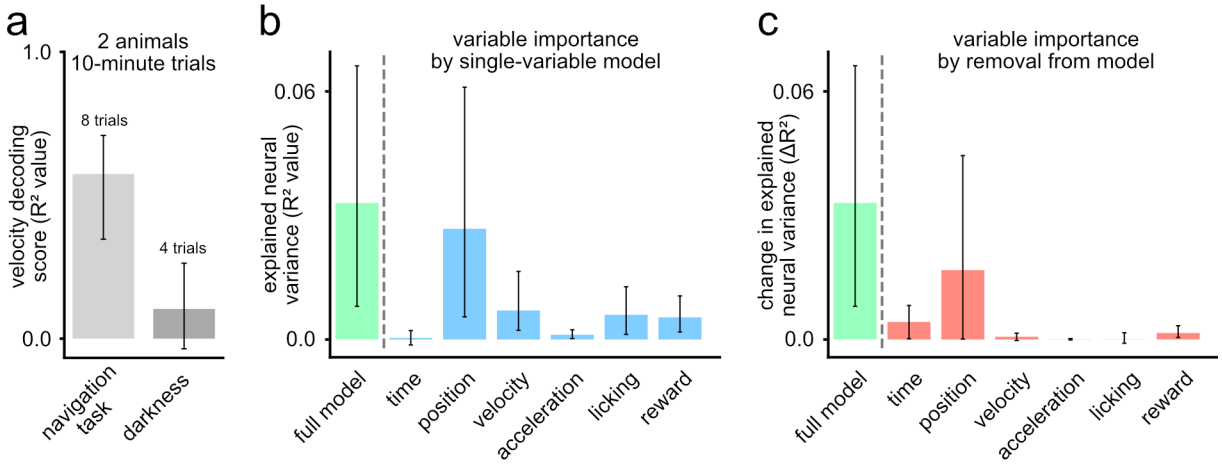


Fig. 2.11 Imaged CA1 populations do not intrinsically encode information about velocity, but behavioural stereotyping results in predictive power. **(a)** Decoder accuracy for prediction of animal velocity based on neural activity in CA1, shown both for trials performed in the cue rich environment and for trials where the animal is in darkness, based on a Lasso regression. Evaluated on withheld laps, as in Fig. 2.10. Error bars show 10th and 90th percentiles over all trials and 10 shuffles of withheld laps. **(b)** The variance in a population's neural activity during the navigation task explained by a linear model. We show both a full model, which uses all kinematic variables simultaneously, and single-variable models, which use only one kinematic variable each. Error bars show 10th and 90th percentiles over all 8 trials and 10 shuffles of withheld laps. **(c)** As in (b), but inferring kinematic variable importance by withholding that variable from the fit and inspecting the change in explained variance against the full model.

us to extract (Fig. 2.11.a). Our apparent success at decoding velocity during the navigation task is due to the animal's highly stereotyped trajectories: it runs at the same speed at each track location on each lap, and so a population that encodes position imparts some amount of information about velocity[7].

In an analysis inspired by the methods in Musall et al. (2019), we investigate how well kinematic task variables (such as velocity, acceleration, time elapsed since start of trial, licking of the reward spout) can explain variance in the neural population. This is essentially 'reverse decoding': predicting neural activity from behaviour (see Methods 2.6.4)[8]. First, we fit a model that uses all kinematic variables simultaneously and note a hypothetical maximum explained variance in the neural activity. We then compare this both to models that only use each kinematic variable as a sole explanatory variable (Fig. 2.11.b), and to models that use

[7]We will briefly discuss in Chapter 4 the strong existing evidence for the modulation of firing rates *within* place fields by running speed. This suggests that decoding velocity is possible with a decoder that can unpick the conjunction between speed and place, but that velocity is not decodable in isolation (such as when the animal is lost in the darkness condition).

[8]The original analysis in Musall et al. (2019) uses sophisticated Generalised Linear Models for the fits. We make do with simpler regressions over hand-crafted features, as our objective is to quickly infer relative importance of kinematic variables rather than fully explain activity in the population.

all kinematic variables except a sole withheld explanatory variable (Fig. 2.11.c). By taking these together, we can infer the extent to which a given kinematic variable is encoded by the population[9]. For example: some amount of neural activity is explained by a model that uses *only* velocity, but since *adding* velocity to the model adds no explanatory power, its relationship to activity is better captured by other variables.

Position on the track is by far the greatest explanatory variable. Interestingly, elapsed time alone has effectively no explanatory power, but when combined with other variables improves the explanatory power of the overall model: the inclusion of time effectively allows the model to partially capture activity changes taking place during the course of a trial.

Velocity, acceleration, and licking, which are all volitional behaviours from the perspective of the animal, have a very limited direct contribution to the activity in the population. From a perspective of 'information leakage', such as increased motion artefacts during periods of acceleration, this is reassuring: we can confirm that the populations we are imaging primarily encode location. When we design the mapping from neural state onto actions for closed-loop BMI in the following chapter, we will need to bear in mind that these states correspond primarily to positions.

2.4.3 Remarks on end-to-end BMI latency

We have previously described the importance of feedback in closed-loop BMI. How long does it take to go from a neuron firing in the field of view to the consequences of that action manifesting on the screen? The answer requires a brief description of the software architecture of `ReTiCaDe`, which runs several components asynchronously from each other (Fig. 2.12.a).

When a neuron fires, its increase in fluorescence will appear in the next frame: assuming this takes place in a random part of the image at a random time, it will take 17 ms on average until the relevant pixels are scanned. A worst case scenario would entail the neuron firing just moments after the region of the image had been swept, needing to wait 33 ms for the scan to return to this location. The microscope readout is fed to an imaging computer as a raw stream of samples at each pixel location (in this case, 3 consecutive samples at each pixel location). As this stream arrives, we marshal the samples into frames and write them into a shared memory block accessible by other processes.

This shared memory is polled at 30 Hz by the signal processing and decoding process, which ingests the latest complete frame. Any current frame 'under construction' as data arrives is ignored until the next poll. As this process runs entirely asynchronously from the process streaming data from the microscope, frames written into shared memory have an average wait time of 17 ms before being picked up.

[9] This requires caution: including multiple explanatory variables that are very closely correlated with each other will cause this analysis to become ill-conditioned.

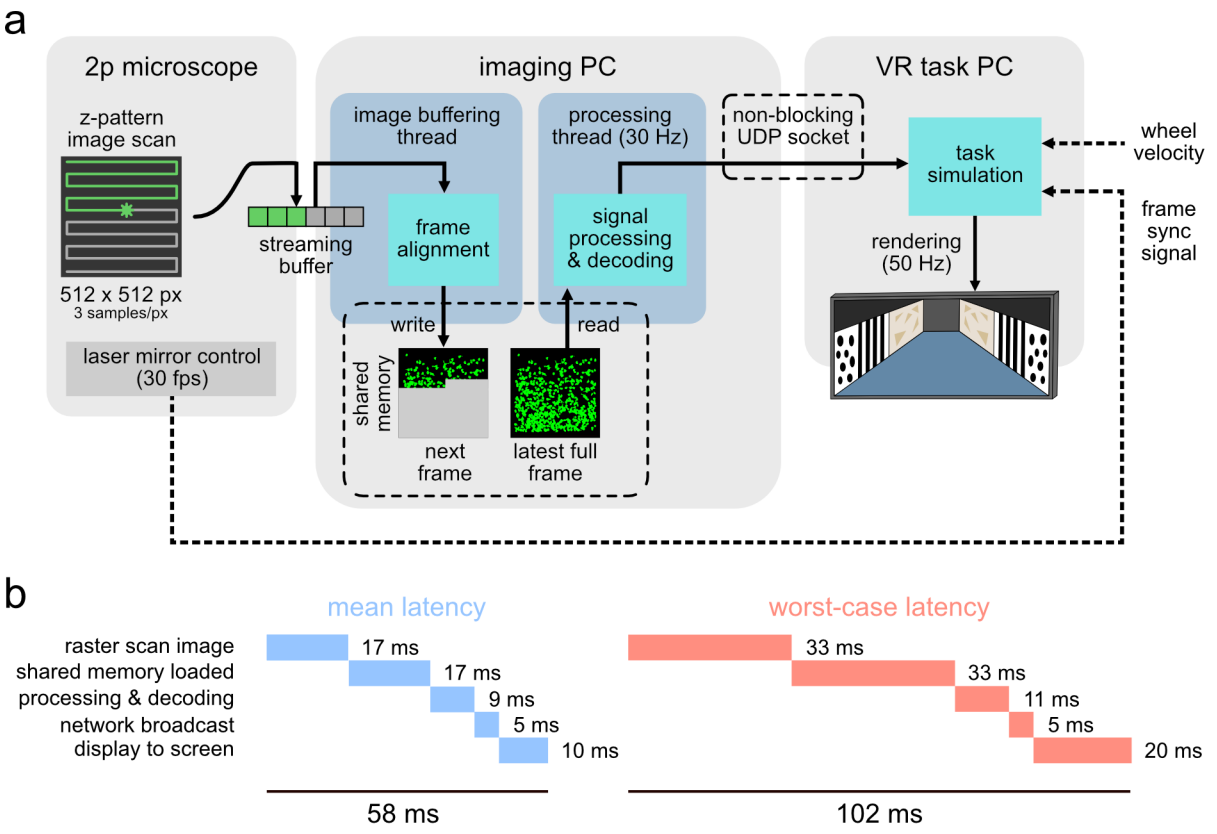


Fig. 2.12 Architecture of the optical BMI software minimises response latency to neural activity. (a) Schematic overview of BMI implementation illustrating the asynchronous execution contexts and propagation of data through the system. A direct measurement of the galvanometer used to control the microscope laser bypasses the BMI processing to ensure image frame alignment with VR task simulation during offline analysis. **(b)** Waterfall diagrams of estimates of the mean and worst-case latencies for the propagation of a Calcium transient from an imaged neuron to an update visible on the VR screen.

After signal processing and decoding are complete, the computer performing imaging and decoding sends a command over a local network to a second computer responsible for rendering the visuals of the VR environment. Rendering takes place at 50 Hz, and so once again we have an average wait time (10 ms, in this instance) before the outcome of a command can be displayed to a screen.

Taken together, we estimate the mean latency of the entire process at slightly under 60 ms (Fig. 2.12.b). We provide as a comparison a hypothetical worst-case propagation time, approximately 100 ms, in which each data handover between processes takes place at each most inopportune moment. By taking high-frequency measurements of the resonant galvanometer used to control the microscope mirror, we can detect spikes in the current mirror control current that correspond to the frame reset times, which allows us to confirm propagation time through the imaging PC and ensure frames are correctly aligned post-

experiment[10]. We find that these propagation times (which exclude the image raster scan and the display-to-screen components) take place within the space of two imaging frames (i.e. between 33 and 66 ms).

In practice, operating system-level scheduling can desynchronise the relative timing of these processes, so there is some variance in propagation latency. The operating system can also pause processes, sometimes for long enough that a frame must be 'skipped' by a downstream component. When this does happen, the component detecting a skipped frame will simply reuse its last valid input and record the occurrence of a skipped frame. These are relatively rare: in a typical 10-minute recording with 17,994 frames, a single-digit number of frames will have been skipped.

These latency estimates are important: when an animal eventually controls progress through a decoder in closed-loop, long delays between cause (neural activity) and effect (actions taken by the BMI) make the control more difficult to learn. We have described latencies that fall well within rodent reaction times (there is considerable variance across task designs, but Goldstein et al. (2022) and Wang and Krauzlis (2018) both report behavioural response times to visual stimuli in the region of ~400 ms), and therefore posit that the visual feedback on behavioural timescales is comparable to the manipulation of a wheel or lever to control a VR task. We conclude that both average latency and spikes in latency will have negligible effects for our experimental results describing closed-loop experiments.

2.5 Discussion

2.5.1 Open-loop evaluation: a green light for closed-loop trials

We have described a Calcium-imaging based BMI design that reads out from place cells in CA1 and is accurate in open-loop, runs in real-time, and uses a straightforward underlying decoder model. This point has previously been reached before for 1-photon imaging of animals navigating real environments (Chen et al., 2023; Sun et al., 2022). Ensuring that we achieve equivalent results in our VR task is a necessary stepping stone on the way to investigating the interaction between closed-loop BMI usage and neural activity in CA1. We claim that the results in this chapter suggest a stable technical foundation from which to stage such investigations. However, we take the opportunity in the following sections to outline some of the design's rougher edges and the technical limitations that stand between it and more ambitious closed-loop experiment designs.

[10]An acknowledgement: the hardware and computer scripts for such alignment were the work of Marius Bauza as part of an earlier, separate project.

2.5.2 Suitability for cross-day decoding

While part of the appeal of imaging CA1 is the visibility of plasticity on short timescales, we aspire to eventually move to reusing decoders from one day to the next. This is not viable with our current methodology. In the experiments described in the following chapters, we manually align the imaging FoV on each day to match the FoV from the previous day. This is adequate for performing post hoc registration to identify the subset of ROIs in the population that match from one day to the next, but the fraction of neurons we preserve from one day to the next is low, as we are stringent with our confidence requirements on identifying matches (see Section 4.6.1). Crucially, this alignment is not sufficiently precise to reuse a decoder from a previous day.

We are once again burdened by the same technical limitation that made adopting the pixel feature space a pragmatic design choice: it is non-trivial to consistently and automatically identify the same ROIs from a short segment of imaging data. An existing and promising ex silico approach involves using two different fluorescent indicators simultaneously on different emission channels (e.g. a red and a green channel), with one channel imaging neural activity and the other serving as an alignment guide using anatomical features (Cheong et al., 2018; Grewe et al., 2010). This complicates the surgery, the imaging hardware, and would require automating alignment of the FoV from one day to the next. Nevertheless, this seems like a natural next step to better leverage the virtues of Calcium imaging, and would have the highly desirable side effects of obviating the need for the pixel feature space (which somewhat obfuscates the mapping between activity and decoder weight) and improving the number of neurons we are able to track across days.

2.5.3 Complexity of the navigation task

Our VR task takes place on a 1D track. Not only is this corridor devoid of choices between paths that might lead to different experiences of the environment, but the animal only ever runs forward. It therefore only ever experiences the visual cues on the track as part of the same sequence.

As a piece of software, `ReTiCaDe` is agnostic both of underlying imaged brain region and task design. It is modular at both the level of interprocess communication (readily adaptable to communicate with other models of incoming image data or VR experiment configurations) and at the level of data processing (we provide pipeline and decoding stages that can be composed and tuned *à la carte*). It is therefore extensible to more complex tasks from a purely technical standpoint.

Taking a more holistic view, highly constrained tasks like the 1D track can paper over one of the current major limitations of optical BMIs. Calcium imaging must make trade-offs in its imaging sessions: at higher laser power, signal-to-noise ratios are higher, and with longer

sessions, we get more training data; but both of these two choices result in greater risk of photobleaching or otherwise damaging the population. In the case of training decoders on neural representations in CA1, animals must repeatedly visit all relevant locations in an environment to provide a suitable training corpus. Repeated traversals of the 1D track ensure that all locations in the environment are visited and well-sampled. Full, open-field navigation based on an optical CA1 BMI is challenging. Task design of this nature is appropriate either to alternative imaging techniques that can leverage longer recording sessions (e.g. Lai et al. (2023), who require 30-40 minute recording sessions for a 1m×1m virtual environment), or would require further iteration on the signal processing side of optical BMIs to maintain decoding accuracy at lower laser powers. If we broaden our horizons beyond CA1, we could look to entirely different brain regions that do not rely on sparse representations of space (e.g. decoding heading velocity and direction rather than position, as is typical in most BMI designs for cursor control or navigation).

Despite these limitations, there is nevertheless a fertile middle-ground: clever task design can restrict exploratory freedom and enforce greater coverage of the environment during the collection of training data. Maze structures with clear decision points (e.g. Johnson and Redish, 2007), or virtual analogues to mazes with reconfigurable structures such as the honeycomb maze (Wood et al., 2018), provide the opportunity to decode navigational choices between different routes in real-time.

In a research landscape in which BMIs are used to control computer cursors (Gilja et al., 2012), produce speech (Willett et al., 2021), manipulate virtual prosthetic arms (Gouéytes et al., 2022), make navigational choices (Sorrell et al., 2024), or perform open-field navigation (Lai et al., 2023), our task hardly seems eye-catching. Nevertheless, we hope to demonstrate in the following chapters that we have much to learn from this unassuming 1D track.

2.6 Methods

2.6.1 Signal processing

Real-time dF/F filter

The conventional form for a dF/F filter applied to a neuron ROI is:

$$F'(t) = \frac{F(t) - \overline{F}}{\overline{F}}$$

Where $F(t)$ is the ROI fluorescence in frame t and $\overline{F}$ is the mean fluorescence of that ROI over the course of an entire session. However, as the BMI requires a causal filter (one that has no knowledge of future values), *and* because we desire a degree of low-pass filtering, we instead

use the following formulation (see Sorrell et al. (2024) for an example of precedent):

$$F'(t) = \frac{F_{\text{fast}}(t) - F_{\text{slow}}(t)}{F_{\text{slow}}(t)}$$

Where $F_{\text{fast}}(t)$ and $F_{\text{slow}}(t)$ are both exponentially smoothed versions of F:

$$F_{\text{fast}}(t) = \alpha_{\text{fast}} F(t) + (1 - \alpha_{\text{fast}}) F_{\text{fast}}(t-1)$$
$$F_{\text{slow}}(t) = \alpha_{\text{slow}} F(t) + (1 - \alpha_{\text{slow}}) F_{\text{slow}}(t-1)$$

The constants α_{fast} and α_{slow} are set to 0.3 and 10^{-3} respectively. These can be understood in terms of a time-constant τ, the number of seconds it would take for an impulse to decay to $1/e$ of its original value, by use of the sampling frequency f_s via:

$$\alpha = 1 - e^{\frac{-1}{\tau f_s}}$$

Which corresponds in time-constants of $\tau_{\text{slow}} = 33.3$ s and $\tau_{\text{fast}} = 0.09$ s: a fairly responsive band-pass filter.

The relatively long time constant for $F_{\text{slow}}(t)$ means that, if initialised at zero, the filter must be 'burned in' on a stream of frames before producing adequately filtered data. In offline analysis, we burn in the filter by computing an initial state for $F_{\text{slow}}(t)$ and $F_{\text{fast}}(t)$ by ingesting $F(t)$ in reverse time order. For online, closed-loop BMI trials, we initialise both $F_{\text{slow}}(t)$ and $F_{\text{fast}}(t)$ to the value of $\overline{F}$ evaluated over the data used to train the decoder.

Butterworth time-domain band-pass filter

As a point of comparison to the real-time dF/F filter, we use a Butterworth band-pass filter with lower cut-off frequency 0.033 Hz, upper cut-off frequency 10 Hz, and order 2, using a reference implementation from `scipy` (Virtanen et al., 2020).

Spatial image filtering

The initial spatial low-pass filter applied prior to motion correction is a 2D Gaussian smoothing with a standard deviation parameter σ equivalent to 9.6μm (accounting for the image resolution at the point the filter is applied). The spatial band-pass filter applied after temporal filtering is a 2D Difference-of-Gaussian filter parameterised by $\sigma_{\text{lower}} = 20\mu$m and $\sigma_{\text{upper}} = 4\mu$m.

Non-rigid motion correction

The non-rigid motion correction algorithm is based on the `scikit-image` implementation of the Iterative Lucas-Kanade optical flow correction algorithm (Lucas and Kanade, 1981; van der Walt et al., 2014). The maximum window evaluated around each pixel has a radius equivalent to 28μm. We evaluate only a single warping, as our displacements are typically small (and subsequent iterations must be traded off against a substantial computational cost). Each motion-corrected frame is aligned against a reference frame. For closed-loop BMI decoding, we construct a reference frame from the mean pixel-wise fluorescence over the 10-minute trial used as training data for the decoder.

2.6.2 Classification-based decoding

All classifiers use existing implementations from the `scikit-learn` package (Pedregosa et al., 2011).

Class balancing

To account for uneven occupancy of spatial bins, training data classes are balanced prior to model fitting by weighting samples to ensure an equal total sample weight in each bin.

Linear SVM

We use an L1 penalty to provide a sparse regularisation. This discourages allocation of non-zero weights to features that provide minimal information about track location. The hyperparameter controlling the strength of the regularisation (denoted C in most texts) is set to 0.05 when using neuron ROIs as features and to 0.2 when using the pixel feature space. These values were selected by a sweep of the parameter-space and evaluating performance on unseen laps across all offline evaluation sessions, and are subsequently fixed for use in the closed-loop BMI experiments described in the next chapter.

Naïve Bayes

We assume feature likelihood to be Gaussian, i.e. that for each feature x_i and observed in class y we can approximate $P(x_i|y)$ as:

$$P(x_i|y) \sim N(\mu_y, \sigma_y^2)$$

Where μ_y and σ_y are estimated from the training data. The sample weighting for class balancing is equivalent to forcing $P(y)$ to take a uniform distribution, and so a model prediction $\hat{y}$

based on an observation of N features takes the form:

$$\hat{y} = \arg\max_{y} \left(\sum_{i}^{N} \log P(x_i | y) \right)$$

Multi-layer perceptron

We determined the network size, parameterised only by the size of the single hidden layer, through manual trial and error, and selected the best-performing MLP as the comparison point to the SVM-based method. The presented MLP uses a tanh activation function, uses a single hidden layer, is fully connected, uses early stopping, and is trained with gradient-descent (Adam). It uses L2 regularisation, with the strength of the regularisation (denoted α in most texts) determined by 5-fold cross-validation.

2.6.3 Regression-based decoding

All regressors, used in decoding quantities as continuous variables rather than as discrete classes, also use existing implementations from the `scikit-learn` package (Pedregosa et al., 2011).

Decoding velocity from CA1

We fit a Lasso regression (a linear regression with an L1 penalty) to predict running velocity from neuron ROI fluorescence. The regularisation strength is determined by 5-fold cross-validation within the training data. Running velocities are averaged over a 7-frame wide window centred on the current frame, corresponding to a smoothing with a boxcar filter of width ~230 ms. The ROI traces are dF/F filtered identically to their use in evaluating classifiers. As with the classification evaluations, we split traversed laps of the track into training and testing data with a 75%-25% split.

Decoding multiple kinematic variables from PPC

We train four models to predict x-positions, y-positions, forward velocities, and heading direction. Neuron traces are filtered using the same dF/F filter as used for the CA1 data. We evaluate the R^2 values of the models for each kinematic variable based on a 75%-25% train-test split of maze attempts, repeating 10 shuffles of the train-test split as with the CA1 classifier evaluations, over 5 sessions (1 session for each of 5 animals) and report the mean R^2 value. The first model is a simple linear regression minimising the squared error of prediction. The second model is a ridge regression (a linear regression with an L2 penalty), and its regularisation strength is determined by 5-fold cross-validation over the training data, evaluated at 20 points spaced logarithmically over the interval $[10^{-3}, 10^{3}]$. The third model is

an SVM with a linear kernel with its regularisation parameter C also determined via the same cross-validation, using a squared epsilon-insensitive loss for an L2 penalty. The fourth model is an MLP regressor with a single hidden layer containing 10 units—this size selected by trial and error to perform well across the entire set of animals (which have varied session duration and ROI counts).

2.6.4 Explanatory importance of kinematic variables

Kinematic variable preprocessing

We expand position from a single variable representing y-coordinate along the track to 10 variables via the following kernel for each decoding bin location:

$$y_{\text{bin index}} = e^{-\left(\frac{\text{bin centre}-y}{\text{bin width}}\right)^2}$$

We smooth velocity by averaging over a 7-frame long window (3 frames either side of the current frame), and then compute the acceleration variable as a numerical derivative of this filtered velocity (thereby mitigating issues associated with numerically differentiating a noisy signal). The raw 'reward' signal is zero everywhere except at the frames corresponding to reward delivery, where it takes the value 1. The reward variable we use in the fit is a convolution of this signal with a Gaussian kernel with a standard deviation equivalent to 20 frames. The raw lick signal is also a binary signal, zero everywhere except on frames corresponding to a lick of the reward spout. The lick variable we use in the fit is a convolution of this signal with a Gaussian kernel with a standard deviation equivalent to 10 frames. Time elapsed since the start of the trial requires no additional processing or filtering. Prior to fitting, all kinematic variables used as input features are subject to scaling to enforce zero mean and unit standard deviation.

Model fitting

In the case of the 'full model', we fit and evaluate one model using all of the kinematic variables as input simultaneously. In the case of single-variable models, we fit and evaluate one model per kinematic variable. In the case of single-omission models, we fit and evaluate a model using all kinematic variables *except* the variable under consideration.

We use a ridge regression to attempt to predict the activity of each neuron in the session (as expressed by dF/F filtered traces), once again using a 75%-25% train-test split and shuffling the nature of this split 10 times, ensuring that the random shuffles are identical across choice of kinematic variable under consideration. We record, for each shuffle of split within each session, the mean explained variance over all neurons. We report this directly as a measure

of variable importance for single-variable models. For the importance in single-omission models, we report ΔR^2: the difference between the score of the full model and the single-omission model.

2.6.5 Neuron ROI contribution to pixel feature space

To estimate the contribution of a neuron ROI to decoding within a given session, we iterate over each pixel in the neuron's footprint (at a resolution of 512x512 pixels) to determine the proportion of overlap with each feature in the feature space (32x32 pixels). We calculate a 'raw' weight for each bin for each neuron ROI by performing an average of the weights in the feature space, with the averaging weighted by the proportion of overlap with the ROI. We then normalise these raw weights across neuron ROIs within each class (spatial bin on the track) by dividing by the standard deviation of the raw weights within that class.

As an important side note: directly comparing weights, whether they are inferred neuron weights or weights in the pixel space, requires normalisation of the input features. In typical machine learning literature, this would be handled by forcing features to have unit variance and zero mean. Here, we rely instead on the dF/F filter in the signal processing pipeline, which effectively normalises each feature space pixel by its mean fluorescence.

2.6.6 Surgical procedures, animal training, and imaging hardware

Acknowledgement: This section is adapted from an article currently in preparation reporting our BMI work. The following details were summarised by Hinze Ho for that article.

Mice

All experimental procedures were conducted in accordance with the UK Animals (Scientific Procedures) Act 1986 (ASPA), following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB). All procedures were performed by trained personnel authorised under Home Office Personal and Project licences.

Mice were housed in conventional plastic cages (16cm × 27cm × 18cm) and were group-housed where possible. Water and food were supplied ad libitum unless otherwise specified. All mice were kept on a 12:12 hour light:dark cycle in an animal facility with regulated temperature (21-23 °C) and humidity (50-60%). All mice were male.

Viral injection and device implantations

Mice were anaesthetised with 1-3% isoflurane in O_2 and placed in a stereotaxic frame (David Kopf Instruments, Model 1900) in all surgical procedures. The scalp was shaved and swabbed with iodine before incisions were made to expose Bregma, from which coordinates were

measured. Analgesic (Metacam, 0.05 mg/10g body weight) and antibiotics (Baytril, 0.05/10g) were administered via subcutaneous injection. Body temperature was maintained at around 35°C and breathing was monitored throughout surgery. Ointment was applied to protect the eyes from dehydration and scarring. Mice were placed in a heated chamber for 1-2 hours following surgery to aid recovery before transfer back to their home-cage. Post-operative care, which generally lasted 5 days, provided Metacam and Baytril orally and monitored body weight.

A head-plate was used for head-fixing the mice during 2-photon imaging. The head plate consisted of a custom-made aluminium piece with a central opening for imaging, flanked by two 'wings' which could be fastened with screws to fixed points relative to the microscope objective during experiments. Outside of experiments, the central opening was kept covered by a screw cap to keep the imaging cannula clean. For head-plate implantation, a circular patch of scalp was excised and the crown of the skull exposed. Viral injection was made in the right hemisphere, and two contralateral sites of craniotomy were drilled for insertion of jeweller's screws to provide additional support for the head-plate. Opaque dental cement (Superbond C&B) was applied to permanently attach the head-plate to the skull. A Nanofil needle and syringe controlled by a micropump were employed for viral injection. A craniotomy of approx. 1mm was made atop the injection site with the following coordinates: AP: 2.0 mm posterior from Bregma; ML: 1.8 mm from midline; DV: 1.3 mm beneath dura (Dombeck et al., 2010). To express Calcium indicator in the dorsal hippocampus stratum pyramidale, 200nl of titre 2.3×10^{13} GC/mL AAV1-syn-jGCaMP8f-WPRE-SV40 (Zhang et al., 2020) (Addgene 162376-AAV1) diluted 1:1 in artificial cerebrospinal fluid was injected at 30 nl/min, followed by a 10-minute wait to minimise backflow before retracting the needle. Exposed skull was covered with Kwik-cast and the head-plate covered with a screw cap.

One week after viral injection, a second surgery implanted the imaging cannula (a stainless steel tube with outer diameter 2.7 mm, inner diameter 2.3 mm and height 2.5 mm). A 2.5 mm diameter cover glass sealed one end of the tube and was attached using UV curable adhesive (Norland Products). A craniotomy of approx. 2.7 mm was made using a trephine drill centred around the injection site. Dura was excised and the cortex overlaying the hippocampus was slowly removed by aspiration using a suction pump. The tissue was regularly irrigated with sterile saline solution to stem bleeding and maintain tissue integrity. The outer layer of the corpus callosum tracts were revealed but generally spared. The alveus was not exposed and remained undamaged. Once all bleeding ceased, a small drop of transparent Kwik-Sil silicone adhesive (WPI) was applied to the centre of the excavation and a cannula was immediately inserted into the craniotomy, pressing the cover-glass against the tissue above the hippocampus. Opaque dental cement was used to permanently attach the cannula to the skull.

Training

Following post-operative recovery, mice were handled and habituated on the styrofoam wheel in the VR setup for two days. Mice were then placed on food restriction and their health and body weight were checked daily. The body weight was kept at no lighter than 85% of the original weight. Mice were then subjected to head-fixing, initially for several minutes, with duration increasing in subsequent sessions until they were able to be head-fixed for 30-40 minutes. While head-fixed, mice were encouraged to run on the wheel and received rewards for doing so, working towards completing at least 100 laps of a fixed track distance within a 40-minute session. During this habituation, the monitors displayed a dark screen. The reward location on the track was randomly distributed to prevent the mice from developing anticipatory behaviour based on distance, as well as to encourage consistent exploratory licking of the reward spout. The initial track distance was 50 cm, increasing to 100 cm, 200 cm, and 400 cm each time the mouse could run the 100 target laps within the 40-minute duration. Once a mouse could complete 100 laps of the 400 cm track, the animal was introduced to the VR environment (screens no longer dark, reward at fixed location) on the following session. This habituation and training programme took an average of 1-2 weeks to complete.

When first introduced to the cue-rich VR track, mice displayed scattered licking along the track, as they were unfamiliar with the environment and reward location. Following several days of exposure, mice began to exhibit behaviour associated with reward anticipation, exemplified by improved licking specificity and deceleration as the animal approached the reward site. A mouse was deemed well-trained when anticipatory licking behaviour became stable and the running velocity profiles were comparable on two consecutive days. Typically, it took mice 5-7 days to become an expert.

2-photon imaging hardware

The imaging microscope was supplied and built by Brucker Corporation, including a Ti:sapphire excitation laser (Chameleon Ultra II, Coherent) with a frequency of 920 nm and 60-70 mW average power at FoV, and mounted on an antivibration table (Thorlabs Inc.). We used Prairie View 5.6 (Brucker Corporation) for software control of the microscope, either manually through the GUI or programatically via a bespoke interface within `ReTiCaDe`. We blocked light leakage from the VR screens with a removable shield of black polyurethane-coated nylon fabric fitted to the head plate. Imaging sites were coarsely located using epifluorescence prior to more exact alignment using 2-photon imaging.

Chapter 3

Closed-loop BMI performance during a simple navigation task

Acknowledgement

The closed-loop BMI experiments described in this chapter were performed in collaboration with Hinze Ho, who performed all animal handling, training, and surgeries, within Julija Krupic's lab at the Department of Physiology, Development, and Neuroscience at the University of Cambridge.

Chapter overview

This chapter describes an experimental paradigm for using a hippocampal BMI in closed-loop. We focus specifically on the behaviour of animals during BMI-controlled trials, outlining factors affecting BMI performance of the navigation task described in the previous chapter. We explore the following main observations:

- BMI usage decouples the navigation task from locomotion on the wheel and reveals a new distinct neural representation of the environment
- BMI decoders trained on locomotion-based traversals of the VR track have poor closed-loop decoding accuracy, despite high accuracy when evaluated offline
- BMI decoders trained on traversals when locomotion and VR motion are not causally related have good closed-loop decoding accuracy and perform the task well
- Closed-loop BMI performance improves both with experience and with attentiveness

Additionally, we lay out an experimental schedule that runs over the course of approximately four weeks, including additional control trials designed to isolate BMI-specific phenomena.

3.1 Introduction

3.1.1 Thought experiment: pairing actions with neural states

The story so far: we have an accurate open-loop decoder, and we are curious to see what happens in a closed-loop setting. But how do we take a decoded position and turn it into a navigation action? Lai et al. (2023) approach this by averaging a decoded position over the course of several seconds and smoothly transporting the animal to that location, framing this as 'volitional control' over an imagined next location. As a first effort for closed-loop control, we are going to attempt something simpler: to construct a scenario in which the animal will succeed at the navigation task as long as it maintains a consistent neural representation of the VR track.

We illustrate this with a thought experiment. Suppose an animal is tasked with navigating through a maze: at each decision point, there is a correct action that will take the animal closer to the goal (Fig. 3.1.a). This pairing of location with navigation action forms a policy[1] that solves the maze. If we were to construct a BMI that had advance knowledge of the correct pairings between position and action, then this BMI would be able to reproduce correct trajectories through the maze by decoding position at each timestep and applying the correct navigation action for that position. Hypothetically, we might expect that this encourages the population to retain features of the neural representation critical for decoder accuracy: the neural population has an incentive to preserve its encoding, as task success becomes yoked to neural activity.

Our VR task is not a maze with choices, but a simple linear track. Nevertheless, the stereotyped parabolic trajectory that expert animals use to solve this navigation task—to run quickly when far away from the reward, to slow down ahead of the reward, and to accelerate quickly once the reward is obtained—allows us to formulate this task as a pairing between position on the track and velocity (Fig. 3.1.b). Our closed-loop BMI experiment uses the imaging and decoding procedures described in the previous chapter to decode a coarse position, one of ten spatial bins, in real time. This position is then converted into a velocity through a 'target trajectory' policy, which mimics the ideal stereotyped trajectory mice take along the track (Fig. 3.1.c, see Methods 3.7.1). Given our accurate decoder, we expect a fairly straightforward reproduction of the parabolic stereotyped trajectories using BMI control, provided that the animal maintains its representation of the track.

This approach does not give the animal freedom to explore using the BMI: it merely requires the animal to pay attention to the navigation task and keep its representation of the track intact. It does, however, open up opportunities to investigate a conditioning-like

[1] Borrowing the language of reinforcement learning, we would formulate the mapping of a state s to some action a as the 'policy' $\pi(a|s)$.

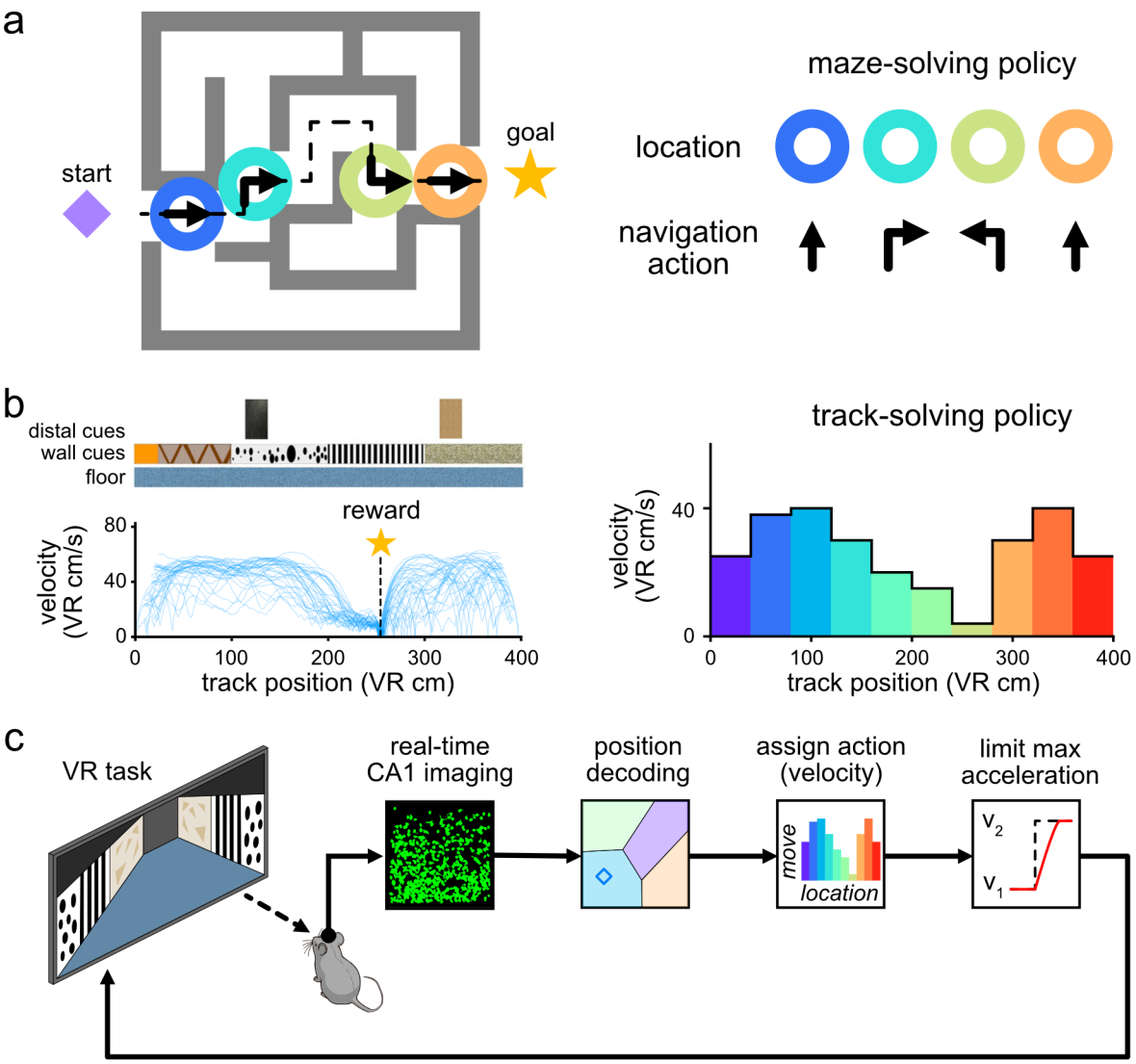


Fig. 3.1 Solutions to navigation problems can be expressed by a pairing of locations to movement actions. **(a)** In a maze with a starting location and goal location, the maze can be correctly traversed by reproducing the correct navigation choice at each decision point. With knowledge of the correct navigation actions, a BMI that accurately decodes locations and selects the corresponding navigation actions will produce successful trajectories through the maze. **(b)** The stereotyped trajectories of velocity on the 1D VR track correspond to an association between locations and actions. The animal's learned solution to the task, its behavioural policy, is to modulate its speed based on its current location on the track. **(c)** Using the BMI framework described in Chapter 2, we can decode the animal's perceived track location from CA1 hippocampus in real-time. By translating this into the correct forward velocity for the decoded location, we produce a closed-loop BMI task that the animal will succeed at for as long as the neural representation of the track can be accurately decoded.

reinforcement of the representation, and allows us to decouple the animal's running on the wheel (or other physical manipulation of a controller) from travel while maintaining a causal relationship between neural activity in CA1 and task success—the animal is not a passive passenger, and still has a degree of agency over its progress through the VR track.

3.1.2 Disastrous initial closed-loop decoding

This simple closed-loop thought experiment fails spectacularly in practice. An animal performs the VR navigation task for 10 minutes, completing an adequate (20+) number of laps of the track and exhibiting expert behaviour. We train a decoder on this data while the animal experiences a break from the task in darkness. We then use this decoder as part of the BMI control scheme to govern the animal's progress along the track in another 10-minute trial (Fig. 3.2.a).

During the BMI trial, the animal continues to pre-lick in anticipation of the reward, and it even continues to manipulate the wheel in a way that would result in correct task performance if the wheel were still responsible for progress along the VR track (Fig. 3.2.b). Despite this compelling behavioural evidence that the animal understands its location on the track, the VR velocity trajectories produced by the BMI indicate extremely poor decoding: they bear no resemblance to the neat parabolic behaviour of expert task performance.

How is it that a decoder that was so accurate in open-loop completely collapses in closed-loop? Rather than a technical failure, the change appears inherent to the neural population: when the loop is closed, the representation changes. Consider the example neuron in Fig. 3.2.c: this neuron has a place field when the animal uses locomotion to navigate along the track, but it appears entirely absent during BMI usage. However, if we disable the BMI in the middle of the trial and allow the animal to control its progress using locomotion, the place field returns on the very next lap. This change in representation takes place quickly and appears reversible. This is reminiscent of remapping, and we hypothesise that there is some BMI-specific representation that does not match the previous locomotion representation.

If that hypothesis is correct, then it is not a huge leap of faith to imagine that this BMI-specific representation might be sufficiently stable for us to fit a decoder to it and recover successful BMI performance of the navigation task. We can extend our session structure to produce a 'second-generation' decoder by using the attempted traversals of the track using the initial BMI as training data (Fig. 3.2.d). This works: when the animal performs BMI trials with this second-generation decoder, the target trajectory is followed well and laps are executed consistently (Fig. 3.2.e).

A description of what happens within the neural population during BMI remapping awaits in Chapter 4, along with clues as to the nature of this apparently context-sensitive remapping. In this chapter, we concern ourselves first with exploring the factors affecting closed-loop

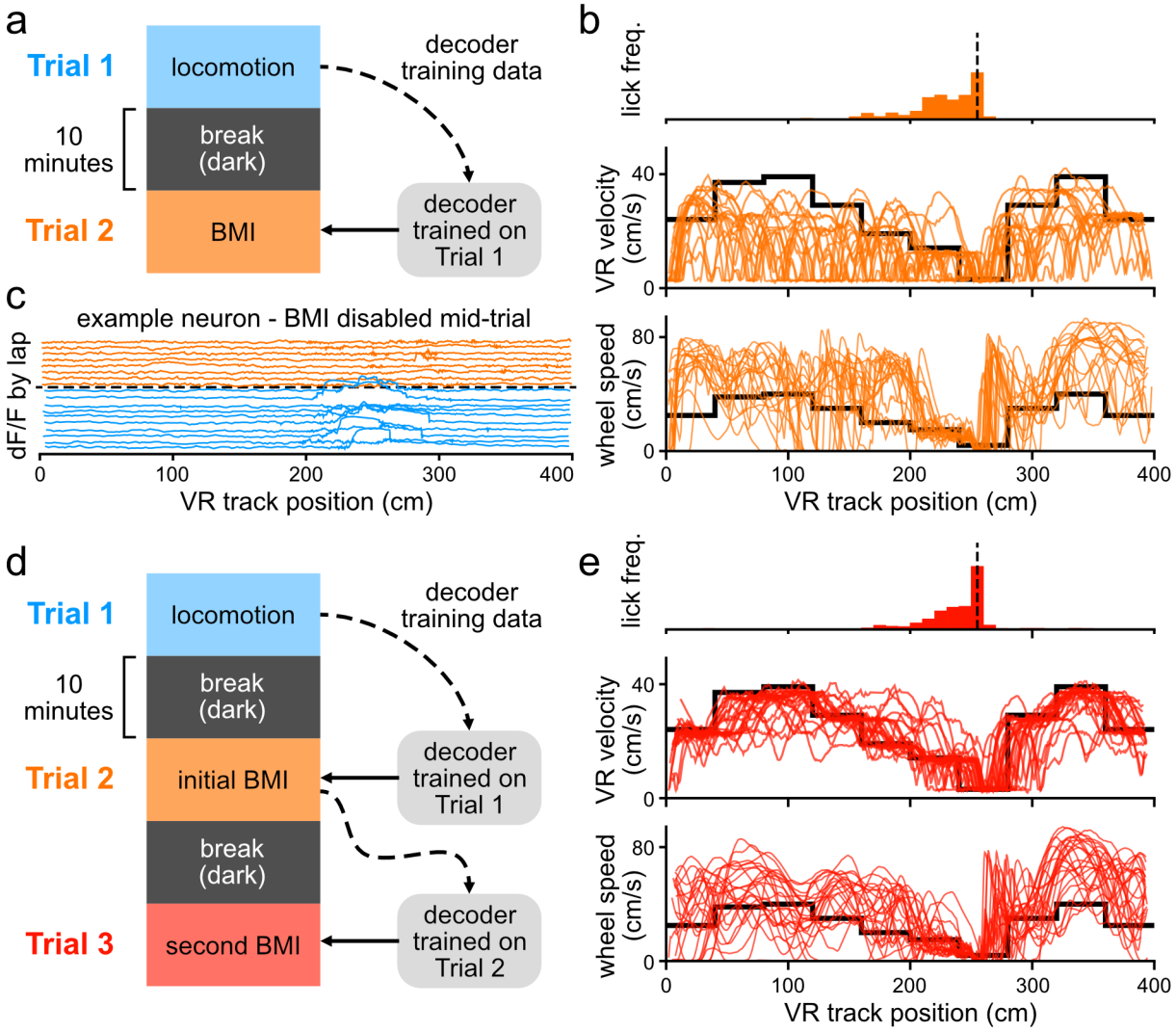


Fig. 3.2 Closed-loop BMI using a decoder trained on locomotion traversals of the environment fails to reproduce stereotyped trajectories. **(a)** An example session structure consisting of two 10-minute trials separated by a 10-minute break. The animal's first trial, in which it traverses the environment by running on the wheel (locomotion), is used as training data for a BMI decoder. The BMI decoder is used to control progress on the track in the second trial. **(b)** A histogram of the animal's lick locations across all laps of the track during the BMI trial, as well as the animal's velocity traversing the environment (each trace shows one lap) and the speed at which it runs on the wheel as a function of track position. The bold line indicates the target velocity profile that would be imposed by a perfectly accurate BMI decoder. **(c)** Lapwise dF/F traces for a single example neuron. Here, the initial laps are performed using BMI (orange traces), but the control modality is swapped from BMI to locomotion (blue traces) midway through the trial (dashed line). **(d)** An extension to the session structure in (a), wherein the attempted traversals of the track performed using BMI in Trial 2 are used to train a second decoder, which is subsequently used in Trial 3. **(e)** Behavioural plots as in (b), shown for Trial 3 in which the decoder has been trained on attempted BMI traversals from Trial 2.

decoding accuracy and demonstrating that this discrepancy between decoders trained on locomotion and trained on (poor) BMI traversals is highly consistent.

3.2 Closed-loop BMI experiments and control sessions

3.2.1 Session structure and control modality trial types

From this point onwards, all the experiments relating to closed-loop hippocampal BMI use a common structure: we run sessions of four 10-minute trials, each spaced by 10 minutes in darkness. Each trial contains multiple laps of the VR track. The first and last trials serve as reference trials, while the inner second and third are a probe condition (Fig. 3.3.a). This structure allows us to investigate both which changes in representation are reversible (remappings corresponding to representations specific to BMI or locomotion) and which changes are persistent (alterations to an existing cognitive map).

The environment cues, reward structure, and trial length are always unchanged. The only variable that changes between trials is the 'control modality': the means by which the animal traverses the environment. There are three major categories of control modality: locomotion, BMI, and autopilot (Fig. 3.3.b).

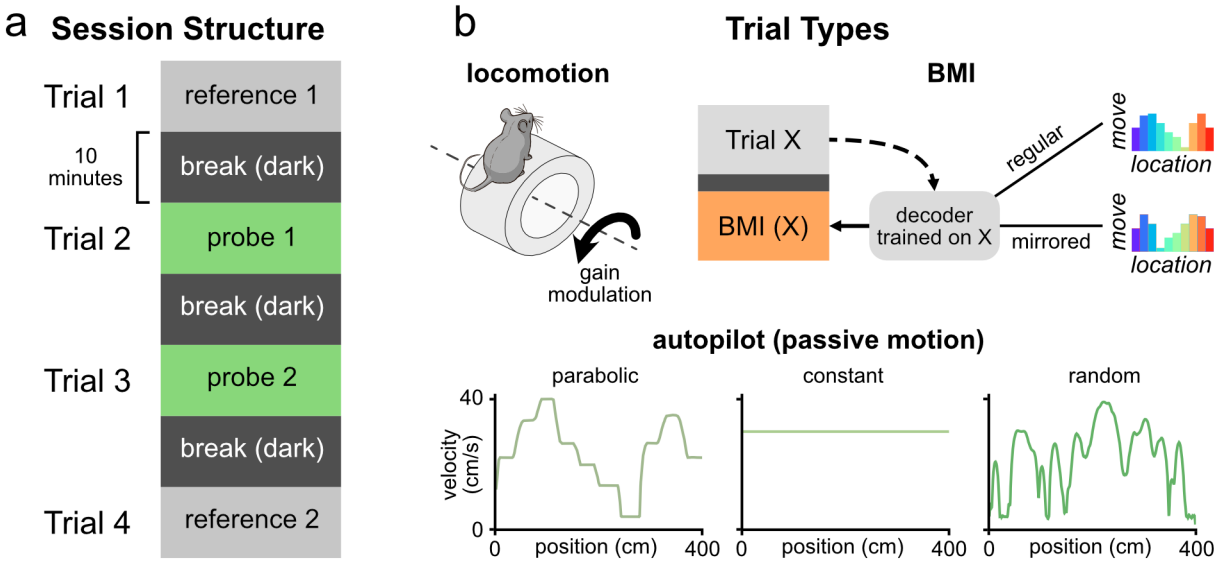


Fig. 3.3 Overview of recording session structure. **(a)** Sessions consist of four 10-minute trials each spaced 10 minutes apart by breaks in darkness. The initial and final trials serve as a reference condition, while the inner trials are probe trials. **(b)** The major trial types are named according to which control modality is used to navigate the track during the probe trials: locomotion (including a gain-modulation variant in which the gain of the wheel is modified), BMI (including a 'mirrored' variant in which the target velocity profile is inverted), and autopilot in which the animal has no control over its progress (including three variants).

In locomotion trials, the animal makes progress along the VR track by running on the wheel. We typically use these for the reference trials at the start and end of each session. A variant of locomotion trials (covered in more detail in Section 4.5) allows us to vary the 'gain' of the wheel, altering the relationship between the animal's running and the distance traversed on the VR track—this eventually functions as a control against effects introduced between BMI and locomotion control modalities that could correspond instead to discrepancies in distance travelled and distance run.

In BMI trials, the neural activity of the population determines progress along the VR track. The decoder for any given BMI trial is always trained on the immediately preceding trial, and is labelled as such (for example: a trial controlled by a BMI whose decoder was trained on trajectories from Trial 2 would be labelled 'BMI 2'). As previously described, BMI control maps a decoded position to a velocity via a target trajectory policy. We use the same target trajectory (shown earlier in Fig. 3.1.b, see Methods 3.7.1) for all animals, even if some animals naturally run faster or slower. This trajectory was hand-constructed to mimic the real stereotyped trajectory of animals. In the 'BMI-mirror' variant (covered in more detail in Section 3.5.1), the target trajectory is mirrored about the centre-point of the track to deliberately disrupt the stereotyped running trajectory.

In autopilot trials, the animal has no control over its progress along the VR track. We use three variants of these trials. These are control trials that allow us to compare the effects of BMI usage on the neural population separately from the effect of removing locomotion-based control of movement. In 'constant' autopilot, the animal travels at a constant 30 cm/s under all circumstances. In 'parabolic' autopilot, we force the animal to follow the target trajectory for BMI trials, which emulates the trajectory an animal would take with a hypothetical 100% accurate decoder. In 'random' autopilot, the animal's forward velocity is selected pseudorandomly (see Methods 3.7.1), which produces varied traversals of the track from one lap to the next that are unpredictable from the perspective of the animal.

3.2.2 Session descriptions

These trial types based on control modalities are combined into types of 4-trial sessions, summarised in Table. 3.1. We show some typical trials illustrative of animal behaviour during these various session types in Fig. 3.4 and Fig. 3.5. We will return to these session types as they become relevant to our argument—this section will be useful to refer back to—but take the opportunity here to make both some general comments about behaviour and explain the rationale for each session type.

	Trial 1	Trial 2	Trial 3	Trial 4
all locomotion	locomotion	locomotion	locomotion	locomotion
BMI	locomotion	BMI (1)	BMI (2)	locomotion
autopilot-BMI	random speed	BMI (1)	BMI (2)	random speed
mirrored BMI	locomotion	mirror BMI (1)	mirror BMI (2)	locomotion
random speed	locomotion	random speed	random speed	locomotion
constant speed	locomotion	constant speed	constant speed	locomotion
parabolic speed	locomotion	parabolic speed	parabolic speed	locomotion
gain modulation	locomotion	modulated gain	modulated gain	locomotion

Table 3.1 Session types are defined by their composition of trial types. All sessions except for the autopilot-BMI session use locomotion for the reference condition in Trial 1 and Trial 4.

'All locomotion' sessions

These sessions act as a baseline, with the probe trials replaced by additional reference locomotion trials. They allow us to assess a baseline rate of turnover in representation that takes place without any experimental manipulations. By construction, wheel speed and VR velocity are an exact match during any locomotion trial (Fig. 3.4.a). Inspecting the velocity traces both within Fig. 3.4.a and the locomotion trials within other sessions in Fig. 3.4 and 3.5, we see that the animals do indeed perform highly stereotyped traversals of the environment. However, while each animal is individually consistent with its own stereotyped strategy, and while there are features common to all of the traversals (e.g. early acceleration, anticipation of the reward site), there are variations from individual to individual. For example, some animals achieve their peak velocities before the reward site, and others after the reward site on each lap. We draw particular attention to the fact that not all animals reach the same peak running speed.

BMI sessions

The bulk of sessions are BMI sessions, in which the animal first performs the navigation task with locomotion. This is used to train an initial decoder, which is used for closed-loop BMI in Trial 2. Decoder accuracy is typically poor in Trial 2, but this is nevertheless used to train a second decoder, which is used for closed-loop BMI in Trial 3.

In Fig. 3.4.b we show an example of a mouse that is particularly fast during its locomotion trials, running at a higher mean velocity during locomotion in the fast segments of the track than the target velocity profile imposed by the BMI in Trials 2 and 3. Instances such as these have two potential confounds: (1) it is possible that, rather than BMI usage, it is a change in the animal's perceived gain, where the animal feels like it must rotate the wheel more times to travel the same distance, that is responsible for remapping; (2) animals running more laps provide a greater amount of training data for the decoder. We will address the former

with a control experiment (gain modulation, see Section 4.5), and the latter with a thorough explanation of factors affecting closed-loop decoder accuracy later in this chapter.

Autopilot-BMI sessions

In autopilot-BMI sessions, the reference trials are replaced with random speed autopilot trials. This means that the initial decoder, used for closed-loop BMI in Trial 2, has been trained on a passive motion-based representation of the track, rather than one based on locomotion. As with the BMI sessions, the decoder is retrained on Trial 2 to produce a second-generation decoder for use in Trial 3.

Various autopilot sessions

The three types of autopilot session, visible in Fig. 3.5.a-c, use passive motion probe trials. These will serve as an important control in the next chapter for reasoning about which changes in representation take place due to the loss of a useful path integration signal (manipulation of the wheel is no longer informative about future location during both closed-loop BMI and passive motion) and which changes take place because BMI has granted this neural population causal control over navigation actions.

We highlight in these examples that the animal will remain engaged and attentive during autopilot sessions: it continues to manipulate the wheel in an attempt to produce the correct, stereotyped behaviour that we would expect in locomotion trials. This attentiveness is also visible in their licking profiles. While it appears that they are late for the reward location in constant speed and random speed autopilot trials, this lateness is due to the animal's movement being forced along the track while it is attempting to remain stationary and lick the reward spout.

Mirrored BMI sessions

The continued 'correct' manipulation of the wheel during non-locomotion trials leaves us with another confound during BMI usage: the animal may be emulating path integration and thereby keeping its hippocampal bearings by manipulating the wheel during BMI, even if such manipulation does not directly control progress along the VR track. Fortunately, we find that this is not universally true: in Section 3.4.1, we will see that some animals do perform BMI well while entirely decoupling their wheel speed from their VR velocity. To supplement these observations, we also perform a handful of speculative experiments that use mirror BMI trials to force a decoupling of wheel motion from VR progress. These have additional utility in planning prospective future work, as they allow us to assess the stability of representations when the trajectory imposed by the BMI runs contrary to the trajectory desired by the animal. We show some example behavioural traces from such a session in Fig. 3.4.d.

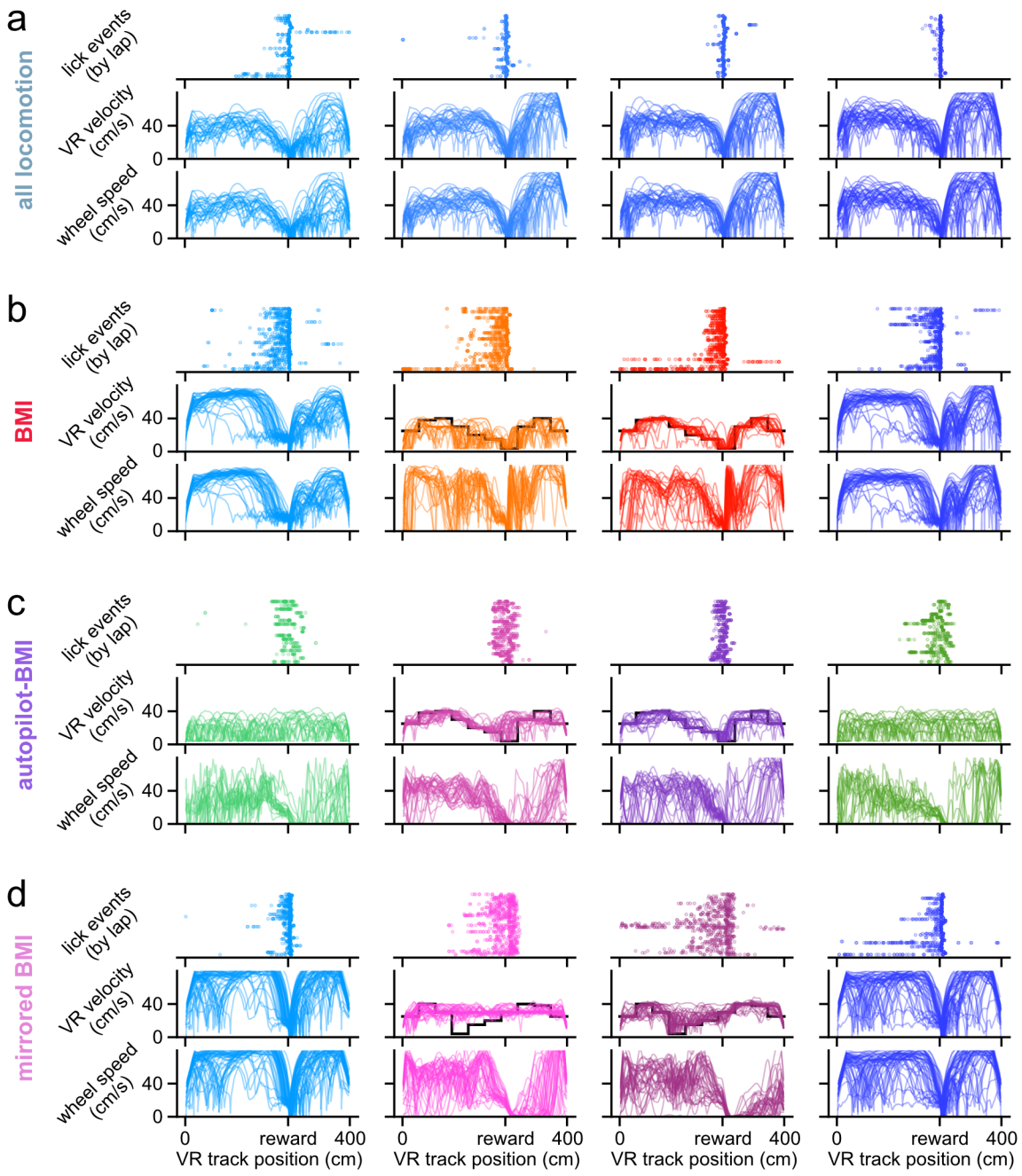


Fig. 3.4 Example animal behaviour across session and trial types. **(a)** The licking events, velocity profiles, and wheel running speed profiles for all 4 trials in an example all-locomotion session. **(b)** As in (a) for an example BMI session. **(c)** As in (a) for an example autopilot-BMI session. **(d)** As in (a) for an example mirrored BMI session. The target velocity policy is shown by a bolded black line in all trials where BMI is used to control the VR velocity.

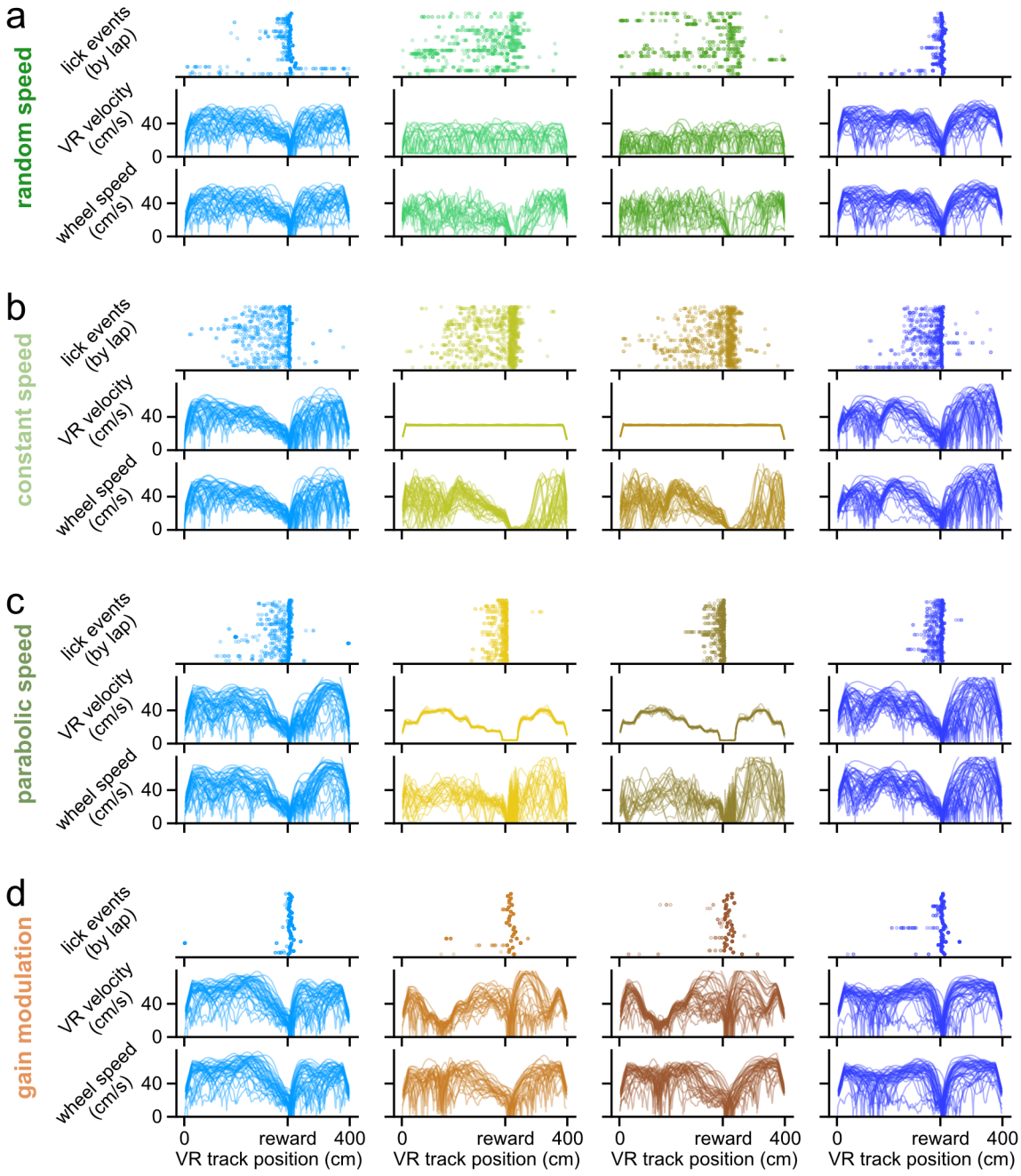


Fig. 3.5 Example animal behaviour across the control sessions and trial types. As in Fig. 3.4 for **(a)** an example random speed autopilot session, **(b)** an example constant speed autopilot session, **(c)** an example parabolic speed autopilot session, and **(d)** an example gain-modulation session. These session types are designed to serve as controls for any changes in representation induced by BMI usage.

Gain modulation sessions

Gain modulation sessions use probe trials which distort the gain of running on the wheel as a function of location on the track. We draw the reader's attention to the disparity between the wheel speed and the VR velocity (Fig. 3.5.d) during probe trials, corresponding to regions of suppressed or amplified gain. We will return to these sessions in the next chapter in Section 4.5, where they are used as part of a control.

3.2.3 Experimental schedule

With our outline of session types and structure complete, we can now introduce the 11 individual animals used in this investigation. We take particular care to describe the order in which animals experienced sessions, both because past experiences strongly influence hippocampal representations and because we are curious to observe any learning or familiarisation with BMI usage on longer timescales (Fig. 3.6.a). All of these sessions are described after animals reached expert performance on the locomotion version of the VR navigation task, with dates relative to the first structured 4-trial session. We timed sessions to coincide with a 4 week-long window in which the indicator was well-expressed (usually starting ~40 days after injection) and provided a steady number of neuron ROIs in each imaging FoV (Fig. 3.6.b).

There are two main groups of 5 animals each: those that immediately perform a bulk of BMI sessions (a minimum of 5) and subsequently experience autopilot sessions, and those that experience autopilot sessions first, then perform a bulk of autopilot-BMI sessions, and finally perform BMI sessions. We take care to indicate the handful of sessions that were either started and then aborted for technical reasons (e.g. bubble obscuring the FoV or partial data corruption) or discarded due to behavioural issues (in all instances, these belonging specifically to MH356 and its refusal to run more than 10 laps in the final reference trial), as these are experiences that may affect representations during the sessions that follow.

MH188 is a special case that falls outside of these two groups. It was data from this animal, and its fellow cohort animal MH187 (which is not part of the main experiments described here), that we presented in the previous chapter when evaluating open-loop decoder accuracy. Unlike the other animals, not all MH188's BMI experiences are shown in Fig. 3.6.a, as it was involved in the refinement of the protocol and technical system tests on the BMI software. For this reason, we omit MH188 from analyses that concern any changes from session to session. However, we choose to include its structured 4-trial sessions for completeness where valid to do so (bringing our total number of valid BMI sessions to 50).

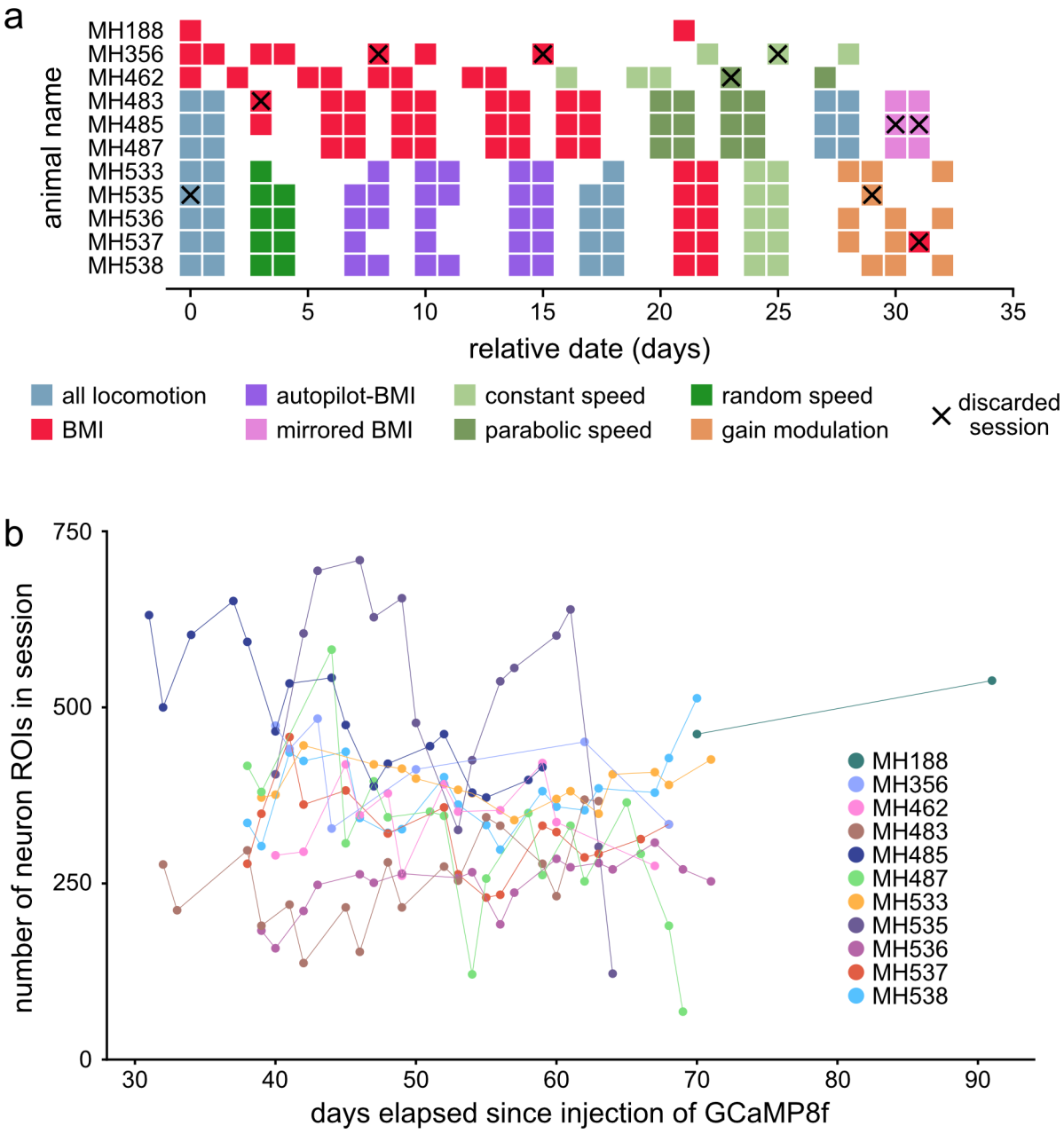


Fig. 3.6 Overview of experimental schedule for each animal. **(a)** An overview of the experimental schedule on a per-animal basis for each of the 11 animals, outlining which sessions ran on which days. Sessions that partially ran but needed to be discarded, either due to imaging issues or due to low animal engagement (refusal to run during reference trials), are marked with a cross. **(b)** The number of neuron ROIs identified in each session from (a) for each animal, highlighting the imaging time window following the injection of GCaMP8f.

3.3 Representational incompatibility between locomotion and BMI contexts

3.3.1 Retraining of decoders produces good closed-loop performance

In our anecdotal introduction, we described closed-loop BMI performance that was poor when trained on a locomotion trial but good when trained on a BMI trial. We can begin to quantify this discrepancy in performance by considering confusion matrices for the decoders (see Methods 3.7.3). We first consider these confusion matrices aggregated over all 50 BMI sessions both for BMI 1, which uses a decoder trained on locomotion, and for BMI 2, which uses a decoder trained on BMI 1 (Fig. 3.7.a and 3.7.b respectively). The difference is stark. The confusion matrix for BMI 2 has a relatively clean diagonal, indicating a decoder output that roughly matches the animal's true position (sometimes erring one bin either side of the true bin, but not biased towards decoding a leading or lagging bin). In contrast, the confusion matrix for BMI 1 indicates a strong propensity to mispredict in favour of the bin containing the reward location or, failing this, to predict an entirely random location on the track[2]. This discrepancy between decoder accuracies suggests some level of incompatibility between the neural representation of the track when the animal traverses the environment with BMI and the representation when the animal traverses using locomotion.

As a contrasting observation, if instead we train the decoder for BMI 1 on traversals made using a random speed autopilot (as in the autopilot-BMI sessions), we find that closed-loop BMI accuracy is immediately at parity with performance during BMI 2 (Fig. 3.7.c-d). While we will eventually discover in Chapter 4 that the autopilot traversals and BMI traversals do not produce identical neural representations of the track, they are at least compatible with each other: any changes to the representation when transitioning between autopilot and BMI do not interfere with decoding accuracy, and must therefore take place in the null-space of the decoder weights.

This phenomenon, conflicting representations of the environment corresponding to a distinct 'locomotion context' and a 'BMI context', is highly consistent. When we consider the confusion matrices for individual sessions (Fig. 3.8.a), we see the same pattern as we do in the aggregate over all sessions. We quantify the improvement between decoders trained on the locomotion context and those trained on the BMI context as a clear jump in decoder scores[3] (Fig. 3.8.b). In contrast, no such change in performance exists when moving from decoders trained on autopilot control to decoders trained on BMI control (Fig. 3.8.c-d).

[2]With sufficient squinting, a very faint diagonal is visible, as are some pockets of better-than-chance decoding at the start and end of the track, but the decoding situation is obviously dire.

[3]These scores have a minor alteration from the open-loop scoring scheme used in the previous chapter to protect against a potential sampling bias, see Methods 3.7.3.

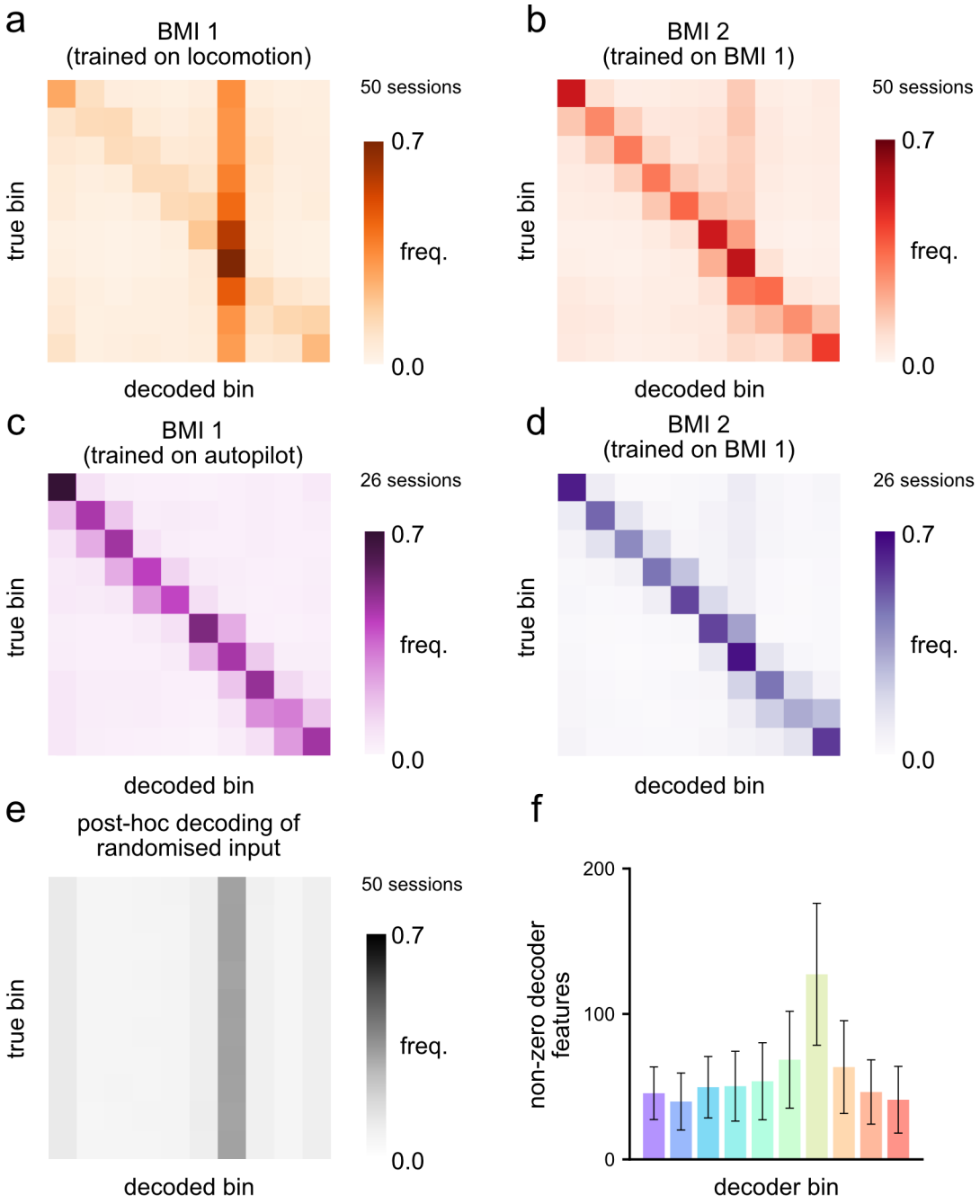


Fig. 3.7 Decoder confusion matrices for closed-loop BMI trials. **(a)** A confusion matrix, normalised by ground truth and aggregated over all 50 BMI 1 trials during BMI sessions, visualising closed-loop BMI decoding accuracy. Each bin decoded by the BMI is compared against the animal's true location when that bin was decoded. **(b)** As in (a), for the same sessions, but showing accuracy during BMI 2 trials. **(c)** As in (a), but for the 26 autopilot-BMI sessions. **(d)** As in (b), but for the 26 autopilot-BMI sessions. **(e)** A confusion matrix produced offline by replacing the input to decoders trained on locomotion trials with Gaussian noise. Averaged over the 50 sessions from (a). **(f)** The mean number of features in the pixel feature-space with non-zero weights for each decoder bin, shown for the decoders trained on locomotion used in (a) and (e). Error-bars are +/- standard deviation over all 50 decoders.

Despite the clear improvement in performance between BMI 1 and BMI 2 in BMI sessions, the reader will observe considerable variance in decoder scores from session to session (Fig. 3.8.b)—including some rare instances of locomotion-trained decoders performing well (which we address in Section 3.4.2). Why is it that some sessions demonstrate excellent BMI performance, and others merely adequate performance? In the following sections, we will attempt to explain the variance in decoder scores in terms of how attentively the animal engages with the task, the extent of its past experience with BMI, and surgery-specific factors such as the number of neurons visible in the FoV.

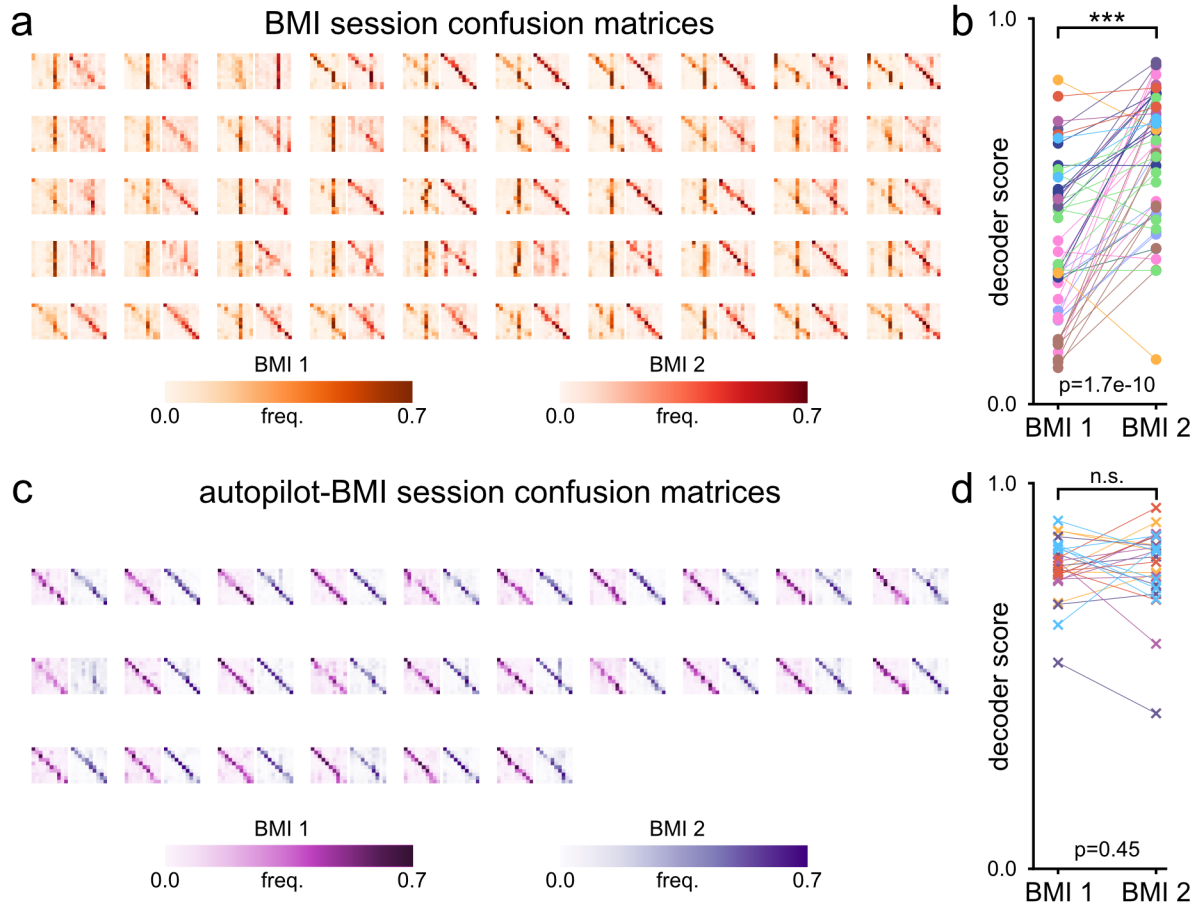


Fig. 3.8 BMI decoders trained on non-locomotion traversals of the track perform better in closed-loop. **(a)** Individual closed-loop confusion matrices for BMI 1 (orange) and BMI 2 (red) trials for all BMI sessions. **(b)** Changes in decoder scores between BMI 1 and BMI 2. Colour indicates animal identify, as per Fig. 3.6. Paired t-test without correction for animal identity. **(c)** As in (a), for the autopilot-BMI sessions. **(d)** As in (b), for the autopilot-BMI sessions.

3.3.2 Over-prediction of the reward location

The failure case of BMI decoders trained on locomotion trials and subsequently used in closed-loop is to over-predict the bin containing the reward site. While it *is* the most frequently occupied bin in the training data due to the animal stopping at this location to receive the reward, this frequency bias should have been avoided by balancing classes during training. Is the cause of the over-prediction due to an artefact of decoder training, or to a more fundamental neural process such as the animal 'imagining' the reward location during its BMI 1 trial?

The answer lies in a well-known feature of populations in CA1 hippocampus: reward locations are salient and are therefore over-represented in the population (shown in our own work in Section 4.3, or in prior art such as Hollup et al., 2001), which will produce an inherent structural bias in our decoder even if classes are balanced during training. If we visualise this in terms of a space of neural activity through which we need to draw a decision boundary, the volume of space occupied by neural activity corresponding to the spatial bin containing the reward site is relatively large. If we take a decoder trained on locomotion and subsequently evaluate its open-loop decoding performance while replacing all of the decoder inputs with Gaussian noise (see Methods 3.7.3), we find that it produces the same over-prediction of the reward site as when a locomotion-trained decoder is used in closed-loop (Fig. 3.7.e). We can further validate this explanation by inspecting the weights allocated to the feature space, which shows the over-representation explicitly: more features have non-zero weights for the bin containing the reward site (Fig. 3.7.f). We take the shape of this misprediction error as an early hint that the differences between representations in the BMI and locomotion context are close to a remapping-like change, in which one representation looks almost like random noise to the other.

As an engineering-centric aside: it is, in principle, possible for us to compensate for this type of structured misprediction. For example, we could downsample from the feature space to produce a more homogeneous distribution of feature space weight by bins. This would change the misprediction from one that is structured to one that is distributed across the entirety of the confusion matrix.

3.4 Factors affecting closed-loop BMI performance

3.4.1 Attentiveness

Animals manipulate the wheel during BMI trials in a manner that would produce correct trajectories, an obvious sign of engagement with the task. This nevertheless produces a tricky confound: as we know, path integration is critical to the formation of cognitive maps,

and it is conceivable that by manipulating the wheel at the correct points the animals can effectively emulate path integration (and indeed may not even realise that manipulating the wheel is no longer the cause of their motion). In practice, during BMI 2 trials, where closed-loop decoders are accurate and solve the navigation task well, different animals exhibit a variety of behaviours: some will manipulate the wheel with an extremely close match to the BMI-produced trajectories (Fig. 3.9.a), producing high correlations between their running speed on the wheel and the VR velocity produced by the BMI (Fig. 3.9.b), while others have a complete disconnect between wheel speed and VR velocity (Fig. 3.9.c-d). Even during BMI, it is rare that the animals remain stationary on the wheel for long. Wheel-running is a naturalistic behaviour for mice (Meijer and Robbers, 2014), and so complete stoppage is likely to require explicit training (or a stronger incentive, such as increased resistance of the wheel).

Higher correlations between VR velocity and wheel speed correlate with high closed-loop decoder scores during BMI 2 trials (Fig. 3.9.e). However, much of the variance remains unexplained, and there are ample instances of good decoder performance despite the wheel speed and VR velocity being entirely decoupled. Additionally, it is difficult to produce situations with high correlation between wheel speed and VR velocity yet low decoder scores, as this would require the animal to match its manipulation speed to the almost-random motions produced by an inaccurate decoder, rather than the stereotyped behaviour it adopted during its own training.

Fortunately, we have another way to measure engagement with the task: whether or not the animal is licking in anticipation of the reward. A simple pre-licking score (see Methods 3.7.2) correlates just as well with closed-loop decoder score as the correlation between VR velocity and wheel speed (Fig. 3.9.f). Reasoning that it is possible for an animal to be attentive to the task but not run on the wheel, or indeed for an experienced animal to be attentive but not lick with as much anticipation, we combine the pre-licking score and wheel manipulation into an attentiveness metric that is high if *either* of these two measures of engagement is high (and therefore low only if both are low, see Methods 3.7.2). This metric explains considerable variance in the BMI 2 decoder score, both in autopilot-BMI sessions and in BMI sessions (Fig. 3.9.g).

We take this observation as plausible initial evidence that it is active engagement with the navigation task that governs good closed-loop BMI performance. Animals are able to achieve high decoder scores (and therefore have underlying stable representations) in spite of the decoupling between wheel speed and VR velocity, and we therefore rule out the possibility that those animals that perform well in BMI are somehow manipulating the wheel in agreement with VR velocity to produce a facsimile of a path integration signal that stabilises or enriches their representations of the track in CA1.

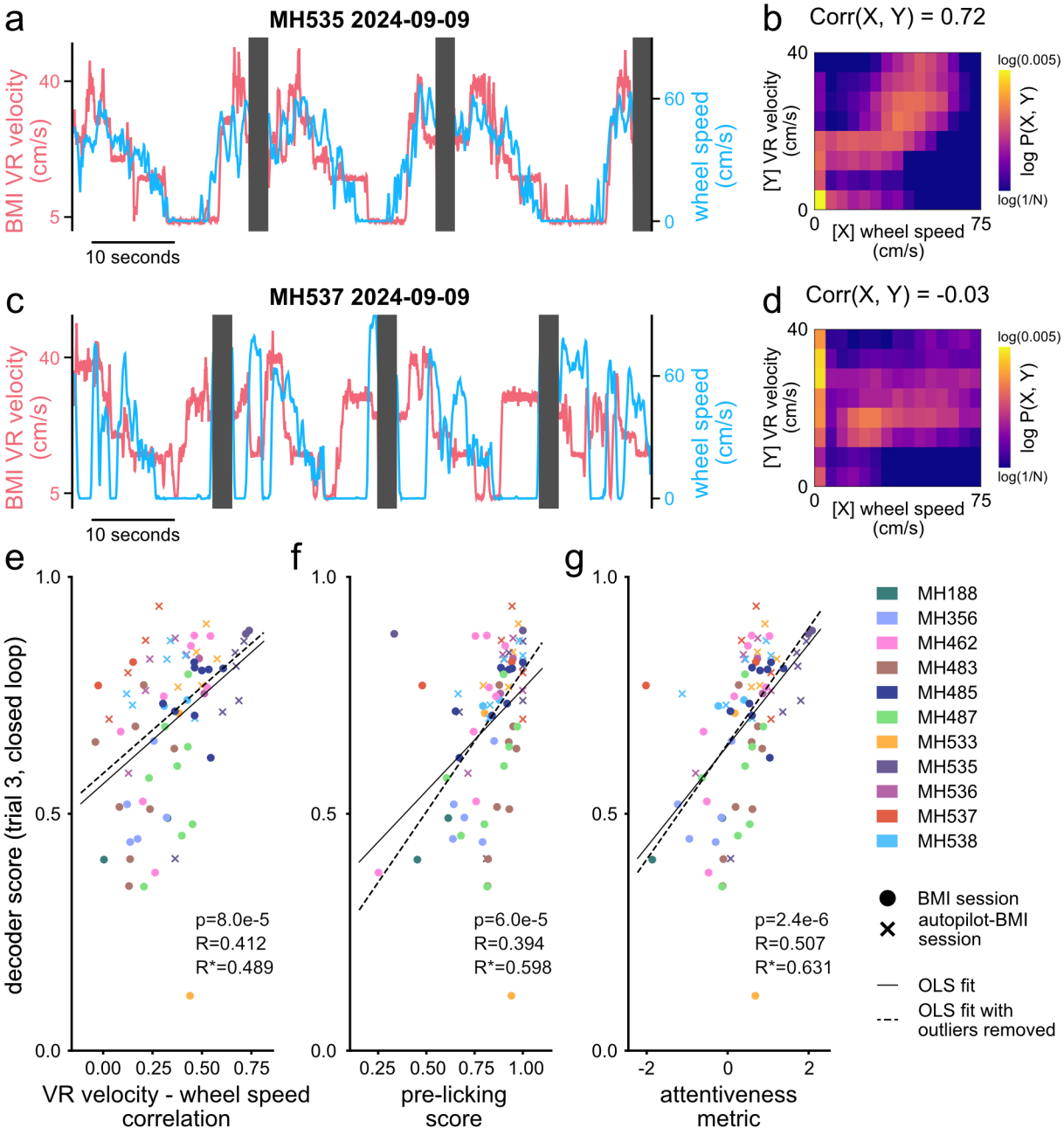


Fig. 3.9 Task attentiveness is an excellent predictor of BMI decoder accuracy after decoder retraining. **(a)** An example time segment of a BMI 2 trial showing both the animal's VR velocity as produced by BMI and its wheel running speed. Grey rectangles indicate inter-lap darkness. **(b)** An occupancy heatmap visualising the relationship between VR velocity and wheel speed over the course of the entire trial shown in (a). **(c)** As in (a), but illustrating a session in which the animal exhibits minimal correlation between VR and wheel velocities. **(d)** As in (b), shown for the trial visualised in (c). **(e, f, g)** Decoder scores during BMI 2 of both BMI (circles) and autopilot-BMI sessions (crosses), shown as related to the correlation between VR velocity and wheel speed, the animal's licking behaviour, and an attentiveness metric combining both of these measures. R-values correspond to OLS fit and R*-values correspond to OLS fit with outliers removed. p-values are corrected for animal identity and trial type with a mixed linear model.

3.4.2 Experience and place cell count

BMI sessions

We might expect that the number of ROIs visible in a session is a good predictor of decoder performance during the BMI 2 trial in BMI sessions: the decoder input has access to more information and a greater degree of redundancy. This is not the case (Fig. 3.10.a), at least for the range of population sizes we image and the spatial resolution at which we classify (a population with zero neurons would obviously be incapable of producing a decoder more accurate than chance, and we discuss later the decoding implications of access to population sizes an order of magnitude or two larger). What *is* an excellent predictor of performance is the number of place cells present in the population, which is not the same from one session to the next (Fig. 3.10.b). We find too that BMI performance improves with experience (Fig. 3.10.c), and underlying this improvement is a gradual increase in the fraction of the imaged population that are place cells (Fig. 3.10.d). While some animals may have better or worse decoder accuracy than others, owing to systematic differences in imaging quality or their general inclination to engage with the task, the increase in the fraction of place cells that appear during Trial 3 in BMI sessions is universal. We will not stretch to the tempting claim that animals are 'learning to use the BMI', but we do claim that at minimum this represents a habituation to BMI that allows a gradual enrichment of the BMI-specific representation of the track.

Improvements in decoder accuracy, increased attention during BMI, and experience are all necessarily correlated with each other in a way that makes it difficult to unpick whether increasing decoder accuracy causally increases attention, or vice-versa. It is possible that an animal gradually learns that attending to the task during BMI, which would generate a richer representation of the track, results in desirable outcomes. Equally, if the representation were enriched through some other process, thereby improving BMI performance, it is possible that attentiveness improves because desirable outcomes take place more frequently.

Autopilot-BMI sessions

We turn to the autopilot-BMI sessions for a comparison, inspecting decoder scores during both BMI 1 and BMI 2 trials. Unlike the cohort of animals described above that experience a long stretch of BMI sessions, this cohort experience a long stretch of autopilot-BMI sessions. Their decoders are high-scoring in closed-loop from the very first day, and demonstrate minimal improvements in score over time (Fig. 3.11.a-b). They have less pronounced increases in population place cell fraction during BMI (Fig. 3.11.c-d).

As an aside, we point out that while the autopilot-BMI decoder scores appear at first-glance to be insensitive to the number of place cells in the session (Fig. 3.11.e-h), this is likely a

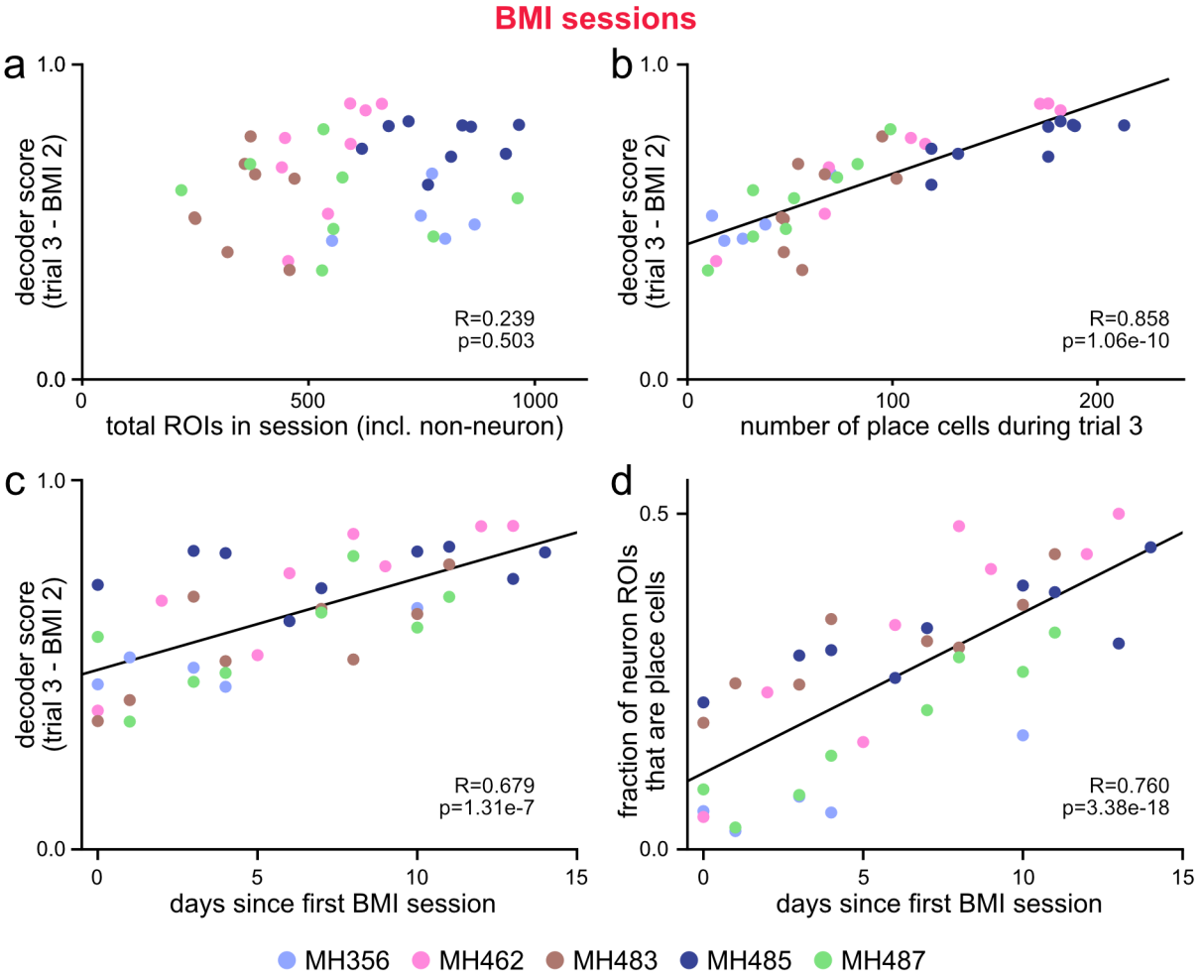


Fig. 3.10 Decoder accuracy improves with number of place cells. Increasing BMI experience results in more place cells in the same population. Decoder score during BMI 2 trials as a function of **(a)** the number of ROIs identified in the session, **(b)** the number of ROIs that are place cells in that session, and **(c)** the number of days elapsed since the animal's first experience with a BMI session. **(d)** The density of place cells within the neural population as a function of the number of days elapsed since the animal's first experience with BMI. All R-values and p-values are based on a linear fit and corrected for animal identity.

sampling artefact: there are no instances of having very few place cells, unlike during the BMI sessions above, and so all of the datapoints exist in a regime where there are sufficient place cells to decode well. More importantly: why is it that autopilot-BMI sessions are immediately rich in neurons with place fields, and yet BMI sessions take time for these to emerge?

We speculate that because the autopilot-BMI animals stop experiencing locomotion trials altogether for several consecutive sessions, and because these animals had experienced the random speed condition prior to the BMI condition, they both started with a 'better' representation of the track in non-locomotion conditions and had no need to maintain two separate representations of the track for each mode of active travel. In contrast, the cohort experiencing primarily BMI sessions needs access to both representations of the track on each day for task success. To address this more fully, we need a few more details concerning the performance of closed-loop decoders trained on locomotion trials, and to understand how this differs for the cohort of animals experienced with the autopilot-BMI protocol.

Performance of closed-loop decoders trained on locomotion trials

The neural representation of the track during BMI 1 in BMI sessions enriches over time (Fig. 3.12.a)—this is hardly unexpected, as it is training decoders on these representations that gives us accurate closed-loop BMI performance during BMI 2. However, despite this enrichment, closed-loop decoder scores remain constant in addition to being poor (Fig. 3.12.b). This in itself is a critical point for distinguishing BMI and locomotion based representations of the track as being separate: the population is *not* retaining more of the representation from locomotion to achieve its greater number of place cells, but rather producing them orthogonal to the locomotion representation.

As was the case for BMI 2 trials, attentiveness remains important for decoder accuracy (Fig. 3.12.d). While the session-to-session fluctuation of ROI count for the same animal is of little consequence to decoder accuracy during BMI 1, we do generally find that those animals with more ROIs achieve better decoder accuracies than those with fewer (Fig. 3.12.e).

When we consider animals experienced with BMI trials (a minimum of 5 previous sessions containing BMI trials, to mitigate discrepancies caused by gradual improvement in decoder accuracy), we make the following observations (Fig. 3.12.c):

- Animals habituated to the autopilot-BMI sessions (group B) have significantly better decoder scores during BMI 1 than animals habituated to BMI sessions (group A)
- Decoder scores during BMI 2 are not significantly different between groups A and B
- Group B still exhibit an increase in accuracy between BMI 1 and BMI 2 when BMI 1 is trained on locomotion trials

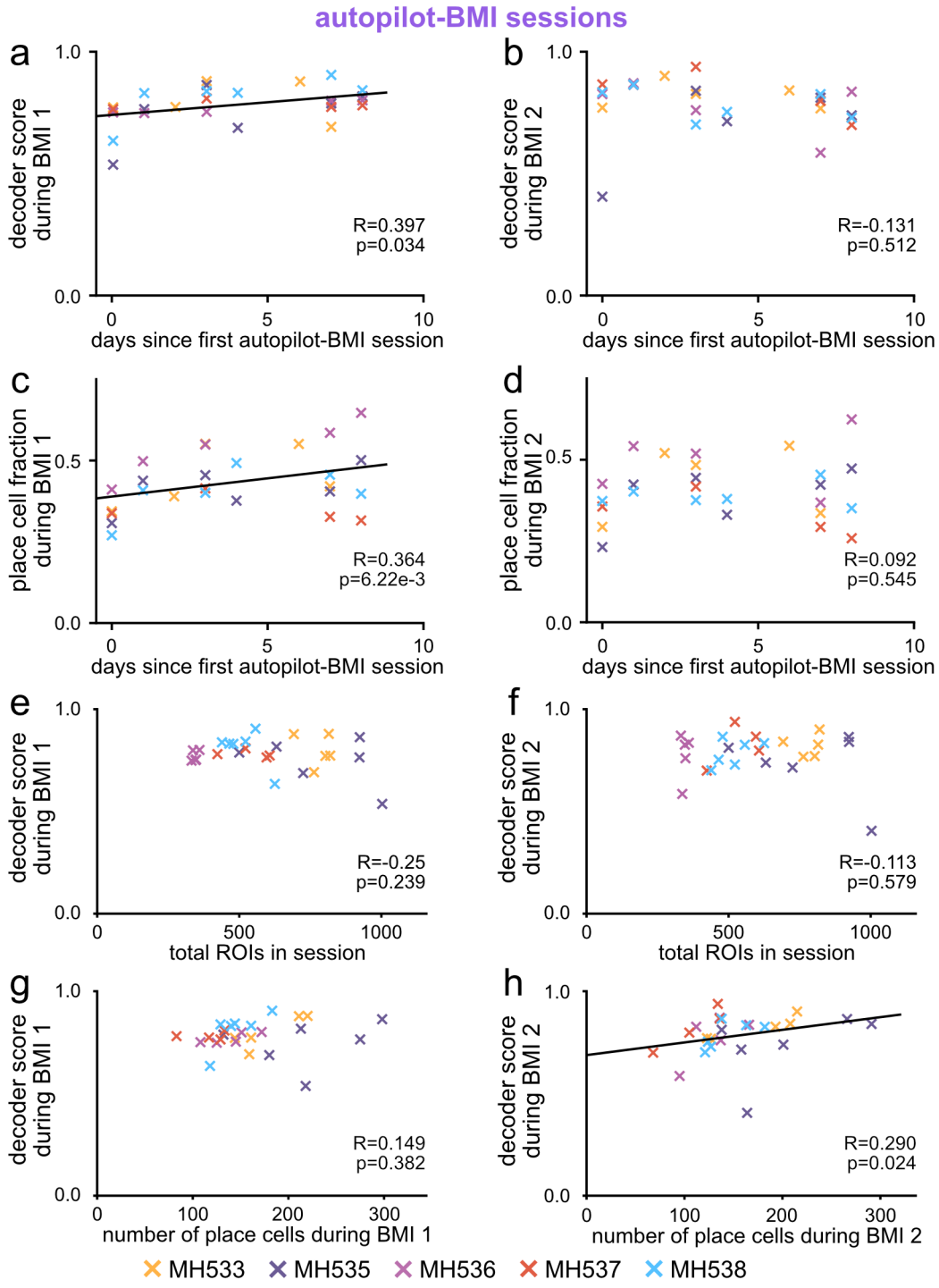


Fig. 3.11 Decoder accuracy improvements for animals undergoing autopilot-BMI trials are limited. **(a-h)** The same quantities as in Fig. 3.10, shown for both BMI 1 and BMI 2 trials during autopilot-BMI sessions. Note that these p-values are *not* corrected for multiple comparisons.

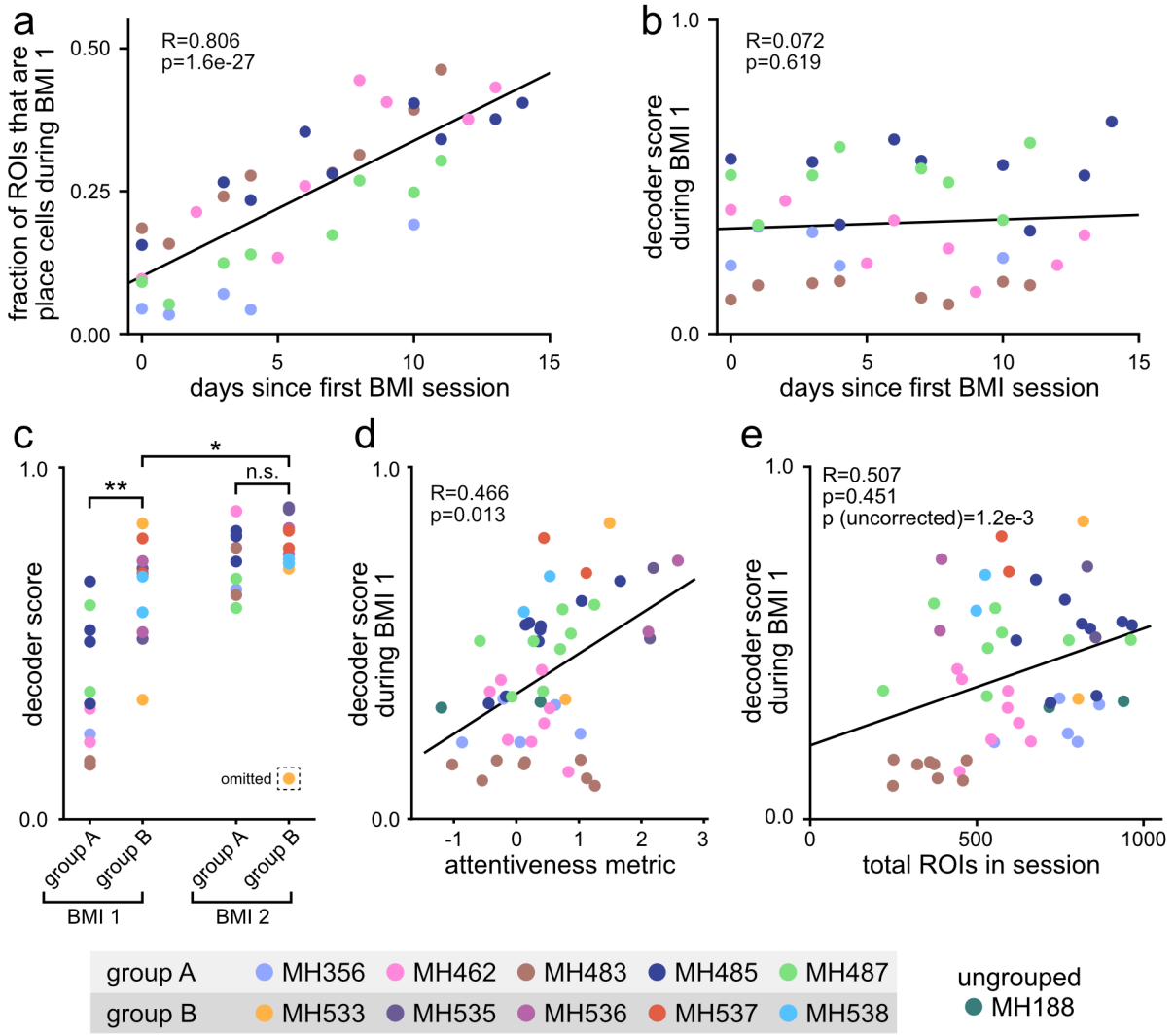


Fig. 3.12 Factors affecting performance for decoders trained on locomotion traversals. **(a)** Fraction of population exhibiting place cells and **(b)** decoder accuracy during BMI 1 trials within BMI sessions as a function of experience. p-values are corrected for animal identity. **(c)** A comparison of decoder accuracy during BMI sessions for two groups of animals: A, with extensive experience of BMI sessions, and B, with extensive experience of autopilot-BMI sessions. Group B have more accurate decoders during BMI 1 ($p = 3.6 \times 10^{-4}$, independent single-sided t-test), though decoders still improve in accuracy from BMI 1 to BMI 2 for this group ($p = 2.01 \times 10^{-2}$, independent single-sided t-test). Comparison shown for animals experienced with BMI only: sessions must have taken place at least 10 days after first experience with a BMI trial. **(d)** Decoder accuracy during BMI 1 trials for all animals and all BMI sessions as a function of attentiveness (linear fit, p-values corrected for animal identity). **(e)** As in (d), shown as a function of total ROIs in session.

Bringing all of this together, we suggest that BMI and autopilot representations of the track are distinct from each other, hence the improvement in group B between BMI 1 and BMI 2, but more mutually similar to each other than the locomotion-specific representation of the track. We find our evidence that past experiences shape closed-loop BMI performance particularly compelling. We observe *both* a gradual session-to-session increase in BMI performance in one type of session *and* a difference in BMI performance based on which type of session an animal is habituated to.

We allow ourselves to speculate that the animals in group A, which perform the task both with locomotion and with BMI, must maintain two separate representations of the track simultaneously. In the next chapter, we will explore how the BMI and locomotion representations interact (and potentially interfere) with each other: BMI usage alters the locomotion-based representation of the track, making it difficult for these representations to be entirely orthogonal to each other. We will eventually compare this with the idea, proposed in the BMI work of Losey et al. (2024), of a 'negotiated equilibrium' within populations that need to maintain different representations for different purposes.

3.5 Remarks on alternative experimental protocols

3.5.1 Altering the target trajectory

We described in an earlier section that some animals altered their behaviour during BMI trials, entirely decoupling their wheel manipulation from their velocity of traversal, while others did not. By altering the policy that the BMI applies to decoded states, we can deliberately force such a decoupling by altering the target trajectory. Running these experiments allows us to further verify that good decoder accuracy is possible in spite of the complete removal of the animal's ability to locate itself via path integration.

As an example of such a forced decoupling, we can mirror the target trajectory around the centrepoint of the track (refer back to Fig. 3.4.d for sample trajectories). This not only decouples VR velocity from wheel speed, but may even juxtapose them entirely. We ran only four mirrored BMI sessions—too few for statistically robust conclusions, but enough to satisfy an initial curiosity. The first trial of these sessions is a locomotion trial, and we observe the same phenomenon as in regular BMI sessions: decoder performance is poor when trained on locomotion, but accurate once trained on BMI 1 (Fig. 3.13.a). We find the decoder scores during BMI 2 in mirrored BMI sessions are largely comparable to the scores in BMI 2 in BMI sessions and autopilot-BMI sessions, despite minimal or negative correlations between wheel speed and VR velocity (Fig. 3.13.b).

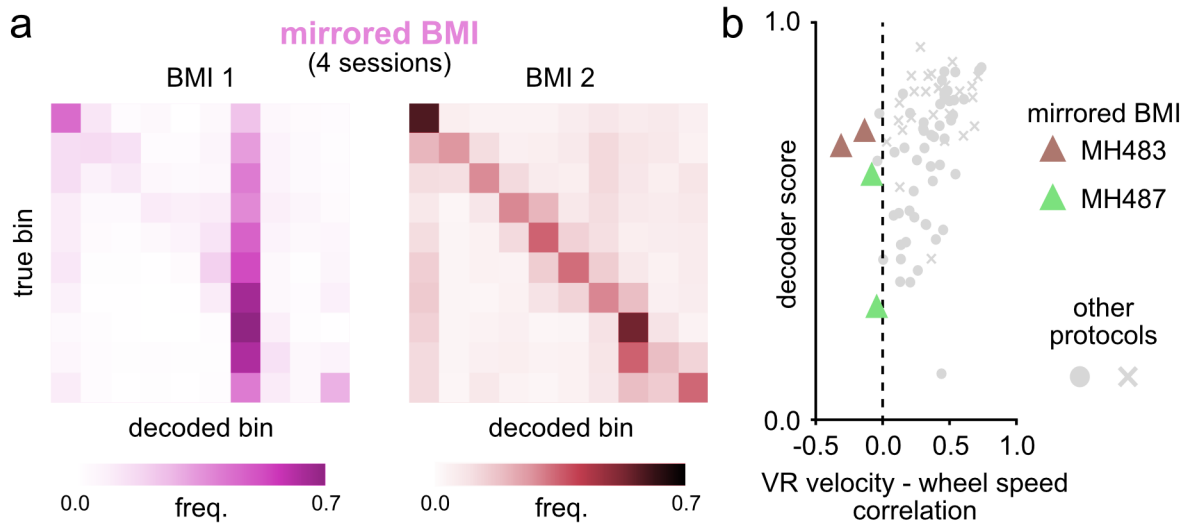


Fig. 3.13 Altering the target velocity profile forces a decoupling between VR velocity and wheel speed. **(a)** Confusion matrices, as in Fig. 3.7, averaged over all mirrored BMI sessions. **(b)** Decoder score as a function of the correlation between wheel speed and VR velocity during the BMI 2 trial of mirrored BMI sessions (coloured) in contrast to other session types in which BMI is used (greys, all animals shown).

3.5.2 Boiling the frog: gradually introducing BMI control

The change between locomotion and BMI contexts is binary and sudden. If the change in control modality took place more gradually, would we see the same jump in hippocampal representation, or would we avoid remapping? Could we incrementally introduce an animal to BMI as an alternative to retraining the decoder to learn a BMI-specific representation?

What we report below is closer to an anecdote than a rigorous investigation, but nevertheless gives some useful intuition for the problem-space. The animal MH187, which had good open-loop decoder accuracy (and was the other animal used in addition to MH188 in the previous chapter for technical validation of the BMI design), adopted a scared and apprehensive posture when we attempted a closed-loop BMI trial—the only animal to do so in our many BMI experiments. We aborted the experiment out of concern for the animal's wellbeing. Subsequently, we devised a 'frog-boiling experiment', in which the VR velocity is a weighted mixture of both running on the wheel and BMI control (Fig. 3.14.a). Over the course of the session, we gradually increased the weighting of the BMI control signal until 100% of the velocity was controlled by BMI (Fig. 3.14.b).

We trained a single decoder on the initial locomotion trial and used it throughout the session. As the weighting of the BMI signal increased, we found that decoder accuracy gradually dropped (Fig. 3.14.c). The natural turnover in representation as a function of elapsed time means that we expect some drop in accuracy from the start of the session to the end of the session, but we can partially account for this by comparing the similarity of the

first locomotion trial to both the final BMI trial and the subsequent final locomotion trial. We perform a crude comparison by training an open-loop decoder on the first locomotion trial and by evaluating it on both the final BMI trial and the final locomotion trial. We find that in each of the three sessions, evaluating the decoder on the final BMI trial (scores: 0.44, 0.50, 0.39) produced lower scores than evaluations on the final locomotion trial (scores in the same session order: 0.46, 0.57, 0.55), indicating a greater representational similarity between the two locomotion trials. This is a preliminary suggestion, far from conclusive evidence, that remapping takes place even if the animal is introduced to the BMI context gradually.

In light of a still-ongoing debate over whether gradually changing visual cues produces gradual changes in hippocampal representation (as in Leutgeb et al., 2005) or an inflection point triggering a sharp change in representation (as in Wills et al., 2005), we are curious to know whether the transition between the locomotion representation and the BMI repre-

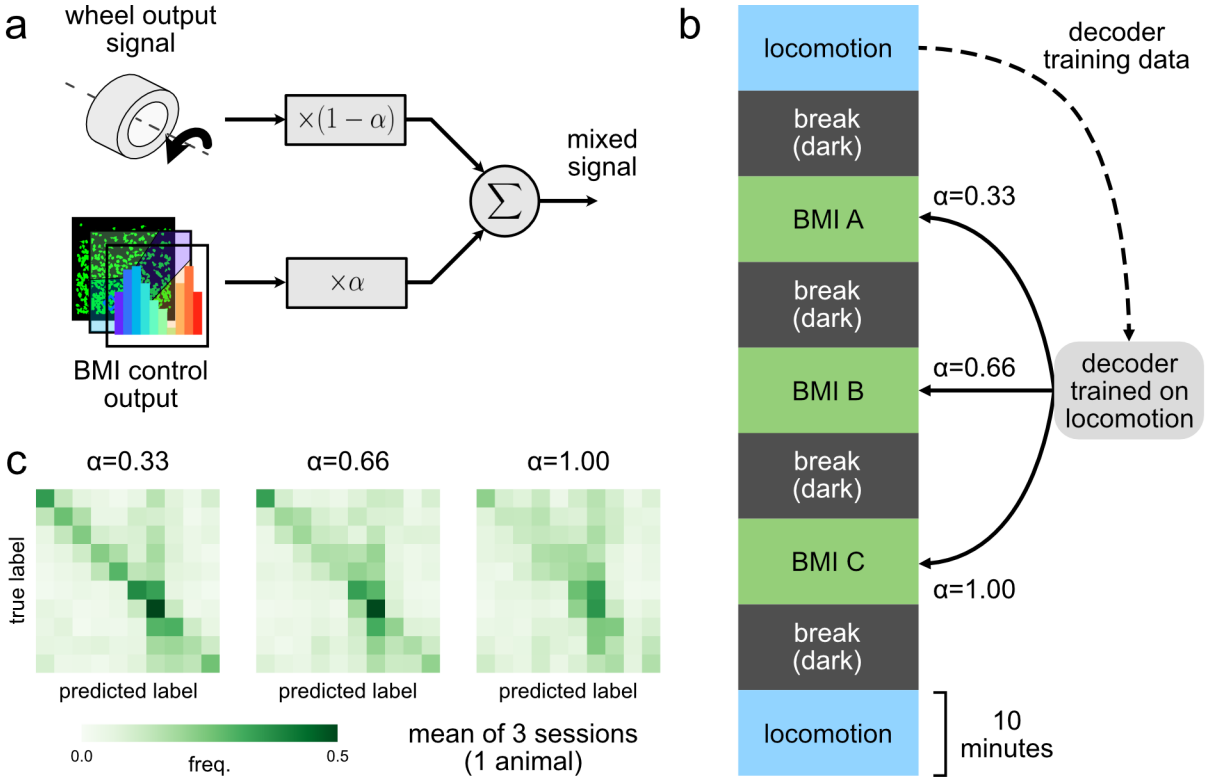


Fig. 3.14 Gradually increasing BMI contribution to VR velocity does not result in good closed-loop task performance. **(a)** A mixed control signal that produces a weighted average of locomotion and BMI control signals, controlled by parameter α to govern the relative weighting of BMI control to locomotion control. **(b)** A session structure consisting of two reference locomotion trials and three BMI trials. A single decoder is trained based on the initial locomotion trial and used for all BMI trials, while α is gradually increased from one BMI trial to the next. **(c)** Confusion matrices for each value of α, as in Fig. 3.7, averaged over 3 sessions for a single animal.

sentation is itself gradual or inflective. This is a question that the data from our preliminary experiments leaves us ill-equipped to answer, but the following chapter will at the very least reveal the existence of representations that appear to blend the BMI and locomotion contexts. This suggests that gradual transitions are possible, as mixed representations can exist, but makes no suggestion as to whether or not they occur in practice.

3.6 Discussion

3.6.1 Open-loop and closed-loop BMI are not equivalent

If there is a single message to take home from this chapter, and perhaps even this thesis, it is this: there is no guarantee that a decoder that is accurate in open-loop evaluations will translate to a BMI that performs well in closed-loop. In this chapter, we provided two contrasting examples of initialising a BMI from non-BMI training data. Decoders trained on locomotion trials perform poorly in closed-loop BMI, while those trained on passive motion trials perform well. In the broader context of BMI design, how can we predict what kinds of training data collection are ill-conditioned for eventual closed-loop control? We propose that the main alternative to simple trial-and-error is a mechanistic consideration of how populations form and maintain stable representations. Had we held more closely to our 'neural sympathy' tenet from the outset of this project, we might indeed have anticipated the initial failure of closed-loop decoding.

In the case of our own experiments, we speculate that the representation in the locomotion context collapses during closed-loop BMI because the change in control modality causes expectations of the representation to be violated. If manipulating the wheel no longer generates the expected forward motion[4], then the cognitive map loses its internal coherence. In the void left behind, the CA1 population forms a new representation that does not depend on relating wheel manipulation to progress along the VR track. On the other hand, when the closed-loop BMI is initialised from a passive motion trial, the representation learned by the BMI has no such prior dependency on actions taken, and so no global remapping of the representation needs to take place. We will elaborate on this hypothesis in the following chapter, with a more detailed consideration of the changes that take place in the CA1 representations.

In addition to fundamental differences in representation, this chapter has shown two further factors relevant to closed-loop BMI performance that open-loop evaluations fail to capture. The existence of such factors is known to the BMI literature (e.g. Chase et al., 2009), but it bears repeating that we can now explore them with 2-photon imaging in mice rather than in electrophysiology with primates. The first is attention: while periods of inattentiveness

[4] Put more formally: if the animal finds itself in some state s_1 and performs an action a_1 which previously caused state transition to s_2, it no longer experiences the state transition to s_2.

will cause degradation in the quality of a representation in a way that is obvious even in open-loop decoding (Pettit et al., 2022), a poorly-constructed decoder can itself lead to inattentiveness and form a vicious cycle. The second is the obvious improvement, from one session to the next, in closed-loop BMI performance. We present this as a gradual enrichment of the BMI-specific representation due to habituation rather than deliberate learning within a BMI-specific context, but this is nevertheless a phenomenon that can only be explored with closed-loop trials.

3.6.2 Remapping during BMI (or lack thereof) in existing work

In Lai et al. (2023), the only other example of population-scale closed-loop BMI based on CA1 hippocampus, there is no retraining of decoder based on activity during BMI trials, and yet the authors demonstrate successful performance of tasks with BMI. There are obvious methodological differences between our work and their experiments: we use mice, they use rats; we use a 1D VR track, they use a 2D VR arena; we motivate with sweetened soy milk, they use water deprivation; we use 10 minutes of training data, they use 30-40 minutes; and so on. While it is possible that some combination of these differences is responsible for the discrepancy in our observations, we suggest that the most relevant factor is the difference in how decoder output is transformed into actions taken. Lai et al. (2023) use a 3-second moving average of decoded position, moving the animal's VR coordinates smoothly and gradually to that location. Recalling that, in our own results, there were many instances of BMI decoders trained on locomotion that still performed better than chance (Fig. 3.8.a), and that the different task in Lai et al. (2023) moves the reward randomly around their arena (which would result in random rather than structured mispredictions), we find it entirely plausible that they were successfully able to decode position adequately for closed-loop control despite remapping associated with a BMI context by averaging over a sufficiently long window. Our work makes it easier to spot such a remapping: while we decode a (discretised) position, we do not transport the animal to the decoded position, and so discrepancies between the ground truth and the decoded position caused by remapping are visible at a glance. We claim, therefore, that our results are entirely compatible with Lai et al. (2023), and that this provides a strong hint that the remapping operation, while potent, must leave some above-chance fragment of the representation intact when moving from locomotion to BMI. This latter should not be surprising: our animals, after all, do not appear lost, and continue to attempt to perform the task as their training dictates.

Sorrell et al. (2024), who describe a PPC-based closed-loop optical BMI, also find no need to retrain their decoder based on a BMI-specific context. While this is an entirely different brain region, and not one prone to remapping in the same context-sensitive manner as CA1, the authors do observe differences between representations of the task during BMI and during

locomotion: disagreements between their BMI controller's output and the animal's intention, as inferred by its manipulation of the ball via which it is accustomed to navigating through their VR maze, are separately encoded in population activity. We find it noteworthy that these PPC experiments also interleave maze traversals of different control modalities, with one BMI traversal for each two performed with locomotion, which may hinder the development of an entirely separate BMI-specific context.

The existence of subspaces of neural activity specific to BMI performance of tasks is not entirely novel (Ganguly et al., 2011; Schroeder et al., 2022), but these observations are based in motor cortex, where the activity is (usually) analysed by some form of dimensionality reduction. The interpretable encoding of place cells, where activity of individual neurons can be tied directly to encoding of the environment, therefore affords us some mechanistic insight into why this takes place—which is exactly what we explore next.

3.7 Methods

3.7.1 Mapping of decoded location to VR velocity

The VR track is divided into 10 spatial bins. Each of these bins has an allocated 'target' speed that, together, create a velocity profile that mimics expert stereotyped performance of the task. This profile was hand-crafted following inspection of the trajectories of an initial cohort of mice, and used as the target trajectory for all BMI and autopilot-BMI trials. The profile is visible as a bolded line in Fig. 3.2.b, Fig. 3.2.c, and Fig. 3.4.b-c, and is explicitly shown with respect to the decoder bins in Fig. 3.1.b. The highest velocity among all bins is 40 cm/s and the lowest is 4 cm/s, with this latter set to a non-zero value to prevent the animal becoming stuck at the reward site.

The BMI maps neural activity onto a forward velocity by first decoding the identity of the most likely spatial bin given a single imaging frame and then looking up the velocity from the target trajectory corresponding to this bin. For example: if the decoder decodes the 3$^{\text{rd}}$ bin on the track, corresponding to the region between 80 cm to 120 cm, then the BMI will apply a forward velocity through the VR track of 40 cm/s.

The decoding takes place at 30 Hz, allowing for velocity changes at this same frequency. We enforce a maximum acceleration between frames in order to avoid jittery, jumpy motion (which risks frightening or disorientating the animals). This maximum acceleration is set to 80 cm s^{-2}, approximately twice the measured maximum acceleration of the animals running on the wheel.

Random speed autopilot trials

The random speed autopilot trials are used by the cohort of animals consisting of MH533, MH535, MH536, MH537, and MH538. The trajectories that are displayed for these animals are generated by taking the traversals produced during BMI 1 (i.e. poor decoder performance) from the earlier cohort of animals (MH356, MH462, MH483, MH485, and MH487) and temporally shifting these traversals by a random offset (between 10% and 90% of the trial duration, uniform sampling). This ensures that the distribution of forward velocities and the number of laps experienced by the animals during autopilot trials matches the distribution of velocities from typical BMI 1 trials. No animal was shown the same 'replay' more than once, preventing the animals from learning any temporal structure of the passive motion.

3.7.2 Behavioural metrics

Visualisation of behaviour traces

When showing multiple laps on the same axes, as in Fig. 3.2, Fig. 3.4, and Fig. 3.5, the wheel velocity and VR velocity traces are smoothed with a boxcar filter (zero-padded between laps, 500 ms window) for ease of viewing. This smoothing is not applied when the laps are visible separately, so unsmoothed laps are visible in Fig. 3.9.

VR velocity - wheel speed correlation

We compute the Pearson correlation coefficient between the VR velocity and the locomotion wheel speed. We exclude from this calculation the periods of time between laps when the screen is in darkness, even though the BMI signal is still produced and the animal is free to manipulate the wheel between laps.

Pre-licking score

We categorise each licking event, based on its track position in cm, into one of three categories: anticipatory, post-reward, and erroneous.

$$\text{lick type} = \begin{cases} \text{post-reward} & \text{if } 250 \leq \text{position} < 300 \\ \text{anticipatory} & \text{if } 200 \leq \text{position} < 250 \\ \text{erroneous} & \text{otherwise} \end{cases}$$

For each lap of the track, we calculate the fraction of anticipatory licking events among both anticipatory and erroneous licking events.

$$\text{lap pre-licking score} = \frac{\#\text{anticipatory}}{\#\text{anticipatory} + \#\text{erroneous}}$$

The 'pre-licking score' for a trial is the mean of this fraction over all laps.

Attentiveness metric

We compute the z-score of the VR velocity - wheel speed correlation over all considered trials. We also compute the z-score of the pre-licking score over the same trials. The attentiveness of metric for a trial is the larger of these two scores for that trial.

3.7.3 Closed-loop decoder evaluation

Confusion matrices

We use confusion matrices as a tool to visualise the accuracy of a decoder. We evaluate them over the course of an entire trial by comparing the output of the decoder, corresponding to a decoded spatial bin, to the true spatial bin at each imaging frame. Each row represents occurrences of the ground truth (that is, instances of the animal being inside of each spatial bin) and each column represents occurrences of a particular decoder outputs (that is, instances of the decoder classifying neural activity as belonging to each spatial bin). Values in individual cells represent relative rates of occurrence. We normalise the confusion matrices by the ground truth, which enforces that the sum of each row within the matrix is unity. For example, the cell in the 3^{rd} row and the 5^{th} column corresponds to the fraction of times the decoder classified the neural activity as belonging to the 5^{th} spatial bin when the animal was located in the 3^{rd} bin. A perfectly accurate decoder would have non-zero cells along only the main diagonal, while a decoder trained on random noise would generate a roughly uniform matrix.

Decoder scores

We score how well each spatial bin is decoded from 0.0 to 1.0. This value represents the fraction of samples that are classified correctly when the animal's track position is within that bin, where 'correctness' requires decoding either the exact current bin or one of its immediate neighbours (N.B. the initial and final bin are not considered mutual neighbours, despite the circular geometry of the task, due to the inter-lap interval spent in darkness). The overall decoder score is the median of the bin scores. A convenient mental shortcut (though not

strictly an equivalence) for the decoder score is a measure of how well the confusion matrix conforms to the shape of a tridiagonal matrix.

This scoring strategy differs slightly from the previous chapter, where the open-loop decoder score was a mean over all samples, corresponding to the probability that a random sample from the evaluation dataset would be correctly classified to within one neighbouring bin. The change from mean over samples to median over bins addresses an issue that arises only in closed-loop: the decoder output controls which spatial bins are sampled, which can create a feedback loop resulting in over-sampling or under-sampling specific bins. For example, a consistent decoding error in one spatial bin can cause an animal to spend a disproportionate amount of time in that bin, which results in over-sampling of poorly decoded regions of the track. Conversely, if the decoder is accurate in a bin that maps to a fast velocity, the animal spends less time in the well-decoded bin, which results in under-sampling of well-decoded regions of the track.

Random decoder input validation

We probe the response of a decoder to random noise as a means of uncovering structural bias towards decoding specific bins. To perform this, we take a trained decoder and the closed-loop BMI trial in which it was used, and retroactively replace each pixel of the input—the feature space as derived from the imaging FoV—with Gaussian noise ($\mu = 0.0$, $\sigma = 1.0$), resampled on every frame. We then evaluate the open-loop accuracy of this BMI by attempting to decode from this random input, and use its classification attempts to produce a confusion matrix.

3.7.4 Tests of statistical significance and effect size

All independent t-tests and paired t-tests use reference implementations from the `scipy.stats` library (Virtanen et al., 2020). Significance (p-values) for fixed effect parameters in linear models that account for animal identity, session type, or both animal identity and session type simultaneously use reference implementations of mixed linear models from the `statsmodels` library (Seabold and Perktold, 2010), which implement the solver proposed in Lindstrom et al. (1988) and use the Wald z-test as the underlying significance test, unless otherwise stated. We visualise linear fits to these same datasets with Ordinary Least Squares (OLS) fits, which do *not* account for animal identity or session type. Quoted R-values are based on the OLS fits. In the one instance we perform outlier removal (in Fig. 3.9), we perform an initial OLS fit and then remove all samples with residuals above the 95^{th} percentile.

Chapter 4

Changes in hippocampal activity induced by BMI usage

Chapter overview

This chapter describes neuron-level and population-level changes in activity that take place when an animal changes between controlling its navigation via running on a wheel and navigating via closed-loop BMI. The main experimental observations described in this chapter are that:

- There are distinct representations of the environment depending on whether or not locomotion is used to control navigation
- Non-locomotion representations of the environment exhibit a greater reliance on changes in visual cues
- Motion through the environment that is unpredictable for the animal results in severely degraded representations
- BMI usage causes components of the 'BMI context' to linger in the representation when an animal is returned to locomotion-based control, resulting in a superposition of multiple representations
- Control experiments that use passive motion do have a representation of the track that is distinct from the locomotion context, but they do not induce this superposition effect
- The distinct representations associated with control modality persist between sessions spaced several days apart
- High weights allocated to neurons by the decoder are correlated with the superposition effect both within sessions and across sessions

4.1 Introduction

4.1.1 Representations sensitive to modes of travel

In the previous chapter, we described decoders with high accuracy in open-loop that performed poorly during closed-loop BMI. We attributed this to a fundamental difference in how the hippocampus represents the same environment under locomotion and BMI conditions. This chapter deals with explaining and quantifying these changes in representation.

In our VR environment, the animal has no access to useful vestibular, auditory, or olfactory cues to help it navigate. The only useful sources of information are visual landmarks (the cues on the walls) and some idiothetic cues (primarily manipulation of the wheel, but also optical flow). While visual cues can be useful to determine location in isolation, path integration is only informative relative to some existing and previously known location in space. Our crude mental model of neural activity in CA1 therefore incorporates a feedback loop: a recent location is combined with path integration signals to produce a predicted location, which is then reconciled with incoming visual cues to determine the current location (Fig. 4.1.a). We take running on the wheel, which is a navigationally relevant action taken by the animal to traverse the environment, to be the signal of foremost importance to path integration.

The feedback loop is useful for sustaining a coherent representation of space that is robust to minor changes within an environment. Typically, small localised changes to the sensory experience within an environment cause only partial remapping in the vicinity of those changes, as path integration can partially compensate for missing sensory inputs—indeed, many place cells even persist when the lights are turned off in an environment (Markus et al., 1994; Quirk et al., 1990) or when cues are removed (O'Keefe and Speakman, 1987). For example, if we were to remove all visual cues in a small segment of our VR track, we would expect only minor changes in firing rate of neurons in the vicinity of the cue disruption[1]. However, during BMI, manipulation of the wheel stops being a useful predictor of future states. Not only is this disruption not spatially localised, but we speculate that it leads to a cascading failure via the feedback loop: a poorly predicted future state leads to subsequently worse prediction errors, and the representation rapidly collapses (Fig. 4.1.b).

Despite this collapse, we achieved good closed-loop performance when we retrained the decoder, even with VR traversal velocity and wheel manipulation speed decoupled from each other. This implies the existence of a second representation, distinct from the locomotion-based representation of the track, constructed in a context in which path integration is an unreliable signal, and this is indeed what we find: cells firing in new but consistent locations (Fig. 4.1.c).

[1] This has been experimentally validated to apply on the very VR track used for our BMI experiments as part of ongoing (but currently unpublished) work by Hinze Ho, Kou Huang, and Julija Krupic. The exact effect is more nuanced, with animals learning to rely more on path integration with experience.

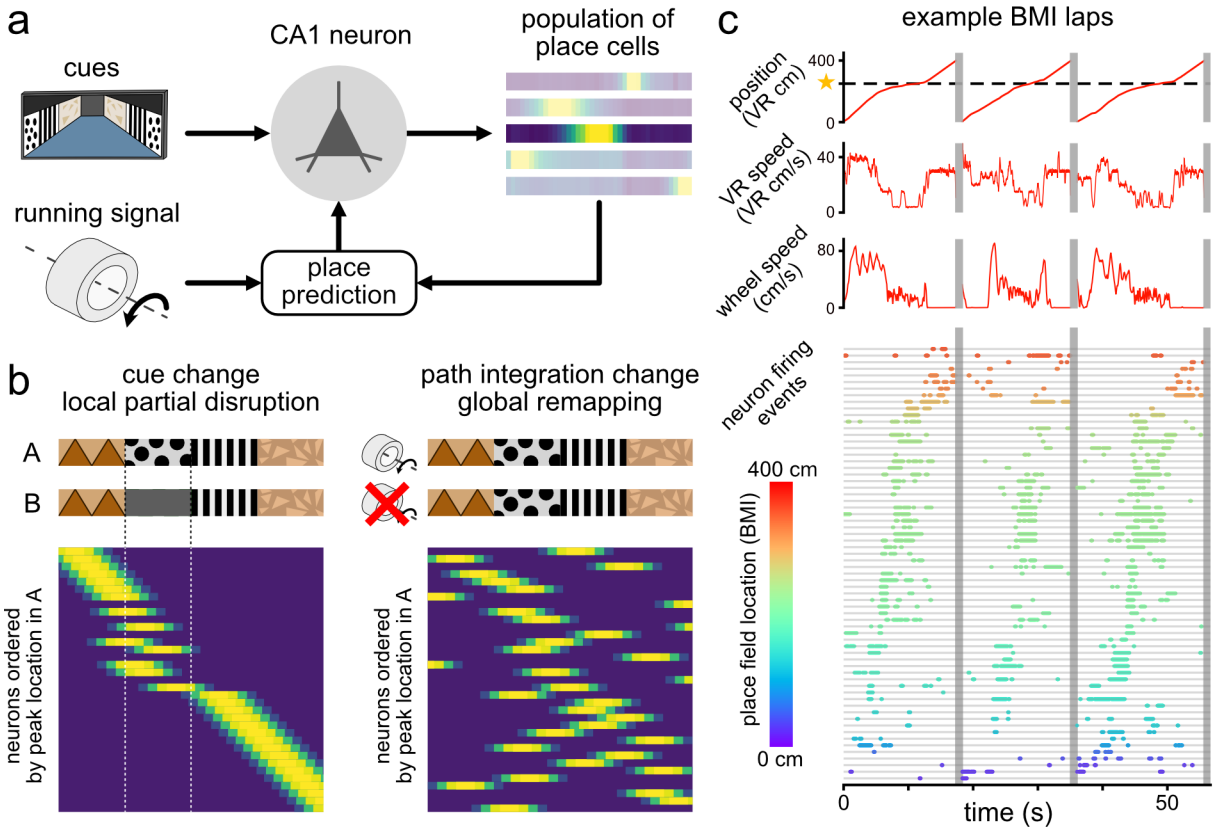


Fig. 4.1 Neural representations of environments can be stable despite uninformative path integration signals. **(a)** Individual CA1 neurons receive signals related to external sensory input (visual cues) and conjunctions of actions taken to navigate through the environment (manipulation of a wheel) with the output of other CA1 neurons. This generates representations of environments that are robust to local disruptions of visual cues. If the path integration signal becomes misleading the compounding errors from the feedback loop risk collapse of the representation. **(b)** A cartoon of population ratemaps with neurons ordered by their peak firing location during some initial condition A and showing activity during some probe condition B. The first probe condition shows a localised cue change, which results in a local and partial disruption of the representation of the track. The second probe condition shows a decoupling between locomotion and travel through the environment, which results in a global remapping affecting the entirety of the track. **(c)** Three example traversals of the track during a BMI 2 trial of a BMI session, showing decoupling of the wheel speed (a path integration signal) and the VR velocity. The lower plot shows neuron firing events, with each row corresponding to an individual neuron (only neurons with place fields during this trial shown, rows ordered by place field location). Despite the decoupling, the neural population successfully establishes cells that consistently fire within specific regions of the track.

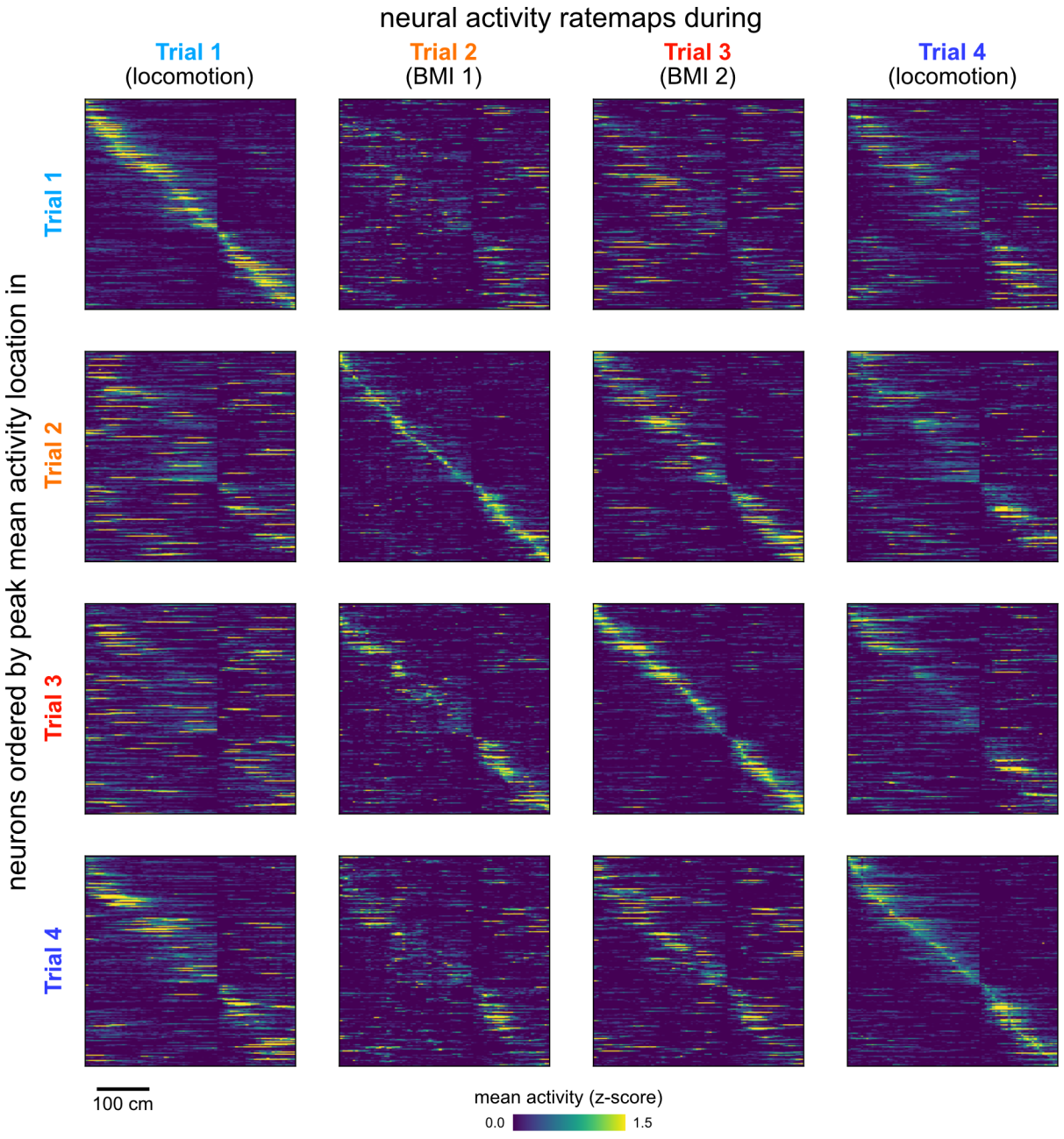


Fig. 4.2 A population of neurons takes on a superposed representation of the environment after BMI usage. Population ratemaps for each of the trials within an example BMI session (352 neurons). Neural activity within each trial is shown by each column of the matrix, while each row orders the neurons according to the location on the track with the largest ratemap value for each trial in turn.

4.1.2 Does BMI usage induce persistent changes in representation?

So far, all the points in our hypothesis apply to passive motion (the autopilot condition) just as well as to BMI control of travel. We also described in the previous chapter that BMI decoders trained on passive motion performed well without the need to retrain, suggesting similar representations of the track. The difference between BMI and passive motion is the causal control of navigation by the animal's neural activity. While the path integration signal becomes unreliable in both cases, the animal retains some degree of agency during BMI: if the representation in the imaged population is held stable, then the animal produces stereotyped trajectories that lead to predictable experiences and rewards. We have already seen that the animal's attention or engagement with the task is a critical component of successful BMI performance. In contrast, autopilot trials demand no attention or engagement from the animal. In this chapter, we will find that the change between the locomotion condition and the passive motion condition produces purely reversible changes in population activity: when the animal is returned to the locomotion condition, its representation reverts to the locomotion-based representation. The BMI representation is 'stickier': we will find instead that a return to locomotion from the BMI condition preserves features of both the original locomotion representation and the BMI representation, resulting in neural activity that appears to be a combination of the two contexts. We call this 'superposition' of the representations.

As a preview of the phenomenon, which we quantify later in this chapter, we show example population ratemaps from a single BMI session in Fig. 4.2. It is immediately obvious that the representation in Trial 1 (the initial locomotion trial) is dissimilar to Trials 2 and 3 (the BMI trials) and similar to Trial 4 (the final locomotion trial). Meanwhile, Trials 2 and 3 have mutually similar representations: the BMI context. What we find surprising is the representation in Trial 4, which bears a resemblance to *both* the initial locomotion representation of the track *and* the BMI representation. Inspecting sample traces from the same session (Fig. 4.3), we find individual cells with place fields that are obviously dependent on how the environment is traversed. Other neurons maintain stable fields. However, some aspect of the superposition of representations appears visible even at the level of individual neurons, with certain place cells responding at locations corresponding to place fields from both contexts[2]. The visible distinction between 'before BMI' and 'after BMI' suggests a change in the existing representation—the result of an interaction between BMI and plasticity, rather than merely a response to changed inputs to the system.

[2]For these illustrative examples of superposition, we have deliberately selected responses where the fields in the BMI context and original locomotion context take place either side of a fixed element of task structure (reward delivery and inter-lap darkness, respectively) to avoid confounds associated with varying travel speeds.

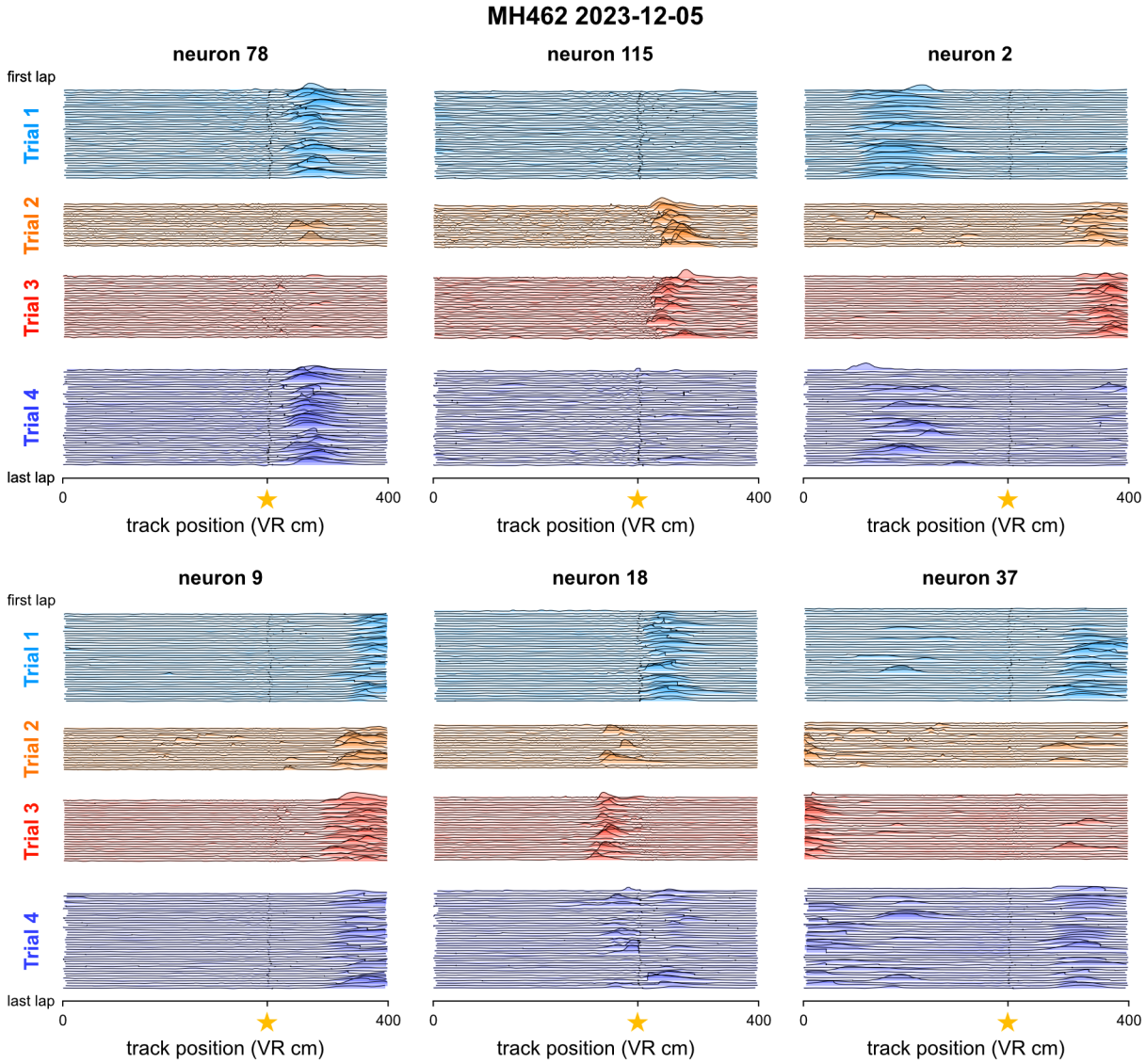


Fig. 4.3 Sample neuron responses during a BMI session. The dF/F traces of six example neurons, all from the same BMI session as Fig. 4.2, selected to showcase a variety of responses induced by the BMI condition. The traces are shown lapwise from top to bottom as a function of position on the VR track and are smoothed with a low-pass filter (2 Hz cut-off) for ease of viewing. Some neurons form fields in locations that are specific to BMI or to locomotion, while others are stable. Superposition at the level of individual neurons is shown by neurons 18 and 37: within the final locomotion trial, they fire both in the place fields from the first locomotion trial and the place fields from the BMI trials.

4.1.3 Background: interactions between cues and path integration

Closed-loop BMI presents a unique experimental condition for the manipulation of representations in CA1. Before proceeding to our own results, we outline here some existing, well-known results about the interaction between path integration and visual cues to help distinguish which of our own observations are novel and unique to BMI usage.

Early disorientation studies, in which animals are subjected to vestibular disruption in darkness prior to introduction to an environment, describe an increased influence of visual cues on activity in CA1 for disorientated animals (Knierim et al., 1995). This suggests an ability to actively compensate for disruption to path integration signals. Equally, the hippocampus appears able to learn how to distinguish which visual cues in a scene are reliable landmarks from those that are mobile or transient (Chakraborty et al., 2004; Jeffery and O'Keefe, 1999)—but can it do the same for path integration signals?

Because path integration necessarily deals with a change in location within an environment, we must make a distinction between signals relating to path integration becoming uninformative (e.g. vestibular cues in VR) and the underlying movement itself becoming unpredictable (e.g. randomly timed displacements). When animals are fully in control of their own motion, their navigation actions both lead to expected outcomes and are represented in path integration signals. When animals are moved passively, either on virtual tracks (Chen et al., 2013) or in real environments through use of physical carts (Terrazas et al., 2005), integration of a running signal is no longer informative, but the sequence of experiences may remain predictable if the passive motion is consistent. This is the case in our constant speed and parabolic speed autopilot trials: every traversal of the track is identical, and so it is easy to predict the change from one state to the next. In contrast, because the forward velocity is randomly selected at each instant during our random speed autopilot trials, and therefore because no two laps are the same, the sequence of experiences on the track is unpredictable.

Existing work has already established that moving from active to passive navigation can, for a sizeable subset of place cells in a population, extinguish or displace place fields that are strongly tied to visual cues (Chen et al., 2013; Terrazas et al., 2005). Extinctions or displacements in these experiments appear reversible: returning to active navigation restores the place fields to their previous locations. Representations under active navigation appear relatively robust to changes in 'gain' (Chen et al., 2013; Jayakumar et al., 2019), which controls how far any given stride transports the animal through the environment. Under passive motion, populations can maintain coherent representations despite changes in the gain of optical flow (Madhav et al., 2024). We interpret this flexibility in the path integration system as a suggestion that representations in CA1 attempt to encode the abstract intention to move, rather than a hardwired, numerically precise step-counter.

BMI navigation fits into this picture as a mode of navigation that obviates the running path integration signal and yet retains a requirement to cognitively participate in the task. While it would be entirely ridiculous to suggest that mice understand they are navigating by maintaining a particular neural state, the structure of the BMI task results in a neurally self-fulfilling prophecy: when the representation behaves consistently, the trajectories through the environment become more predictable, which we hypothesise results in a reinforced representation.

4.2 Quantifying place cell properties

4.2.1 Properties of individual neurons

We begin with an overview of how we quantify the activity of individual neurons within imaging sessions. The neuron in Fig. 4.4 displays a cut-and-dry example of a reversible modality-specific response: it has a place field at one location on the track during locomotion trials, and at an entirely different location during BMI trials.

The dF/F traces (Fig. 4.4.a) can be transformed either into neuron events, a binary view of time periods in which the neuron is 'on' or 'off' (as shown earlier for an entire population of place cells in Fig. 4.1.c), or into ratemaps which describe the regions of the track in which the neuron is active (Fig. 4.4.b). We combine these two transformations, first using the ratemap to identify candidate field regions and then using event frequency within those regions as a measure of consistency, to automatically derive place field locations and dimensions within each trial (Fig. 4.4.c, see Methods 4.8.1). We use an overlap metric to identify which fields are shared between pairs (example matching fields indicated in Fig. 4.4.c, see Methods 4.8.1). Finally, we have at our disposal the inferred decoder weights allocated to this neuron by the decoder training process (recalling Methods 2.6.5). Unsurprisingly, this particular neuron is allocated large weights for the region in which its training data shows a place field, and is zero elsewhere (Fig. 4.4.d-e)—a neat illustration of why decoder retraining is necessary in BMI sessions. These measures, place field spans, decoder weights, and ratemaps, are the basic per-neuron properties we make use of in this chapter[3].

[3]Readers who are veterans of the hippocampal literature will note the absence of a frequently used spatial mutual information metric originating from Skaggs et al. (1992). This was originally designed for electrode-based recordings that resolve individual spikes. Adaptations of this metric for Calcium imaging exist (e.g. as used in Sheintuch et al., 2020). We choose to omit this metric on the basis that it is extremely sensitive to choice of quantisation, can spuriously mark randomly firing neurons as highly informative, and would require high confidence in spike deconvolution from Calcium traces to provide a trustworthy bit/spike measure.

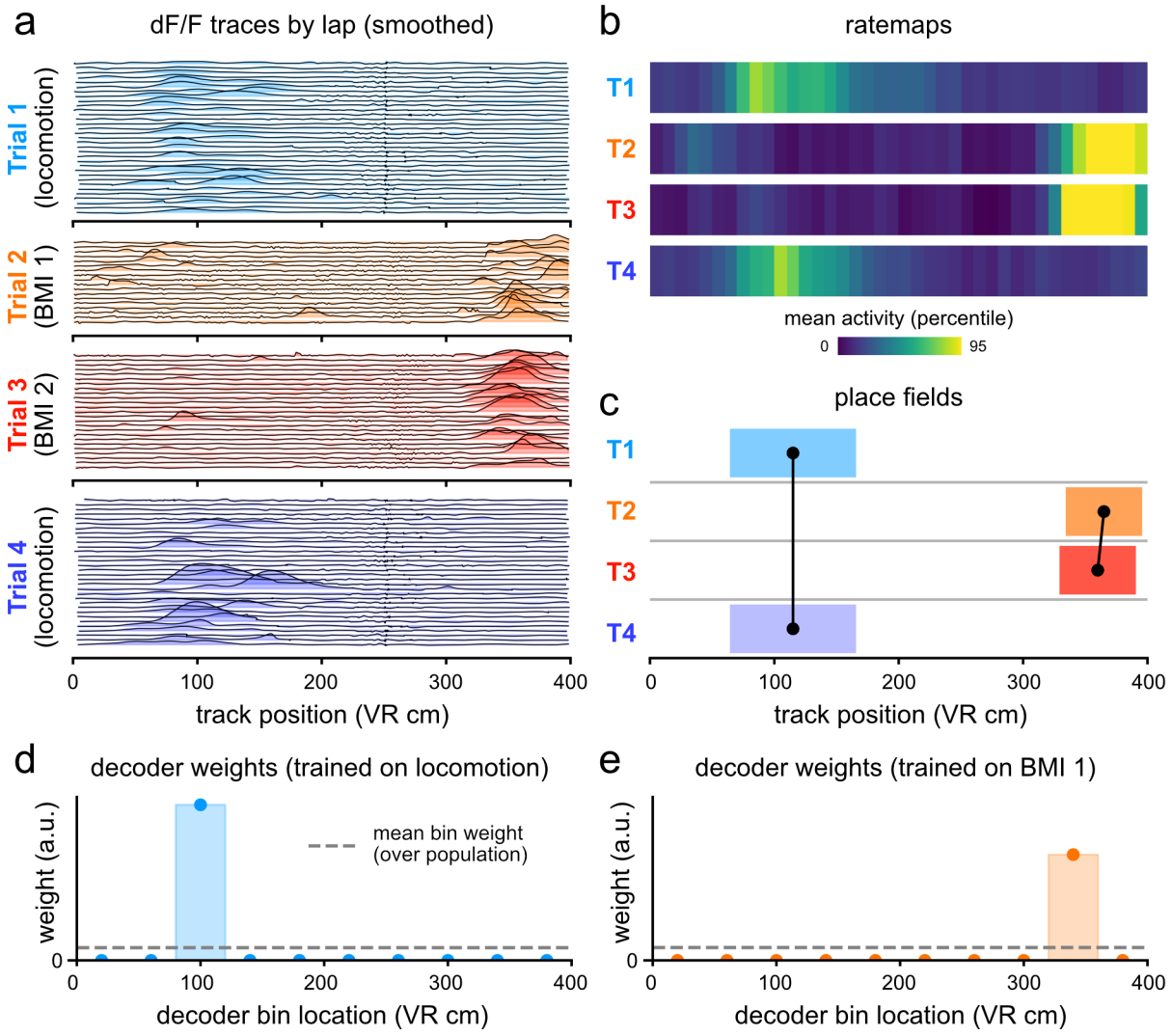


Fig. 4.4 Overview of an example neuron's activity during a BMI session. **(a)** dF/F traces for each lap of the track (top: first lap, bottom: last lap), smoothed with a low-pass filter (cut-off frequency 2 Hz) for ease of viewing. **(b)** Ratemaps of activity for each trial in the session, discretised to 5 cm bins. **(c)** Automatically identified place fields in each trial. Links show place fields automatically identified as matching between trials. **(d, e)** The decoder weights associated with this neuron for decoding each of the 10 spatial bins used for the closed-loop BMI trials.

4.2.2 Place cell frequency across various control modalities

We find that approximately one third of the neurons in our imaged populations have place fields during locomotion trials (Fig. 4.5.a). This is in agreement with existing literature, with the caveats that the fraction is sensitive to the choice of place field identification method (Grijseels et al., 2021) and that individual animals do exhibit mean fractions greater or lower than this value. We ascribe this animal-to-animal variance to minor differences in the exact location of our imaging site within CA1 for each animal and to the behavioural personalities of the individual mice.

In the last chapter, we saw how the fraction of the population that are place cells[4] increased with experience during BMI trials. Here, we show that the place cell fraction also changes from trial to trial within a session. We show these fractions for random speed sessions in Fig. 4.5.b. In the probe trials of these sessions, in which progress along the VR track is unpredictable from the point of view of the animal, we find a steep fall-off in the number of neurons participating in representations in the environment.

To account for the heterogeneity from animal to animal in the baseline fraction of the population that are place cells, we summarise changes in place cell counts relative to the session mean rather than with an absolute value (Fig. 4.5.c-h). The representation of the track is relatively poor during random speed trials, as visible both during the random speed sessions, in which these trials are probes, and the autopilot-BMI sessions, in which these trials are used to train the initial BMI decoder.

Additional place cells emerge during the BMI trials of autopilot-BMI sessions (Fig. 4.5.e). This is a change in representation, but the continued high accuracy of the decoder during closed-loop trials suggests that this takes place largely within the decoder's null-space (i.e. neurons allocated zero or low weights by decoder training can form place fields without interfering with the decoder readout). We speculate that this representational enrichment takes place because the BMI laps of the track are more predictable than the random speed trials, rather than some BMI-specific conditioning effect. For comparison: constant speed and parabolic autopilot trials do not experience a fall-off in place cell count (Fig. 4.5.g-h), as these provide more predictable experiences than the random speed condition (Fig. 4.5.f).

As a methodological aside, we point out a small yet consistent dip in the place cell count that takes place in the final trial of sessions, as visible in locomotion sessions, BMI sessions, autopilot-BMI sessions, and parabolic autopilot sessions. We found no evidence that this was caused by imaging artefacts, such as indicator exhaustion, nor can we explain it in simple behavioural terms, such as a reduced number of laps run in the final trial[5]. This dip is not critical to any of the arguments we make in the following sections, but as we begin to point

[4] We use 'place cell' within the context of a single trial within a session as a shorthand for 'neuron exhibiting at least one place field during this trial'.

[5] This dip is also present in our collaborators' electrophysiology recordings and non-VR experiments.

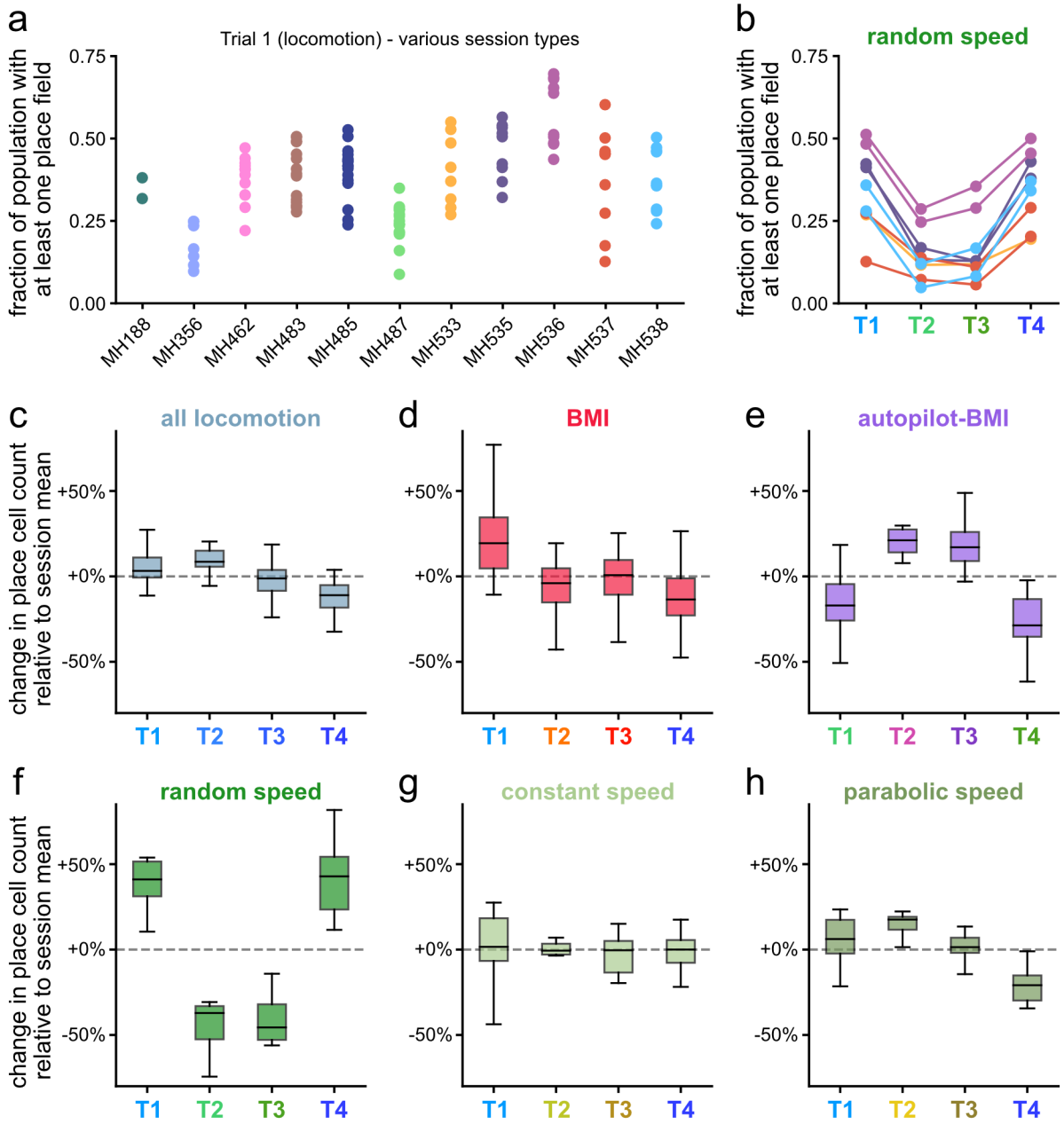


Fig. 4.5 Fractions of neural populations identified as place cells across animals and trial types. **(a)** The fraction of cells in the population with at least one place field during initial locomotion trials. Shown for all animals and all sessions in which Trial 1 used locomotion to navigate the track. **(b)** The same quantity shown for each trial within each random speed autopilot session. Colour of trace corresponds to animal identity from (a). **(c-h)** The fraction of place cells for each trial relative to the mean fraction of place cells in the same recording session, shown for a variety of session types. Box-plots indicate median, interquartile range, and quartiles +/- 1.5 × interquartile range to the nearest datapoint. Fliers omitted.

out ways in which the final trial in sessions experiences superposed representations of early trials following BMI usage, we stress that such superpositions cannot be accounted for merely by an increase of the number of place cells active in the final trial, as the number of these cells slightly *decreases* in practice.

Finally, we consider the second trial during BMI sessions (shown within Fig. 4.5.d). Despite poor BMI performance, these trials do not exhibit a drop-off in place cell count on the same scale as the random speed trials. We recall that the forward velocity distributions experienced in random speed autopilot trials explicitly emulate those experienced during BMI 1 trials (see Methods 3.7.1). While the velocities during BMI 1 are not entirely random—indeed, we have previously seen that there are even a handful of examples of passable decoder scores—we find it intriguing that the place field numbers do not fall off as they do in random speed trials. We considered this same place field fraction during BMI 1 in the previous chapter (see Fig. 3.12), and observed that it increased over time alongside improvements in BMI 2 task performance. We take this as a preliminary hint that representations of the BMI context may be somehow 'stickier' than purely passive motion contexts.

4.3 Increased importance of visual cues

Place fields do not cover the length of the track uniformly. We consider first the distribution of place fields during sessions consisting entirely of locomotion trials (Fig. 4.6.a). The distribution does not change from trial to trial during this type of session. The reward site is particularly salient, as we established in the previous chapter, as is the start of the track and the region corresponding to immediate anticipation of the reward site.

These locations remain salient in BMI sessions during BMI 1 and BMI 2 trials (Fig. 4.6.b). However, during BMI we notice what appears to be small bump in the distribution of place fields starting at the 100 cm position on the track. This bump is also present during the BMI trials of autopilot-BMI sessions, and is even more pronounced during the autopilot trials from those same sessions (Fig. 4.6.c).

The 100 cm position on the track coincides with a change in the visual cues on the walls (Fig. 4.6.d). When we consider each of three types of trial based on control modality—locomotion, BMI, and autopilot—and aggregate these across all sessions in which they take place, we find that the place field coverage (expressed as a density, see Methods 4.8.2) is demonstrably greater during BMI and autopilot as compared to locomotion ($p = 2.2 \times 10^{-9}$ and $p = 8.4 \times 10^{-11}$ respectively, comparison of density at the binned location 20-25 cm onwards from the cue transition, two-sided independent Welch t-test over per-trial densities), with the density following the cue during autopilot trials being slightly higher during the BMI trials ($p = 4.1 \times 10^{-2}$, same measure). We take the anchoring to cues in the trials performed without locomotion as tentative evidence that representations of the track that lack a useful

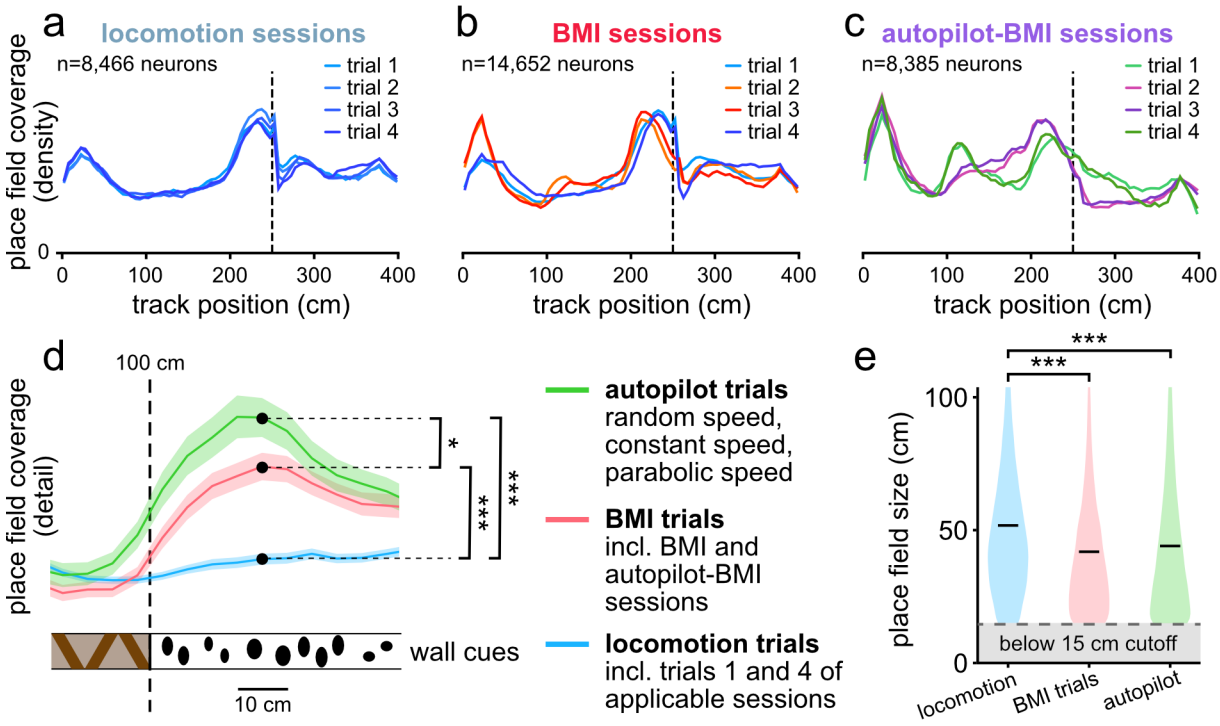


Fig. 4.6 BMI and autopilot representations anchor themselves to visual cues. Relative density of place field coverage as a function of track location, shown for **(a)** locomotion sessions, **(b)** BMI sessions, and **(c)** autopilot sessions. **(d)** A detail of the region of the track in the vicinity of the 100 cm mark, a point of transition between the visual wall cues. Place field coverage traces shown aggregated over trials of the same locomotion type, with confidence intervals showing +/- SEM (estimated via bootstrap sampling of trials, 1,000 shuffles). **(e)** Distribution of place field sizes for each control modality. Ticks indicate mean place field size. Our automatic identification of place field prohibits fields ≤ 15 cm.

path integration signal come to rely more heavily on environmental cues for the formation of a cohesive representation of space, which aligns with existing results (Jeffery and O'Keefe, 1999; Knierim et al., 1995).

Place fields appear larger in locomotion trials than in BMI and autopilot trials (Fig. 4.6.e, $p = 1.8 \times 10^{-40}$ and $p = 1.7 \times 10^{-15}$ respectively, two-sided independent t-test over trials of mean place field size within trial). While it is tempting to claim that this too is somehow related to the absence of path integration, we explain this as a product of a more tame experimental reality: animals traverse the environment faster with locomotion, both in terms of mean lap time and maximum speed, than the autopilot or BMI trials. This results both in a predisposition to larger fields[6] and a smearing of the decaying Calcium transients following

[6]Firing rate within a field is widely agreed to increase with running speed (Geisler et al., 2007; Góis and Tort, 2018; McNaughton et al., 1983). Thresholding in place field identification algorithms may therefore result in apparently larger reported fields. Studies with methodologies that attempt to correct for this explicitly report minimal impact of running speed on field size (Chen et al., 2018), or even field shrinkage at high running speeds (Maurer et al., 2012).

spikes over a greater distance, which together contribute to place fields that are measured to be on average ~20% larger. Discrepancies in velocities between traversal conditions are difficult to eliminate entirely. Terrazas et al. (2005) have previously commented on the near-inevitability of this distortion. In their work, the autopilot-equivalent condition has slightly faster traversals than locomotion, and consequently is the condition with enlarged place fields. We make no specific claims regarding the modulation of place cell sizes by control modality, BMI or otherwise, and intend only to raise the difference in sizes as an analytical caveat.

4.4 Superposition of representations following BMI usage

4.4.1 Representational similarity viewed at population-level

Population vector correlations

We use the mean population vector correlation (PVC) between pairs of trials as a measure of their representational similarity (see Methods 4.8.2). For example, in sessions in which all trials are locomotion-based (Fig. 4.7.a-b), we find that there is a gradual turnover in the representation: temporally adjacent trials (T1 and T2, T2 and T3, T3 and T4) are relatively similar, while trials further apart (e.g. T1 and T4) are less similar. We can perform pairwise comparisons of the PVCs for pairs of trials, allowing statements such as “T1 is more similar to T2 than T2 is to T4" or “T1 is no more similar to T2 than T3 is to T4" (see Methods 4.8.4 for a full description of statistical methodology, including multiple comparisons correction and correction for animal identity). We now systematically step through each of the session types, commenting both on whether the probe trials appear to present a different representation from the reference trials, and on whether the final trial exhibits superposition from the previous trials.

BMI sessions

In BMI sessions (Fig. 4.7.c-d), pairs of representations sharing a control modality (T1-T4 and T2-T3) display high similarity, while pairs of representations of mismatched control modalities (e.g. T1-T2) display relatively low similarity. The exception is the T3-T4 pairing, which exhibits higher similarity than the T1-T2 pairing ($p = 0.0500$ after rounding to 3 significant figures; Bayes Factor 14.7)[7]. This represents a quantifiable impact of the superposition effect across BMI sessions: components of the representation from T3 (BMI 2) remain in T4 (locomotion).

[7] A value this close to the accepted standard for significance warrants comment. We assure the reader that this is the result of an honest effort to handle the statistics rigorously and to correct conservatively for animal identities and multiple comparisons. If we were to be unscrupulous and apply a single-sided paired t-test for whether T3-T4 > T1-T2 with these corrections removed, the p-value would appear as low as 8.88×10^{-6}.

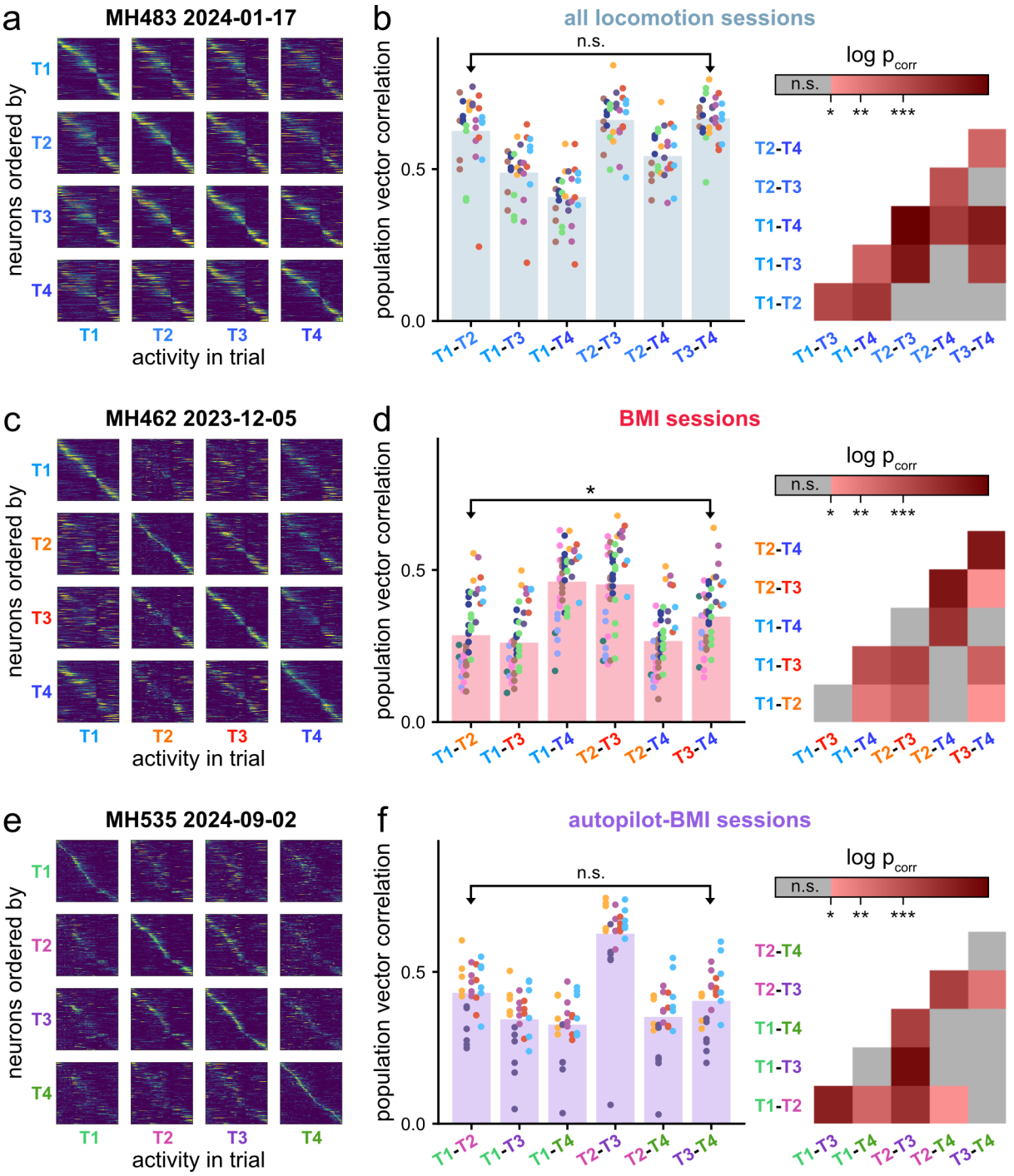


Fig. 4.7 Locomotion-based representations of environments exhibit a clear difference in representation before and after BMI usage. **(a)** A population ratemap for a typical 'all locomotion' session. **(b)** The mean population vector correlation between each pair of trials. Each point represents a single session (colour indicates animal identity), shown for all 'all locomotion' sessions. The right matrix compares the PVCs between each pair of trials to the PVCs for each other pair of trials for statistical significance, correcting for animal identity and multiple comparisons. Colourbar ticks indicate $p = 0.05$, $p = 0.01$, and $p = 0.001$ significance levels respectively. **(c, d)** As in (a) and (b) for all BMI sessions. **(e, f)** as in (a) and (b), shown for all autopilot-BMI sessions.

However, despite this evidence in favour of superposition, we highlight that the T2-T4 PVC does not appear elevated, just T3-T4. There are several factors at play here. First, these T2-T4 are separated further apart in time, and are therefore subject to more of the inevitable gradual turnover of the representation (as outlined earlier with the locomotion-only sessions). Second, the lower trajectory quality in T2, caused by a poor match between decoder and representation, results in a less predictable experience for the animal, which may reduce place field stability. Third, the decoder weights are different between T2 and T3, and any positive reinforcement of activity that is 'useful' for the decoder will take place during the well-executed laps from T3. This makes any argument that relates high decoder weights and induced superposition—a point we will return to later in this chapter—apply only to those neurons with high weights during T3.

Autopilot-BMI sessions

Autopilot-BMI sessions (Fig. 4.7.e-f) present an interesting case: accurate closed-loop performance suggests no fundamental remapping between T1 (the initial autopilot trial) and the two BMI trials (T2 and T3). We suggest that no remapping takes place because path integration is not a useful input to the representation prior to entering BMI in this session type. However, as we pointed out during the previous section, the more predictable nature of the BMI trials from the perspective of the animal results in an enrichment of the existing representation, which subsequently disappears in T4 once traversals once again become unpredictable. This manifests as a relatively high T2-T3 similarity, with PVCs that are otherwise low. We do not observe the superposition effect here (the PVCs of T1-T2 are very similar to T3-T4), as a superposition would require two distinct representations to superpose with each other, rather than the same representation becoming temporarily enriched during the BMI trials.

Autopilot sessions

The autopilot sessions use locomotion as the initial and final reference trials and autopilot (passive motion) during the two inner probe trials. Just as we described for BMI sessions, these exhibit a remapping during the probe trial: there is a locomotion-based representation of the track, and a different representation in which path integration via wheel running is no longer useful (Fig. 3.5.a-b). This is visible through the relatively high T1-T4 and T2-T3 similarities. We show these PVCs on the same axes for parabolic and constant speed sessions as they are phenomenologically similar: both are traversals of the VR environment in which the animal has no control over its motion, but are nevertheless predictable and consistent experiences. We show random autopilot sessions separately, as we have previously explained that these exhibit a large drop in place cell count during the probe trials (Fig. 3.5.c-d).

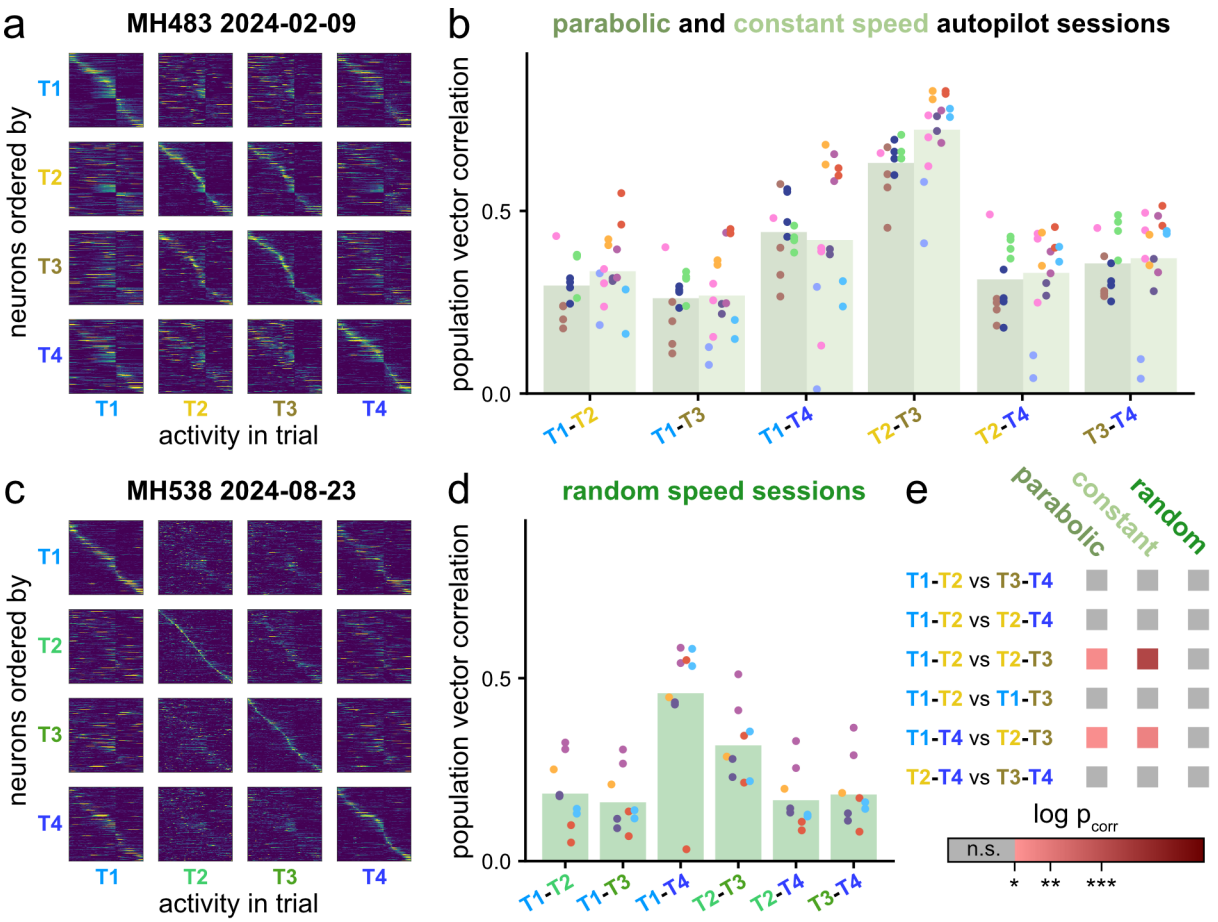


Fig. 4.8 The superposition effect appears absent in autopilot-based control sessions. **(a)** A population ratemap for a typical parabolic autopilot session. **(b)** Mean population vector correlations between pairs of trials during all constant speed and parabolic autopilot sessions. **(c)** As in (a), shown for a random speed autopilot session. **(d)** As in (b), shown for all random speed autopilot sessions. **(e)** Selected statistical significances, as in Fig. 4.7, of population vector correlation differences between different choices of pairs of trial, shown for all autopilot session types.

Across all autopilot sessions, we find no significant difference between the T1-T2 and the T3-T4 representational similarities for any of the autopilot session types. We take this as an absence of evidence for superposition as measured by PVC at the level of populations (Fig. 3.5.e). We acknowledge that our total number of autopilot sessions is smaller than the number used for BMI, and that they are fragmented across three session types. However, even if we aggregate over all autopilot session types (for a total of n=37), we fail to find significant evidence of superposition as measured by representational similarity from T1-T2 to T3-T4 ($p = 0.471$, even after removal of the repeated comparisons correction).

Superposition appears unique to BMI sessions as measured by population metrics

We take this as strong evidence that there is a fundamental difference in the behaviour of the neural population between BMI sessions and autopilot sessions. We claim that the superposition effect appears, when measured by population-wide metrics, to be unique to the BMI condition. While the removal of the path integration signal explains reversible changes in representation common to both the BMI and autopilot contexts, superposition corresponds to persistent changes that must be explained by some other mechanism. We speculatively offer two candidate explanations. First, it is possible that a direct interaction between the BMI decoder weights and neurons induces a reinforcement of representation that persists even after removal from the BMI context. Second, we entertain the possibility that CA1 somehow encodes 'agency', or that its representations are otherwise influenced by the ability to exert control over navigational outcomes. To attempt to prise apart these explanations, we will next inspect the responses of individual neurons and their interactions with decoder weights.

4.4.2 Inspecting individual place cell behaviour

Grouping responses into archetypes

So far, we have taken an arm's length view of the differences in the representations of BMI, autopilot, and locomotion contexts by considering only population-wide metrics. But what does the superposition effect look like at the resolution of individual neuron responses? It is conceivable that the superposed representation is constructed from the contributions of two distinct subsets of the population. Each subset could contain its own context-sensitive place cells, which form place fields either in BMI or in the initial locomotion trial, with each subset becoming partially[8] active during the final locomotion trial. It is also conceivable that the response of individual neurons is altered, such that they have place fields in multiple locations simultaneously during the final locomotion trial, with place field locations drawn from both the initial locomotion and the BMI contexts.

We categorise the responses of all neurons in BMI sessions according to a set of rules that define response 'archetypes' (Fig. 4.9, see Methods 4.8.3). These rules are based exclusively on where place fields are located during each of the four trials within a session, and the archetypes themselves are heuristic: we intend for them to give a general sense of tuning changes in the population, rather than some fundamental description of how activity is organised in CA1.

[8] As a reminder, we earlier observed that the total fraction of neurons with place fields does not increase following BMI usage (Fig. 4.5.d). This means that it is not possible for the entirety of the place cells from each context to be active in the final locomotion trial—partial contributions would be necessary to remain consistent with our observations.

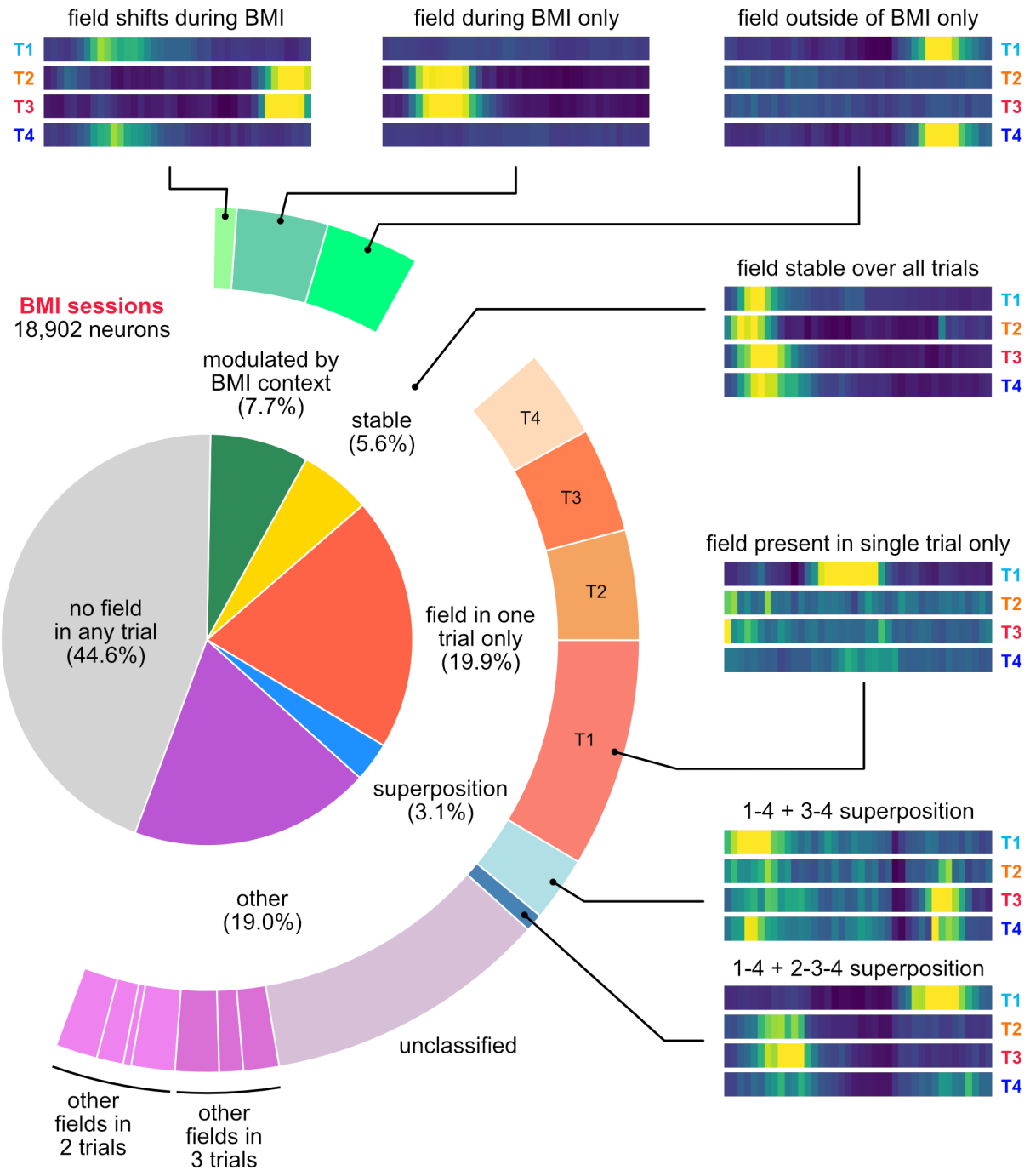


Fig. 4.9 Archetypes of changes in place cell responses during and following BMI usage. A categorisation of every neuron ROI identified across all BMI sessions. Archetype and sub-archetype identification is based on changes in place field location across the four trials within a session. Each sub-archetype is accompanied by the ratemaps for a manually selected neuron that well-exemplifies typical behaviour in that category.

A relatively small proportion of neurons (5.6% of the entire population) are stable: they maintain their place field across all four trials in a BMI session. A slightly larger portion (7.7%) are modulated by the probe condition. We break these into three sub-archetypes: neurons with fields in a consistent location but only in locomotion trials, neurons with fields in a consistent location but only in BMI trials, and neurons that have fields in two different locations—one for each condition. These are tidy examples of reversible changes in representation specific to the BMI or locomotion contexts.

Many neurons exhibit transient fields (~20%), with a field occurring only in one trial and their activity not spatially localised otherwise. We find it interesting that these are predominantly fields in the initial locomotion trial—though we will shortly see that this is an effect common to other session types, even in controls consisting of only locomotion trials (see Fig. 4.10.a), and is therefore likely to be a separate effect related to the first encounters with the track on each day of the experiment.

Finally, we come to patterns of activity that help us account for the superposition effect. Certain neurons exhibit a field at some location A during the initial locomotion trial and a field at a different location B in the BMI context. We categorise neurons into the superposition archetype if they have place fields in common with both A and B during the final locomotion trial. The two sub-archetypes, which depend on whether the field is present (T1-T4 + T2-T3-T4) or absent (T1-T4 + T3-T4) during the first of the two BMI trials, account for 3.1% of observed neurons. The former of these two sub-archetypes is less common: the field must survive the greater turnover from experience or elapsed time, and it has stringent criteria for its identification (an additional field in T2 matching locations in both T3 and T4 while not matching any in T1).

Many neurons (~45%) form no fields at any point during a session, and many neurons form fields shared across several trials in ways not specific to context (although we will demonstrate later in Fig. 4.11.a that there are far fewer of these than the context-sensitive pairs) or in patterns that entirely escape our classification rules.

Archetypes in other session types

We now move to consider what responses look like in other session types in terms of these archetypes. Our first point of comparison is the collection of sessions consisting entirely of locomotion trials (Fig. 4.10.a). Compared to the BMI sessions, these do not show evidence of a second representation during the two inner trials: there are *zero* instances (across over 10^4 neurons) of a place field appearing at one location in T1-T4 and another location in T2-T3, and few examples of fields that appear uniquely in the T1-T4 or T2-T3 pairs. Many more of the neurons show stable fields (~15%), and fewer have transient fields that appear in only a



Fig. 4.10 Archetypes of changes in place cell responses for other session types. The same categorisation rules as used in Fig. 4.9, as applied to **(a)** sessions in which all trials are locomotion, **(b)** autopilot-BMI sessions, **(c)** parabolic and constant speed autopilot sessions (merged by raw aggregation), and **(d)** random speed autopilot sessions.

single trial. A small portion of neurons are classified as exhibiting superposed fields, which raises the possibility that these emerge by chance—a possibility we will control for shortly.

We turn next to the autopilot-BMI sessions (Fig. 4.10.b). These have a similar number of stable fields and single-trial fields to the BMI sessions, but a far lower incidence of superposed fields. We repeat our previous explanation for fewer superposed fields: we claim that there is only one representation 'in use' during these sessions, and so we do not expect an effect that arises from the interference of two representations. From our earlier observations, we know that there are fewer place cells during the random autopilot trials in T1 and T4 during these sessions: this explains the large quantity of neurons that exhibit a field in the same location in T2 and T3 only.

In the autopilot sessions using either constant speed or parabolic speed profiles as the probe trials (which we show merged here, but in the interest of rigour will also show separately in the analysis of the next subsection), we see an overall distribution of neuron archetypes that is largely similar to BMI sessions: some cells are modulated by the context, some are stable, and approximately a fifth exhibit transient fields. This is consistent with our explanation that there are two representations of the track, one that depends on locomotion and one that does not. There are two main differences: superposition, of which we see less than in BMI sessions (2.2% vs 3.1%), and neurons exhibiting no field at all, of which we also see fewer (~35% vs ~45%). As a fraction of the neurons that form fields at all, superposition is even less prevalent than the 2.2% figure suggests.

For completeness, we show also the random autopilot sessions (Fig. 4.10.d). These have extremely few stable fields (1.7%), but many fields common between the two reference locomotion trials T1 and T4. This is consistent with our observation that fields are impaired when environment traversal is unpredictable.

Superposition is unique to BMI sessions as measured by individual responses

So far, we have considered the relative occurrence of each of the response archetypes only on aggregate—a raw sum over all sessions, and without accounting for the varied number of ROIs visible for each animal (as previously shown in Fig. 3.6.b). We suspect that neurons falling under the superposition archetype are more frequent in BMI sessions, and so we now consider archetype distributions on a session-by-session basis. In this context, we must account for discrepancies in the fraction of neurons that engage in forming a representation from animal to animal (as described previously in Fig. 4.5), which we accomplish by reporting relative occurrences of archetypes as a fraction of the total number of neurons that form fields in any trial within a given session, rather than as a fraction of the complete fraction of the population. This is an important methodological detail, as it leads to numerically larger population fractions than quoted in the previous section.

Before considering neurons exhibiting superposition, we first consider neurons that have a single place field that is shared across exactly two trials during BMI sessions (Fig. 4.11.a). This passes an initial sanity-check: there are relatively many fields specific to the BMI context or the locomotion context, and relatively few between contexts. Those few neurons that do share a field between contexts are more likely to occur between temporally successive trials. We find that the numbers of neurons sharing fields between T1-T2 and T3-T4 are very similar. This is an important detail: it suggests that the high representational similarity between T3 and T4 during BMI sessions (as measured by PVC) cannot be explained by T3-T4 simply having more neurons that share exactly one field than T1-T2. We stress that this classification excludes neurons that share a single field across three trials (such as T2-T3-T4), and also highlight that neurons sharing a field in exactly T2-T4 are very rare. We conclude that in order to support superposition at the level of a population, some richer dynamics must be at play, with neurons either forming fields in multiple places via superposition or otherwise retaining fields across T2-T3-T4 more readily than across T1-T2-T3.

We consider next the relative occurrence of neurons classified as exhibiting superposition across session types, and find that BMI sessions exhibit the greatest amount of superposed fields (Fig. 4.11.b). We compare the occurrence of such superposed fields to a random shuffling of trial order (Fig. 4.11.c, see Methods 4.8.3) and find that only BMI sessions exhibit a greater number of superposed fields than randomised shuffles at a statistically significant level ($p = 1.32 \times 10^{-2}$, single-sided one-sample t-test). With the exception of the random speed autopilot sessions (which have insufficient samples to draw statistically sound conclusions from), we find that all non-BMI session types score negatively on this metric: for these sessions, chance arrangements of the trial order generate *more* occurrences of superposition than the real data. The reader will note the high variance in the fraction of superposed neurons relative to chance in the BMI sessions. We fail to explain this with animal identity, experience, BMI performance, or behavioural metrics. We attribute this variance instead to methodological limitations: the ad hoc archetype categorisation scheme is far from perfect, and this is compounded by its reliance on an automatic place field identification scheme with inherent sensitivity to unrelated factors. Taken together with certain sessions having far smaller total ROI counts, which may result in single-digit numbers of neurons in the superposition archetype, the measure is altogether very noisy.

Decoder weights are high for neurons exhibiting superposition of fields

While we have demonstrated a superposition-like effect both with a population-level metric (mean PVC) and by categorising neurons into archetypes, we have not yet considered the element that distinguished BMI sessions from autopilot sessions: the use of a decoder for closed-loop control. We show in Fig. 4.11.d, aggregated over all BMI sessions, the distribution

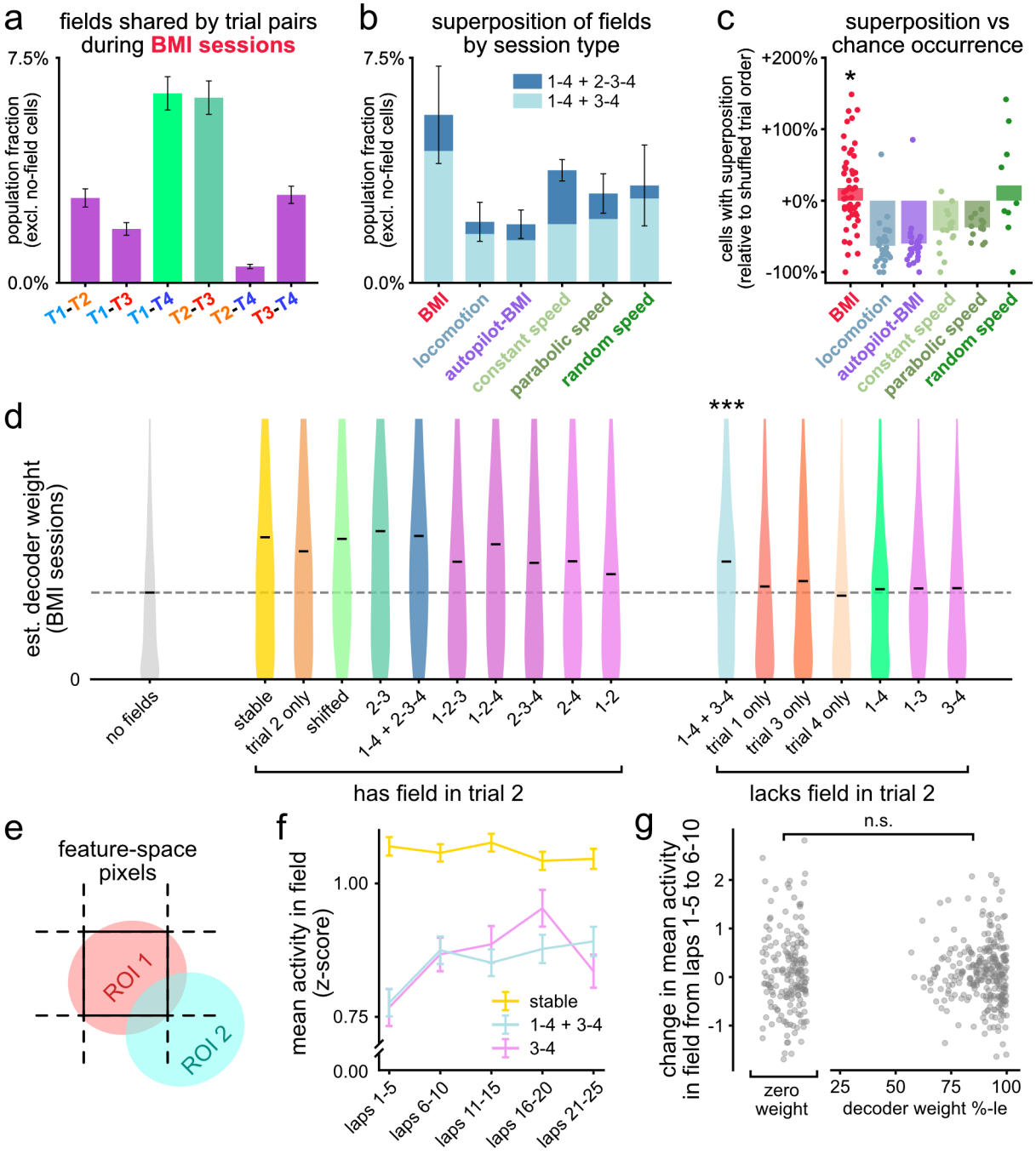


Fig. 4.11 Neurons exhibiting superposition have relatively high decoder weights. **(a)** For BMI sessions, the fraction of place cells that have fields shared between the given pairs of trials only. Error-bars show +/- SEM over sessions. **(b)** For each session type, the fraction of place cells that fall into the superposition archetype. Error-bars show +/- SEM over sessions. **(c)** The number of superposition instances in populations compared to random shuffling of the trial order within a session. Points correspond to individual sessions. **(d)** Violin-plot of mean decoder weight (over all bins) distributions for neuron archetypes for decoders trained on Trial 2. BMI sessions only. [*Caption continues...*]

Fig. 4.11 [*Continued*] **(e)** Neurons without any fields can end up with residual inferred weights by overlap with other neuron ROIs in the decoder's pixel feature-space. **(f)** Mean activity within place fields averaged over 5 lap intervals during Trial 3 for stable neurons, 1-4+3-4 type superposition neurons, and neurons sharing a field between exactly T3 and T4. Aggregated over all BMI sessions. Error-bars show +/- SEM over neurons. **(g)** Change in mean activity within place fields (z-score) from laps 1-5 to laps 6-10 for 1-4+3-4 type superposition neurons, shown as a function of decoder weight percentile (as measured within each session). Each point corresponds to one neuron (n=460). No significant difference between the mean increases of zero-weight neurons compared to weighted neurons as assessed by independent t-test.

of decoder weights for each archetype of neuron for the decoder trained on T2 and used for control during T3. We group these archetypes into three categories: those neurons with no fields (which do not always have zero-weight, due to the potential overlap in the decoder feature-space from adjacent neurons with meaningful information about position—see cartoon in Fig. 4.11.e), those neurons with a field in T2 (and we therefore expect to be allocated relatively high weights), and those neurons without a field in T2 (and we therefore expect to have weights comparable to neurons with no place fields). We find that our expectations hold—with one exception: those neurons that exhibit superposition with fields in T1-T4 and T3-T4 but lacking a field in T2. These have higher weights than other neuron archetypes without a field in T2 ($p = 1.13 \times 10^{-8}$, Mann-Whitney U-test, see Methods 4.8.4). We take this as further evidence that participation of the population in the closed-loop control of the navigation task via the BMI's decoder is necessarily involved in the superposition effect.

The T1-T4+T3-T4 superposition neurons increase their firing rate within their place fields relatively early during T3 (firing rate increase from laps 1-5 to 6-10, $p = 1.40 \times 10^{-3}$, paired t-test across all 460 neurons). In contrast, neurons with a stable field throughout the session exhibit no such increase in firing rate (Fig. 4.11.f). At first glance, this might suggest a purely classical conditioning effect, whereby neurons that happen to end up with 'stray' weights from the decoder training process elevate their firing rates to produce better BMI outcomes, but this hypothesis does not hold water. The neurons that have a field shared between only T3 and T4 do not have elevated decoder weights (Fig. 4.11.d) and undergo the same increase in firing within their eventual place fields (Fig. 4.11.f), and so we ascribe this increase in firing rate to 'natural' field formation rather than conditioning. To corroborate this point, there appears to be no relationship between decoder weight and magnitude of firing rate increase between laps 1-5 and 6-10 for the T1-T4+T3-T4 superposition neurons (Fig. 4.11.g). Putting our observation more simply, elevated decoder weights do not cause a sudden increase in firing early in T3, but somehow contribute to the retention of a field from T3 to T4 specifically for neurons with a history of activity earlier in the session. We will eventually return to the

role of decoder weights when we consider their impact on the retention of fields from one session to the next, rather than from one trial to another within a session.

Reconciling response archetypes with PVC-based similarity

We quantified our claim of superposed representations in BMI sessions at the level of populations using mean PVC, noting that measured between T3 and T4 this is 21.4% greater than when measured between T1 and T2 (referring back to Fig. 4.7.d). We have now shown concrete examples of individual neurons that exhibit clearly superposed examples of fields, but these account only for 3.1% of the entire population—a number insufficient to entirely account for the change in PVC.

We can coarsely quantify how much these superposed neurons contribute to the change in PVC. If we omit all neurons falling under the superposition archetype from the PVC calculation, we find that the T1-T2 mean PVC is unchanged (to within less than 0.3%), but the T3-T4 mean PVC is now only 17.2% greater. This means that the superposition archetype accounts only for one fifth of the change in the PVC metric. Where do the other four-fifths come from?

Our archetype classification is stringent. It leaves a large swath of the population with unclassified responses (see Fig. 4.9). Our system is designed to highlight clear-cut examples of population behaviour, and so it relegates inconveniently ambiguous patterns of field formation to the 'unclassified' category. However, stripping out the 'unclassified' neurons from the PVC calculation by the same process as used for the superposition neurons above does not meaningfully change (i.e. ≤ 1 percentage point) the increase in mean PVC between T1-T2 and T3-T4. While the 'unclassified' category may well contain internal subsets that undergo a change in response that contributes to superposition as measured by PVC, it itself is not categorically biased towards these neurons.

We deliberately omitted neurons sharing a field in exactly T2-T3-T4 with no field in T1 when defining the superposition archetype. These neurons can conceivably be viewed as a superposition of the locomotion and BMI representations with the locomotion representation simply being blank, but they make for far less compelling examples without two place fields active simultaneously in different locations. If these neurons are excluded from the PVC calculation, we can determine that they also contribute approximately one fifth of the increase in the mean PVC between T1-T2 and T3-T4.

Removing each other archetype of neurons from the PVC calculation yields no further explanatory gains. This means that our archetype-based categorisation system explains, in total, approximately 40% of the increase in PVC as measured by exclusion of categories. For a set of heuristic categories designed to be illustrative, we believe this is a satisfactory starting point. We entertain too the possibility that some component of the superposition effect might modulate all neurons and thereby defy any explanation based purely on categorisation.

4.5 Revisiting path integration with gain modulation

We have one final confound to address before we can be satisfied that the superposition effect is induced by BMI usage: what if the so-called 'BMI-context' corresponds to a situation in which the animal retains causal control over its navigation, but with a changed relationship between wheel running and VR traversal? After all, if we cast our minds back to the previous chapter, we recall that some animals had running speeds that do not exactly match the target trajectory imposed by the BMI (see Fig. 3.4.b for an example in which the animal runs much faster than the target velocity).

While we partially addressed this in the previous chapter by demonstrating instances of adequate decoder accuracy in spite of decoupling between wheel speed and BMI traversal

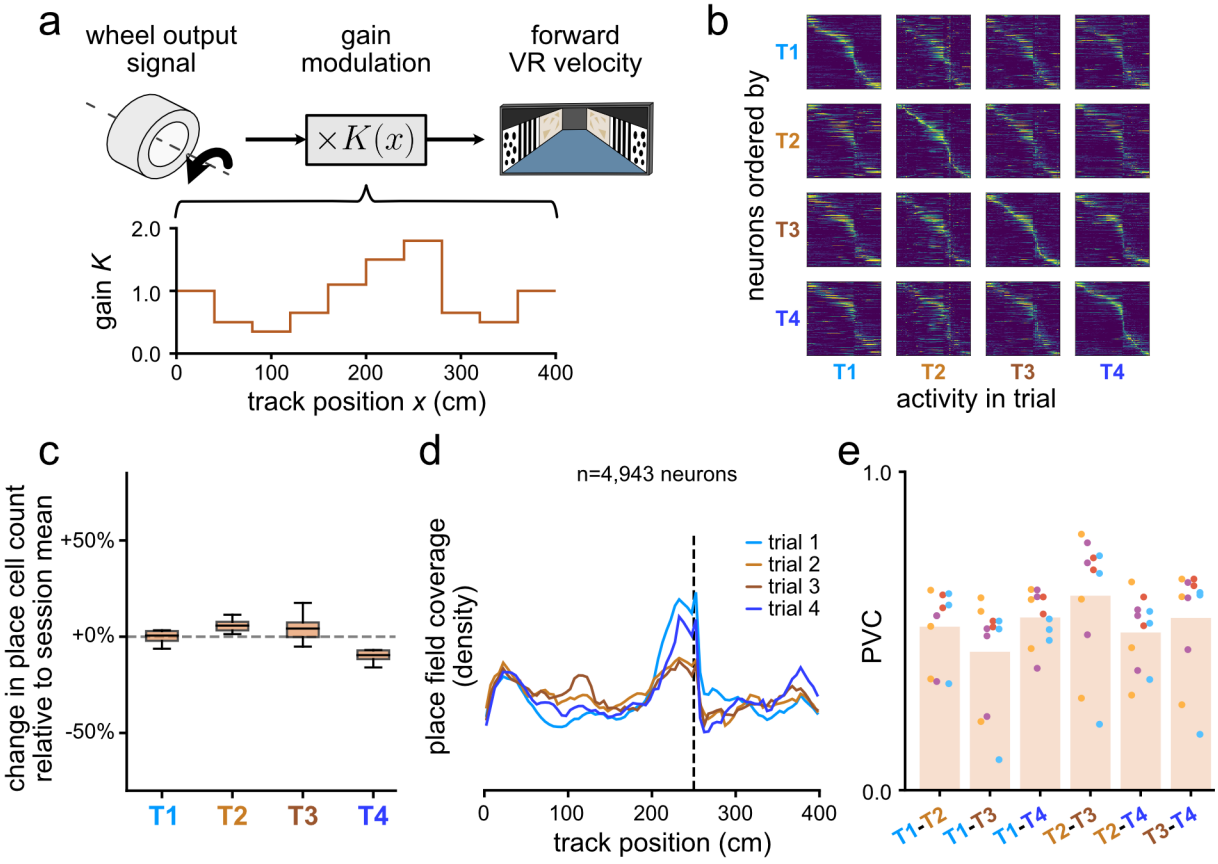


Fig. 4.12 Locomotion under altered gain condition causes no remapping or superposition effects. **(a)** Schematic of the gain modulation control modality. Different portions of the track experience different forward velocities for the same rotation of the wheel. **(b)** Example population ratemap for a gain modulation session, as in Fig. 4.7.a. **(c)** Relative changes in place cell count for gain modulation sessions, as in Fig. 4.5.c. **(d)** Place field coverage aggregated over all gain modulation sessions, as in Fig. 4.6.a. **(e)** Mean PVCs between pairs of trials within gain modulation sessions, as in Fig. 4.7.b.

velocity, we can now complete this argument in terms of representational similarity using gain modulation trials. These trials alter the ratio between rotation of the wheel and distance travelled on the VR track. Rather than a constant gain alteration, we attempt to more aggressively disrupt the reliability of the path integration system by altering the gain as a function of track location. Using the same 40 cm bins used for decoder classification, we implement a gain profile that attenuates speed in the region of fast running and amplifies speed in the region of slow running (Fig. 4.12.a, for traces of velocity profiles during such trials, refer back to Fig. 3.5.d in the previous chapter).

Visual inspection of the population ratemaps during sessions using gain modulation trials as probes reveals no global remapping (Fig. 4.12.b), which is in agreement with the result in Chen et al. (2013). Changes in the number of place cells between trials during gain modulation sessions (Fig. 4.12.c) closely mirror the 'all locomotion' sessions (Fig. 4.5.c). We do not see strong evidence of a more cue-based representation of the track when considering place-field coverage (despite what appears to be a small bump during Trial 3 only), though we do see what appears to be a local alteration near the reward site (Fig. 4.5.d). Finally, quantifying representational similarity using the PVC metric suggests that no superposition-like changes take place, given that T1-T2 are as similar to each other as T3-T4. We conclude from this control experiment that the persistent changes in neural responses that take place following BMI usage cannot be explained by changes in gain, and believe that this strengthens our case that these persistent changes are induced by experience with an alternative and cognitively engaged means of traversing the environment.

4.6 Representations on cross-session timescales

4.6.1 Tracking neuron ROIs from session to session

We complete our description of the changes in representation that take place during and following BMI usage with a consideration of the changes in representation that take place between sessions. We use `CellReg` (Sheintuch et al., 2017) to align the `suite2p`-derived ROI footprints from one session to another by computing a static translation and rotation of the FoV (Fig. 4.13.a, see Methods 4.8.5). We then use `CellReg`'s model of matched-ROI confidence to identify ROIs with matching footprints between sessions, setting an operating point that yields a 95% true-positive rate (Fig. 4.13.b). We are subsequently able to track certain neuron ROIs across days (Fig. 4.13.c). This registration works well for sessions that are close together in time, with the probability of registering the same neuron between a pair of sessions falling off for sessions further apart in time (Fig. 4.13.d). Consequently, neurons which can be identified in a handful of sessions are common, while those that can be tracked across many sessions (10+) are rare (Fig. 4.13.e).

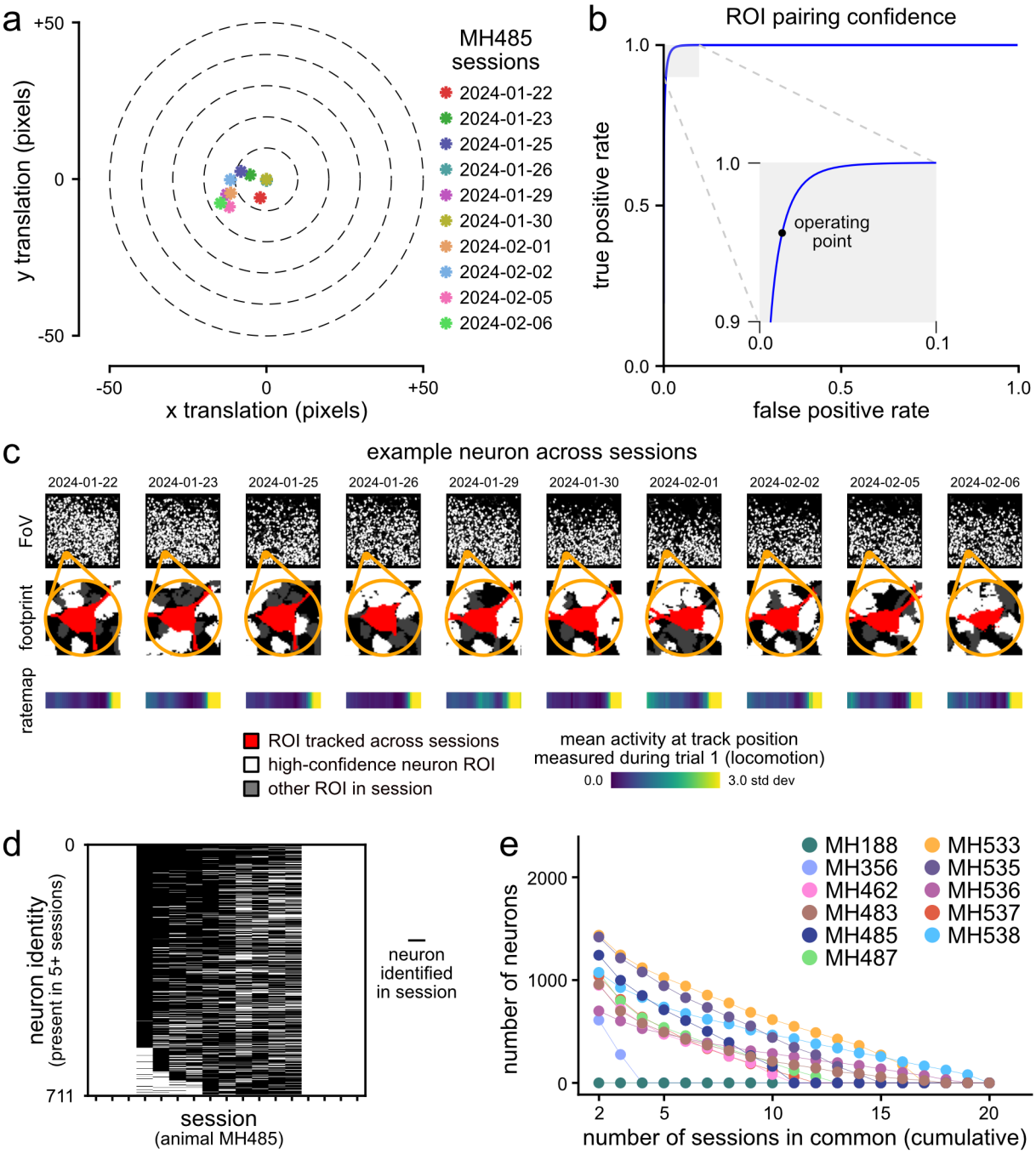


Fig. 4.13 Technical overview of cross-day cell registration. **(a)** Example FoV translations relative to a reference session as used for cross-day ROI alignment. **(b)** Estimated receiver operating characteristic curve for identification of ROI pairs between all sessions from (a), showing our choice of operating point (95% true positive). **(c)** An example of an individual neuron tracked across all sessions in (a), including location in FoV, ROI footprint, and activity ratemap during the initial locomotion trial in each session. **(d)** Neuron presence matrix for all sessions of MH485, including only neurons tracked across at least 5 sessions. Empty columns indicate sessions that failed alignment. **(e)** For each animal, the number of neurons identified between sessions that are visible in at least the given number of sessions.

4.6.2 Context-specific representations persist between sessions

Many neurons have consistent responses from one BMI session to the next. We find examples of neurons with a single place field that is entirely stable across trials and sessions, neurons with consistent preference for either the BMI or the locomotion context, and neurons whose firing location is consistently modulated by control modality[9] (Fig. 4.14.a). Representations in CA1 change both as a function of time and of experience (Elyasaf et al., 2024; Geva et al., 2023; Khatib et al., 2023; Ziv et al., 2013) and our experimental schedule does not run BMI sessions on an exactly daily cadence (referring back to Fig. 3.6 in the previous chapter). We therefore restrict our quantitative analysis to pairs of BMI sessions that take place exactly one day apart in order to avoid confounds associated with inequivalent amounts of BMI experience between comparison points.

Stability of the distinction between contexts

We turn once again to mean PVC as a tool for quantifying representational similarity between pairs of trials (see Methods 4.8.5, which details session pair inclusion and accounting for the imbalances between the number of ROIs that match between session pairs), evaluating each trial on some day D against each trial on the following day $D+1$ (Fig. 4.14.b). There is an immediately obvious structure to this comparison matrix: any pair of trials of the same trial type has a higher PVC than any pair of trials with mismatched trial types. We interpret this as stability of the distinction between BMI and locomotion contexts, and this is consistent with the view that the distinction is explained mechanistically by the presence or absence of path integration signals.

A more subtle feature of this matrix is a broken symmetry about the main diagonal: when we compare the PVC between T4 on day D and T1 on day $D+1$ (PVC: 0.473, standard deviation within bootstraps: 1.47×10^{-2}), we find it greater than the PVC between T1 on day D and T4 on day $D+1$ (PVC: 0.395, standard deviation within bootstraps: 2.48×10^{-2}; comparison significance $p = 3.3 \times 10^{-3}$). This means that the representation at the end of day D is more similar to the representation at the start of day $D+1$ than vice versa. We interpret this as the experience of BMI trials driving a change in representation: the PVC between trials with no BMI trials performed between them is higher than for those trials separated by BMI trials.

For this interpretation to withstand scrutiny, we must take into account that *any* experience with the track, BMI or otherwise, has the potential to drive some change in representation. We offer as comparison the same evaluation performed on the pairs of constant autopilot and parabolic autopilot sessions for the same animals. Over the 5 such session pairs

[9] We offer, entirely qualitatively, the observation that the neurons with place fields that are modality-sensitive in different locations are prone to being classified as superposition on some days and as modulated by modality with a shifted field on others. This highlights that our archetype classification scheme is not entirely robust.

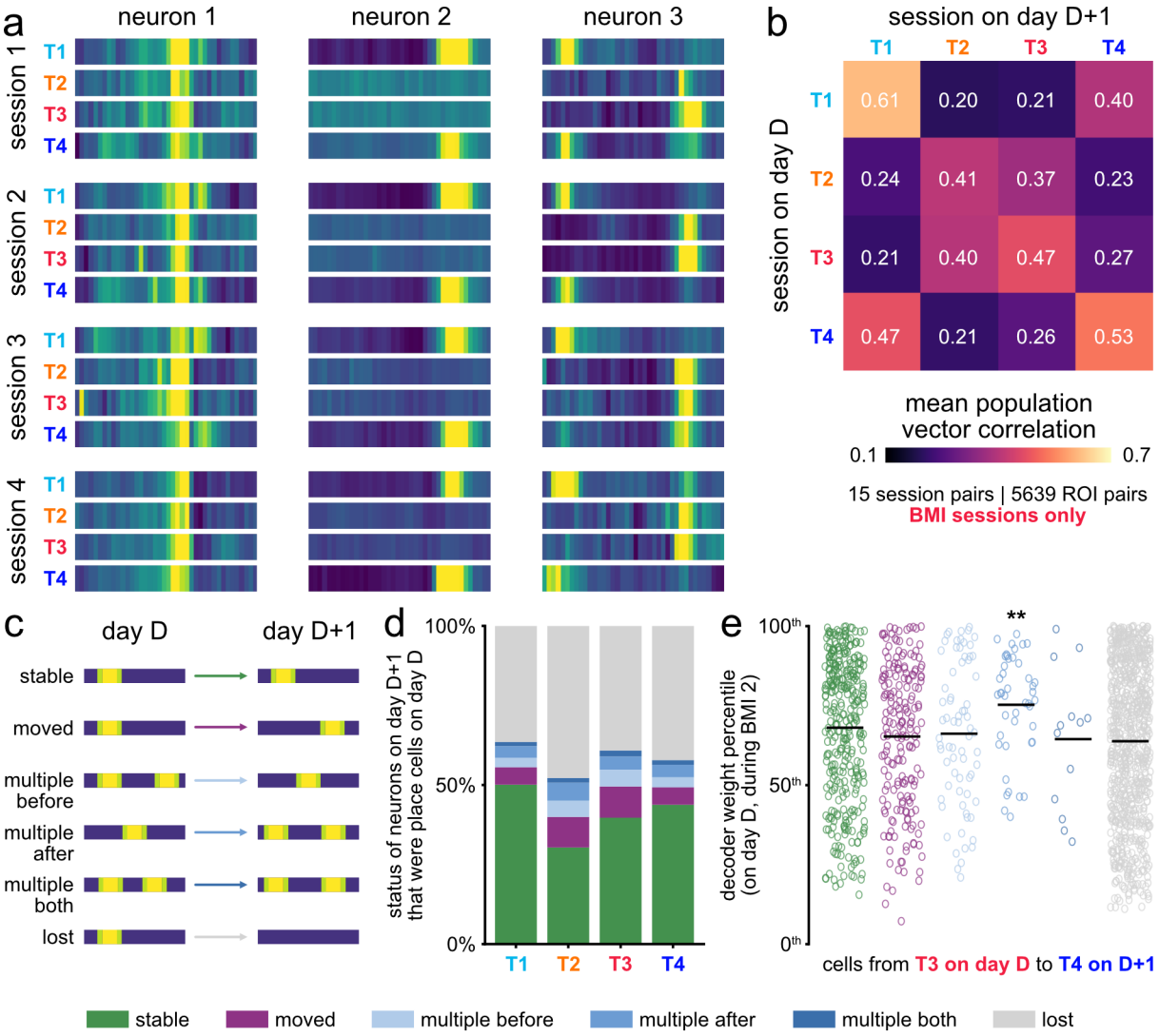


Fig. 4.14 Locomotion and BMI contexts remain distinct between sessions. **(a)** Ratemaps for three example neurons with stable context-specific and context-agnostic fields from the same animal over the course of four consecutive sessions. **(b)** Mean population vector correlation between trial pairs from BMI sessions exactly one day apart. **(c)** Cartoon of six categories describing the status of a neuron with a field on some day D. **(d)** Distributions of these categories for BMI sessions exactly one day apart, comparing T1 on day D to T1 on day $D+1$, T2 on day D to T2 on day $D+1$, and so on. **(e)** Decoder weight percentiles for the decoder used on day D, shown for each category of field change when comparing T3 on day D to T4 on day $D+1$. Neurons with a single place field on day D during BMI and multiple place fields during T4 on day $D+1$ have significantly higher decoder weights (One-way ANOVA: $p = 6.0 \times 10^{-3}$, independent two-sided t-test against population: $p = 4.4 \times 10^{-3}$).

that meet our inclusion criteria (1,773 matched neurons), we do not see the same symmetry breaking: the two PVCs are very similar (absolute difference 7.1×10^{-4}, proportional difference 0.0020%)[10], especially when compared to the change visible for the BMI session pairs (absolute difference 7.8×10^{-2}, proportional difference 16%). We consider also the pairs of 'all locomotion' sessions, with the caveat that we include all animals (and not just the cohort experiencing a majority of BMI sessions) to bring us from 3 to 10 valid session pairs (3,085 matched neurons). These also display a broken symmetry (absolute difference 4.8×10^{-2}, proportional difference 12%), albeit less pronounced than for the BMI sessions.

Given the small sample size, the small difference in effect size, and that the animal identities differ between the comparisons, we would be remiss to present this as strong direct evidence for BMI-accelerated changes in representation. Nevertheless, we allow ourselves to speculate that an active role in navigating through environments drives changes in the representations, whether by BMI or by locomotion, and that a passive role does not.

Our final observation about the PVC comparison matrix for BMI sessions pertains to the initial locomotion trial, T1. The similarity of T1 on day D to T1 on day $D+1$ is the largest across the entire matrix, which seems to run contrary to the notion that experience drives changes—should we not expect T4 on day D to be more similar to T1 on $D+1$, as these trials are adjacent? While we find this result initially surprising, we recall from Fig. 4.9 that T1 has the highest incidence of unique place fields. This suggests that T1's position as the first trial of a day grants its representation some unique quality—perhaps a novelty signal, an anticipation signal, or the absence of fatigue—that accounts in part for its high self-similarity between days. Indeed, we also find that for pairs of trials in both autopilot sessions and 'all locomotion' sessions, the highest value in the cross-day PVC comparison matrix remains T1 to T1. This is consistent with our earlier observations that T1 has the highest incidence of unique place cells in locomotion and autopilot session types (Fig. 4.10.a,c).

Quantifying changes in place field between sessions

The high self-similarity of T1 forces us to consider whether the superposition effect is a purely transient phenomenon induced by BMI on short timescales, or whether it has effects that persist between sessions. As with the single session analysis earlier in this chapter, we will approach this by building an intuition through assigning categories to neurons and evaluating their relative frequencies.

To compare a pair of trials that span two sessions, we consider the neurons with at least one field on day D and categorise the status of that same neuron's tuning on day $D+1$ (Fig.

[10] As each session lasts 70 minutes, the comparison between PVCs is necessarily comparing trials that do not take part exactly the same number of hours apart. We use the observation that these particular PVCs are very similar to highlight that active navigation of the environment is a far more important factor than the discrepancy in elapsed time.

4.14.c, see Methods 4.8.5)[11]. We show these in Fig. 4.14.d for changes in fields for the same trial number between two sessions (i.e. comparing T1 to T1 between days, T2 to T2 between days, and so on). This gives us a general sense of the stability of representation for each trial. T1 is the most stable, with approximately 50% of its place cells retaining their field at a consistent location. The next largest category for T1 is fields that disappear from one session to the next (~36%). Few fields move in T1, and the number of cells with multiple fields is nearly balanced over time: the fraction of cells transitioning from one field to multiple fields (3.6%) is close to the fraction of cells transitioning from multiple fields to a single field (3.1%).

Fields in the BMI context are less stable, especially in T2, which retains only ~30% of its fields in consistent locations between sessions. This suggests a faster rate of representational turnover, an observation that is consistent with the lower cross-session PVC values from the previous section. For the post-BMI locomotion context in T4, we find minimal evidence of a net increase over time in the number of neurons with multiple fields (multiple fields before and single field after: 3.2%, multiple fields after and single field before: 4.0%). This suggests that there is no cumulative build-up of superposed fields over the course of several sessions.

However, despite no build-up of such fields[12], we find tentative evidence that the decoder weights allocated to neurons do demonstrate an interaction with superposition on cross-session timescales. In Fig. 4.14.e, we consider the decoder weights associated with the categories of fields between T3 on day D with T4 on day $D+1$. Those neurons with a single field during T3 on the first day and multiple fields during T4 on the second day, which is a loose approximation of those neurons that show superposition on the second day, have considerably higher decoder weights than all other categories—a phenomenon reminiscent of the effect we described within individual sessions in Fig. 4.11.d. This relationship is not necessarily causal: while it is appealing to think that decoder weights on one day might directly influence which neurons exhibit superposition on subsequent days, we must also consider that the neural activity on which decoders are trained from one day to the next share part of a representation in common. If the same neurons are prone to superposition, then the high decoder weights predictive of superposition on day D will also be predictive of superposition on day $D+1$.

If the same neurons are indeed prone to superposition across days, then is this a matter of a learned relationship? Or could we instead explain their predisposition through an input to CA1 that is absent from passive motion, yet common to the BMI and locomotion conditions?

[11] The rationale for taking exclusively neurons with a field present on D is to avoid interference from another time-dependent effect: we described in the previous chapter that the fraction of neurons that are place cells during BMI trials increases over time.

[12] Whether it is even possible to accumulate superposed fields is in question. If activity in CA1 is subject to regularisation to enforce sparseness of representation (McKenzie, 2018), this would drive a gradual turnover in representation (Ratzon et al., 2024). See Lee et al. (2020) for a counterargument suggesting that within the heterogeneity of CA1 populations some neurons may be well-suited to multiple fields within the same spaces.

4.7 Discussion

4.7.1 Are representations modulated by agency?

We have described fundamentally different representations of a visually identical linear track environment. This is not an observation without precedent: earlier work has induced new representations of visually identical cue sets by manipulating the animal's experience prior to cue exposure (Tanila, 1999), has found representations influenced by differing task demands within the same environment (Zutshi et al., 2025), and has even observed multiple maps of the same space coexist on short timescales with no apparent prompting or manipulation (Sheintuch et al., 2020). In our case, these representations can be swapped between at will by controlling how the animal traverses the environment. The main feature that distinguishes locomotion trials from either autopilot or BMI trials is that the animal's manipulation of the wheel no longer produces expected outcomes of movement through the environment, causing the representation of these latter trial types to rely more heavily on visual cues.

In the case of autopilot trials, the change in representation is entirely reversible: returning the animal to a condition in which it can navigate by wheel manipulation restores the previous locomotion-based representation. This appears largely in agreement with existing VR-based studies of passive motion (Chen et al., 2013). However, we reserve judgement as to whether this remapping effect would be as pronounced in a real environment, where a variety of additional sensory cues are available to supplement and stabilise the representation.

Our main novel observation is that BMI usage causes some component of the BMI representation to leak into the locomotion representation, resulting in a superposition of the two representations. We offer two interpretations of this effect. The first interpretation frames CA1 hippocampus as a sequence predictor. The structure of the task during BMI is such that the maintenance of a consistent neural representation of the track produces consistent traversals, while deviating from that representation produces more varied traversals that are unpredictable. While this is not direct classical conditioning of the population, which would encourage field retention by directly reinforcing their elevated firing, it creates an experimental dynamic where representations are self-reinforcing by causal responsibility for consistent experiences. By construction, the representation learned by the decoder is more stable than any other representation of the track during BMI trials. This interpretation proposes that this indirect reinforcement of the representation results in its partial retention when the animal is returned to locomotion.

The second interpretation frames our observations in terms of agency: the ability for the animal to influence outcomes. During BMI trials, and unlike autopilot trials, the animal is causally in control of its motion. While we choose to entirely sidestep the philosophical questions about whether an animal understands that it is able to navigate, we do seriously

entertain the idea that repeated exposure to external outcomes that consistently coincide with internally-driven neural states will result in patterns of activity that could reasonably be called 'representations of agency'. Such signals need not originate in CA1 to explain our results, they need only to eventually route to these populations. Existing work has already established that individual cells within CA1 exhibit heterogeneous responses to signals originating outside CA1, such as whether visual cues are local or distal (Knierim, 2002; Shapiro et al., 1997; Tanila, 1999), or how closely tied they are to sensory landmarks or path integration signals (Anderson and Jeffery, 2003; Geiller et al., 2017; Gothard et al., 2001). If some subset of a given CA1 population is more sensitive to incorporating a hypothetical agency signal, we can speculatively explain why superposition takes place at the level of individual neurons. When agency-sensitive neurons form new fields during BMI, these fields can persist after a return to locomotion, as the animal retains its agency despite this transition from one control modality to another.

As a point of comparison, Losey et al. (2024) describe a BMI experiment in macaque motor cortex where the mapping between neural activity and cursor control of a reach task is deliberately perturbed from a familiar to an unfamiliar one. After the animal learns the new mapping over the course of hundreds of trials, the original familiar mapping is reinstated. There is evidence of residual 'memory traces' of the new mapping during subsequent use of the original original mapping, and these are largely confined to the null-space of the original mapping. These are persistent representation changes that appear phenomenologically similar to our results, and are presented in the context of a 'negotiated equilibrium' between the requirements for both neural mappings to achieve task success. While a promising parallel to our own findings, this motor cortex BMI task requires explicit and laborious learning by the population to conform to particular new patterns, rendering it potentially closer to a conditioning experiment than our own. We challenge a purely conditioning-based explanation of our own data on two counts. First, while neurons with superposition are predisposed to larger decoder weights, there are many examples of such neurons entirely within the null-space of the decoder. Second, the autopilot-BMI sessions do not demonstrate activity in the post-BMI autopilot trial that suggests any kind of representational leak from the BMI trials: it appears that actively navigating through the environment is a requirement to observe the superposition effect.

The two interpretations, sequence prediction and agency signals, are not mutually exclusive: if an animal knows that it can act, it should have a model of the sequence of events that will result from those actions (Buzsáki et al., 2014). Nevertheless, we believe it plausible to disambiguate these two interpretations through adjustments to our experimental paradigm. For example, our experiments trained decoders on an entire FoV, with the philosophy that reading out from as many neurons as possible would increase the likelihood of successful BMI performance. However, as we established during the previous chapter, some animals

with excellent indicator expression and many neurons visible reached a plateau where higher ROI counts led to negligible increases in decoder accuracy[13]. We propose instead to deliberately restrict the BMI decoder to use one half of the FoV only, thereby providing an in vivo control population in the other half of the FoV. Measuring whether this control population experiences superposition at the same rate as the decoded population would allow us to take the first steps towards ruling out or confirming an indirect reinforcement effect, and in the case that both interpretations have explanatory merit, to quantify their relative importance.

4.7.2 Leveraging the strengths of optical BMI

The proposed 'trick' of splitting the FoV in a way that provides a consistent probe and control population from one day to the next is an example of a potential advantage of optical BMI over electrophysiology-based alternatives. This is a convenient point to critically reflect on our closed-loop optical BMI, the first of its kind applied to CA1 hippocampus, and determine how well we have leveraged its advantages. In candour, the main 'selling points' of our optical BMI—interpretable decoder weights and the accurate tracking of cells from session to session—featured only minimally in this chapter. Our findings and analysis focus primarily on a phenomenon that is prominently visible within single sessions, and we expect to be readily reproducible with more traditional BMI designs. Nevertheless, knowledge of these phenomena is of critical relevance to future hippocampal BMI designs.

The full corpus of data from our experiments, which spans a total of 158 recording sessions in the four-trial format, is still under ongoing analysis. Of particular interest is the second main cohort of animals, which experienced primarily autopilot-BMI sessions. These have much better cross-day registration of neurons (referring back to Fig. 4.13.e), and the animals do not need to simultaneously maintain two different representations of the track. This presents an excellent opportunity for future work that delivers on the strengths of optical BMIs to validate decoders that can be adapted for re-use from one session to the next.

4.7.3 A remapping that is global, but not total

The remapping that takes place between contexts is *global*, in that it affects all portions of the track, but it is not *total*: we reported that approximately 5% of the neurons in our imaged populations had place fields that were stable across contexts during BMI sessions. A visual inspection of the population ratemap for an individual session, as illustrated earlier in Fig. 4.2, makes it clear why a decoder trained on locomotion fares poorly with the BMI context. Nevertheless, the animal itself continues to attempt to perform the task, manipulating the wheel at an appropriate velocity and licking in anticipation of its reward. We must then either

[13] This applies at the 40 cm spatial bin accuracy. If we demanded a more fine-grained classification of the track, reaching the plateau in decoder score would require more neurons.

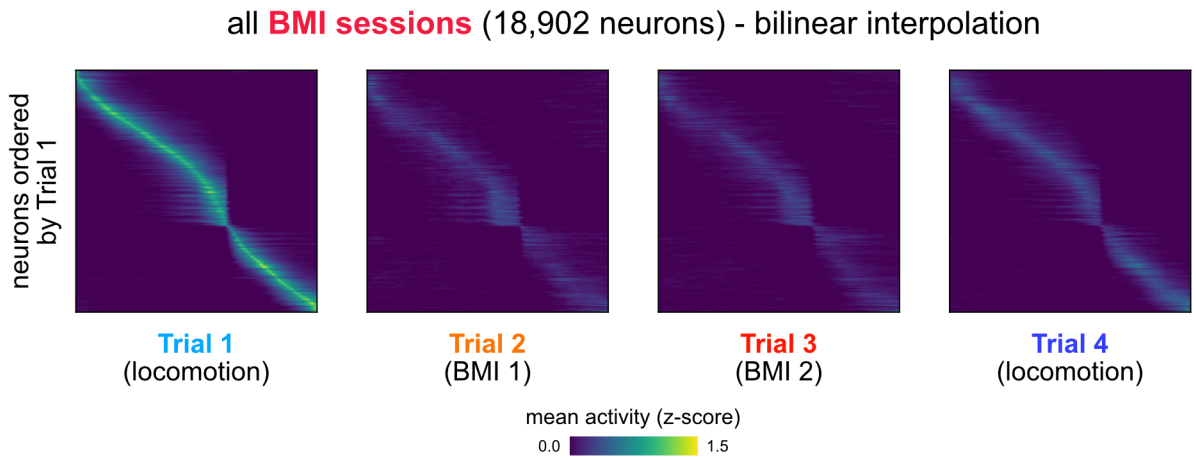


Fig. 4.15 Remapping between contexts does not entirely erase the underlying representation when viewed over a large population. Population ratemap synthesised from all BMI sessions across all animals, ordered by peak activity during Trial 1. The images use bilinear interpolation to merge multiple neurons into single pixels, as the vertical resolution of the figure is insufficient to render the ~20k neurons.

believe that this task structure does not rely on hippocampal function—something that would be extremely inconvenient for the field of hippocampal study—or that whatever remains of the representation is sufficient for the animal to keep its bearings[14].

While our imaging is restricted to observing a population of at most ~1,000 neurons, the brain itself is not. If we synthetically construct a view of activity for a much larger population by concatenating the ratemaps of each individual session, we find that the representation of the track remains distinguishable at every point despite the change in context from locomotion to BMI (Fig. 4.15). In addition to providing an explanation for how animal behaviour remains intact following remapping, this view of activity is heartening from the perspective of BMI design: perhaps achieving stable readouts robust to a variety of operating contexts is merely a matter of scale. However, while simply increasing the number of imaged neurons and naïvely averaging might allow a decoder's accuracy to survive one change in representation in the form of a remapping event, such a strategy will eventually fall prey to continued changes in representation. If only some fraction of neurons survive with their place field intact from one task performance to the next (taking our ~50% fields marked as stable from one day to the next from Fig. 4.14 as an example), it will not take long for successive changes in representation to accumulate to the point where the original representation is unrecognisable—even should we somehow be able to image on the scale of the entire hippocampus simultaneously. This

[14]We point to Jeffery et al. (2003) as an early historical example of this same dichotomy. Their task cannot be learned in animals with hippocampal lesions, and yet task performance remains consistent despite apparently widespread remapping of hippocampal populations.

forms the starting point for the next chapter: how should neural readouts, BMI or otherwise, approach the maintenance of a stable readout despite ongoing change?

4.8 Methods

4.8.1 Quantification of neuron properties

Processing of neuron ROI fluorescence timeseries

We process neuron ROI fluorescence traces by first removing contamination from neuropil using a standard weighting coefficient (Pachitariu et al., 2018; Zong et al., 2022).

$$F_{\text{corr}} = F_{\text{ROI}} - 0.7 \times F_{\text{neuropil}}$$

In practice, F_{neuropil} takes very small values and the impact of this step is negligible for our datasets, as our surgical procedure aspirates cortex within the implanted cannula up to the alveus (the surface of the hippocampus), resulting in minimal interference from neuropil. We apply a dF/F filter to the corrected trace, using the formulation and parameters previously described in 2.6.1.

Ratemaps

As we use Calcium imaging rather than electrophysiology-based recordings, our ratemaps do not directly signify a mean firing rate at a location measurable in Hz, but are instead a proxy for mean activity at that location. To compute ratemaps, we discretise the track into 5 cm bins. A raw ratemap value is simply the mean value of a neuron ROI's dF/F trace within that bin. When visualising multiple ratemaps spanning multiple trials, we perform a normalisation: the dF/F trace is z-scored using the mean and standard deviation over all trials shown for that ROI.

Place field identification

We identify place fields for neuron ROIs according to the following steps, which are an adaptation to our task and dataset of the 'Combination Method'[15] for place field identification described in Grijseels et al. (2021). The first step of this method requires identifying periods of 'Calcium events', which are spikes in Calcium activity defined as:

- Starting at the point in time at which activity exceeds the upper threshold T_{upper}

[15] An acknowledgement: Kou Huang of the Krupic Lab evaluated a variety of place field identification techniques on other datasets acquired from the same 1D VR task during locomotion-based sessions for a separate project, and his efforts provided an informed decision when selecting this method.

- Ending at the point in time at which activity falls below the lower threshold T_{lower}
- Lasting for at least 3 frames (~100 ms)

The two threshold are defined as:

$$T_{\text{upper}} = 3 \times T_{\text{critical}}$$
$$T_{\text{lower}} = 0.5 \times T_{\text{critical}}$$

Where T_{critical} is chosen for a neuron ROI by slicing a trial timeseries into 2-second windows, evaluating the standard deviation of the dF/F activity in each window, and taking the median of these standard deviations over all 2-second windows.

The second step requires determining candidate regions for place fields. We compute a ratemap for the ROI at a resolution of 5 cm. We then define cut-off threshold Y such that $Y = 0.25 \times (p_{max} - p_{25})$, where p_{25} is the 25th percentile of the ratemap bins and p_{max} is the largest ratemap bin value. Candidate place fields are contiguous regions of the ratemap exceeding the cut-off threshold Y, at least 15 cm in length, and no more than 200 cm in length. We apply two initial criteria to determine whether a candidate place field should be retained:

- The mean ratemap value within the candidate field must exceed 4× the mean ratemap value outside of all candidate fields for the ROI.
- The 'activity ratio' must exceed 25%. The activity ratio is the proportion of time classified as a Calcium event while the animal is within the spatial bounds of the place field.

Finally, we perform a comparison against shuffled data. This breaks the neuron ROI's dF/F timeseries into blocks of 120 frames (~5 seconds) and shuffles the order of these blocks, while keeping the timeseries for the animal's position unaltered. We perform the same place-field identification procedure as above, and reject any candidate fields that do not exceed an activity ratio greater than 95% of 1,000 evaluations on the shuffled data.

It is common in other works performing place field identification to remove samples in which the animal is stationary (typically defined as moving at < 5 cm/s). We choose not to remove these samples, as our experimental setup effectively has two decoupled velocities: traversal of the VR environment, and the speed of running on the wheel.

Field matching

When comparing place fields of the same neuron between two trials, we say that two fields match if they satisfy both of the following criteria:

1. the length of track overlapped by the two fields is at least half of the length of the smaller of the two fields

2. the field centres are no more than 15 cm apart

4.8.2 Quantification of population-level properties

Relative change in place cell counts

For n_i place cells in the i^{th} trial within a single session, the proportional change in place cell count relative to the session mean is simply:

$$\text{proportional } \Delta n_i = \frac{n_i - \overline{n}}{\overline{n}}$$

Where $\overline{n}$ is the mean number of place cells over all trials in that session.

Place field coverage

To estimate place field coverage within a single trial, we first discretise the track at the same resolution as place field identification (5 cm bins) and mark each of these bins as having initially zero coverage. For each neuron in the trial, we increment each bin's coverage value at each bin overlapped by that neuron's place fields. After all neurons have contributed to the coverage map, we normalise the coverage map to have unit area so that it represents density.

Population vector correlation

PVC is a measure of similarity of population activity between a pair of trials at a given location. We compute the ratemap for each trial and perform z-score normalisation for each neuron (as described above). This produces an $N \times M$ matrix for each trial (N neurons, M spatial bins). The PVC at a given location, corresponding to the j^{th} spatial bin, is the Pearson correlation coefficient between the two j^{th} columns of these matrices. In this work, we do not make use of PVC at any specific location, but instead use the mean PVC over all bins as an overall measure of similarity between representations.

4.8.3 Archetypes of context-specificity

Definition of archetypes

We apply the following criteria to automatically categorise neurons into archetypes based on differences between place field presence and location between an initial and final reference trial and two inner probe trials. These criteria are applied sequentially, halting once a neuron has been categorised.

1. Complete absence of place fields in all trials: mark as 'no field'

2. Exactly one trial contains place fields: mark as appropriate 'single trial only'
3. All trials have at least one place field matching at least one place field in every other trial: 'stable'
4. Trials 1 and 4 share a field, and no other trials share a field: 'modulated by context, reference-specific'
5. Trials 2 and 3 share a field, and no other trials share a field: 'modulated by context, probe-specific'
6. Trials 1 and 4 share a field and Trials 2 and 3 share a field, but no fields are shared between any other pair of trials: 'modulated by context, shifted field'
7. Trials 1 and 4 share a field and Trials 2, 3, and 4 share a field: 'superposition, 1-4+2-3-4'
8. Trials 1 and 4 share a field and Trials 3 and 4 share a field: 'superposition, 1-4+3-4'
9. Other scenarios in which exactly two trials share a field: 'other, pair'
10. Other scenarios in which a subgroup of three trials all share a field pairwise with each other trial in that subgroup, and the remaining trial shares no fields with any trials: 'other, triad'
11. Mark as 'unclassified'

Evaluating chance-levels of archetype occurrence

Within each session, we perform 1,000 shuffles of the order of trial occurrence. We derive the number of cells falling into each archetype (as per the criteria above) for each shuffle. If the number of chance occurrences for a given archetype is K_{chance}, calculated by the mean over all shuffles, and the number of observed occurrences is K_{observed}, we use the following metric as a measure of occurrence relative to chance within a session:

$$\text{proportional } \Delta \text{ archetype occurrence} = \frac{K_{\text{observed}} - K_{\text{chance}}}{K_{\text{chance}}}$$

4.8.4 Tests of statistical significance and effect size

Pairwise PVC comparisons

Within a given session type, we compare each mean PVC (corresponding to the similarity between a pair of trials) to each other PVC via a two-sided t-test. We treat animal identity as a within-subjects factor. We correct for repeated comparisons using a Bonferroni correction.

This is a conservative correction (i.e. it increases the likelihood of false negatives). For m total comparisons, this correction is equivalent to scaling the threshold for significance by a factor of $1/m$. Our implementation handles this by multiplying all p-values by m and keeping the significance threshold fixed at 0.05. In each instance, we ran a one-way repeated measures ANOVA (with animal identity as a within-subjects factor) before proceeding to the pairwise comparisons to verify the presence of significant differences between trial pairings. Both of these tests are implemented using the existing `Pingouin` software library (Vallat, 2018). Where reported, the Bayes Factor is estimated with `Pingouin`, which internally relies on the methods from Rouder et al. (2009).

Decoder weight comparisons

We compare the decoder weights for neurons classified as belonging to the T1-T4+T3-T4 superposition archetype (population A) to all other place cells without a field in T2 (population B, consisting of the concatenation of archetypes T1 only, T3 only, T4 only, T1-T4, T1-T3, and T3-T4). The distributions of A and B are not normally distributed, and so we make use of a non-parametric test. We apply a Mann-Whitney U-test to test whether A is stochastically greater than B, using a reference implementation from the `scipy.stats` software library (Virtanen et al., 2020).

4.8.5 Identification of neuron ROIs between sessions

CellReg-based tracking

We use `CellReg` (Sheintuch et al., 2017) to identify ROIs that are highly likely to correspond to the same neuron from one session to the next. We configure `CellReg` to use only rigid (rotation + translation) FoV transformations to align ROI footprints. Sessions that do not have an adequate rigid transformation (usually caused by subtle out-of-plane misalignment) are discarded entirely from cross-day registration. Cell identity matches are based on the cross-correlation of the ROI footprints with a correlation threshold of 0.65. Each registration process requires a reference session to serve as a baseline for aligning the FoV of each other session. We select this reference session from those near the middle of the animal's imaging schedule and use manual trial and error to minimise the number of sessions that fail FoV alignment.

Population vector correlation across days

We consider only the cohort of animals that experience a majority of BMI sessions (MH356, MH462, MH483, MH485, MH487) and only BMI sessions spaced exactly one day apart. There are no instances of three days of consecutive imaging for any given animal, and there are

therefore no instances of a session being included in multiple pairs. The experimental schedule described in the previous chapter produced 17 such pairs, of which we consider the 15 that contain ≥100 neurons matched between both sessions.

We account for discrepancies in the number of neurons registered between sessions by concatenating ratemaps for each day (D and $D+1$) within the session pair. This produces a merged ratemap for each trial on each day containing 5,639 neuron ROIs for each day. We evaluate the PVC of this merged ratemap as described in 4.8.2.

We perform bootstrapping over session pair inclusion (10,000 iterations of sampling with replacement from the 15 session pairs), reporting the standard deviation over these bootstraps of individual PVC values as a quantification of variability. To establish whether one PVC is significantly larger than another (for which we make only one comparison, and so no multiple-comparisons correction is needed), we report the fraction of bootstraps in which the PVC with the smaller mean exceeded the PVC with the larger mean as a simple quantification of significance.

We repeat this analysis for the same animals for constant speed and parabolic speed autopilot sessions (same animals, 1,773 neurons total). We also repeat it for the 'all locomotion' sessions, but include all cohorts of animals (resulting in 3,085 matched neurons) due to insufficient matched neurons across days in this session type for the first cohort—acknowledging that this weakens the strength of our conclusions.

Place field changes between sessions

We make use of the following sequential rules in order to categorise the status of neurons in a specific trial on day $D+1$ that were place cells (i.e. had one or more place fields) during a specific trial on day D:

1. If there are zero fields on day $D+1$, mark as 'lost'
2. If there is more than one field on day $D+1$ and a single field on day D, mark as 'multiple fields after'
3. If there is more than one field on day $D+1$ and more than one field on day D, mark as 'multiple fields both'
4. If there is more than one field on day D, mark as 'multiple fields before'
5. At this point, the only possibility is that each day has exactly one field. If those fields match (according to the matching criteria in 4.8.1) mark as 'stable'
6. Mark as 'moved'

Chapter 5

Unsupervised adaptive correction of representational drift

Acknowledgement

The analysis, figures, and several discussion points described in this chapter are adapted from *Heavy-tailed statistics of cortical representational drift are advantageous for stabilised downstream readouts* (Micou and O'Leary, 2024). The only new material not covered by this pre-print pertains to Fig. 5.10 and Fig. 5.11, in which we extend the analysis to the data collected from our BMI experiments.

Chapter overview

This chapter deals with a fundamental question in representational drift: how can downstream consumers of the information encoded within an upstream population cope with ongoing changes in how that information is represented? We focus specifically on compensating for drift without access the ground truth. In brief, we:

- Describe an entirely unsupervised scheme for detecting and correcting for changes in tuning in terms of a didactic toy model
- Use our model to illustrate that many stochastic processes at the level of neurons can generate equivalent rates of drift at the level of entire populations, but that those processes favouring sudden 'jumps' in tuning are easier to correct for
- Predict that the advantages of these jumps are particularly relevant at high drift rates
- Investigate three in vivo datasets from V1, PPC, and CA1, and determine that they contain a mixture of tuning change statistics that favours sudden retuning
- Find that the preference for jump-like retuning is stronger in more plastic populations

5.1 Introduction

5.1.1 Motivating unsupervised adaptive decoding

At the end of the previous chapter, we described how the seemingly disparate BMI-specific and locomotion-specific representations of the track did, in fact, share a common basis for the representation of the environment (Fig. 4.15), but that this was only visible when considering a population two orders of magnitude greater than our individual recording sessions. A decoder averaging over such a large population would be able to recover good accuracy in spite of the remapping, but this decoder would nevertheless be stuck with detrimental weights corresponding to neurons that have retuned in the new context. If such changes were to accumulate over time, the decoder would gradually lose accuracy—that is, unless it implemented some mechanism to determine which neurons were no longer part of the representation and which neurons had newly joined the representation. In essence, this is what we did when we retrained our decoder on the BMI representation of the track and subsequently achieved good closed-loop performance. The way we retrained our decoder, however, is cheating: we had access to the ground-truth of the animal's position on the track during the BMI trials. We described, back in the opening paragraphs of this thesis, the importance of designing adaptive decoders that could compensate for changes in tuning without the need for supervised retraining.

BMI engineers are not the first to encounter this design problem. Individual neurons within populations have no direct access to a ground truth. Instead, cortical populations must build up a representation of the external world purely through statistical regularities in upstream activity. We speculate, then, that these populations must be well-suited to the task of determining which changes in upstream activity correspond to noisy observations, and which correspond to fundamental changes in representation.

We take representational drift as an ideal minimal example to study how the brain might tackle such adaptive decoding. In experimental observations of drift, imaged populations encode the same information on every day, at the same fidelity, and using the same general strategies of encoding—but the role each neuron plays in its representation changes over time. After repeated drift, the accumulation of nonstationary changes in representation will eventually drive a fixed decoder to an accuracy no better than chance. However, an adaptive decoder capable of determining how neurons have been altered by drift may remain accurate indefinitely. We focus specifically on unsupervised adaptation by considering statistical regularities in activity (Fig. 1.3, from the introduction to this thesis, provides a helpful illustration).

Drift is a process that is gradual at the level of entire populations: a representation on one day will look similar on the following day. The crux of the argument in this chapter is that a variety of stochastic processes at the level of individual neurons can generate drift that

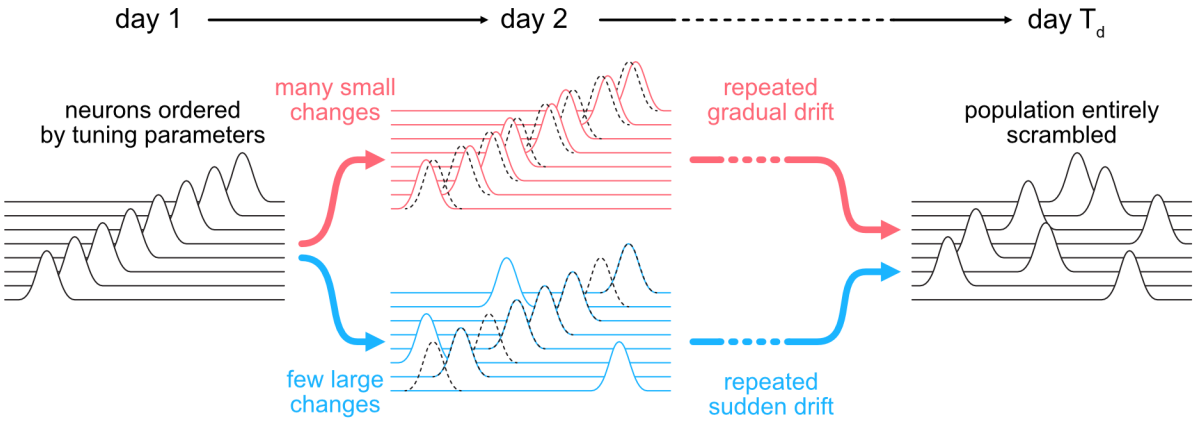


Fig. 5.1 A variety of underlying stochastic processes at the level of individual neurons can produce drift at the level of a population code. For example, a drift paradigm in which all neurons experience gradual drift in their tuning parameters will eventually scramble a population code beyond recognition. However, this could also be achieved through a 'sudden drift' random process that generates large changes in the tuning of a different random subset of neurons on each day.

appears gradual at the level of populations (Fig. 5.1), but that these are not equally easy for a downstream readout to adapt to. We first present a toy model that contrasts two different types of drift that degrade a fixed linear readout at exactly the same rate, but that have different underlying neuron-level statistics. The first model, which we call 'gradual', slowly alters the tuning of each neuron in the population (as is common in computational models of drift, see Rule and O'Leary (2022) for an example). The second model, which we call 'sudden', instead makes dramatic changes in tuning to a random subset of the population. We describe why this second model is advantageous, and argue that real, in vivo brains would benefit from these 'heavy-tailed' distributions for neuron retuning. Finally, we perform tentative analysis of two existing drift experiments and revisit our own hippocampal BMI experiments to determine the extent to which this prediction holds.

5.1.2 Existing representational drift datasets

We have access to two existing and well-studied datasets that observe representational drift, both consisting of 2-photon Calcium imaging recordings in mice.

The first of these is from Driscoll et al. (2017), which contains recordings from the PPC during performance of a T-maze task. This experiment will be familiar to the reader: it is the same experiment design we used for evaluating open-loop decoders on PPC (see Section 2.2.3). It consists of a long corridor broken into two sections terminating in a decision point at which the animal must turn left or right. The first section displays wall cues which indicate whether a left or right turn will be rewarded, and the second section has blank walls so that

the turn direction must be remembered. We have also previously reproduced ratemaps from Driscoll et al. (2017) as an illustration of drift (Fig. 1.2.a). These capture the firing rate of neurons as a function of distance along the corridor (not unlike place cells) over the course of several days and indeed show a gradual turnover in the representation. The original work quantifies drift both via explicit paramterisation (whether the neurons have preferences for left or right turns, where their peak of activity is located along the track, and whether or not neurons gain or lose peaks in firing location from one day to the next) and by verifying the gradual degradation of a fixed decoder over time.

The second dataset is from Marks and Goard (2021), which contains recordings from the V1 region of the visual cortex during exposure to movies both of natural scenes and of oriented gratings. While neuron firing rates are easy to parameterise in terms of the orientation of a grating, alternative methods become necessary for the movie stimulus. Instead, the authors quantify drift by determining whether or not neurons are mutually tuned to similar stimuli by measuring the correlation between pairs of neurons within the population, and are thereby able to construct schematics of functional connectivity and track changes in this connectivity over time to infer rates of drift. The original work eventually concludes that drift rates are substantially faster for the movie stimulus than the grating stimulus.

These experiments differ dramatically in the rates at which they observe drift in the population. Drift in the PPC is visible from one day to the next, while drift in V1 takes place from week to week. While some of this discrepancy in drift rates is necessarily due to vastly different task design and choices of method to quantify drift, we attribute a component of the discrepancy to underlying differences in the plasticity of the populations, consistent with the expectation that more plastic populations drift faster (Driscoll et al., 2022; Micou and O'Leary, 2023; Rule et al., 2019).

5.2 Modelling adaptive decoding

5.2.1 Adaptive decoding via continual tuning parameter inference

The objective of an adaptive decoder is to produce an accurate readout from the population in spite of drift. Each day, the adaptive decoder updates its estimates of how individual neurons are tuned to stimuli. The adaptive decoder should not 'cheat' with knowledge of what takes place outside of the neural population: it must instead use changes in the statistical regularities of the population to infer changes in the tuning of individual neurons.

Concretely, we consider a model neural population of N neurons in which ideal responses $\mathbf{x} \in \mathbb{R}^N$ to a given stimulus $\mathbf{s} \in \mathbb{R}^K$ can be parameterised by P tuning parameters per neuron. This allows us to express the complete tuning of the population as $\boldsymbol{\theta} \in \mathbb{R}^{N \times P}$. Individual neurons, whose ideal firing responses are subject to independent random noise ξ, have their

responses defined by an activation function a common to all neurons in the population that relates tuning parameters to activity:

$$x_i = a(\mathbf{s}, \theta_i) + \xi \tag{5.1}$$

On some day t of an experiment, we take M samples of neural activity in response to stimulus exposures ($\mathbf{X}_t \in \mathbb{R}^{N\times M}$ and $\mathbf{S}_t \in \mathbb{R}^{K\times M}$ respectively). For a given set of tuning parameters $\boldsymbol{\theta}$, we can use the activation function a, a prior distribution for stimulus exposure, and a model of the noise process for ξ to construct the function g, an approximation[1] of the likelihood of M observations of neural activity $\mathbf{X}$ that does not depend on knowledge of the stimulus values $\mathbf{s}$.

$$p(\mathbf{X}|\boldsymbol{\theta}) \approx g(\mathbf{X}, \boldsymbol{\theta}) \tag{5.2}$$

The objective of our model adaptive decoder is to maintain an accurate estimate of tuning in the population, $\widehat{\boldsymbol{\theta}}_t$, despite ongoing drift and without knowledge of $\mathbf{s}$. Naïvely, we might attempt to recover $\widehat{\boldsymbol{\theta}}_t$ from $\mathbf{X}_t$ directly by finding the parameters that maximise $g(\mathbf{X}, \boldsymbol{\theta})$. However, because the activation function a is common to all neurons, the solution-space for maximising $g(\mathbf{X}, \boldsymbol{\theta})$ has symmetries that render the solution-space degenerate. Because a describes all neurons as interchangeable, a set of parameters maximising $g(\mathbf{X}, \boldsymbol{\theta})$ could be shuffled and assigned to entirely different neurons without affecting the solution. Without their history, neurons lack identity.

By considering the tuning of neurons on the previous day, $\boldsymbol{\theta}_{t-1}$, we are able to break these symmetries. We use a stochastic model of how drift perturbs the population from one day to the next, $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$. This produces a maximum a posteriori update rule for detecting tuning changes in the population.

$$\widehat{\boldsymbol{\theta}}_t = \arg\max_{\boldsymbol{\theta}} \mathscr{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}_t) \tag{5.3}$$

$$\text{where} \quad \mathscr{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}_t) = \log\left(p(\mathbf{X}_t|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}_{t-1})\right) \tag{5.4}$$

Treating drift as a stochastic process that is independent of neuron identity and in which each neuron drifts independently from other neurons, we can express $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$ in terms of a per-neuron drift prior f.

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) = \prod_i^N f(\theta_{i,t}, \theta_{i,t-1}) \tag{5.5}$$

[1] In practice, as it is biologically implausible for a downstream neuron attempting to stabilise a readout to remember every observation of upstream activity, we base this approximation on summary metrics such as the correlations between the activities of upstream neurons.

We then combine our model for observed activity g with our description for drift f and express the daily update rule for $\widehat{\boldsymbol{\theta}}$ as:

$$\widehat{\boldsymbol{\theta}}_t = \underset{\boldsymbol{\theta}}{\arg\max} \left(\log g(\mathbf{X}_t, \boldsymbol{\theta}) + \sum_i^N \log f(\theta_i, \widehat{\theta}_{i,t-1}) \right) \tag{5.6}$$

5.2.2 Exploiting statistical regularities in a simulated population model

As an illustrative example of this adaptive decoding technique, we consider a simple model of a neural population that encodes the direction and the magnitude of a stimulus: a system that is reminiscent of in vivo populations both of head-direction coding in the hippocampal formation and the response of populations in the visual cortex to grating orientations (Fig. 5.2.a). The population responds to a stimulus that corresponds to a point within the unit circle defined by a magnitude and direction (i.e. $K = 2$). During simulated experiments, these stimulus values are both drawn from uniform distributions at each sample:

$$\mathbf{s} = \begin{pmatrix} r \\ \phi \end{pmatrix} \sim \begin{pmatrix} U(0,1) \\ U(0,2\pi) \end{pmatrix} \tag{5.7}$$

Neurons each have a single tuning parameter ($P = 1$), their preferred stimulus angle, and are subject to additive Gaussian noise.

$$x_i = a(r, \phi, \theta_i) + \xi \tag{5.8}$$

$$\text{where} \quad \xi \sim N(0, \sigma^2) \tag{5.9}$$

$$\text{and} \quad a(r, \phi, \theta_i) = \begin{cases} r\cos^2(\theta_i - \phi) & \text{if } 0 \leq \angle(\theta_i, \phi) \leq \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

Where for the convenience of dealing with the circular geometry of this example, we have defined $\angle(\psi_1, \psi_2)$ as the magnitude of the smallest angle between the angles ψ_1 and ψ_2.

On the first day, we 'cheat' by allowing the model to see to the ground-truth values of the stimulus to initialise the values of $\widehat{\boldsymbol{\theta}}_{t=1}$. These stimulus values are not available to the decoder on subsequent days: it then has access only to observations of population activity and knowledge of its own estimate of tuning on the previous day (Fig. 5.2.b). Effectively, we treat the initial learning as a *fait accompli*, and are instead concerned with how well our adaptive decoder is able to retain an accurate estimate of tuning over time.

To construct an expression for $g(\mathbf{X}, \boldsymbol{\theta})$, we consider pairs of neurons and the correlation of their activities within a session. Neurons with similar tuning values are more likely to be highly correlated with each other (Fig. 5.2.c). The correlation between a pair of neurons with some difference $\Delta\theta$ between their preferred firing angles is a probability distribution governed

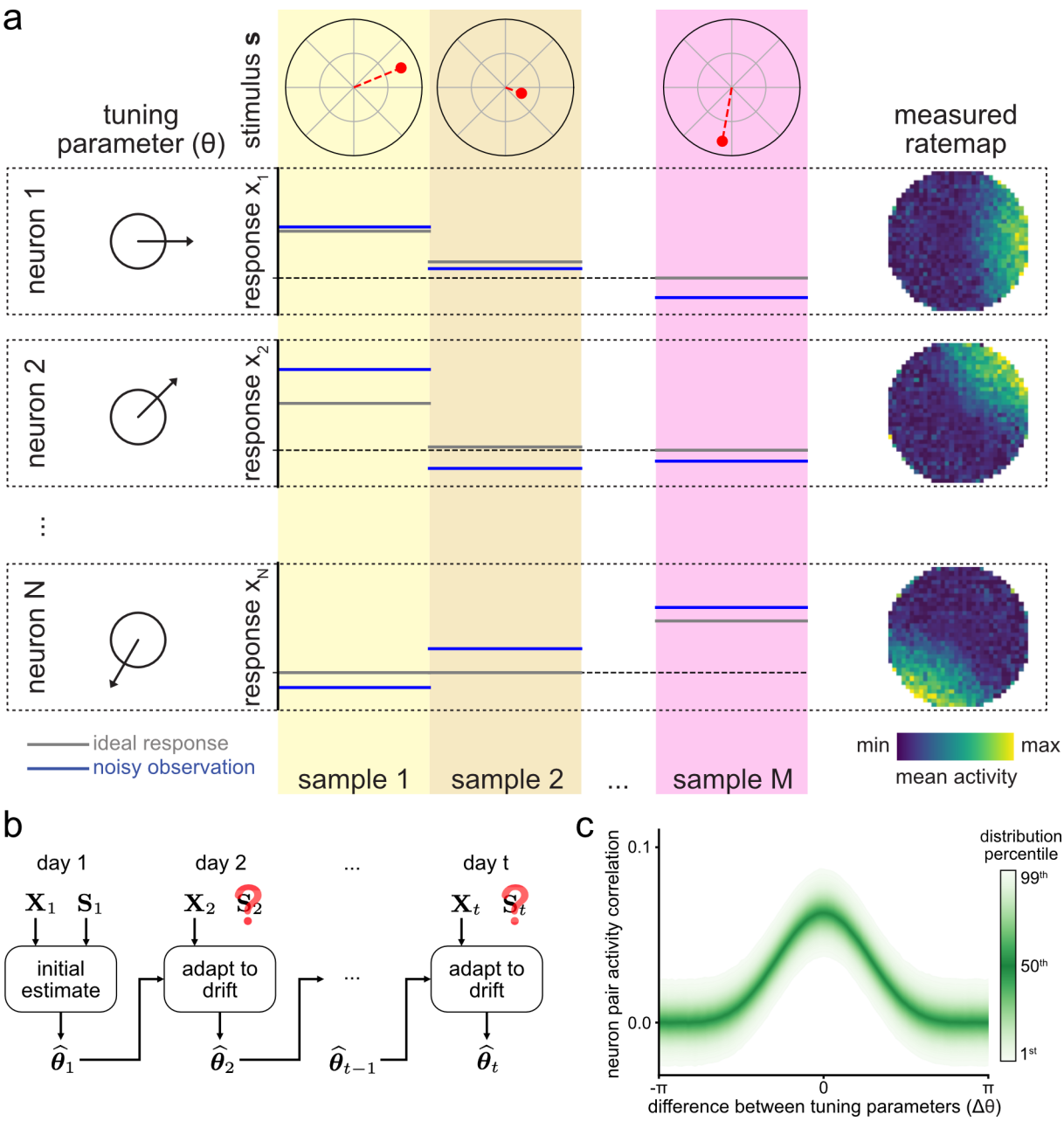


Fig. 5.2 A toy model of stimulus encoding for exploring adaptive decoding. **(a)** A model of a neural population that encodes the orientation and magnitude of a stimulus. Each neuron in the population has a tuning parameter that describes its preferred orientation. The population is sampled multiple times during a session, with the activity of each neuron independently affected by noise at each sample. **(b)** Knowledge of the stimulus ground-truth from an initial experimental session is used to initialise an estimate of the tuning parameters that describe the relationship between stimulus and population activity. Subsequently, and with no further reference to the stimulus ground-truth, an adaptive decoding scheme attempts to keep these estimates accurate while the underlying tuning parameters drift. **(c)** Visualisation of the probability distribution for the value of the correlation between a pair of neurons as a function of the difference between their tuning parameters (simulation parameters $M = 10^4$ and $\sigma = 1.0$).

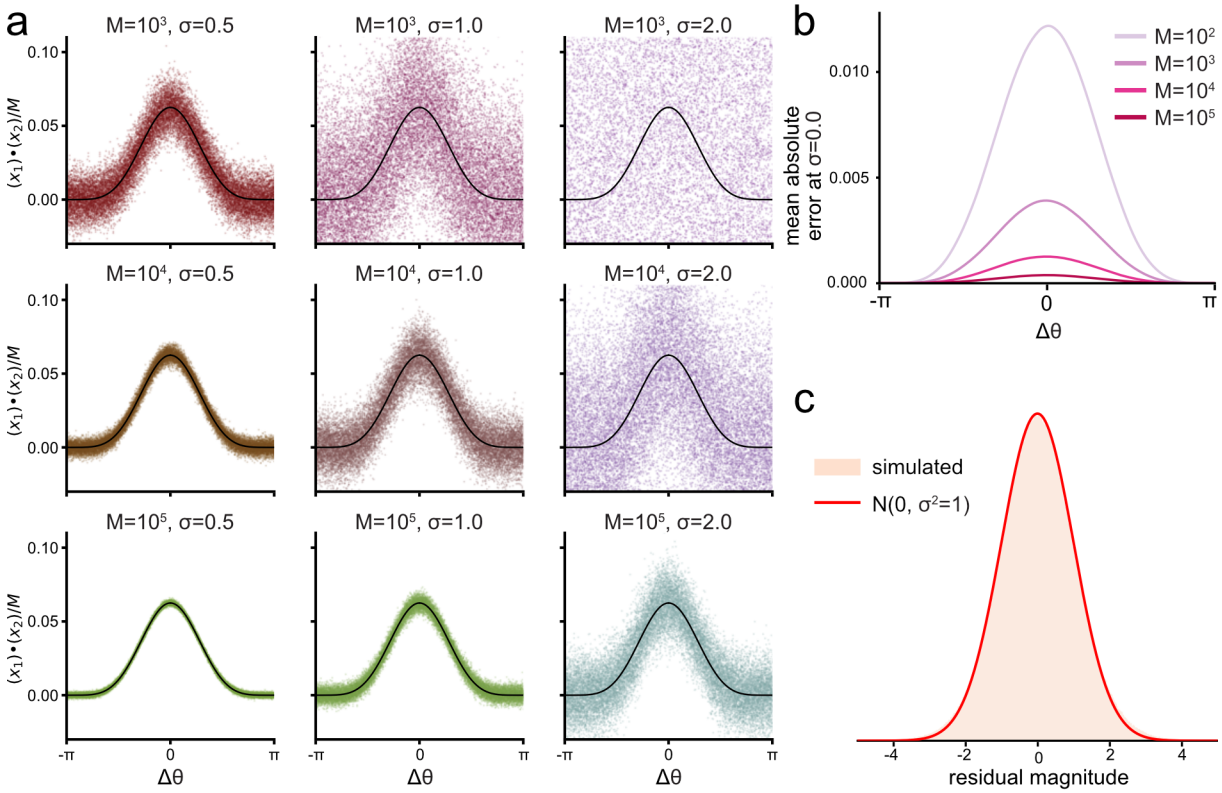


Fig. 5.3 Statistics of simulated neuron pair correlations closely match analytical forms. **(a)** Visualisations of the distribution of correlations between neuron pairs separated by an angular distance $\Delta\theta$ for a variety of observation noise levels σ and total numbers of samples M. The black trace shows the approximation for the expected value of the correlation. **(b)** The mean absolute error between the approximation for the expected value of the pairwise correlation and its actual value as computed by a Monte Carlo simulation, shown in the absence of any observation noise ($\sigma = 0$) as a function of angular distance between the two neurons. **(c)** Distribution of the error residual, the difference between the expected value of the pairwise correlation and its actual value, in the presence of observation noise ($\sigma = 1.0$), presented with a comparison to the normal distribution, simulated at $M = 10^3$.

by two underlying random processes: the M random samples observed on any given day, and the observation noise ξ perturbing those observations (Fig. 5.3.a).

Taking M to be sufficiently large (Fig. 5.3.b), the expected value of the correlation between two neurons of a (positive) angular separation $\Delta\theta$ is well-approximated by:

$$E(\Delta\theta) = \frac{1}{6\pi}\frac{1}{16}\left(3\sin 2\Delta\theta + 2(\pi - \Delta\theta)(\cos 2\Delta\theta + 2)\right) \tag{5.11}$$

The distribution of the residual between the expected value and the measured value is dominated by the additive noise ξ and is therefore normally distributed (Fig. 5.3.c). This allows a simple expression for the likelihood of the correlation a pair of neurons $\mathbf{x}_i$ and $\mathbf{x}_j$ with tuning

parameters θ_i and θ_j observed M times within a session:

$$p(\mathbf{x}_i, \mathbf{x}_j | \theta_i, \theta_j) \approx h(\mathbf{x}_i, \mathbf{x}_j, \theta_i, \theta_j) \tag{5.12}$$

$$= \frac{\sqrt{M}}{\sigma\sqrt{2\pi}} e^{-\frac{M(\mathbf{x}_i \cdot \mathbf{x}_j / M - E(\angle(\theta_i, \theta_j)))^2}{2\sigma^2}} \tag{5.13}$$

Treating the probabilities of pairwise correlations as independent, which is an approximation of $p(\mathbf{X}|\boldsymbol{\theta})$, we can then construct an expression for $g(\mathbf{X}, \boldsymbol{\theta})$:

$$g(\mathbf{X}, \boldsymbol{\theta}) = \prod_i^N \prod_j^N h(\mathbf{x}_i, \mathbf{x}_j, \theta_i, \theta_j) \tag{5.14}$$

When solving the component of $g(\mathbf{X}, \boldsymbol{\theta})$ for the optimisation of $\mathscr{L}$ for $\widehat{\boldsymbol{\theta}}_t$, this amounts to minimising the sum of squared differences between observations of the correlations between neuron pairs and their expected values $\boldsymbol{E}$ as parameterised by $\boldsymbol{\theta}$.

Useful properties of the toy model

This slightly contrived model of a population has several noteworthy properties that make it a convenient illustrative example. First, it is easy to relate changes in the tuning parameters to the accuracy of a decoder of stimulus orientation. The mean squared error of the parameter estimate, $||\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t||^2 / N$, is directly equivalent to the expected squared error of a simple decoder that outputs a predicted orientation based on a circular average of tuning parameters weighted by firing rate. In this context, we can therefore talk about parameter estimate error and adaptive decoder error interchangeably.

Secondly, the need to break symmetries using a drift prior is obvious: while $g(\mathbf{X}, \boldsymbol{\theta})$ helps determine the angular differences between neuron tuning parameters, it provides no information about the absolute orientation of any tuning parameter. Knowledge of how the population was tuned on an earlier day, as well as a drift prior describing how that tuning is likely to change, is required to anchor the system to an absolute frame of reference.

Finally, the function $g(\mathbf{X}, \boldsymbol{\theta})$ is smooth for fixed $\mathbf{X}$ by construction, making its contribution to the likelihood $\mathscr{L}$ (as previously described in equation 5.6) well-conditioned for solvers based on gradient ascent.

5.2.3 Contrasting statistics of drift at the level of individual neurons

We consider two simple but quite opposite models for drift in individual neuron tuning: a 'gradual' model, in which all neurons undergo relatively small tuning changes from one day to the next, and a 'sudden' (or 'heavy-tailed') model, in which most neurons retain their tuning

while a small subset retune by making relatively large changes to their tuning (Fig. 5.1, Fig. 5.4.a).

In the gradual drift model, each parameter θ_i moves around the unit circle as if buffeted by Brownian motion: it is described by an Ornstein-Uhlenbeck process. The new value of a tuning parameter measured on a subsequent day is parameterised by ρ, the expected absolute distance drifted.

$$\theta_{i,t}|\theta_{i,t-1} \sim N(\theta_{i,t-1}, \rho^2) \tag{5.15}$$

$$f_{\text{gradual}}(\theta_{i,t}, \theta_{i,t-1}) = \frac{1}{\rho\sqrt{2\pi}} e^{-\frac{(\angle(\theta_{i,t}, \theta_{i,t-1}))^2}{2\rho^2}} \tag{5.16}$$

In the sudden drift model, each neuron undergoes a spontaneous retuning instead of gradually adjusting its tuning parameter. On any given day, it either retains its previous tuning parameter or selects an entirely new value. This distribution is parameterised by the spontaneous retuning probability α.

$$f_{\text{sudden}}(\theta_{i,t}, \theta_{i,t-1}) = \begin{cases} \frac{\alpha}{2\pi} + (1-\alpha)\delta(\theta_{i,t} - \theta_{i,t-1}) & \text{for } 0 \le \theta_{i,t} < 2\pi \\ 0 & \text{otherwise} \end{cases} \tag{5.17}$$

Where δ denotes the Dirac delta function.

To allow a fair comparison between sudden and gradual drift models, we define the drift rate of two models to be equivalent when the repeated application of drift scrambles the output of a fixed decoder in the same expected number of days T_d.

$$\text{drift rate} = \frac{1}{T_d} \tag{5.18}$$

In the case of our illustrative example, we quantify the change in an individual neuron's tuning between day 1 and day t as the angular distance between tuning parameters $\angle(\theta_1, \theta_t)$. As drift progresses, the expected value of the change in tuning approaches the expected value between the initial tuning parameter θ_1 and one chosen entirely at random θ_r.

$$\lim_{t\to\infty} E[\angle(\theta_1, \theta_t)] = E[\angle(\theta_1, \theta_r)] = \frac{\pi}{2} \tag{5.19}$$

We describe a population as 'scrambled on day T_d' if its similarity to the original population is reduced to within an ϵ of chance, i.e. it satisfies:

$$\frac{E[\angle(\theta_1, \theta_r)] - E[\angle(\theta_1, \theta_{T_d})]}{E[\angle(\theta_1, \theta_r)]} < \epsilon \tag{5.20}$$

Which allows us to generate an equivalence between the parameters ρ and α of the two drift models (Fig. 5.4.b). We choose $\epsilon = 0.05$, noting that the equivalence is not sensitive to this choice of parameter for small ϵ (Fig. 5.5.a-b). While the two models of drift provide indistinguishable long-term behaviour, the differences in short-term, transient behaviour have the potential for material impact on our adaptive decoding strategy.

5.2.4 Remarks on solving the optimisation

Solving the optimisation for the gradual drift model

We use gradient ascent to solve for the most likely values of $\widehat{\boldsymbol{\theta}}_t$.

$$\mathscr{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}_t) = \sum_i^N \sum_j^N \log h(\mathbf{x}_i, \mathbf{x}_j, \theta_i, \theta_j) + \sum_i^N \log f(\theta_i, \widehat{\theta}_{i,t-1}) \tag{5.21}$$

The left summation depends only on the neuron activation model and the right summation depends only on the drift model. In this simple example, we can find analytical forms of the Jacobian and the Hessian, which makes solving via gradient ascent faster. This makes use of two important properties of h. First, it is symmetrical over i, j:

$$h(\mathbf{x}_i, \mathbf{x}_j, \theta_i, \theta_j) \equiv h(\mathbf{x}_j, \mathbf{x}_i, \theta_j, \theta_i) \tag{5.22}$$

Second, in the special case when $i = j$, all choices of θ_i are equivalent.

$$\frac{\partial}{\partial \theta_i} \left(h(\mathbf{x}_i, \mathbf{x}_i, \theta_i, \theta_i) \right) = 0 \tag{5.23}$$

Combining these properties and taking the derivative of the log-likelihood with respect to a single parameter:

$$\frac{\partial \mathscr{L}}{\partial \theta_i} = 2 \sum_{j \neq i}^N \left(\frac{\partial}{\partial \theta_i} \left(\log h(\mathbf{x}_i, \mathbf{x}_j, \theta_i, \theta_j) \right) \right) + \frac{\partial}{\partial \theta_i} \left(\log f(\theta_i, \widehat{\theta}_{i,t-1}) \right) \tag{5.24}$$

And then, to produce the Hessian:

$$\frac{\partial^2 \mathscr{L}}{\partial \theta_i \partial \theta_j} = \begin{cases} \frac{\partial^2}{\partial \theta_i^2} \left(\log f(\theta_i, \widehat{\theta}_{i,t-1}) \right) & \text{if } i = j \\ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left(\log h(\mathbf{x}_i, \mathbf{x}_j, \theta_i, \theta_j) \right) & \text{otherwise} \end{cases} \tag{5.25}$$

We initialise of the parameters for the gradient ascent is to use their values from the previous day, $\widehat{\boldsymbol{\theta}}_{t-1}$. Complete expressions and their derivations are available in the supplementary material of Micou and O'Leary (2024).

Exact solution to the optimisation for the sudden drift model

The discontinuity in f_{sudden} means that it is ill-conditioned for a gradient ascent-based solution applied to the entire parameter space. Instead, we consider a solution that decomposes the problem into smaller components that can be individually solved by gradient ascent (and, in the subsequent section, a technique for approximating this solution for large populations).

For each neuron, there are two possible cases: either the neuron keeps its tuning parameter, or it retunes. In the first case, trivially $\widehat{\theta}_{i,t} = \widehat{\theta}_{i,t-1}$. In the second case, the neuron needs a tuning parameter update *and* should be withheld from use in estimating the parameters of other neurons, as $\widehat{\theta}_{i,t-1}$ is no longer a good estimate of $\theta_{i,t}$. To solve, we perform two steps:

- Infer which neurons have kept their tuning parameter intact
- Use this subset of neurons to update the parameters of the retuned neurons

Let the set A contain the neurons which have retained their tuning parameters. A neuron i is a member of this set if and only if $u_i = 1$, where:

$$u_i(\mathbf{X}, \widehat{\boldsymbol{\theta}}_{t-1}) = \begin{cases} 1 \text{ if } P(\mathbf{X}|\text{changed})P(\text{changed}) < P(\mathbf{X}|\text{changed}')P(\text{changed}') \\ 0 \text{ otherwise} \end{cases} \tag{5.26}$$

In the concrete example for the sudden drift model, the above condition can be written as:

$$\int_0^{2\pi} g(\mathbf{x}_i, \psi, \widehat{\boldsymbol{\theta}}_{t-1}) \frac{\alpha}{2\pi} \mathrm{d}\psi < g(\mathbf{x}_i, \widehat{\theta}_{i,t-1}, \widehat{\boldsymbol{\theta}}_{t-1})(1-\alpha) \tag{5.27}$$

Where $g(\mathbf{x}_i, \psi, \widehat{\boldsymbol{\theta}}_{t-1})$ is the probability of the observations $\mathbf{x}_i$ given a tuning angle ψ. This allows us to express the update rule as:

$$\widehat{\boldsymbol{\theta}}_t = \arg\max_{\boldsymbol{\theta}} \mathscr{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}) \simeq \begin{bmatrix} z_0(\theta_0, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}) \\ \dots \\ z_i(\theta_i, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}) \\ \dots \\ z_N(\theta_N, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}) \end{bmatrix} \tag{5.28}$$

Where:

$$z_i(\theta_i, \widehat{\boldsymbol{\theta}}_{t-1}, \mathbf{X}) = \begin{cases} \widehat{\theta}_{i,t-1} & \text{if } i \in A \\ \arg\max_{\boldsymbol{\theta}} \sum_{j \in A} \log h(\mathbf{x}_i, \mathbf{x}_j, \theta_i, \widehat{\theta}_{j,t-1}) & \text{otherwise} \end{cases} \tag{5.29}$$

Noting that the use of the approximately equal symbol, as this solution discards the small amount of information available in the pairwise mutual tuning between neurons not in A.

Approximate solution to the optimisation for the sudden drift model

In practice, evaluating the condition exactly for u_i becomes computationally intractible for large M due to resolution limitations of floating point number representations: it requires the product of very many small numbers. Circumventing this limitation in the typical fashion—a sum of logarithms—requires a closed form of the integral of g in terms of the tuning parameters, which we have no guarantee of being able to procure for even relatively straightforward task formulations. We instead choose to make an approximation in determining set membership of A.

We can trivially evaluate $\log g(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}_{t-1})$ for each neuron in the population and are therefore able to rank each neuron in the population by conformity to expectation. For a drift rate parameterised by α and a number of neurons in the population N, the number of neurons that drifted is binomially distributed. We can select a probability threshold q (by using the CDF of the binomial distribution) that results in us excluding the bottom Q neurons from A by ranking.

The choice of q is a hyperparameter. If q is large (e.g. 0.95), many of the neurons excluded from A are likely to still have accurate estimates of $\theta_{i,t}$. For small N and high drift rates (large α), this can exclude a sizeable portion of the population and reduce the accuracy of tuning parameter estimates. Conversely, if q is small (e.g. 0.5), neurons that require re-estimating will be included in A and not have their parameters updated. In all simulations shown in this work, we set $q = 0.9$.

This approximate solution has the potential to disadvantage an adaptive decoder compensating for sudden drift when compared to one compensating for gradual drift. Indeed, we will see artefacts corresponding to spurious corrections of well-tuned neurons at very high levels of observation noise σ that can render this adaptive decoder *worse* than a fixed decoder (Fig. 5.4.c). However, given our choice of neuron population sizes *and* our observation in the following section that sudden drift outperforms gradual drift in spite of this disadvantageous approximation, we take our use of this approximation to be immaterial to our conclusions.

Implementation of solvers

The gradient ascent components of these solutions are implemented using the `scipy`'s library's (Virtanen et al., 2020) `optimize.minimize` function (in the case of the multivariate optimisation for the gradual case) and the `optimize.minimize_scalar` function (in the case of the independent scalar optimisations for the sudden case). The underlying solver method is based on `L-BFGS-B` (Liu and Nocedal, 1989), with solution bounds for each parameter θ_i set to -2π to 2π and only remapped to the range $-\pi$ to π after convergence.

5.2.5 Sudden drift is advantageous for adaptive decoding

In simulations that apply the adaptive decoding scheme to our illustrative model, there is an obvious discrepancy between gradual and sudden drift in the difficulty of distinguishing statistical fluctuations from nonstationary changes. On the very first day of drift, the adaptive decoder can accurately infer changes in the tuning parameters of the population for both types of drift under low observation noise (parameterised by σ). However, as σ increases, adaptive decoders make larger errors when inferring tuning parameters subject to gradual drift (Fig. 5.4.c), as it is harder to distinguish gradual drift (a nonstationary change in tuning) from a statistical fluctuation (the added noise ξ).

We proceed with applying the adaptive decoding scheme from one day to the next, on each day inferring tuning parameters based on the previous day's tuning parameter estimates. Adaptive decoders remain accurate long after fixed decoders perform no better than chance. However, errors compound, and this eventually degrades the readout of adaptive decoders (Fig. 5.4.d).

As the number of neurons in the population increases, adaptive decoders maintain their accuracy for longer for two reasons (Fig. 5.4.e). First, while additional neurons require estimating an increasing number of parameters, the introduction of those neurons provide more neuron pairs with which to narrow the distribution of $p(\mathbf{X}|\boldsymbol{\theta})$. The number of additional parameters to estimate grows with N, but the number of pairs available grows with N^2, which reduces the likelihood of poorly estimating the tuning of neurons. Second, additional neurons provide redundancy that is relevant over the course of many days of drift: it is statistically inevitable that, eventually, the tuning parameter of an individual neuron will be poorly estimated. With greater redundancy in the population, it is easier to notice an individual poor tuning parameter estimate on some later day, allowing subsequent correction of the estimate and forestalling compounding errors. Sudden drift has an advantage not only because it introduces smaller errors under noisy observations from one day to the next (Fig. 5.4.c), but also because it can more effectively make use of redundancy to fix error compounding (note the flattening gradient in Fig. 5.4.d and the better scaling with population size in Fig. 5.4.e), as from the perspective of the adaptive decoder a neuron with a low likelihood of its estimated tuning parameter is no different from a neuron that has spontaneously retuned.

Unsurprisingly, both gradual and sudden drift become easier to correct for under lower rates of drift, as the distribution of changes in tuning is narrower. Nevertheless, for the same population size and the same rate of drift, sudden drift is always easier to adapt to (Fig. 5.4.e). The learning that takes place within real neural populations may necessitate gradual changes, but supporting those gradual changes would require either larger populations or slower changes to achieve error rates equivalent to what is achievable with sudden tuning changes.

This presents a trade-off for the brain: if a learning rule requires updates applied to many neurons simultaneously, populations need either to learn more slowly or to a use more redundant encoding. We therefore predict that the brain will favour sudden tuning changes over gradual ones to support learning with smaller populations in the interest of efficiency.

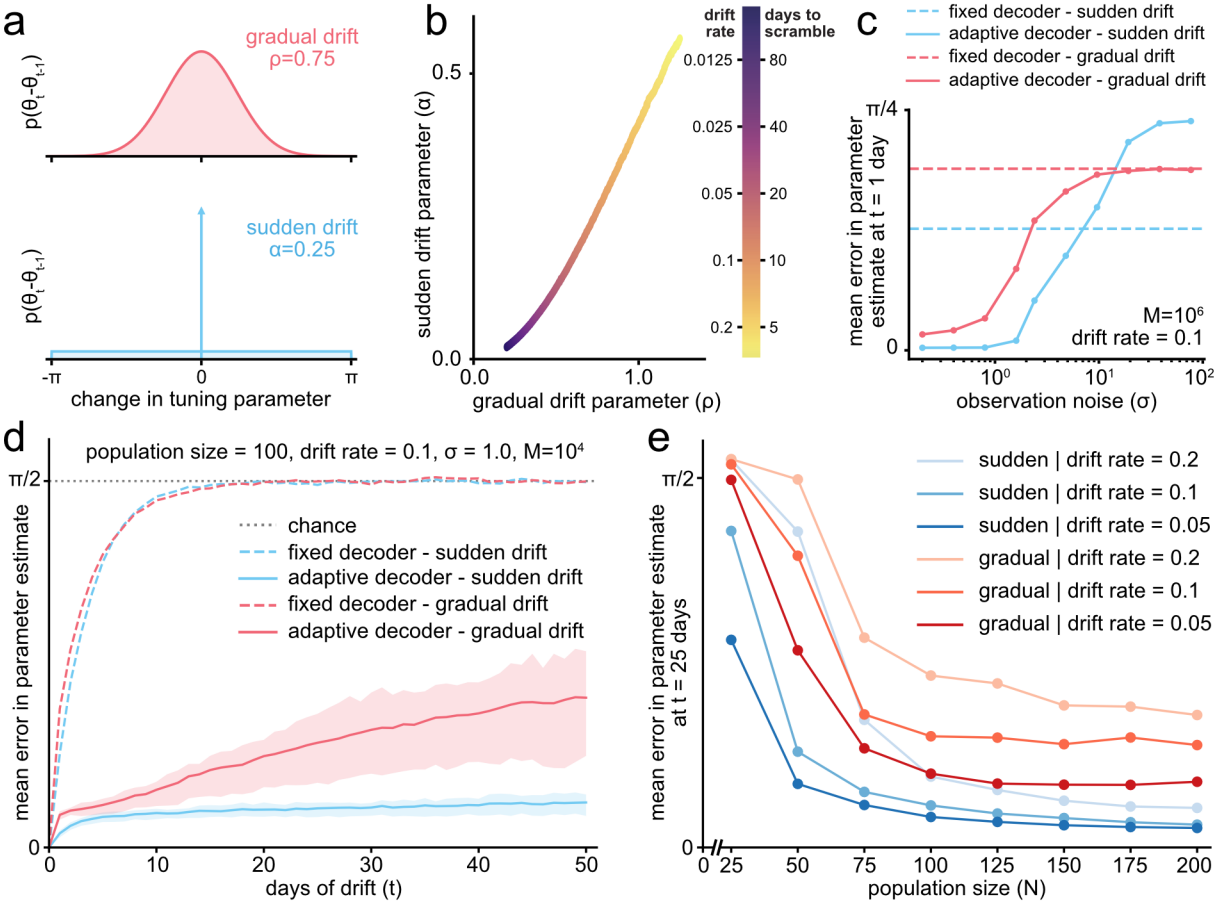


Fig. 5.4 Sudden drift is easier to compensate for than gradual drift. **(a)** The two example models of drift can be expressed as probability distributions describing the change in each neuron's tuning parameter from one day to the next. **(b)** Equivalence between the two drift models determined by the expected number of days required to drive the error of a fixed decoder to within 5% of chance. **(c)** The resulting error in tuning parameter estimates after the first day of drift, shown as a function of the observation noise σ for both drift models. **(d)** Comparison of the accuracy of parameter estimates between fixed and adaptive decoders for both sudden and gradual drift models, both using drift rates of 0.1. Shaded regions show the interquartile range over 100 simulations. **(e)** The estimation error after 25 days shown for a variety of drift rates (all sufficiently high to scramble the population before 25 days) and for a variety of population sizes. Each datapoint represents the mean of 100 simulations.

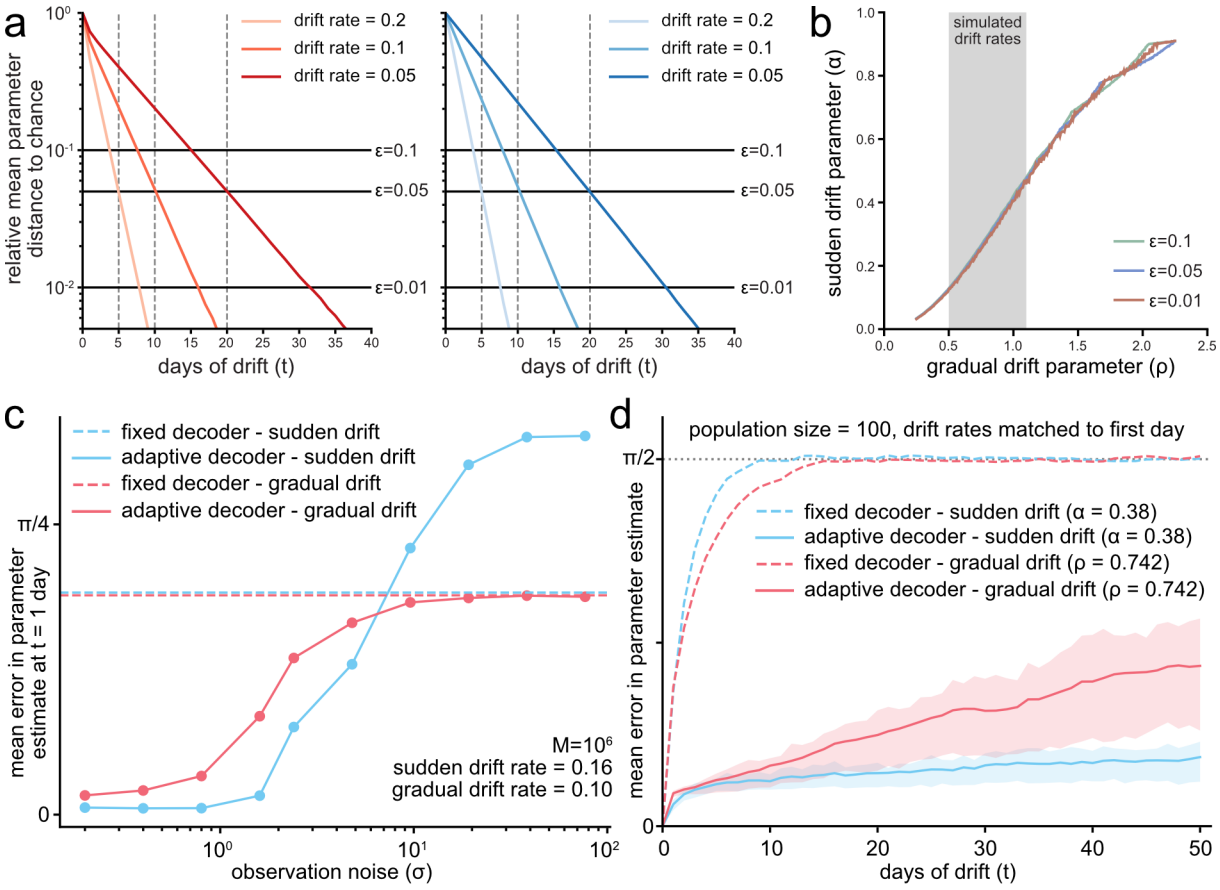


Fig. 5.5 Sudden drift retains its advantage even with less generous equivalences between sudden and gradual drift rates. **(a)** The approach of tuning parameters to their chance values over the course of several days of drift, mean of 10^6 simulations. Shown for the gradual drift model (left, red) and the sudden drift model (right, blue). Drift rates in the legend listed for $\epsilon = 0.05$. **(b)** The equivalence mapping between the gradual and sudden drift parameters for multiple values of ϵ, highlighting the range of gradual drift rates used in the simulations of this chapter. **(c)** As in Fig. 5.4.c, but using an alternative drift equivalence between sudden and gradual drift. This equivalence induces the same amount of error in a fixed decoder after the first day of drift. **(d)** As in Fig. 5.4.d, but using the alternative drift equivalence from (c).

Remarks on the equivalence of drift rates

Before we move on to explore the consequences of these observations, we seek to rule out the possibility that they are based on a false equivalence between rates of drift across the two types of drift. We recall first that we based our drift rate equivalence on scrambling a population's encoding to within some ϵ of chance, and that we arbitrarily picked $\epsilon = 0.05$. We find that our generation of the drift rate equivalence is not sensitive to this parameter for small ϵ (Fig. 5.5.a), and that the equivalences generated hold across the range of drift rates described in our simulations across a choice of ϵ (Fig. 5.5.b).

Next, we consider that our drift equivalence is based on an evaluation over a long timescale: this may grant sudden drift an unfair advantage, as the Brownian motion-like exploration of parameter space by gradual drift is more likely to return to previous values. Indeed, using gradual drift to scramble the population beyond recognition in the same number of days as the sudden drift induces more error in a fixed decoder after the first day of drift (as visible in Fig. 5.4.c). Nevertheless, even if we redefine the equivalence between sudden and gradual drift to cause identical degradation in a fixed decoder after the first day of drift (Fig. 5.5.c), resulting in sudden drift models completely scrambling populations faster than gradual ones, we still find the same advantages of sudden over gradual drift (Fig. 5.5.d).

5.3 Is in vivo drift sudden or gradual?

Our computational results imply that neural populations with downstream readouts can facilitate unsupervised drift compensation by making representational changes stark. We expect that this is especially relevant for fast drift rates. The use of sudden jumps in tuning would allow populations to support faster changes in representation—that is, faster learning—or to support changes at the same rate but with a smaller population size. Given these potential advantages, we expect the statistics of in vivo populations to reflect a preference for heavy-tailed drift, and that this preference is more likely to arise in more plastic, faster-drifting populations.

5.3.1 Distinguishing between drift types

Distributions of neuron pair correlations

Before we turn to real data, we will refer back to our toy model to illustrate how we might distinguish between sudden and gradual drift in in vivo datasets without relying on an explicit parameterised relationship between stimulus and activity. Suppose that we have picked a pair of neurons from a population and measured on two consecutive days, D_1 and D_2, the correlations C_{D_1} and C_{D_2} between the activities of the neurons in the pair. We define the scaled change in correlation for that neuron pair as:

$$\Delta C = \frac{C_{D_2} - C_{D_1}}{C_{D_2}} \text{ for } C_{D_2} \geq C_{\min} \tag{5.30}$$

Most neurons are not correlated with each other, and so when we visualise distributions of such scaled changes in correlation (as in Fig. 5.6.a-b), we include only neuron pairs with some minimum correlation $C_{\min}$ on D_2. We also exclude such pairs from our fits, as neurons that are not correlated with each other have small values of C_{D_2}, a term which features in the

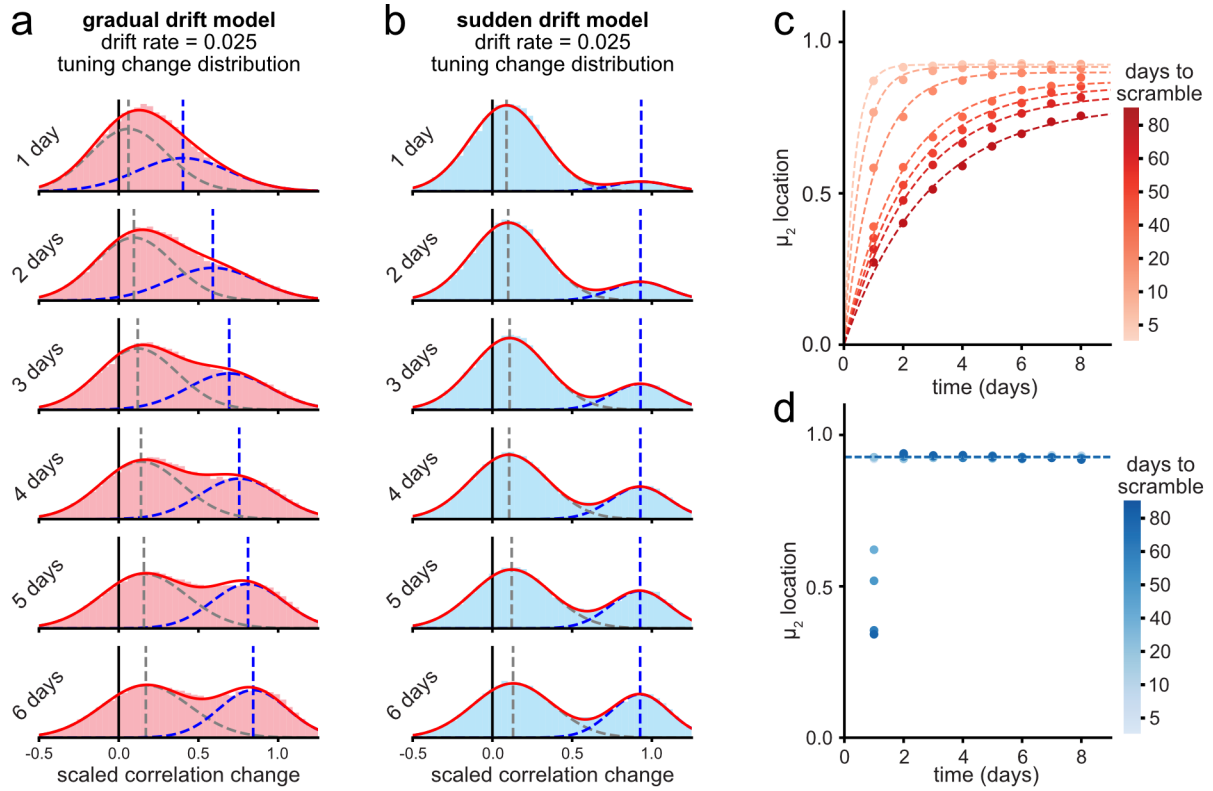


Fig. 5.6 Populations undergoing gradual or sudden drift can be distinguished by considering changes in pairwise neuron statistics. **(a)** Histograms showing the evolution of the distribution of changes in scaled correlations between neuron pairs in the gradual model of drift over the course of several days. The fit is a 2-component mixture of Gaussians, with the overall fit shown as a red trace, and each underlying component and its mean shown as dashed lines (grey: lower component, blue: upper component). **(b)** As in (a), but shown for the sudden model of drift at an equivalent drift rate. **(c)** Location of the upper component mean (μ_2) in the mixture of Gaussians fit for the gradual drift model shown against time for a range of drift rates. The dashed line indicates a fit to the exponential approach approximation. **(d)** As in (c), but shown for the sudden model of drift and using a constant value approximation.

denominator of ΔC and therefore risks amplifying noise. This inclusion criterion introduces a positive bias to the distributions.

When we consider such distributions from sessions a single day apart, most measured changes in pairwise correlation are driven by statistical fluctuations (our observation noise level σ in the toy models), which produces the component of the distribution near $\Delta C = 0.0$. In the case of sudden drift (Fig. 5.6.b), we can see a second distinct component of the distribution far from the origin: these are pairs of neurons that were previously uncorrelated with each other and have now become correlated. Over time, as drift is repeated over the course of several days, more and more of the distribution migrates to this second component corresponding to retuning from the first component corresponding to stable tuning. In contrast, for gradual drift (Fig. 5.6.a), it is harder to distinguish changes in correlation that are due to

statistical fluctuation from ones that are due to underlying retuning. The second component of the distribution, corresponding to larger nonstationary changes in tuning, emerges only gradually over time. This is the core intuition for why sudden drift is advantageous for an adaptive decoder. We stress that, after sufficiently many days of drift, both types of drift tend to the same final shape of distribution (an observation we shall return to in Fig. 5.10.d), as both gradual and sudden drift will eventually scramble a population entirely. However, by considering the transient period in which these distributions differ, we have a coarse means by which to distinguish the underlying statistics of drift.

Fitting 2-component mixtures of Gaussians

We model these distributions as a mixture of two component Gaussian distributions, with one component describing noise-driven fluctuations and the other describing drift-driven changes in tuning. We fit the distributions using EM as implemented by the `scikit-learn` library (Pedregosa et al., 2011) for the means, standard deviations, and relative weights of the mixture components. By decomposing the overall fits into their components, we can see a clear distinguishing feature: the component with the larger mean, μ_2, is stable for sudden drift (Fig. 5.6.d), while it asymptotes for gradual drift (Fig. 5.6.c). For gradual drift, we can approximate μ_2 as a function of the elapsed time in days t:

$$\mu_{2,\text{gradual}}(t) = c\,(1 - e^{-bt}) \tag{5.31}$$

While for sudden drift, μ_2 is simply independent of time:

$$\mu_{2,\text{sudden}}(t) = c \tag{5.32}$$

The parameters b and c correspond respectively to a factor associated with drift rate and a factor associated with the mean correlation between randomly chosen pairs of neurons.

We expect in vivo data to display a mixture of both sudden and gradual drift, and so when we consider the behaviour of μ_2 in real datasets in the following section, we will describe it as a weighted mixture of the two models above, parameterised by a mixing ratio a:

$$\mu_{2,\text{composite}}(t) = a\mu_{2,\text{gradual}}(t) + (1-a)\mu_{2,\text{sudden}}(t) \tag{5.33}$$

In all cases, we find the values for a, b, and c that minimise the MSE between measured values of μ_2 and the model subject to the following constraints:

$$0 \le a \le 1, \quad b \ge 0, \quad 0 \le c \le 1 \tag{5.34}$$

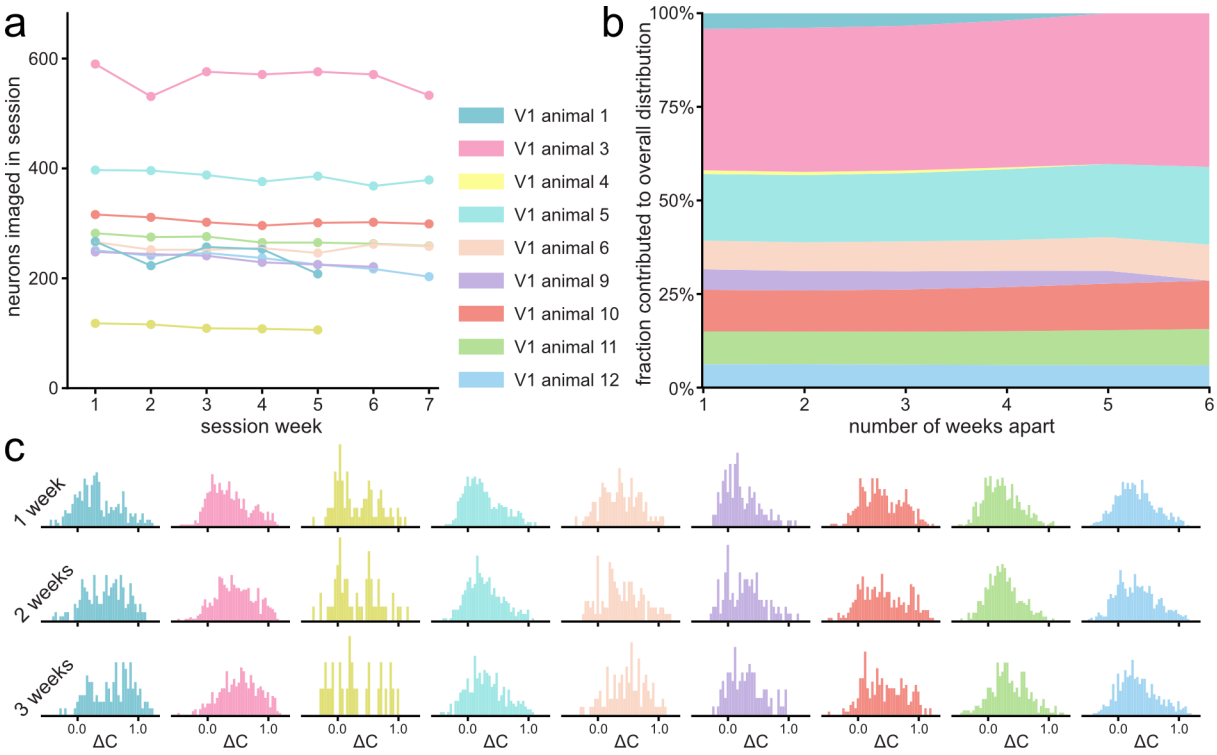


Fig. 5.7 Overview of the V1 drift dataset. **(a)** The number of regions of interest (ROIs) labelled as neurons in each recording session for each animal of the Marks & Goard (2021) V1 dataset. **(b)** The proportional contribution of each animal to the aggregated distribution of scaled changes in pairwise correlation. **(c)** Distributions of changes in pairwise correlation for the individual animals, shown for 1, 2, and 3 sessions apart.

5.3.2 Pre-processing of the in vivo datasets

V1 dataset

The Marks and Goard (2021) dataset comprises recordings from 12 animals imaged weekly for 5-7 weeks, typically registering ~250 neurons in a given session (Fig. 5.7.a). We omit mice 7 and 8 from our analysis, as their imaging schedules deviate from the schedules common to the other animals. We also omit mouse 2 as, uniquely, its surgical procedure introduced two separate imaging windows. We include only neurons labelled as having a registration quality of 3 or higher (unambiguously identified across sessions). Our analysis makes use only of the experimental protocol which exposes animals to movies of naturalistic scenes, choosing to omit the protocol showing oriented gratings, as they exhibit a far lower rate of drift.

To calculate the activity correlation between neuron pairs during each session, we convolve the raw Calcium traces with a Gaussian window (standard deviation equivalent to 250 ms). We subsequently use these smoothed traces as the basis for correlation between neuron pairs, which we calculate using a Pearson correlation coefficient. When calculating the

scaled correlation change, we set the minimum correlation for inclusion of a neuron pair as $C_{\min} = 0.5$.

When visualising the distribution of correlation changes for experiments some T weeks apart, we include all pairs of sessions in the dataset that are separated by that interval. For example, if $T = 2$, we would include week 1 paired with week 3, week 2 paired with week 4, week 3 paired with week 5, and so on. While correlations between neuron pairs can only be evaluated within an individual animal, we aggregate all neuron pair correlations from all animals for each time separation T. As described above, we use a mixture of Gaussians fit to compute μ_2 for each time separation T. We use bootstrap sampling of the distribution (1,000 iterations per distribution) to estimate the SEM of μ_2.

The disparity in the number of recorded neurons for each animal means that some animals contribute more pairs than others to the aggregated distribution (Fig. 5.7.b). We do not perform any weighting to correct for this. The relative contribution of each animal remains stable across values of T, and so we are not concerned that any conclusions we draw are due to variance across individual animals masquerading as a change over time. The shapes of the scaled correlation changes across individual animals look, at a glance, to be mutually similar (Fig. 5.7.c), but we pursue no analysis of individual animals in this work.

PPC dataset

The Driscoll et al. (2017) dataset comprises of recordings from 5 animals (m01 through m05). We include only neurons that are identified between sessions at the highest level of confidence as labelled in the source dataset. Animals m02 and m05 have a lower total number of neurons identifiable between sessions (Fig. 5.8.a) and experience a more rapid decline in the number of neurons identifiable between sessions as the sessions become separated by longer time periods. Consequently, while we include all animals in our analysis, we find that m01, m03, and m04 are more represented in the aggregated data. Nevertheless, the proportion of neuron pairs contributed by each animal remains relatively stable (Fig. 5.8.b). As with the V1 dataset, we find that consideration of individual session pairs within animals yields too few neuron pairs to meaningfully fit distributions (Fig. 5.8.c), and we will once again aggregate over all sessions and animals for a given time separation in days.

To obtain traces for neuron pairs to correlate, we first convolve spike times (which are available in the original dataset and were produced by deconvolving Calcium traces) with a Gaussian window (standard deviation equivalent to 450 ms). The design of the T-maze experiment from which this dataset is collected includes an inter-trial interval in which the animal is kept in darkness for 2 seconds, and we deliberately choose to include neural activity from this interval. We set $C_{\min} = 0.5$, the same choice as the V1 data.

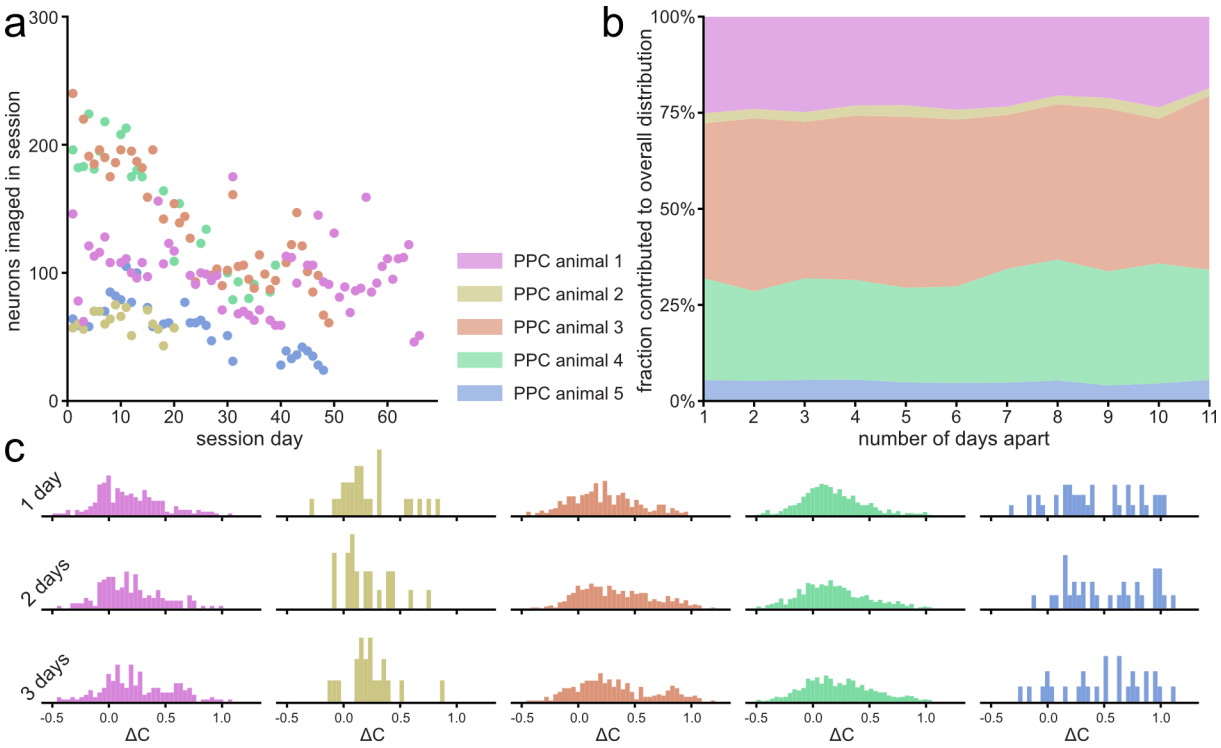


Fig. 5.8 Overview of the PPC drift dataset. **(a)** The number of regions of interest (ROIs) labelled as neurons in each recording session for each animal of the Driscoll et al. (2017) PPC dataset. **(b)** The proportional contribution of each animal to the aggregated distribution of scaled changes in pairwise correlation. **(c)** Distributions of changes in pairwise correlation for the individual animals, shown for 1, 2, and 3 sessions apart.

5.3.3 More plastic in vivo populations have more sudden drift

Inspecting the distributions of correlation changes for individual sessions reveals very little about the nature of drift in these populations: it is hard to get a complete picture of distributions when we rely not only on neurons being present in recordings from one day to the next, but on *pairs* of neurons surviving between recording sessions. Nevertheless, when we aggregate over all animals and sessions, we find that the V1 dataset, which drifts slowly over the course of weeks, demonstrates a gradual increase over time of μ_2 (Fig. 5.9.a), reminiscent in part of the distributions generated by gradual drift. In contrast, the PPC dataset, which drifts over the course of several days, has a more stable value of μ_2 closely reminiscent of the distributions generated by sudden drift (Fig. 5.9.b).

Under further scrutiny, drift in both of these experiments is necessarily a mixture of sudden and gradual drift (Fig. 5.9.c-d). By inspecting how the composite fit balances contributions from the underlying sudden and gradual drift models, we find that for the PPC experiments the composite fit attributes 73.6% (SEM=7.85%, estimated via bootstrap) of the changes in the population to the sudden change model, compared to 57.8% (SEM=8.35%) for the

V1 experiments. Simply put, the PPC drift is 'more sudden' than the V1 drift ($p = 0.034$, single-sided t-test over 1,000 bootstrapped fits).

This is obviously an extremely coarse analysis, and takes into account only two studies that differ in many more ways than the underlying plasticity of their populations. While it is hardly strong evidence for a design trade-off between rate of learning and the ability for downstream components to adapt to changes in representation, it nevertheless provides an early indication that this line of reasoning may have merit. More importantly, populations in both of these studies display a significant component of drift that appears mediated by jumps in tuning, which runs contrary to models of drift that rely on gradual tuning updates.

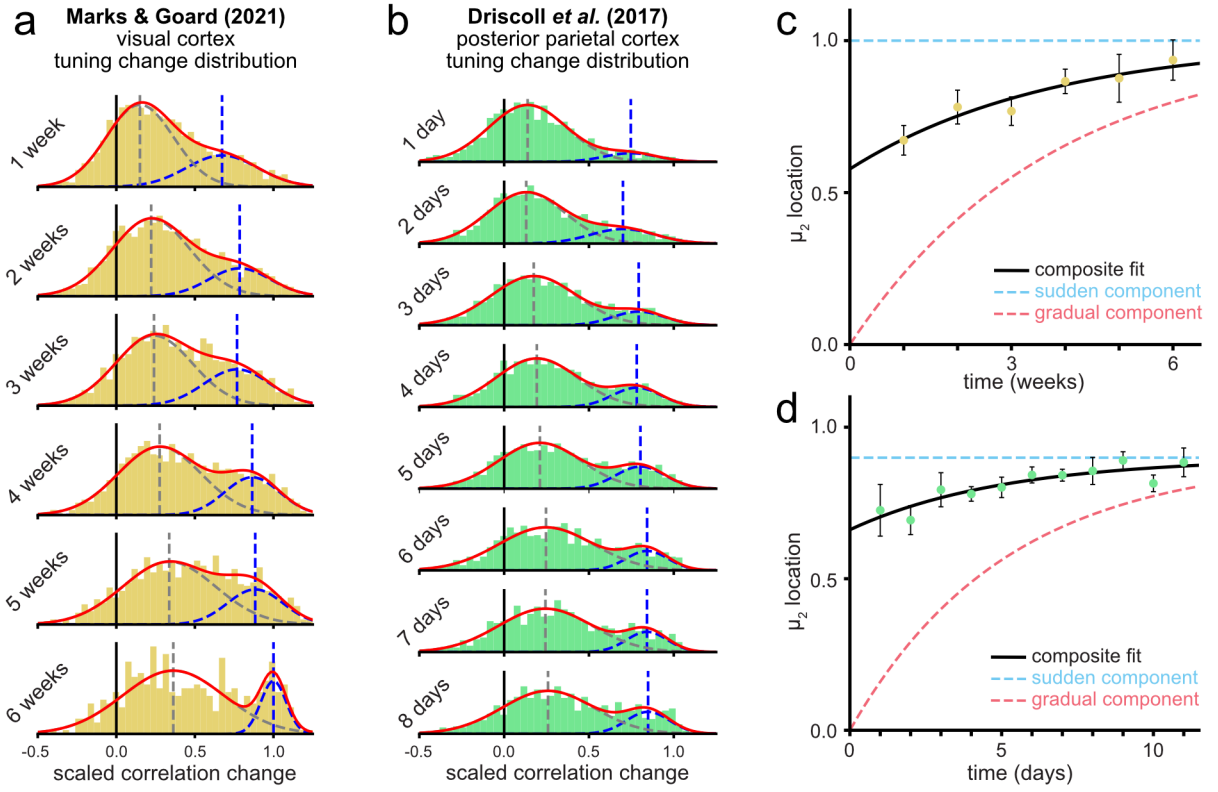


Fig. 5.9 in vivo drift favours sudden changes, especially in a faster-drifting population. **(a)** Distributions of the scaled correlation change, as in Fig. 5.6, but shown for in vivo data from the Marks and Goard (2021) dataset comprising of recordings from the visual cortex during exposure to movies of naturalistic scenes. Shown with a 2-component mixture of Gaussian fit. **(b)** As in (a), but shown for in vivo data from the Driscoll et al. (2017) dataset comprising recordings from the PPC during a T-Maze task. **(c)** Location of the upper mean (μ_2) in the mixture of Gaussians fit for the Marks and Goard (2021). Error bars indicate +/- SEM as estimated via bootstrap. The fit is a weighted combination of gradual and sudden drift models, with each component rendered as a dashed line. **(d)** As in (c), but shown for the Driscoll et al. (2017) dataset.

5.4 Revisiting cross-session tuning in our BMI experiments

5.4.1 Distribution of correlation changes between neuron pairs in CA1

We now take the opportunity to revisit our own CA1 BMI experiments to determine whether the statistics of tuning changes are compatible with our hypothesis that sudden retuning offers an advantage for highly plastic populations. We must stress that the BMI experiments were not designed as an investigation into drift, and that this dataset has several features that make it imperfectly suited to address the sudden or gradual nature of drift in CA1 hippocampus. Not least of these issues is that these sessions may not correspond to the steady-state of an already-mastered task: we have already demonstrated that the BMI trials within the sessions show an enrichment over time, and that this representation can bleed-out and superpose with the locomotion-based representation of the track. We refrain, therefore, from even calling these tuning changes 'drift', but still seek to satisfy our curiosity as to whether these tuning changes are sudden.

We recall from Fig. 4.14 the previous chapter, in which we considered the explicit changes in tuning for sessions spaced exactly one day apart for the cohort of animals performing consecutive BMI sessions, that approximately half of the neurons with single place fields in the initial locomotion trial retained that place field on the following day. Approximately 40% of the neurons lost their field entirely, and the remaining neurons exhibited either substantial movement of their field (of a sufficient magnitude to no longer classify as the same field) or displayed multiple fields. While this on-off behaviour of neurons dropping in and out of the representation is a straightforward example of sudden drift, the definition of 'same field' between two sessions is generous, and may hide more subtle gradual drift in the population at large.

We prepare distributions of scaled correlation changes for our recordings from CA1 using the same methods we applied to the PPC and V1 datasets. We consider only the initial locomotion trials of BMI sessions. Calcium traces for each ROI are smoothed via dF/F filtering and then convolved with a Gaussian kernel with a standard deviation of 4 frames (~130 ms) for additional smoothing before evaluating the correlation matrix. As previously, we use $C_{\min} = 0.5$. These distributions are shown in Fig. 5.10.a, with the locations and movement of μ_2 from the mixture of Gaussian fits shown in Fig. 5.10.b. The timing of sessions (referring back to Fig. 3.6) results in a paucity of sessions spaced 2 or 5 days apart, and discrepancies in how well we were able to track ROIs across several days result in an overrepresentation of animal MH485 (Fig. 5.10.c).

The immediate jump of μ_2 to near its final value and a weighting of 72.4% (SEM=22.4%) towards the sudden component of the composite fit suggests a high rate of sudden changes. While this observation would support our hypothesis, we find the analysis unsatisfactory: the

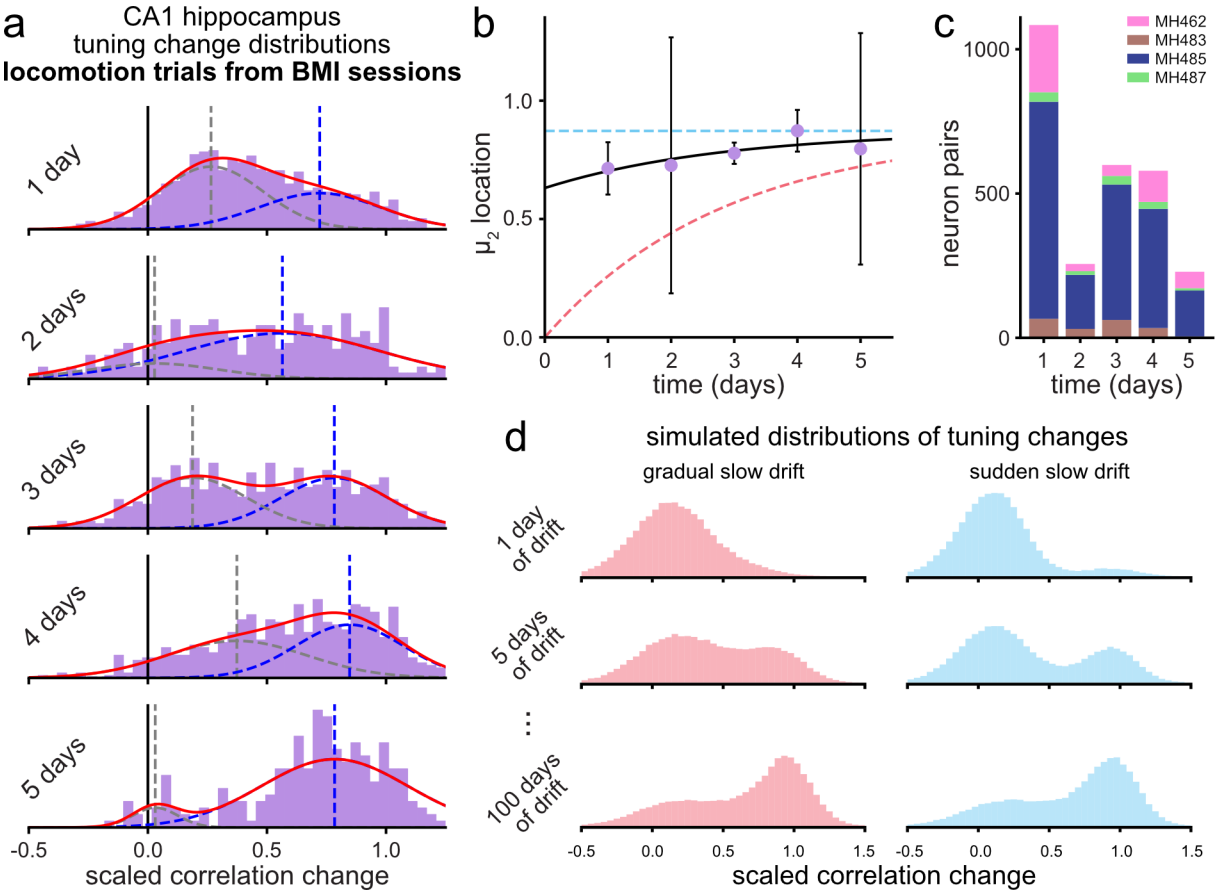


Fig. 5.10 Distribution of scaled correlation changes for CA1 populations suggest sudden jumps in tuning, but are questionable from a methodological standpoint. **(a)** Scaled correlation change distributions with a 2-component mixture of Gaussians fit for the initial locomotion trials from BMI sessions. **(b)** Location of the upper mean μ_2 as a function of elapsed time between sessions. Error bars indicate +/- SEM, estimated via bootstrapping. **(c)** The contribution of neuron pairs from each animal at each time spacing between sessions. **(d)** After many iterations of drift, or indeed after few iterations of very fast drift, the ΔC distributions generated by simulated gradual and sudden drift become indistinguishable.

SEM is high and this indicates a considerable sensitivity to which neuron pairs are included. Furthermore, as we have previously described, the shape of these distributions converges for gradual and sudden drift after sufficiently many applications of drift (Fig. 5.10.d). Our method of distinguishing between gradual and sudden tuning is therefore poorly suited to very high rates of tuning change.

5.4.2 Place field displacement distributions indicate sudden changes

Unlike the V1 dataset, in which it is simply unfeasible to parameterise the relationship between an individual neuron's activity and the current frame displayed in a naturalistic movie, and unlike the PPC dataset, in which neurons encode a variety of features about the task at hand,

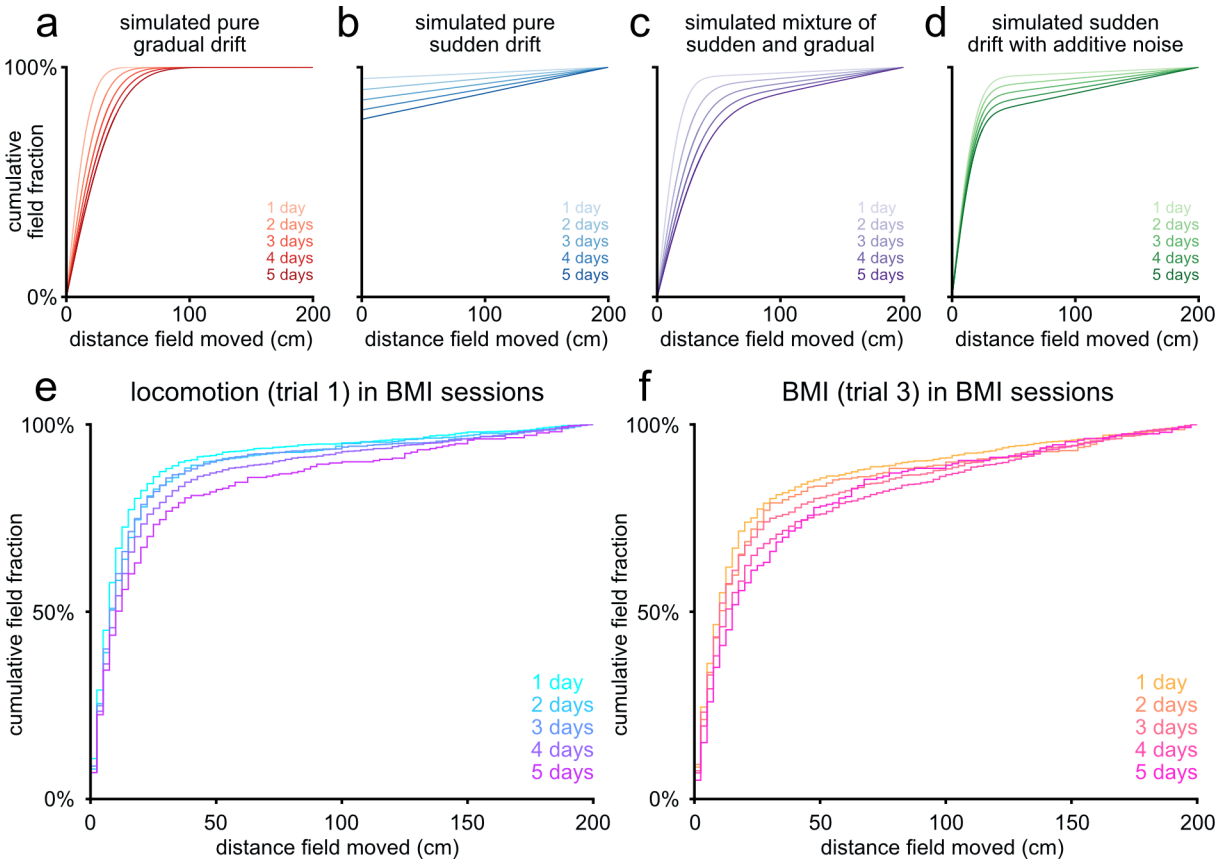


Fig. 5.11 Displacements of place field location over time suggest sudden, jump-like retuning. **(a)** Cumulative distributions of absolute displacement of a place field over the course of 5 simulated days under a gradual drift model of field movement (10^6 simulated cells, $\rho = 15$ cm/day). **(b)** As in (a), for a sudden drift model (retuning probability $\alpha = 0.05$). **(c)** As in (a), for a drift model that applies both the gradual and sudden drift from (a) and (b) simultaneously. **(d)** As in (b), but with observations of the place field centres perturbed by Gaussian noise ($\sigma = 15$ cm). **(e)** Cumulative distributions showing the absolute displacement of place field centres during Trial 1 (locomotion) between BMI sessions, shown for a variety of time separations between sessions. **(f)** As in (e), but shown for Trial 3 (closed-loop BMI).

we are able to parameterise the tuning of neurons in CA1 in terms of place field locations. This gives us another lens through which to view the statistics of tuning changes in CA1: the statistics of place field displacement.

We consider first an in silico version of the 400 cm long VR track used in the BMI experiments. We model only the statistics of movement of the centre of a place field for the same neuron over the course of repeated drift. We treat the geometry of the track as circular, as the animal wraps around to the start of the track after completing a lap, resulting in a maximum field displacement of 200 cm between sessions. This formulation is essentially identical to the simulated drift in the preferred angle θ from our toy head-direction system model, except that the units are in cm rather than radians. In gradual drift, these place field

centres move in a Brownian motion-like pattern, with overall statistics characterised by a Gaussian distribution of the distance moved by each place field centre in the population (Fig. 5.11.a). In sudden drift, this distribution has a distinctly different shape (Fig. 5.11.b): neurons either keep their field centre in the same location (0 displacement), or have their field move to an entirely new location (a uniform distribution over the entire track). As before, we acknowledge that in vivo drift is likely to be a combination of gradual and sudden drift. A composite distribution, in which neurons are subject to both a daily normally-distributed displacement from their previous location and some small chance of remapping entirely, combines the two features of these distributions (Fig. 5.11.c): the characteristic diagonal line of sudden drift, and the characterstic lateral distribution shifting of gradual drift. We are careful to draw a distinction between this composite distribution, which blends gradual and sudden drift, and a distribution generated by sudden drift under observation noise (Fig. 5.11.d). When sudden changes are the dominant form of drift in the population, the distinction becomes challenging: neurons remap before they have time to accumulate small displacements from gradual drift.

We then qualitatively compare these general shapes of distribution to those extracted from place field locations in our in vivo data[2]. We show these both for the initial locomotion trials from BMI sessions (Fig. 5.11.e) and for the second BMI trial in those same sessions (Fig. 5.11.f). Sudden changes in tuning feature prominently in both cases, with apparently greater turnover in the BMI case—this latter is not surprising, as we have previously seen that learning-like enrichment of the representation takes place with increasing BMI experience (Fig. 3.10) and weaker retention of fields across days in the BMI context (Fig. 4.14). Gradual changes are more difficult to identify with any confidence: if such changes take place, they are subtle and challenging to distinguish from noisy observations. We conclude—taken together with our earlier results that neurons drop in and out of the representation between sessions—that sudden changes in tuning are the predominant form of representation in our observations of highly plastic CA1 hippocampus populations.

5.5 Discussion

5.5.1 Drift is gradual at the level of populations, not neurons

Our computational simulations illustrate the advantages of populations mediating changes in representation through stark and sudden changes in the tuning of individual neurons. We are quick to highlight that the advantages of sudden changes in neuronal tuning do not preclude the existense of gradual changes: on the contrary, we demonstrate that it is possible to adapt

[2]A methodological detail: for situations in which a neuron has multiple fields, we take the smallest distance between all pairs of fields across days to avoid missing gradual changes.

to such gradual changes, and that at least some component of experimentally observed drift is attributable to these gradual processes. However, to support such gradual changes, the population undergoing gradual changes must make concessions: it must either achieve higher redundancy by employing a larger number of neurons to encode the same information, or it must drift more slowly, or it must rely on some external feedback system to compensate for drift.

The statistical process that we use to model heavy-tailed drift, a jump of a neuron's tuning to a new tuning completely independent of its previous state, is a convenient illustration of the advantages of sudden changes in representation at the level of individual neurons. However, it does not represent an optimal retuning rule for minimising error in a downstream readout. It is possible to construct other drift processes that can better perform this role via distributions with even heavier tails. For example, we might imagine that upon retuning a neuron is forced to adopt a new tuning state that is sufficiently different from its previous state. This 'active avoidance' of previous tuning values makes it even easier to distinguish statistical fluctuations from nonstationary changes.

The relevance of the advantages of heavy-tailed changes in tuning extends beyond representational drift. Other nonstationary changes in the encoding of information, such as the incorporation of new information during learning, face the same challenge of making themselves visible through the noise of day-to-day fluctuations. If we entertain the view that some component of drift is in fact driven by the repeated application of some continual learning rule, we open up the possibility of deriving constraints on the learning rules implemented by the brain through measurements of drift. While it is experimentally challenging to attribute new behaviours or newly learned information to changes in in vivo physiology, we speculate that—if measured—such changes would also demonstrate stark, heavy-tailed statistics.

5.5.2 Causes of sudden jumps in in vivo tuning

We are not the first to describe jump-like changes in tuning in experimental data during the performance of mastered, stable tasks. These have been explicitly highlighted in CA1 hippocampus[3], described as step-like changes (Zheng et al., 2024), and in the motor cortex (de Laittre and MacLean, 2024), described as the heavy tail of otherwise stable tuning distributions.

While it is appealing to take these reports as direct support for the mediation of changes in tuning via sudden jumps to facilitate stable downstream readouts, we must be cautious: there are other potential reasons for sudden changes. Step-like or otherwise abrupt changes in the weights of neurons in the innermost hidden layers of deep linear artificial neural networks are

[3] Some of the earliest measurements of drift, as recorded in CA1 hippocampus and reported in Ziv et al. (2013), seem to exhibit a mixture of both gradual and sudden changes, but no explicit distinction was made between them in the original work.

dynamics that emerge necessarily from training with backpropagation (Saxe et al., 2014)—and even simple learning rules can produce jump-like adjustements in parameter-space (Löwe et al., 2024). Equally, explanations rooted in cell biology, rather than abstract learning rules, might explain why we observe sudden changes: in addition to obviously localised and stark effects like neurogensis or cell death, dendritic lifetimes in regions such as CA1 are short—on the order of 1-2 weeks (Attardo et al., 2015)—and the ensuing rates of synaptic turnover do not necessarily cause gradual changes in the tuning of their associated neurons (Ziv and Brenner, 2018).

Our prediction, that the populations that drift faster are 'more sudden' in their drift, is borne out by our consideration of real data; but these are only two (three, if we are so bold as to count our own CA1 data) datapoints that derive from experiments involving different tasks, recorded under different conditions, and measure from brain regions with myriad potential physiological differences. We claim only, therefore, that our predictions present no obvious contradictions with current evidence. Going forward, we hope for a specific scrutiny in experimental measurements of drift with regards to the statistics of change on the level of neurons rather than populations. This will further inform how downstream readouts can maintain their accuracy, which is of relevance not only to the mechanisms of drift, but also to the design of adaptive BMI decoders.

5.5.3 Implications for adaptive BMI decoders

The two key components of the implementation of our adaptive decoder are an accurate drift prior and a description of how neurons tune to a particular kind of stimulus in a population. Our illustrative toy model deliberately uses drift priors, activation functions, and stimulus exposure distributions that are all symmetric. Introducing asymmetries in these priors has the potential to provide additional information to downstream populations—for example, by having some neurons that are more stably tuned than others or by having frequent stimuli that can serve as reference points—but would also substantially complicate solving the optimisation problem. We do not claim that real, in vivo downstream populations explicitly perform Bayesian optimisation to account for changes in tuning, but we do posit that it is physiologically plausible for a downstream population to make use of these key prior distributions when implementing an adaptive decoding strategy: the same brain regions across individuals within a species encode similar information, using similar encoding strategies, and experience drift with similar statistics. Readouts across populations that connect between brain regions therefore have access, from an evolutionary standpoint, to reasonable priors about their upstream neural activity—and this may be sufficient to implement adaptive decoding strategies.

Could we, as BMI designers, then make use of these same priors? The formulation of an adaptive decoder in this chapter is obviously a toy: it requires a parameterisation of neuron tuning in terms of variables external to the population, and requires that such a parameterisation is a very good description of how neural activity relates to the outside world. Despite this, a broader interpretation of the same core idea, which is to update how neurons are relevant to decoding by considering the similarity of their activity to other neurons within the population, is potentially extensible to decoder design.

We have determined from our consideration of in vivo data that real neural populations exhibit jumps in tuning. We know, too, that real populations exhibit considerable redundancy. We therefore propose, as future work, investigation into a scheme of unsupervised adaptive decoding applicable to real neural data that pools neurons with similar activity together into multi-neuron ensembles. Just as with the toy model, this scheme would 'cheat' by initialising a decoder with training data. However, instead of training the decoder directly on neurons, we propose training the decoder on the collective activity of these ensembles. On subsequent days, the decoder weights can remain fixed, while the individual neurons' membership in ensembles are updated based on similarity of activity to other neurons.

This proposed scheme represents a bridge between an already-widespread technique, supervised learning applied to training data collected daily and used to relearn the projection of high-dimensional neural data onto a stable lower-dimensional manifold (Degenhart et al., 2020; Gallego et al., 2017; Karpowicz et al., 2025; Willett et al., 2021), and an appreciation—a neural sympathy—for the statistics of drift that emerge from the design constraints on tuning changes within the brain. Together, these offer a potential path forward towards unsupervised adaptation of BMI decoders to neural variability.

Chapter 6

Discussion

6.1 Summary of contributions

The work in this thesis has ostensibly been motivated by current and future engineering problems in BMI design. We approached these problems not by proposing state-of-the-art clinical hardware implementations or sophisticated decoder architectures, but by pursuing an understanding of the changes in representation that take place during and between instances of BMI usage. This means that most of our main contributions, which we summarise below, are far closer to basic science than concrete design recommendations. In addition to enabling novel BMI-based experiments, our results are relevant to a broader neuroscientific audience: we have experimental evidence of new representational phenomena in the hippocampal formation and explicit predictions regarding the statistics of drift.

Our first main contribution is technical: we implemented an optical BMI based on 2-photon Calcium imaging. Our framework is reusable and potentially suitable across a variety of brain regions, but its design focused on decoding from neural populations in CA1 hippocampus. We demonstrated that this BMI decodes from populations of hundreds of neurons simultaneously in real time on consumer-grade computer hardware. It achieves high decoding accuracy from short spans of training data and allows for rudimentary interpretation of the interactions between individual neurons and their associated decoder weights.

Our second contribution is a novel type of closed-loop BMI experiment. We described a paradigm in which the neural representation of an environment, rather than a specific motion or stimulus, needs to be held consistent to provide accurate BMI task performance. This paradigm explicitly separates the decoding of an interpretable neural state from the policy that dictates how that state is mapped an action during closed-loop task performance. This distinction allowed us to diagnose deteriorations in closed-loop task performance due to underlying changes in representation.

Applying this closed-loop BMI experimental paradigm to CA1 hippocampus, we found that representations of a virtual environment differ depending on whether or not the environment was traversed by running on a wheel. The extent of this remapping is global and dramatic, and our observations supplement a wealth of existing evidence for the incorporation of locomotion-based path integration signals into populations in CA1. This is a poignant reminder that accuracy in open-loop decoding does not translate to closed-loop BMI performance.

Despite both being conditions without running-based travel, we found differences between BMI-based navigation and purely passive motion. BMI-usage induces changes in representation that persist even when animals return to locomotion-based navigation. These changes are additive, such that the post-BMI representation appears as a superposition of the pre-BMI and BMI representations. This effect is unique to experimental sessions that mix locomotion and BMI-based traversals of the environment. We also found that BMI-based representations of the same virtual environment improve with time and experience, leading to gradually improved performance of closed-loop BMI task execution.

Our final contribution deals with the statistics of tuning changes in representations. We demonstrated, by means of a computational model, that representations that drift by making sudden jumps in their tuning are easier for downstream readouts to adapt to than representations that gradually retune all neurons simultaneously. We then showed that changes in tuning in real, in vivo data demonstrate a preference for sudden changes in tuning on a per-neuron basis, and that this preference is potentially stronger in neural populations experiencing faster rates of representational turnover.

6.2 Relation to existing work

6.2.1 Representational drift

Many descriptions of drift present a trajectory through a parameter space to describe ongoing changes in neural tuning, leaning on visual metaphors like valleys in a loss landscape or manifolds of optimal task performance (for a few recent examples, see Raman and O'Leary, 2021; Ratzon et al., 2024; Driscoll et al., 2022; Micou and O'Leary, 2023). While we believe these descriptions to be apt, the metaphors mentally evoke trajectories through the landscape that we visualise as 'smooth', and result in mathematical constructions of drift that resemble Brownian motion in parameter space (Pashakhanloo and Koulakov, 2023; Rule and O'Leary, 2022). Some formulations of artificial neural networks trained with backpropagation do indeed demonstrate drift that resembles Brownian motion-like diffusion (Chaudhari and Soatto, 2018), but our modelling, analysis of existing datasets, and emerging experimental

evidence (de Laittre and MacLean, 2024; Zheng et al., 2024) all suggest that in vivo populations demonstrate distinctly heavier-tailed changes in representation.

The role of our contribution is to draw attention to a specific statistical property of in vivo measurements of drift that hints at a particular set of learning rules. The heavy tail of the distributions that govern retuning suggests a scheme for updating representations in a way that does not disrupt downstream consumers of the representation without resorting to explicit feedback signals or side-channel communication. If we entertain the notion that drift is, at least in part, a result of continual learning, then our observations offer a concrete constraint on models of the learning rules applied within the brain: they should generate sudden jumps.

We are keen to draw a distinction between sudden changes in the tuning of neurons and sudden changes in the behaviour of animals. Sudden changes in behaviour are readily observable in experimental settings (Rosenberg et al., 2021), and these necessarily correspond to sudden and widespread changes in neural activity (Durstewitz et al., 2010; Shen et al., 2018). This is sometimes framed in terms of 'sudden insight' (Löwe et al., 2024), which corresponds to a dramatic improvement in task performance. In contrast, observations of drift deal with changes in tuning taking place despite constant task performance. Subtle changes in behaviour remain an extremely important factor in experimental observations of drift, as they can account for much of the apparent changes in population tuning (Liberti et al., 2022; Sadeh and Clopath, 2022). Nevertheless, it is neural activity that drives behaviour, and not the other way around. In Micou and O'Leary (2023), we have previously argued that we should expect some component of the nonstationary statistics of drift to leak into behavioural strategies. But more subtle features, like our 'heavy-tails', risk being drowned out by the redundancy within large populations codes: behaviour can still change smoothly and gradually, even if the mechanisms that enact changes at the level of neurons are gradual.

6.2.2 Representations of navigation actions in CA1

Actions and how they transport us from one place to the next are a fundamental building block for internal navigation models. We find state-action pairings across a variety of modern computational models of hippocampal function (Gershman, 2018; Jensen et al., 2024; Raju et al., 2024; Whittington et al., 2020). Perturbing the means of navigation through an environment provides an alternative experimental axis to perturbations of the environment itself. However, while mice can learn to navigate through the same environment through a variety of different action strategies, such as learning to drive carts (Crawford et al., 2020), it has been historically difficult to experimentally decouple these alternative means of traversing the environment from differences in the sensory experience of manipulating those means of traversal. The hippocampal BMI in this work preserves the sensory experience of wheel

manipulation, yet retains a causal relationship between the animal's neural activity and its performance of a navigation task.

Our control experiments with passive motion reaffirm existing evidence of the incorporation of locomotion signals into place field activity within CA1 (Chen et al., 2013; Terrazas et al., 2005), while our controls that modulate the gain of the wheel mid-session produce no remapping—also in agreement with existing work (Jayakumar et al., 2019). These observations are consistent with a view of the hippocampus that treats active navigation actions as critically important for representations.

The superposition effect induced by BMI usage provides a concrete and novel result from a unique experimental condition. This is an opportunity to refine and challenge existing computational models that attempt to emulate hippocampal learning. We are tentatively optimistic that models like clone-structured cognitive graphs are well-suited to express multiple means of travel through the same environment (Dedieu et al., 2019; George et al., 2021), especially in light of their recent success in capturing the evolution of representations of multiple subtly different environments (Sun et al., 2025). Furthermore, the gradual enrichment across sessions of the BMI context represents a learning effect on an entirely different timescale, which fits into a more contemporary picture of learning the hippocampal formation: a mixture of learning rules both fast and slow (Liao and Losonczy, 2024).

Our results have led us to speculate that the presence or absence of agency within an environment influences neural populations in CA1. We are careful to distance this work from similar-sounding phenomena, such as a 'sensation of agency' in humans (often supplemented by fMRI recordings, see David et al. (2008) for a review), or reports from the psychology literature of conditioning patterns that alter behaviour such that animals take no action even when action is possible (Seligman, 1972). These phenomena exist at an entirely different layer of reasoning, and it would be an overreach to tie these effects to changes in place cell behaviour in the hippocampus. We merely report that the neural substrate for cognitive maps appears permanently altered by experiences in which the animal is causally responsible for how those experiences unfold. Any discussion of mechanism, whether this is an internally-generated agency signal or a property that is intrinsic to the learning rules acting on CA1, stands to be better informed by future work designed specifically to distinguish such hypotheses, for which we make some suggestions below. Finally, we feel compelled to reiterate a previously discussed caveat: our experimental results are based on head-fixed VR, which represents only a narrow slice of the sensory and behavioural experiences typical to animal experience in real environments (Minderer et al., 2016). Translating between real and virtual environments does not always keep hippocampal phenomena intact (Acharya et al., 2016; Aghajan et al., 2015; Ravassard et al., 2013), and any claim of a signal as potentially fundamental as 'agency' would demand evidence that overcomes the shortcomings of VR.

6.2.3 Hippocampal BMI design

The description of results in this thesis corresponds, to the best of our knowledge, to the first instance of a closed-loop optical BMI that decodes from populations in CA1, and follows Lai et al. (2023) as the second instance of any closed-loop BMI in the hippocampal formation that goes beyond conditioning the firing of individual neurons. Nevertheless, the existence of a variety of open-loop feasibility assessments (Chang et al., 2021; Chen et al., 2023; Sun et al., 2022), several of which contain descriptions of bespoke hardware and are not simply post-hoc analyses of existing data, suggests a genuine appetite to implement hippocampal BMI as a tool for scientific inquiry (see Batista, 2020). We expect that the coming years hold a wealth of work that refines these techniques and uses new BMI paradigms for creative experiments. However, despite the obvious success of these open-loop feasibility assessments, none yet appear to have been translated into closed-loop experiments with published results.

We have demonstrated and mechanically justified one potential difficulty for moving to closed-loop from open-loop assessments: the remapping that takes place between manual task performance and BMI task performance. Successful closed-loop BMI that reads out from the hippocampus must compensate for this remapping. We have described two approaches that actively mitigate the remapping: training on passive observation of a task being completed, or retraining the decoder on the BMI condition. Equally, we point to the success of Lai et al. (2023), whose task design is able to leverage a very long averaging window to capture the residual component of the representation that does not remap.

We benefitted from an experimental paradigm that splits the decoding of a neural state from the mapping of that state to an action. Furthermore, that neural state corresponds to an experimentally measurable variable—position on a track—and this made it tractable to identify and explain the root cause behind poor closed-loop performance. Had we adopted a more traditional approach to BMI training, by directly learning the relationship between neural activity and action, it is likely that we would never have made the leap to a protocol that produces successful closed-loop BMI performance. We are optimistic that our observations of a conflict between a locomotion and BMI context is not only a finding that is interesting from the perspective of basic science, but a key technical detail that can pave the way to a proliferation of hippocampal BMI experiments.

6.2.4 BMI stability in clinical settings

We were clear from the outset of this thesis: the hippocampus is not a clinically desirable target for BMI, and optical BMI is not suitable for use in human patients. We argue that while the hippocampus has some idiosyncrasies that are fascinating in their own right, its exceptional plasticity proves an ideal experimental testbed for adaptive decoders—especially when

combined with 2-photon imaging to confidently resolve changes at the level of individual neurons.

Since the outset of our hippocampal BMI work, a variety of promising techniques for decoder adaptation have been validated on offline data (Karpowicz et al., 2025; Ma et al., 2023; Wilson et al., 2023). We know first-hand the risks of assuming that open-loop accuracy implies closed-loop performance (see Chase et al. (2009) for an earlier cautionary tale), but validating these techniques in practice is often prohibitively expensive. The reach tasks in the offline datasets described above are performed by human or macaque participants, and gathering a dataset that comprehensively validates decoder adaptation is an endeavour that spans months. While it would be naïve to assume that strategies that work well in closed-loop for CA1 populations will necessarily work well for populations in the motor cortex, our framework nevertheless offers a cheaper intermediate testing stage. By permitting closed-loop BMI in mice in a brain region with quick representational turnover, we have the option of validating adaptive decoding techniques before moving to more costly trials that are closer to clinical scenarios.

6.3 Future work

6.3.1 Adaptive correction of in vivo closed-loop decoders

The obvious segue from the preceding section is to propose extending our own protocol to support experiments which adapt the decoder from one day to the next in an entirely unsupervised manner. We have already outlined the prerequisite minor technical upgrades to support this paradigm, so we focus here on a bigger picture: maintaining closed-loop BMI performance in the steady state following task mastery with BMI.

Unlike our work so far, we envisage an experimental schedule of daily sessions in which *all* trials are performed with BMI, entirely foregoing passive motion or locomotion-based navigation in order to avoid conflicts between alternate representations of the same environment. During the initialisation stage of the schedule, each of these BMI trials uses a decoder trained with our already-described supervised method, with the very first trial corresponding to a random decoder initialisation (resulting in motion similar to the random speed condition in our existing work). Once the BMI context becomes saturated with place cells (no session-to-session increase in the fraction of the population forming place fields), the schedule moves into an unsupervised stage, in which the initial trial on each day reuses the latest decoder from the previous day. This forms the basis for unsupervised alterations to the decoder weights used in subsequent trials on that day, with the objective of sustaining decoder accuracy for as many consecutive days as possible.

A dataset already described in this work, the cohort of animals which experience many consecutive autopilot-BMI sessions, provides an adequate starting point for offline assessment of candidate adaptive decoding strategies. Such an assessment would be well-positioned to compare existing state-of-the-art unsupervised manifold realignment (Karpowicz et al., 2025) to a model that explicitly parameterises place cell tuning curves and drift priors similar to our work in Chapter 5.

6.3.2 Evolution of hippocampal BMI paradigms

The closed-loop BMI design in our protocol enforced a set policy based on decoding a neural state. This policy replicated an animal's expert task execution if the representation of an environment was held stable. This served as a useful technical validation and offered insight into the roles of path integration and agency in CA1. We explore here some alternative schemes for mapping neural activity onto navigational actions in other VR task designs.

Traditional reversal learning paradigms flip a learned association between stimulus and reward. How quickly a subject adjusts is a convenient measure of cognitive flexibility (Izquierdo et al., 2017). BMI variants of these paradigms, in which control outputs from decoders are mirrored or rotated mid-experiment, can provide compelling evidence of volitional modulation of neural activity (Jarosiewicz et al., 2008; Menéndez et al., 2025), but translating these experiments to hippocampal populations requires some creativity. At navigational decision points, hippocampal activity produces neural trajectories corresponding to prospective routes—notably recorded from CA3 by Johnson and Redish (2007). We envisage a protocol in which animals use BMI to navigate a familiar environment rich in binary decision points. At each decision point, the BMI control strategy requires the animal to generate a strong bias towards one path in its preplay activity to proceed in a given direction. If this BMI task can be executed—which we stress is speculative, and would be an impressive technical achievement in its own right[1]—a reversal learning task becomes possible by flipping the valence of which preplay trace corresponds to which selected trajectory. Should animals be able to quickly learn this inversion, this would strongly support the ideas of volitionally produced hippocampal activity proposed in Lai et al. (2023), while a failure to learn the inversion would imply a strong reinforcement-based conditioning of representations via BMI.

A more immediately tractable line of investigation is to attempt to replicate the protocol we described in this thesis while imaging or decoding from grid cells in MEC. Campbell et al. (2018) outline an experimental protocol using a 1D VR track much like our own gain modulation trials. They describe a sensitivity of grid cells to changes in gain, unlike our observations of place cells, which is consistent with a view of the hippocampal formation

[1] As a cautionary technical note: these preplay trajectories play out on a timescale that is difficult to resolve with Calcium imaging, though it is in principle possible to catch residual traces. An electrophysiology-based imaging approach is likely better-suited to this line of investigation.

that treats MEC as a region rich in self-motion signals. In entirely passive motion situations, the characteristic hexagonal firing patterns of grid cells appear badly disrupted (Winter et al., 2015). If grid cells can retain consistent activity in conditions in which running is decoupled from motion while navigational agency is retained via BMI, this would provide compelling evidence that 'navigation' in rodent hippocampus can be bound to more abstract forms of active motion than merely running—which in turn would suggest a more generic, less spatially-specialised role for the hippocampal formation.

6.3.3 Neural representations of mixed modes of navigation

Our observations support the idea of fundamentally different neural representations for active navigation and passive motion within the same environment. They also suggest that the formation of the cognitive map depends on the mode of navigation, even when the sensory experience of performing those navigation actions is similar. This provides both theoretical avenues of investigation—existing models of learning in the hippocampus typically consider only a single mode of travel—and fertile experimental ground.

From a pragmatic standpoint, VR experimental designs offer an inroad for multiple modes of navigation for an identical environment. Our work compared a running condition, which is naturalistic[2], and a BMI condition, which we assume to be highly unintuitive for the animal. We are curious to consider the representation of environments with physical interfaces for navigation that are more closely matched, such as decision tasks based on forelimb lever manipulation or nose-poking. However, the highly structured nature of these tasks makes it difficult to distinguish representations of sequential memorisation tasks from abstract representations of environments, and the sensory paucity in VR risks representations that are highly orthogonal. This presents an experimentally challenging design space, but with significant potential for insight into whether the hippocampal formation is hardwired for naturalistic navigation, or whether it should be better understood through the lens of abstract relations between arbitrary states and actions.

Bridging into theory requires recreating our existing experimental results in silico. While it might seem natural, for the purposes of modelling, to describe multiple modes of travel in terms of an explicit input variable that simply denotes the current mode of travel, this implies that the animal is able to sensorially distinguish between the modes-something we have taken great experimental pains to avoid. Instead, we require a model in which the animal must infer how its actions relate to navigation. Concretely, passive motion corresponds to a condition in which the hippocampus must passively learn arbitrary transition probabilities, while with active navigation the animal has greater certainty about which states will follow from the

[2] Furthermore, we can imagine a strong case for running signals being given 'special treatment' in navigational systems from physiological (Stranahan et al., 2007) and anatomical (Maurer et al., 2005) perspectives.

current one. This framing of the problem lends itself well to the family of models that derive from hidden Markov models (e.g. George et al., 2021), which we would expect to see implicitly discover multiple modes of travel as a latent variable.

The twist introduced by a BMI mode of navigation involves coupling the output of such a model to exploration of a simulated environment. To mimic our in vivo BMI protocol, we would need to 'decode' a neural state and then enforce a mapping from this state to a navigation action. Is this feedback loop sufficient to disrupt the separation between the contexts and reproduce the superposition effect—the argument that superposition is the intrinsic result of a sequence predictor caught in a self-fulfilling prophecy—or is the injection of an extrinsic agency or mode-of-travel signal a prerequisite for reproducing our experimental observations?

6.4 Closing remarks

It is an exciting time to be an engineer in neuroscience. We are at a point in the field's history where technological advances lead to observations that challenge established wisdom, and this in turn feeds into new inventions. For example, we do not need to look far into the past—just prior to the emergence of long-term population-wide imaging—to find a view of memory that treats expert behaviour as synaptically crystalline.

This thesis sits at an intersection of BMI design, abstract representations of navigation within the brain, and the constraints on how representations evolve over time. Viewed as a component within an organism, the brain plays the role of incorporating new information about a potentially changing external world. New circumstances or new experiences alter neural representations, and in unfamiliar tasks it is the responsibility of the brain to generate novel behaviour. It seems counterintuitive, then, to seek to constrain neural activity to fixed patterns so that it complies to an immutable decoder. We look forward to the development, over the coming years, of neural decoders that are themselves more plastic, their designs shaped by a sympathy for the underlying flexibility of the systems with which they are meant to interact. We hope that our work represents steps taken towards interfaces that are compatible with neural plasticity.

References

Reza Abiri, Soheil Borhani, Eric W. Sellers, Yang Jiang, and Xiaopeng Zhao. A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of Neural Engineering*, 16(1):011001, January 2019. doi: 10.1088/1741-2552/aaf12e.

Lavanya Acharya, Zahra M. Aghajan, Cliff Vuong, Jason J. Moore, and Mayank R. Mehta. Causal Influence of Visual Cues on Hippocampal Directional Selectivity. *Cell*, 164(1): 197–207, January 2016. doi: 10.1016/j.cell.2015.12.015.

Tyson Aflalo, Spencer Kellis, Christian Klaes, Brian Lee, Ying Shi, Kelsie Pejsa, Kathleen Shanfield, Stephanie Hayes-Jackson, Mindy Aisen, Christi Heck, Charles Liu, and Richard A. Andersen. Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science*, 348(6237):906–910, May 2015. doi: 10.1126/science.aaa5417.

Zahra M. Aghajan, Lavanya Acharya, Jason J. Moore, Jesse D. Cushman, Cliff Vuong, and Mayank R. Mehta. Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. *Nature Neuroscience*, 18(1):121–128, January 2015. doi: 10.1038/nn.3884.

Andrew S. Alexander, Jennifer C. Robinson, Holger Dannenberg, Nathaniel R. Kinsky, Samuel J. Levy, William Mau, G. William Chapman, David W. Sullivan, and Michael E. Hasselmo. Neurophysiological coding of space and time in the hippocampus, entorhinal cortex, and retrosplenial cortex. *Brain and Neuroscience Advances*, 4:2398212820972871, November 2020. doi: 10.1177/2398212820972871.

Richard A. Andersen, Tyson Aflalo, and Spencer Kellis. From thought to action: The brain–machine interface in posterior parietal cortex. *Proceedings of the National Academy of Sciences*, 116(52):26274–26279, December 2019. doi: 10.1073/pnas.1902276116.

Michael I. Anderson and Kathryn J. Jeffery. Heterogeneous Modulation of Place Cell Firing by Changes in Context. *Journal of Neuroscience*, 23(26):8827–8835, October 2003. doi: 10.1523/JNEUROSCI.23-26-08827.2003.

Gopala K. Anumanchipalli, Josh Chartier, and Edward F. Chang. Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753):493–498, April 2019. doi: 10.1038/s41586-019-1119-1.

Dmitriy Aronov, Aaron S. Andalman, and Michale S. Fee. A Specialized Forebrain Circuit for Vocal Babbling in the Juvenile Songbird. *Science*, 320(5876):630–634, May 2008. doi: 10.1126/science.1155140.

Alessio Attardo, James E. Fitzgerald, and Mark J. Schnitzer. Impermanence of dendritic spines in live adult CA1 hippocampus. *Nature*, 523(7562):592–596, July 2015. doi: 10.1038/nature14467.

Aaron P. Batista. Brain-computer interfaces for basic neuroscience. *Handbook of Clinical Neurology*, 168:233–247, 2020. doi: 10.1016/B978-0-444-63934-9.00017-2.

Katie C. Bittner, Aaron D. Milstein, Christine Grienberger, Sandro Romani, and Jeffrey C. Magee. Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science*, 357 (6355):1033–1036, September 2017. doi: 10.1126/science.aan3846.

Charlotte N. Boccara, Francesca Sargolini, Veslemøy Hult Thoresen, Trygve Solstad, Menno P. Witter, Edvard I. Moser, and May-Britt Moser. Grid cells in pre- and parasubiculum. *Nature Neuroscience*, 13(8):987–994, August 2010. doi: 10.1038/nn.2602.

Vegard Heimly Brun, Trygve Solstad, Kirsten Brun Kjelstrup, Marianne Fyhn, Menno P. Witter, Edvard I. Moser, and May-Britt Moser. Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, 18(12):1200–1212, 2008. doi: 10.1002/hipo.20504.

Laure Buhry, Amir H. Azizi, and Sen Cheng. Reactivation, Replay, and Preplay: How It Might All Fit Together. *Neural Plasticity*, 2011:203462, 2011. doi: 10.1155/2011/203462.

Daniel Bush, Caswell Barry, Daniel Manson, and Neil Burgess. Using Grid Cells for Navigation. *Neuron*, 87(3):507–520, August 2015. doi: 10.1016/j.neuron.2015.07.006.

György Buzsáki and Edvard I. Moser. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16(2):130–138, February 2013. doi: 10.1038/nn.3304.

György Buzsáki and David Tingley. Space and Time: The Hippocampus as a Sequence Generator. *Trends in Cognitive Sciences*, 22(10):853–869, October 2018. doi: 10.1016/j.tics.2018.07.006.

György Buzsáki, Adrien Peyrache, and John Kubie. Emergence of Cognition from Action. *Cold Spring Harbor Symposia on Quantitative Biology*, 79:41–50, January 2014. doi: 10.1101/sqb.2014.79.024679.

Malcolm G. Campbell, Samuel A. Ocko, Caitlin S. Mallory, Isabel I. C. Low, Surya Ganguli, and Lisa M. Giocomo. Principles governing the integration of landmark and self-motion cues in entorhinal cortical codes for navigation. *Nature Neuroscience*, 21(8):1096–1106, August 2018. doi: 10.1038/s41593-018-0189-y.

Subhojit Chakraborty, Michael I. Anderson, Amir M. Chaudhry, Julie C. Mumford, and Kathryn J. Jeffery. Context-independent directional cue learning by hippocampal place cells. *European Journal of Neuroscience*, 20(1):281–292, 2004. doi: 10.1111/j.1460-9568.2004.03464.x.

Chia-Jung Chang, Wei Guo, Jie Zhang, Jon Newman, Shao-Hua Sun, and Matt Wilson. Behavioral clusters revealed by end-to-end decoding from microendoscopic imaging. *bioRxiv*, April 2021. doi: 10.1101/2021.04.15.440055.

Josh Chartier, Gopala K. Anumanchipalli, Keith Johnson, and Edward F. Chang. Encoding of Articulatory Kinematic Trajectories in Human Speech Sensorimotor Cortex. *Neuron*, 98(5):1042–1054.e4, June 2018. doi: 10.1016/j.neuron.2018.04.031.

Steven M. Chase, Andrew B. Schwartz, and Robert E. Kass. Bias, optimal linear estimation, and the differences between open-loop simulation and closed-loop performance of spiking-based brain–computer interface algorithms. *Neural Networks*, 22(9):1203–1213, November 2009. doi: 10.1016/j.neunet.2009.05.005.

Pratik Chaudhari and Stefano Soatto. Stochastic Gradient Descent Performs Variational Inference, Converges to Limit Cycles for Deep Networks. In *2018 Information Theory and Applications Workshop (ITA)*, pages 1–10, February 2018. doi: 10.1109/ITA.2018.8503224.

Guifen Chen, John A. King, Neil Burgess, and John O'Keefe. How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences*, 110(1):378–383, January 2013. doi: 10.1073/pnas.1215834110.

Guifen Chen, John A. King, Yi Lu, Francesca Cacucci, and Neil Burgess. Spatial cell firing during virtual navigation of open arenas by head-restrained mice. *eLife*, 7:e34789, June 2018. doi: 10.7554/eLife.34789.

Zhe Chen, Garrett J. Blair, Changliang Guo, Jim Zhou, Juan-Luis Romero-Sosa, Alicia Izquierdo, Peyman Golshani, Jason Cong, Daniel Aharoni, and Hugh T Blair. A hardware system for real-time decoding of in vivo calcium imaging data. *eLife*, 12:e78344, January 2023. doi: 10.7554/eLife.78344.

Soon K. Cheong, Wenjun Xiong, Jennifer M. Strazzeri, Constance L. Cepko, David R. Williams, and William H. Merigan. In Vivo Functional Imaging of Retinal Neurons Using Red and Green Fluorescent Calcium Indicators. *Advances in experimental medicine and biology*, 1074:135–144, 2018. doi: 10.1007/978-3-319-75402-4_17.

Cynthia A Chestek, Vikash Gilja, Paul Nuyujukian, Justin D Foster, Joline M Fan, Matthew T Kaufman, Mark M Churchland, Zuley Rivera-Alvidrez, John P Cunningham, Stephen I Ryu, and Krishna V Shenoy. Long-term stability of neural prosthetic control signals from silicon cortical arrays in rhesus macaque motor cortex. *Journal of Neural Engineering*, 8(4):045005, July 2011. doi: 10.1088/1741-2560/8/4/045005.

Breanne P. Christie, Derek M. Tat, Zachary T. Irwin, Vikash Gilja, Paul Nuyujukian, Justin D. Foster, Stephen I. Ryu, Krishna V. Shenoy, David E. Thompson, and Cynthia A. Chestek. Comparison of spike sorting and thresholding of voltage waveforms for intracortical brain–machine interface performance. *Journal of Neural Engineering*, 12(1):016009, December 2014. doi: 10.1088/1741-2560/12/1/016009.

Shuo-Yen Chueh, Yuanxin Chen, Narayan Subramanian, Benjamin Goolsby, Phillip Navarro, and Karim Oweiss. Meta plasticity and Continual Learning: Mechanisms subserving Brain Computer Interface Proficiency. *bioRxiv*, March 2025. doi: 10.1101/2025.03.22.644643.

Kelly B. Clancy, Aaron C. Koralek, Rui M. Costa, Daniel E. Feldman, and Jose M. Carmena. Volitional modulation of optically recorded calcium signals during neuroprosthetic learning. *Nature Neuroscience*, 17(6):807–809, June 2014. doi: 10.1038/nn.3712.

Laura L. Colgin, Stefan Leutgeb, Karel Jezek, Jill K. Leutgeb, Edvard I. Moser, Bruce L. McNaughton, and May-Britt Moser. Attractor-Map Versus Autoassociation Based Attractor Dynamics in the Hippocampal Network. *Journal of Neurophysiology*, 104(1):35–50, July 2010. doi: 10.1152/jn.00202.2010.

Laura Lee Colgin, Edvard I. Moser, and May-Britt Moser. Understanding memory through hippocampal remapping. *Trends in Neurosciences*, 31(9):469–477, September 2008. doi: 10.1016/j.tins.2008.06.008.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, September 1995. doi: 10.1007/BF00994018.

L. Elizabeth Crawford, Laura E. Knouse, Molly Kent, D. Vavra, O. Harding, D. LeServe, N. Fox, X. Hu, P. Li, C. Glory, and Kelly G. Lambert. Enriched environment exposure accelerates rodent driving skills. *Behavioural Brain Research*, 378:112309, January 2020. doi: 10.1016/j.bbr.2019.112309.

M.C. Dadarlat, R.A. Canfield, and A.L. Orsborn. Neural Plasticity in Sensorimotor Brain-Machine Interfaces. *Annual Review of Biomedical Engineering*, 25:51–76, 2023. doi: 10.1146/annurev-bioeng-110220-110833.

Nicole David, Albert Newen, and Kai Vogeley. The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition*, 17(2):523–534, June 2008. doi: 10.1016/j.concog.2008.03.004.

Elizabeth A. de Laittre and Jason N. MacLean. The natural variability of a dexterous motor skill is stably encoded in the cortex of freely behaving mice. *bioRxiv*, 2024.

Antoine Dedieu, Nishad Gothoskar, Scott Swingle, Wolfgang Lehrach, Miguel Lázaro-Gredilla, and Dileep George. Learning higher-order sequential structure with cloned HMMs. *arXiv*, May 2019. doi: 10.48550/arXiv.1905.00507.

Alan D. Degenhart, William E. Bishop, Emily R. Oby, Elizabeth C. Tyler-Kabara, Steven M. Chase, Aaron P. Batista, and Byron M. Yu. Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity. *Nature Biomedical Engineering*, 4(7):672–685, July 2020. doi: 10.1038/s41551-020-0542-9.

Daniel Deitch, Alon Rubin, and Yaniv Ziv. Representational drift in the mouse visual cortex. *Current Biology*, 31(19):4327–4339.e6, October 2021. doi: 10.1016/j.cub.2021.07.062.

Federico Devalle, Licheng Zou, Gloria Cecchini, and Alex Roxin. Representational drift as the consequence of ongoing memory storage. *bioRxiv*, December 2024. doi: 10.1101/2024.06.25.600729.

Jack DiGiovanna, Babak Mahmoudi, Jose Fortes, Jose C. Principe, and Justin C. Sanchez. Coadaptive Brain–Machine Interface via Reinforcement Learning. *IEEE Transactions on Biomedical Engineering*, 56(1):54–64, January 2009. doi: 10.1109/TBME.2008.926699.

Daniel A. Dombeck, Christopher D. Harvey, Lin Tian, Loren L. Looger, and David W. Tank. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neuroscience*, 13(11):1433–1440, November 2010. doi: 10.1038/nn.2648.

Giulia Dominijanni, Solaiman Shokur, Gionata Salvietti, Sarah Buehler, Erica Palmerini, Simone Rossi, Frederique De Vignemont, Andrea d'Avella, Tamar R. Makin, Domenico Prattichizzo, and Silvestro Micera. The neural resource allocation problem when enhancing human bodies with extra robotic limbs. *Nature Machine Intelligence*, 3(10):850–860, October 2021. doi: 10.1038/s42256-021-00398-9.

John E. Downey, Nathaniel Schwed, Steven M. Chase, Andrew B. Schwartz, and Jennifer L. Collinger. Intracortical recording stability in human brain-computer interface users. *Journal of Neural Engineering*, 15(4):046016, August 2018. doi: 10.1088/1741-2552/aab7a0.

Laura N. Driscoll, Noah L. Pettit, Matthias Minderer, Selmaan N. Chettih, and Christopher D. Harvey. Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. *Cell*, 170 (5):986–999.e16, August 2017. doi: 10.1016/j.cell.2017.07.021.

Laura N. Driscoll, Lea Duncker, and Christopher D. Harvey. Representational drift: Emerging theories for continual learning and experimental future directions. *Current Opinion in Neurobiology*, 76:102609, October 2022. doi: 10.1016/j.conb.2022.102609.

Daniel Durstewitz, Nicole M. Vittoz, Stan B. Floresco, and Jeremy K. Seamans. Abrupt Transitions between Prefrontal Neural Ensemble States Accompany Behavioral Transitions during Rule Learning. *Neuron*, 66(3):438–448, May 2010. doi: 10.1016/j.neuron.2010.03.029.

Eva L. Dyer, Mohammad Gheshlaghi Azar, Matthew G. Perich, Hugo L. Fernandes, Stephanie Naufel, Lee E. Miller, and Konrad P. Körding. A cryptography-based approach for movement decoding. *Nature Biomedical Engineering*, 1(12):967–976, December 2017. doi: 10.1038/s41551-017-0169-7.

Sadegh Ebrahimi, Jérôme Lecoq, Oleg Rumyantsev, Tugce Tasci, Yanping Zhang, Cristina Irimia, Jane Li, Surya Ganguli, and Mark J. Schnitzer. Emergent reliability in sensory cortical coding and inter-area communication. *Nature*, 605(7911):713–721, May 2022. doi: 10.1038/s41586-022-04724-y.

Samuel Eckmann, Yashar Ahmadian, and Máté Lengyel. Theta-modulated memory encoding and retrieval in recurrent hippocampal circuits. In *Bernstein Conference*, 2024. doi: 10.12751/nncn.bc2024.016.

Gal Elyasaf, Alon Rubin, and Yaniv Ziv. Novel off-context experience constrains hippocampal representational drift. *Current Biology*, November 2024. doi: 10.1016/j.cub.2024.10.027.

Jens-Bastian Eppler, Thomas Lai, Dominik Aschauer, Simon Rumpel, and Matthias Kaschube. Representational drift reflects ongoing balancing of stochastic changes by Hebbian learning. *bioRxiv*, January 2025. doi: 10.1101/2025.01.05.631363.

Oxana Eschenko, Wiâm Ramadan, Matthias Mölle, Jan Born, and Susan J. Sara. Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. *Learning & Memory*, 15(4):222–228, April 2008. doi: 10.1101/lm.726008.

Ariane S. Etienne and Kathryn J. Jeffery. Path integration in mammals. *Hippocampus*, 14(2): 180–192, 2004. doi: 10.1002/hipo.10173.

Joline M. Fan, Paul Nuyujukian, Jonathan C. Kao, Cynthia A. Chestek, Stephen I. Ryu, and Krishna V. Shenoy. Intention estimation in brain machine interfaces. *Journal of Neural Engineering*, 11(1):016004, January 2014. doi: 10.1088/1741-2560/11/1/016004.

André A. Fenton. Remapping revisited: how the hippocampus represents different spaces. *Nature Reviews Neuroscience*, 25(6):428–448, June 2024. doi: 10.1038/s41583-024-00817-x.

Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I. Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132): 190–194, March 2007. doi: 10.1038/nature05601.

Juan A. Gallego, Matthew G. Perich, Lee E. Miller, and Sara A. Solla. Neural Manifolds for the Control of Movement. *Neuron*, 94(5):978–984, June 2017. doi: 10.1016/j.neuron.2017.05.025.

Juan A. Gallego, Matthew G. Perich, Raeed H. Chowdhury, Sara A. Solla, and Lee E. Miller. Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience*, 23(2):260–270, February 2020. doi: 10.1038/s41593-019-0555-4.

Juan A. Gallego, Tamar R. Makin, and Samuel D. McDougle. Going beyond primary motor cortex to improve brain–computer interfaces. *Trends in Neurosciences*, 45(3):176–183, March 2022. doi: 10.1016/j.tins.2021.12.006.

Karunesh Ganguly, Dragan F. Dimitrov, Jonathan D. Wallis, and Jose M. Carmena. Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nature Neuroscience*, 14(5):662–667, May 2011. doi: 10.1038/nn.2797.

Mona M. Garvert, Raymond J. Dolan, and Timothy E. J. Behrens. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, 6:e17086, April 2017. doi: 10.7554/eLife.17086.

Jeffrey L. Gauthier and David W. Tank. A Dedicated Population for Reward Coding in the Hippocampus. *Neuron*, 99(1):179–193.e7, July 2018. doi: 10.1016/j.neuron.2018.06.008.

Tristan Geiller, Mohammad Fattahi, June-Seek Choi, and Sébastien Royer. Place cells are more strongly tied to landmarks in deep than in superficial CA1. *Nature Communications*, 8(1): 14531, February 2017. doi: 10.1038/ncomms14531.

Caroline Geisler, David Robbe, Michaël Zugaro, Anton Sirota, and György Buzsáki. Hippocampal place cell assemblies are speed-controlled oscillators. *Proceedings of the National Academy of Sciences*, 104(19):8149–8154, May 2007. doi: 10.1073/pnas.0610121104.

Dileep George, Rajeev V. Rikhye, Nishad Gothoskar, J. Swaroop Guntupalli, Antoine Dedieu, and Miguel Lázaro-Gredilla. Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nature Communications*, 12(1):2392, April 2021. doi: 10.1038/s41467-021-22559-5.

Samuel J. Gershman. The Successor Representation: Its Computational Logic and Neural Substrates. *Journal of Neuroscience*, 38(33):7193–7200, August 2018. doi: 10.1523/JNEUROSCI.0151-18.2018.

Samuel J. Gershman, Christopher D. Moore, Michael T. Todd, Kenneth A. Norman, and Per B. Sederberg. The Successor Representation and Temporal Context. *Neural Computation*, 24 (6):1553–1568, June 2012. doi: 10.1162/NECO_a_00282.

Nitzan Geva, Daniel Deitch, Alon Rubin, and Yaniv Ziv. Time and experience differentially affect distinct aspects of hippocampal representational drift. *Neuron*, 111(15):2357–2366.e5, August 2023. doi: 10.1016/j.neuron.2023.05.005.

Maya Geva-Sagiv, Liora Las, Yossi Yovel, and Nachum Ulanovsky. Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation. *Nature Reviews Neuroscience*, 16(2):94–108, February 2015. doi: 10.1038/nrn3888.

Mariana Gil, Mihai Ancau, Magdalene I. Schlesiger, Angela Neitz, Kevin Allen, Rodrigo J. De Marco, and Hannah Monyer. Impaired path integration in mice with disrupted grid cell firing. *Nature Neuroscience*, 21(1):81–91, January 2018. doi: 10.1038/s41593-017-0039-3.

Vikash Gilja, Paul Nuyujukian, Cindy A. Chestek, John P. Cunningham, Byron M. Yu, Joline M. Fan, Mark M. Churchland, Matthew T. Kaufman, Jonathan C. Kao, Stephen I. Ryu, and Krishna V. Shenoy. A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, 15(12):1752–1757, December 2012. doi: 10.1038/nn.3265.

Andrea Giovannucci, Johannes Friedrich, Pat Gunn, Jérémie Kalfon, Brandon L. Brown, Sue Ann Koay, Jiannis Taxidis, Farzaneh Najafi, Jeffrey L. Gauthier, Pengcheng Zhou, Baljit S. Khakh, David W. Tank, Dmitri B. Chklovskii, and Eftychios A. Pnevmatikakis. CaImAn an open source tool for scalable calcium imaging data analysis. *eLife*, 8:e38173, January 2019. doi: 10.7554/eLife.38173.

Gabrielle Girardeau, Karim Benchenane, Sidney I. Wiener, György Buzsáki, and Michaël B. Zugaro. Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, 12(10):1222–1223, October 2009. doi: 10.1038/nn.2384.

Sheridan Goldstein, Lupeng Wang, Kerry McAlonan, Mateus Torres-Cruz, and Richard J. Krauzlis. Stimulus-driven visual attention in mice. *Journal of Vision*, 22(1):11, January 2022. doi: 10.1167/jov.22.1.11.

Katalin M. Gothard, Kari L. Hoffman, Francesco P. Battaglia, and Bruce L. McNaughton. Dentate Gyrus and CA1 Ensemble Activity during Spatial Reference Frame Shifts in the Presence and Absence of Visual Input. *The Journal of Neuroscience*, 21(18):7284–7292, September 2001. doi: 10.1523/JNEUROSCI.21-18-07284.2001.

Dorian Goueytes, Henri Lassagne, Daniel E. Shulz, Valérie Ego-Stengel, and Luc Estebanez. Learning in a closed-loop brain-machine interface with distributed optogenetic cortical feedback. *Journal of Neural Engineering*, 19(6):066045, December 2022. doi: 10.1088/1741-2552/acab87.

Benjamin F. Grewe, Dominik Langer, Hansjörg Kasper, Björn M. Kampa, and Fritjof Helmchen. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature Methods*, 7(5):399–405, May 2010. doi: 10.1038/nmeth.1453.

Christine Grienberger, Andrea Giovannucci, William Zeiger, and Carlos Portera-Cailliau. Two-photon calcium imaging of neuronal activity. *Nature Reviews Methods Primers*, 2(1):1–23, September 2022. doi: 10.1038/s43586-022-00147-1.

Dori M. Grijseels, Kira Shaw, Caswell Barry, and Catherine N. Hall. Choice of method of place cell classification determines the population of cells identified. *PLoS computational biology*, 17(7):e1008835, July 2021. doi: 10.1371/journal.pcbi.1008835.

Zé Henrique T. D. Góis and Adriano B. L. Tort. Characterizing Speed Cells in the Rat Hippocampus. *Cell Reports*, 25(7):1872–1884.e4, November 2018. doi: 10.1016/j.celrep.2018.10.054.

Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, August 2005. doi: 10.1038/nature03721.

Aleesha Hamid and Per Ola Kristensson. 40 Years of Eye Typing: Challenges, Gaps, and Emergent Strategies. *Proc. ACM Hum.-Comput. Interact.*, 8(ETRA):222:1–222:19, May 2024. doi: 10.1145/3655596.

Timothy L. Hanson, Camilo A. Diaz-Botia, Viktor Kharazia, Michel M. Maharbiz, and Philip N. Sabes. The "sewing machine" for minimally invasive neural recording. *bioRxiv*, March 2019. doi: 10.1101/578542.

Christopher D. Harvey, Philip Coen, and David W. Tank. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62–68, April 2012. doi: 10.1038/nature10918.

Demis Hassabis, Carlton Chu, Geraint Rees, Nikolaus Weiskopf, Peter D. Molyneux, and Eleanor A. Maguire. Decoding Neuronal Ensembles in the Human Hippocampus. *Current Biology*, 19(7):546–554, April 2009. doi: 10.1016/j.cub.2009.02.033.

James B. Heald, Máté Lengyel, and Daniel M. Wolpert. Contextual inference in learning and memory. *Trends in Cognitive Sciences*, 27(1):43–64, January 2023. doi: 10.1016/j.tics.2022.10.004.

Espen J. Henriksen, Laura L. Colgin, Carol A. Barnes, Menno P. Witter, May-Britt Moser, and Edvard I. Moser. Spatial representation along the proximodistal axis of CA1. *Neuron*, 68(1):127–137, October 2010. doi: 10.1016/j.neuron.2010.08.042.

Christian Herff, Dominic Heger, Adriana de Pesters, Dominic Telaar, Peter Brunner, Gerwin Schalk, and Tanja Schultz. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in Neuroscience*, 9, 2015.

Riichiro Hira, Fuki Ohkubo, Yoshito Masamizu, Masamichi Ohkura, Junichi Nakai, Takashi Okada, and Masanori Matsuzaki. Reward-timing-dependent bidirectional modulation of cortical microcircuits during optical single-neuron operant conditioning. *Nature Communications*, 5(1):5551, November 2014. doi: 10.1038/ncomms6551.

Leigh R. Hochberg, Mijail D. Serruya, Gerhard M. Friehs, Jon A. Mukand, Maryam Saleh, Abraham H. Caplan, Almut Branner, David Chen, Richard D. Penn, and John P. Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–171, July 2006. doi: 10.1038/nature04970.

Leigh R. Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y. Masse, John D. Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S. Cash, Patrick van der Smagt, and John P. Donoghue. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375, May 2012. doi: 10.1038/nature11076.

Stig A. Hollup, Sturla Molden, James G. Donnett, May-Britt Moser, and Edvard I. Moser. Accumulation of Hippocampal Place Fields at the Goal Location in an Annular Watermaze Task. *Journal of Neuroscience*, 21(5):1635–1644, March 2001. doi: 10.1523/JNEUROSCI.21-05-01635.2001.

Elisa Mira Holz, Loic Botrel, Tobias Kaufmann, and Andrea Kübler. Long-term independent brain-computer interface home use improves quality of life of a patient in the locked-in state: a case study. *Archives of Physical Medicine and Rehabilitation*, 96(3 Suppl):S16–26, March 2015. doi: 10.1016/j.apmr.2014.03.035.

Christian Hölscher, Wolfgang Jacob, and Hanspeter A Mallot. Reward modulates neuronal activity in the hippocampus of the rat. *Behavioural Brain Research*, 142(1):181–191, June 2003. doi: 10.1016/S0166-4328(02)00422-9.

Daisuke Ishikawa, Nobuyoshi Matsumoto, Tetsuya Sakaguchi, Norio Matsuki, and Yuji Ikegaya. Operant Conditioning of Synaptic and Spiking Activity Patterns in Single Hippocampal Neurons. *Journal of Neuroscience*, 34(14):5044–5053, April 2014. doi: 10.1523/JNEUROSCI.5298-13.2014.

Hiroyuki Ito, Soichiro Fujiki, Yoshiya Mori, and Kenji Kansaku. Self-reorganization of neuronal activation patterns in the cortex under brain-machine interface and neural operant conditioning. *Neuroscience Research*, 156:279–292, July 2020. doi: 10.1016/j.neures.2020.03.008.

Alicia Izquierdo, Jonathan L. Brigman, Anna K. Radke, Peter H. Rudebeck, and Andrew Holmes. The neural basis of reversal learning: An updated perspective. *Neuroscience*, 345:12–26, March 2017. doi: 10.1016/j.neuroscience.2016.03.021.

Shantanu P. Jadhav, Caleb Kemere, P. Walter German, and Loren M. Frank. Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory. *Science (New York, N.Y.)*, 336(6087):1454–1458, June 2012. doi: 10.1126/science.1217230.

Beata Jarosiewicz, Steven M. Chase, George W. Fraser, Meel Velliste, Robert E. Kass, and Andrew B. Schwartz. Functional network reorganization during learning in a brain-computer interface paradigm. *Proceedings of the National Academy of Sciences*, 105(49):19486–19491, December 2008. doi: 10.1073/pnas.0808113105.

Beata Jarosiewicz, Anish A. Sarma, Daniel Bacher, Nicolas Y. Masse, John D. Simeral, Brittany Sorice, Erin M. Oakley, Christine Blabe, Chethan Pandarinath, Vikash Gilja, Sydney S. Cash, Emad N. Eskandar, Gerhard Friehs, Jaimie M. Henderson, Krishna V. Shenoy, John P. Donoghue, and Leigh R. Hochberg. Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. *Science Translational Medicine*, 7(313):313ra179–313ra179, November 2015. doi: 10.1126/scitranslmed.aac7328.

Ravikrishnan P. Jayakumar, Manu S. Madhav, Francesco Savelli, Hugh T. Blair, Noah J. Cowan, and James J. Knierim. Recalibration of path integration in hippocampal place cells. *Nature*, 566(7745):533–537, February 2019. doi: 10.1038/s41586-019-0939-3.

Kathryn J. Jeffery. Place Cells, Grid Cells, Attractors, and Remapping. *Neural Plasticity*, 2011(1):182602, 2011. doi: 10.1155/2011/182602.

Kathryn J. Jeffery and John M. O'Keefe. Learned interaction of visual and idiothetic cues in the control of place field orientation. *Experimental Brain Research*, 127(2):151–161, July 1999. doi: 10.1007/s002210050785.

Kathryn J. Jeffery, Alexandra Gilbert, Stephen Burton, and Anna Strudwick. Preserved performance in a hippocampal-dependent spatial task despite complete place cell remapping. *Hippocampus*, 13(2):175–189, 2003. doi: 10.1002/hipo.10047.

Kristopher T. Jensen, Naama Kadmon Harpaz, Ashesh K. Dhawale, Steffen B. E. Wolff, and Bence P. Ölveczky. Long-term stability of single neuron activity in the motor system. *Nature Neuroscience*, 25(12):1664–1674, December 2022. doi: 10.1038/s41593-022-01194-3.

Kristopher T. Jensen, Guillaume Hennequin, and Marcelo G. Mattar. A recurrent network model of planning explains hippocampal replay and human behavior. *Nature Neuroscience*, 27(7):1340–1348, July 2024. doi: 10.1038/s41593-024-01675-7.

Brian B. Jeon, Thomas Fuchs, Steven M. Chase, and Sandra J. Kuhlman. Existing function in primary visual cortex is not perturbed by new skill acquisition of a non-matched sensory task. *Nature Communications*, 13(1):3638, June 2022. doi: 10.1038/s41467-022-31440-y.

Karel Jezek, Espen J. Henriksen, Alessandro Treves, Edvard I. Moser, and May-Britt Moser. Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, 478(7368): 246–249, October 2011. doi: 10.1038/nature10439.

Adam Johnson and A. David Redish. Neural Ensembles in CA3 Transiently Encode Paths Forward of the Animal at a Decision Point. *Journal of Neuroscience*, 27(45):12176–12189, November 2007. doi: 10.1523/JNEUROSCI.3761-07.2007.

James J. Jun, Nicholas A. Steinmetz, Joshua H. Siegle, Daniel J. Denman, Marius Bauza, Brian Barbarits, Albert K. Lee, Costas A. Anastassiou, Alexandru Andrei, Çağatay Aydın, Mladen Barbic, Timothy J. Blanche, Vincent Bonin, João Couto, Barundeb Dutta, Sergey L. Gratiy, Diego A. Gutnisky, Michael Häusser, Bill Karsh, Peter Ledochowitsch, Carolina Mora Lopez, Catalin Mitelut, Silke Musa, Michael Okun, Marius Pachitariu, Jan Putzeys, P. Dylan Rich, Cyrille Rossant, Wei-Lung Sun, Karel Svoboda, Matteo Carandini, Kenneth D. Harris, Christof Koch, John O'Keefe, and Timothy D. Harris. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236, November 2017. doi: 10.1038/nature24636.

Brianna M. Karpowicz, Yahia H. Ali, Lahiru N. Wimalasena, Andrew R. Sedler, Mohammad Reza Keshtkaran, Kevin Bodkin, Xuan Ma, Daniel B. Rubin, Ziv M. Williams, Sydney S. Cash, Leigh R. Hochberg, Lee E. Miller, and Chethan Pandarinath. Stabilizing brain-computer interfaces through alignment of latent dynamics. *Nature Communications*, 16(1):4662, May 2025. doi: 10.1038/s41467-025-59652-y.

Kalman A. Katlowitz, Michel A. Picardo, and Michael A. Long. Stable Sequential Activity Underlying the Maintenance of a Precisely Executed Skilled Behavior. *Neuron*, 98(6):1133–1140.e3, June 2018. doi: 10.1016/j.neuron.2018.05.017.

Alexandra Mansell Kaufman, Tristan Geiller, and Attila Losonczy. A Role for the Locus Coeruleus in Hippocampal CA1 Place Cell Reorganization during Spatial Reward Learning. *Neuron*, 105(6):1018–1026.e4, March 2020. doi: 10.1016/j.neuron.2019.12.029.

Pamela J. Kennedy and Matthew L. Shapiro. Motivational states activate distinct hippocampal representations to guide goal-directed behaviors. *Proceedings of the National Academy of Sciences*, 106(26):10805–10810, June 2009. doi: 10.1073/pnas.0903259106.

Philip R. Kennedy and Roy A. E. Bakay. Restoration of neural output from a paralyzed patient by a direct brain connection. *NeuroReport*, 9(8):1707, June 1998.

Clifford G. Kentros, Naveen T. Agnihotri, Samantha Streater, Robert D. Hawkins, and Eric R. Kandel. Increased Attention to Spatial Context Increases Both Place Field Stability and Spatial Memory. *Neuron*, 42(2):283–295, April 2004. doi: 10.1016/S0896-6273(04)00192-8.

Dorgham Khatib, Aviv Ratzon, Mariell Sellevoll, Omri Barak, Genela Morris, and Dori Derdikman. Active experience, not time, determines within-day representational drift in dorsal CA1. *Neuron*, 111(15):2348–2356.e4, August 2023. doi: 10.1016/j.neuron.2023.05.014.

Sung-Phil Kim, John D. Simeral, Leigh R. Hochberg, John P. Donoghue, and Michael J. Black. Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *Journal of neural engineering*, 5(4):455–476, December 2008. doi: 10.1088/1741-2560/5/4/010.

Tony Hyun Kim and Mark J. Schnitzer. Fluorescence imaging of large-scale neural ensemble dynamics. *Cell*, 185(1):9–41, January 2022. doi: 10.1016/j.cell.2021.12.007.

James J. Knierim. Dynamic Interactions between Local Surface Cues, Distal Landmarks, and Intrinsic Circuitry in Hippocampal Place Cells. *Journal of Neuroscience*, 22(14):6254–6264, July 2002. doi: 10.1523/JNEUROSCI.22-14-06254.2002.

James J. Knierim and Geeta Rao. Distal landmarks and hippocampal place cells: Effects of relative translation versus rotation. *Hippocampus*, 13(5):604–617, 2003. doi: 10.1002/hipo.10092.

James J. Knierim, Hemant S. Kudrimoti, and Bruce L. McNaughton. Place cells, head direction cells, and the learning of landmark stability. *Journal of Neuroscience*, 15(3):1648–1659, March 1995. doi: 10.1523/JNEUROSCI.15-03-01648.1995.

Benjamin J. Kraus, Robert J. Robinson, John A. White, Howard Eichenbaum, and Michael E. Hasselmo. Hippocampal "Time Cells": Time versus Path Integration. *Neuron*, 78(6):1090–1101, June 2013. doi: 10.1016/j.neuron.2013.04.015.

Julija Krupic, Marius Bauza, Stephen Burton, and John O'Keefe. Local transformations of the hippocampal cognitive map. *Science*, 359(6380):1143–1146, March 2018. doi: 10.1126/science.aao4960.

Dharshan Kumaran, Hans Ludwig Melo, and Emrah Duzel. The Emergence and Representation of Knowledge about Social and Nonsocial Hierarchies. *Neuron*, 76(3):653–666, November 2012. doi: 10.1016/j.neuron.2012.09.035.

Chongxi Lai, Shinsuke Tanaka, Timothy D. Harris, and Albert K. Lee. Volitional activation of remote place representations with a hippocampal brain–machine interface. *Science*, 382(6670):566–573, November 2023. doi: 10.1126/science.adh5206.

Jae Sung Lee, John J. Briguglio, Jeremy D. Cohen, Sandro Romani, and Albert K. Lee. The Statistical Structure of the Hippocampal Code for Space as a Function of Time, Context, and Value. *Cell*, 183(3):620–635.e22, October 2020. doi: 10.1016/j.cell.2020.09.024.

Máté Lengyel, Jeehyun Kwag, Ole Paulsen, and Peter Dayan. Matching storage and recall: hippocampal spike timing–dependent plasticity and phase response curves. *Nature Neuroscience*, 8(12):1677–1683, December 2005. doi: 10.1038/nn1561.

Anthony Leonardo and Masakazu Konishi. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature*, 399(6735):466–470, June 1999. doi: 10.1038/20933.

Jill K. Leutgeb, Stefan Leutgeb, Alessandro Treves, Retsina Meyer, Carol A. Barnes, Bruce L. McNaughton, May-Britt Moser, and Edvard I. Moser. Progressive Transformation of Hippocampal Neuronal Representations in "Morphed" Environments. *Neuron*, 48(2):345–358, October 2005. doi: 10.1016/j.neuron.2005.09.007.

Eric C. Leuthardt, Gerwin Schalk, Jonathan R. Wolpaw, Jeffrey G. Ojemann, and Daniel W. Moran. A brain–computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, 1(2):63–71, June 2004. doi: 10.1088/1741-2560/1/2/001.

Colin Lever, Stephen Burton, Ali Jeewajee, John O'Keefe, and Neil Burgess. Boundary vector cells in the subiculum of the hippocampal formation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(31):9771–9777, August 2009. doi: 10.1523/JNEUROSCI.1319-09.2009.

Zheng Li, Joseph E. O'Doherty, Mikhail A. Lebedev, and Miguel A. L. Nicolelis. Adaptive decoding for brain-machine interfaces through Bayesian parameter updates. *Neural Computation*, 23(12):3162–3204, December 2011. doi: 10.1162/NECO_a_00207.

Zhenrui Liao and Attila Losonczy. Learning, Fast and Slow: Single- and Many-Shot Learning in the Hippocampus. *Annual Review of Neuroscience*, 47(Volume 47, 2024):187–209, August 2024. doi: 10.1146/annurev-neuro-102423-100258.

William A. Liberti, Tobias A. Schmid, Angelo Forli, Madeleine Snyder, and Michael M. Yartsev. A stable hippocampal code in freely flying bats. *Nature*, 604(7904):98–103, April 2022. doi: 10.1038/s41586-022-04560-0.

Mary J. Lindstrom, , and Douglas M. Bates. Newton—Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association*, 83(404):1014–1022, December 1988. doi: 10.1080/01621459.1988.10478693.

Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1):503–528, August 1989. doi: 10.1007/BF01589116.

Darby M. Losey, Jay A. Hennig, Emily R. Oby, Matthew D. Golub, Patrick T. Sadtler, Kristin M. Quick, Stephen I. Ryu, Elizabeth C. Tyler-Kabara, Aaron P. Batista, Byron M. Yu, and Steven M. Chase. Learning leaves a memory trace in motor cortex. *Current Biology*, 34(7):1519–1531.e4, April 2024. doi: 10.1016/j.cub.2024.03.003.

Bruce D. Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *7th international joint conference on Artificial intelligence*, page 10, 1981.

Anika T. Löwe, Léo Touzo, Paul S. Muhle-Karbe, Andrew M. Saxe, Christopher Summerfield, and Nicolas W. Schuck. Abrupt and spontaneous strategy switches emerge in simple regularised neural networks. *PLOS Computational Biology*, 20(10):e1012505, October 2024. doi: 10.1371/journal.pcbi.1012505.

Xuan Ma, Fabio Rizzoglio, Kevin L Bodkin, Eric Perreault, Lee E Miller, and Ann Kennedy. Using adversarial networks to extend brain computer interface decoding accuracy over time. *eLife*, 12:e84296, August 2023. doi: 10.7554/eLife.84296.

Manu S. Madhav, Ravikrishnan P. Jayakumar, Brian Y. Li, Shahin G. Lashkari, Kelly Wright, Francesco Savelli, James J. Knierim, and Noah J. Cowan. Control and recalibration of path integration in place cells using optic flow. *Nature Neuroscience*, 27(8):1599–1608, August 2024. doi: 10.1038/s41593-024-01681-9.

Tyler D. Marks and Michael J. Goard. Stimulus-dependent representational drift in primary visual cortex. *Nature Communications*, 12(1):5169, August 2021. doi: 10.1038/s41467-021-25436-3.

Etan J. Markus, Carol A. Barnes, Bruce L. McNaughton, Victoria L. Gladden, and William E. Skaggs. Spatial information content and reliability of hippocampal CA1 neurons: Effects of visual input. *Hippocampus*, 4(4):410–421, 1994. doi: 10.1002/hipo.450040404.

Andrew P. Maurer, Shea R. VanRhoads, Gary R. Sutherland, Peter Lipa, and Bruce L. McNaughton. Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. *Hippocampus*, 15(7):841–852, 2005. doi: 10.1002/hipo.20114.

Andrew P. Maurer, Sara N. Burke, Peter Lipa, William E. Skaggs, and Carol A. Barnes. Greater running speeds result in altered hippocampal phase sequence dynamics. *Hippocampus*, 22 (4):737–747, 2012. doi: 10.1002/hipo.20936.

Sam McKenzie. Inhibition shapes the organization of hippocampal representations. *Hippocampus*, 28(9):659–671, September 2018. doi: 10.1002/hipo.22803.

Bruce L. McNaughton, Carol A. Barnes, and John O'Keefe. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Experimental Brain Research*, 52(1):41–49, 1983. doi: 10.1007/BF00237147.

Bruce L. McNaughton, Francesco P. Battaglia, Ole Jensen, Edvard I. Moser, and May-Britt Moser. Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678, August 2006. doi: 10.1038/nrn1932.

W. Hamish Mehaffey and Allison J. Doupe. Naturalistic stimulation drives opposing heterosynaptic plasticity at two inputs to songbird cortex. *Nature Neuroscience*, 18(9):1272–1280, September 2015. doi: 10.1038/nn.4078.

Johanna H. Meijer and Yuri Robbers. Wheel running in the wild. *Proceedings of the Royal Society B: Biological Sciences*, 281(1786):20140210, July 2014. doi: 10.1098/rspb.2014.0210.

Jorge A. Menéndez, Jay A. Hennig, Matthew D. Golub, Emily R. Oby, Patrick T. Sadtler, Aaron P. Batista, Steven M. Chase, Byron M. Yu, and Peter E. Latham. A theory of brain-computer interface learning via low-dimensional control. *eLife*, 14, May 2025. doi: 10.7554/eLife.106309.1. Publisher: eLife Sciences Publications Limited.

Josh S Merel, Roy Fox, Tony Jebara, and Liam Paninski. A multi-agent control framework for co-adaptation in brain-computer interfaces. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Charles Micou. Reticade: the REal TIme CAlcium imaging DEcoder, April 2022. URL github.com/CharlesMicou/reticade.

Charles Micou and Timothy O'Leary. Representational drift as a window into neural and behavioural plasticity. *Current Opinion in Neurobiology*, 81:102746, August 2023. doi: 10.1016/j.conb.2023.102746.

Charles Micou and Timothy O'Leary. Heavy-tailed statistics of cortical representational drift are advantageous for stabilised downstream readouts. *bioRxiv*, September 2024. doi: 10.1101/2024.09.26.614914.

Matthias Minderer, Christopher D. Harvey, Flavio Donato, and Edvard I. Moser. Virtual reality explored. *Nature*, 533(7603):324–325, May 2016. doi: 10.1038/nature17899.

Kenji Mizuseki, Kamran Diba, Eva Pastalkova, and György Buzsáki. Hippocampal CA1 pyramidal cells form functionally distinct sublayers. *Nature neuroscience*, 14(9):1174–1181, August 2011. doi: 10.1038/nn.2894.

Ida Momennejad and Marc W. Howard. Predicting the Future with Multi-scale Successor Representations. *bioRxiv*, October 2018. doi: 10.1101/449470.

Edvard I. Moser, May-Britt Moser, and Bruce L. McNaughton. Spatial representation in the hippocampal formation: a history. *Nature Neuroscience*, 20(11):1448–1464, November 2017. doi: 10.1038/nn.4653.

Robert U. Muller and John L. Kubie. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 7(7):1951–1968, July 1987. doi: 10.1523/JNEUROSCI.07-07-01951.1987.

Robert U. Muller, Matthew Stead, and Janos Pach. The hippocampus as a cognitive graph. *Journal of General Physiology*, 107(6):663–694, June 1996. doi: 10.1085/jgp.107.6.663.

Simon Musall, Matthew T. Kaufman, Ashley L. Juavinett, Steven Gluf, and Anne K. Churchland. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, 22(10):1677–1686, October 2019. doi: 10.1038/s41593-019-0502-4.

Jan Saputra Müller, Carmen Vidaurre, Martijn Schreuder, Frank C. Meinecke, Paul von Bünau, and Klaus-Robert Müller. A mathematical model for the two-learners problem. *Journal of Neural Engineering*, 14(3):036005, June 2017. doi: 10.1088/1741-2552/aa620b.

Anne Nagelhus, Sebastian O. Andersson, Soledad Gonzalo Cogno, Edvard I. Moser, and May-Britt Moser. Object-centered population coding in CA1 of the hippocampus. *Neuron*, 111(13):2091–2104.e14, July 2023. doi: 10.1016/j.neuron.2023.04.008.

Junichi Nakai, Masamichi Ohkura, and Keiji Imoto. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nature Biotechnology*, 19(2):137–141, February 2001. doi: 10.1038/84397.

Nikhilesh Natraj, Sarah Seko, Reza Abiri, Runfeng Miao, Hongyi Yan, Yasmin Graham, Adelyn Tu-Chan, Edward F. Chang, and Karunesh Ganguly. Sampling representational plasticity of simple imagined movements across days enables long-term neuroprosthetic control. *Cell*, 188(5):1208–1225.e32, March 2025. doi: 10.1016/j.cell.2025.02.001.

Fernando Nottebohm. Auditory Experience and Song Development in the Chaffinch Fringilla Coelebs. *Ibis*, 110(4):549–568, 1968. doi: 10.1111/j.1474-919X.1968.tb00063.x.

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

John O'Keefe. Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, 51(1):78–109, January 1976. doi: 10.1016/0014-4886(76)90055-8.

John O'Keefe and Michael L. Recce. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3):317–330, 1993. doi: 10.1002/hipo.450030307.

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

Marius Pachitariu, Nicholas A. Steinmetz, Shabnam N. Kadir, Matteo Carandini, and Kenneth D. Harris. Fast and accurate spike sorting of high-channel count probes with KiloSort. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Marius Pachitariu, Carsen Stringer, Mario Dipoppa, Sylvia Schröder, L. Federico Rossi, Henry Dalgleish, Matteo Carandini, and Kenneth D. Harris. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. July 2017. doi: 10.1101/061507.

Marius Pachitariu, Carsen Stringer, and Kenneth D. Harris. Robustness of Spike Deconvolution for Neuronal Calcium Imaging. *Journal of Neuroscience*, 38(37):7976–7985, September 2018. doi: 10.1523/JNEUROSCI.3339-17.2018.

EunHye Park, Dino Dvorak, and André A. Fenton. Ensemble Place Codes in Hippocampus: CA1, CA3, and Dentate Gyrus Place Cells Have Multiple Place Fields in Large Environments. *PLOS ONE*, 6(7):e22349, July 2011. doi: 10.1371/journal.pone.0022349.

Farhad Pashakhanloo and Alexei Koulakov. Stochastic Gradient Descent-Induced Drift of Representation in a Two-Layer Neural Network. In *Proceedings of the 40th International Conference on Machine Learning*, pages 27401–27419. PMLR, July 2023.

Angelique C. Paulk, Yoav Kfir, Arjun R. Khanna, Martina L. Mustroph, Eric M. Trautmann, Dan J. Soper, Sergey D. Stavisky, Marleen Welkenhuysen, Barundeb Dutta, Krishna V. Shenoy, Leigh R. Hochberg, R. Mark Richardson, Ziv M. Williams, and Sydney S. Cash. Large-scale neural recordings with single neuron resolution using Neuropixels probes in human cortex. *Nature Neuroscience*, 25(2):252–263, February 2022. doi: 10.1038/s41593-021-00997-0.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

János A Perge, Mark L Homer, Wasim Q Malik, Sydney Cash, Emad Eskandar, Gerhard Friehs, John P Donoghue, and Leigh R Hochberg. Intra-day signal instabilities affect decoding performance in an intracortical neural interface system. *Journal of Neural Engineering*, 10 (3):036004, April 2013. doi: 10.1088/1741-2560/10/3/036004.

Noah L. Pettit, Xintong C. Yuan, and Christopher D. Harvey. Hippocampal place codes are gated by behavioral engagement. *Nature Neuroscience*, 25(5):561–566, May 2022. doi: 10.1038/s41593-022-01050-4.

Mark H. Plitt and Lisa M. Giocomo. Experience-dependent contextual codes in the hippocampus. *Nature Neuroscience*, 24(5):705–714, May 2021. doi: 10.1038/s41593-021-00816-6.

Noeline W. Prins, Justin C. Sanchez, and Abhishek Prasad. Feedback for reinforcement learning based brain-machine interfaces using confidence metrics. *Journal of Neural Engineering*, 14(3):036016, June 2017. doi: 10.1088/1741-2552/aa6317.

Gregory J. Quirk, Robert U. Muller, and John L. Kubie. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 10(6):2008–2017, June 1990. doi: 10.1523/JNEUROSCI.10-06-02008.1990.

Rajkumar Vasudeva Raju, J. Swaroop Guntupalli, Guangyao Zhou, Carter Wendelken, Miguel Lázaro-Gredilla, and Dileep George. Space is a latent sequence: A theory of the hippocampus. *Science Advances*, 10(31):eadm8470, July 2024. doi: 10.1126/sciadv.adm8470.

Rabie A. Ramadan and Athanasios V. Vasilakos. Brain computer interface: control signals review. *Neurocomputing*, 223:26–44, February 2017. doi: 10.1016/j.neucom.2016.10.024.

Dhruva V. Raman and Timothy O'Leary. Optimal plasticity for memory maintenance during ongoing synaptic change. *eLife*, 10:e62912, September 2021. doi: 10.7554/eLife.62912.

Aviv Ratzon, Dori Derdikman, and Omri Barak. Representational drift as a result of implicit regularization. *eLife*, 12, May 2024. doi: 10.7554/eLife.90069.

Pascal Ravassard, Ashley Kees, Bernard Willers, David Ho, Daniel Aharoni, Jesse Cushman, Zahra M. Aghajan, and Mayank R. Mehta. Multisensory Control of Hippocampal Spatiotemporal Selectivity. *Science*, 340(6138):1342–1346, June 2013. doi: 10.1126/science.1232655.

Matthew Rosenberg, Tony Zhang, Pietro Perona, and Markus Meister. Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *eLife*, 10:e66175, July 2021. doi: 10.7554/eLife.66175.

Jeffrey N. Rouder, Paul L. Speckman, Dongchu Sun, Richard D. Morey, and Geoffrey Iverson. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2):225–237, April 2009. doi: 10.3758/PBR.16.2.225.

Michael E. Rule and Timothy O'Leary. Self-healing codes: How stable neural populations can track continually reconfiguring neural representations. *Proceedings of the National Academy of Sciences*, 119(7):e2106692119, February 2022. doi: 10.1073/pnas.2106692119.

Michael E. Rule, Timothy O'Leary, and Christopher D. Harvey. Causes and consequences of representational drift. *Current Opinion in Neurobiology*, 58:141–147, October 2019. doi: 10.1016/j.conb.2019.08.005.

Michael E. Rule, Adrianna R. Loback, Dhruva V. Raman, Laura N. Driscoll, Christopher D. Harvey, and Timothy O'Leary. Stable task information from an unstable neural population. *eLife*, 9:e51121, July 2020. doi: 10.7554/eLife.51121.

Sadra Sadeh and Claudia Clopath. Contribution of behavioural variability to representational drift. *eLife*, 11:e77907, August 2022. doi: 10.7554/eLife.77907.

Patrick T. Sadtler, Kristin M. Quick, Matthew D. Golub, Steven M. Chase, Stephen I. Ryu, Elizabeth C. Tyler-Kabara, Byron M. Yu, and Aaron P. Batista. Neural constraints on learning. *Nature*, 512(7515):423–426, August 2014. doi: 10.1038/nature13665.

Can Sarica, Christopher R. Conner, Kazuaki Yamamoto, Andrew Yang, Jürgen Germann, Melissa M. Lannon, Nardin Samuel, Michael Colditz, Brendan Santyr, Clement T. Chow, Christian Iorio-Morin, David H. Aguirre-Padilla, Stefan Thomas Lang, Artur Vetkas, Cletus Cheyuo, Aaron Loh, Ghazaleh Darmani, Oliver Flouty, Vanessa Milano, Michelle Paff, Mojgan Hodaie, Suneil K. Kalia, Renato P. Munhoz, Alfonso Fasano, and Andres M. Lozano.

Trends and disparities in deep brain stimulation utilization in the United States: a Nationwide Inpatient Sample analysis from 1993 to 2017. *Lancet Regional Health - Americas*, 26: 100599, September 2023. doi: 10.1016/j.lana.2023.100599.

Andrew M. Saxe, James L. McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv*, February 2014. doi: 10.48550/arXiv.1312.6120.

Gerwin Schalk, Jan Kubánek, Kai J. Miller, N. R. Anderson, Eric C. Leuthardt, Jeffrey G. Ojemann, Dave Limbrick, Daniel Moran, Lester A. Gerhardt, and Jonathan R. Wolpaw. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *Journal of Neural Engineering*, 4(3):264–275, June 2007. doi: 10.1088/1741-2560/4/3/012.

Carl E. Schoonover, Sarah N. Ohashi, Richard Axel, and Andrew J. P. Fink. Representational drift in primary olfactory cortex. *Nature*, 594(7864):541–546, June 2021. doi: 10.1038/s41586-021-03628-7.

Karen E. Schroeder, Sean M. Perkins, Qi Wang, and Mark M. Churchland. Cortical Control of Virtual Self-Motion Using Task-Specific Subspaces. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 42(2):220–239, January 2022. doi: 10.1523/JNEUROSCI.2687-20.2021.

Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

Martin E. Seligman. Learned helplessness. *Annual review of medicine*, 23:407—412, 1972. doi: 10.1146/annurev.me.23.020172.002203.

Maryam M. Shanechi, Amy L. Orsborn, Helene G. Moorman, Suraj Gowda, Siddharth Dangi, and Jose M. Carmena. Rapid control and feedback rates enhance neuroprosthetic control. *Nature Communications*, 8(1):13825, January 2017. doi: 10.1038/ncomms13825.

Matthew L. Shapiro, Heikki Tanila, and Howard Eichenbaum. Cues that hippocampal place cells encode: Dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, 7(6):624–642, 1997. doi: 10.1002/(SICI)1098-1063(1997)7:6<624::AID-HIPO5>3.0.CO;2-E.

Mark E. J. Sheffield and Daniel A. Dombeck. Dendritic mechanisms of hippocampal place field formation. *Current Opinion in Neurobiology*, 54:1–11, February 2019. doi: 10.1016/j.conb.2018.07.004.

Liron Sheintuch, Alon Rubin, Noa Brande-Eilat, Nitzan Geva, Noa Sadeh, Or Pinchasof, and Yaniv Ziv. Tracking the Same Neurons across Multiple Days in Ca2+ Imaging Data. *Cell Reports*, 21(4):1102–1115, October 2017. doi: 10.1016/j.celrep.2017.10.013.

Liron Sheintuch, Nitzan Geva, Hadas Baumer, Yoav Rechavi, Alon Rubin, and Yaniv Ziv. Multiple Maps of the Same Spatial Context Can Stably Coexist in the Mouse Hippocampus. *Current Biology*, 30(8):1467–1476.e6, April 2020. doi: 10.1016/j.cub.2020.02.018.

Wangbing Shen, Yu Tong, Feng Li, Yuan Yuan, Bernhard Hommel, Chang Liu, and Jing Luo. Tracking the neurodynamics of insight: A meta-analysis of neuroimaging studies. *Biological Psychology*, 138:189–198, October 2018. doi: 10.1016/j.biopsycho.2018.08.018.

Krishna V. Shenoy and Jose M. Carmena. Combining Decoder Design and Neural Adaptation in Brain-Machine Interfaces. *Neuron*, 84(4):665–680, November 2014. doi: 10.1016/j.neuron.2014.08.038.

Paul G. Shinkman, Charles J. Bruce, and Bryan E. Pfingst. Operant conditioning of single-unit response patterns in visual cortex. *Science (New York, N.Y.)*, 184(4142):1194–1196, June 1974. doi: 10.1126/science.184.4142.1194.

Erez Shor, Pedro Herrero-Vidal, Adam Dewan, Ilke Uguz, Vincenzo F. Curto, George G. Malliaras, Cristina Savin, Thomas Bozza, and Dmitry Rinberg. Sensitive and robust chemical detection using an olfactory brain-computer interface. *Biosensors and Bioelectronics*, 195:113664, January 2022. doi: 10.1016/j.bios.2021.113664.

William Skaggs, Bruce McNaughton, and Katalin Gothard. An Information-Theoretic Approach to Deciphering the Hippocampal Code. In *Advances in Neural Information Processing Systems*, volume 5. Morgan-Kaufmann, 1992.

Marc W. Slutzky, Luke R. Jordan, Todd Krieg, Ming Chen, David J. Mogul, and Lee E. Miller. Optimal Spacing of Surface Electrode Arrays for Brain Machine Interface Applications. *Journal of neural engineering*, 7(2):26004, April 2010. doi: 10.1088/1741-2560/7/2/026004.

Trygve Solstad, Edvard I. Moser, and Gaute T. Einevoll. From grid cells to place cells: A mathematical model. *Hippocampus*, 16(12):1026–1031, 2006. doi: 10.1002/hipo.20244.

Trygve Solstad, Charlotte N. Boccara, Emilio Kropff, May-Britt Moser, and Edvard I. Moser. Representation of Geometric Borders in the Entorhinal Cortex. *Science*, 322(5909):1865–1868, 2008.

Ethan Sorrell, Michael E. Rule, and Timothy O'Leary. Brain–Machine Interfaces: Closed-Loop Control in an Adaptive System. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):167–189, 2021. doi: 10.1146/annurev-control-061720-012348.

Ethan Sorrell, Daniel E. Wilson, Michael E. Rule, Helen Yang, Fulvio Forni, Christopher D. Harvey, and Timothy O'Leary. An optical brain-machine interface reveals a causal role of posterior parietal cortex in goal-directed navigation. *bioRxiv*, November 2024. doi: 10.1101/2024.11.29.626034.

Nicholas A. Steinmetz, Cagatay Aydin, Anna Lebedeva, Michael Okun, Marius Pachitariu, Marius Bauza, Maxime Beau, Jai Bhagat, Claudia Böhm, Martijn Broux, Susu Chen, Jennifer Colonell, Richard J. Gardner, Bill Karsh, Fabian Kloosterman, Dimitar Kostadinov, Carolina Mora-Lopez, John O'Callaghan, Junchol Park, Jan Putzeys, Britton Sauerbrei, Rik J. J. van Daal, Abraham Z. Vollan, Shiwei Wang, Marleen Welkenhuysen, Zhiwen Ye, Joshua T. Dudman, Barundeb Dutta, Adam W. Hantman, Kenneth D. Harris, Albert K. Lee, Edvard I. Moser, John O'Keefe, Alfonso Renart, Karel Svoboda, Michael Häusser, Sebastian Haesler, Matteo Carandini, and Timothy D. Harris. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, 372(6539):eabf4588, April 2021. doi: 10.1126/science.abf4588.

Alexis M. Stranahan, David Khalil, and Elizabeth Gould. Running induces widespread structural alterations in the hippocampus and entorhinal cortex. *Hippocampus*, 17(11):1017–1022, 2007. doi: 10.1002/hipo.20348.

Chen Sun, Wannan Yang, Jared Martin, and Susumu Tonegawa. Hippocampal neurons represent events as transferable units of experience. *Nature Neuroscience*, 23(5):651–663, May 2020. doi: 10.1038/s41593-020-0614-x.

Dechuan Sun, Forough Habibollahi Saatlou, Yang Yu, Ranjith Rajasekharan Unnithan, and Chris French. Utilising hippocampal neuronal calcium activity in mouse CA1 for a multimodal optical brain-computer interface. *bioRxiv*, April 2022. doi: 10.1101/2022.04.26.489497.

Weinan Sun, Johan Winnubst, Maanasa Natrajan, Chongxi Lai, Koichiro Kajikawa, Arco Bast, Michalis Michaelos, Rachel Gattoni, Carsen Stringer, Daniel Flickinger, James E. Fitzgerald, and Nelson Spruston. Learning produces an orthogonalized state machine in the hippocampus. *Nature*, pages 1–11, February 2025. doi: 10.1038/s41586-024-08548-w.

David Sussillo, Paul Nuyujukian, Joline M. Fan, Jonathan C. Kao, Sergey D. Stavisky, Stephen Ryu, and Krishna Shenoy. A recurrent neural network for closed-loop intracortical brain–machine interface decoders. *Journal of Neural Engineering*, 9(2):026027, March 2012. doi: 10.1088/1741-2560/9/2/026027.

David Sussillo, Sergey D. Stavisky, Jonathan C. Kao, Stephen I. Ryu, and Krishna V. Shenoy. Making brain–machine interfaces robust to future neural variability. *Nature Communications*, 7 (1):13749, December 2016. doi: 10.1038/ncomms13749.

Susumu Takahashi and Yoshio Sakurai. Information in small neuronal ensemble activity in the hippocampal CA1 during delayed non-matching to sample performance in rats. *BMC Neuroscience*, 10(1):115, September 2009. doi: 10.1186/1471-2202-10-115.

Heikki Tanila. Hippocampal place cells can develop distinct representations of two visually identical environments. *Hippocampus*, 9(3):235–246, 1999. doi: 10.1002/(SICI)1098-1063(1999)9:3<235::AID-HIPO4>3.0.CO;2-3.

Jeffrey S. Taube, Robert U. Muller, and James B. Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 10(2):420–435, February 1990.

Sarah A. Tennant, Lukas Fischer, Derek L. F. Garden, Klára Zsófia Gerlei, Cristina Martinez-Gonzalez, Christina McClure, Emma R. Wood, and Matthew F. Nolan. Stellate Cells in the Medial Entorhinal Cortex Are Required for Spatial Learning. *Cell Reports*, 22(5):1313–1324, January 2018. doi: 10.1016/j.celrep.2018.01.005.

Alejandro Terrazas, Michael Krause, Peter Lipa, Katalin M. Gothard, Carol A. Barnes, and Bruce L. McNaughton. Self-Motion and the Hippocampal Spatial Metric. *The Journal of Neuroscience*, 25(35):8085–8096, August 2005. doi: 10.1523/JNEUROSCI.0693-05.2005.

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

Anupratap Tomar and Thomas J. McHugh. The impact of stress on the hippocampal spatial code. *Trends in Neurosciences*, 45(2):120–132, February 2022. doi: 10.1016/j.tins.2021.11.005.

Eric M. Trautmann, Sergey D. Stavisky, Subhaneil Lahiri, Katherine C. Ames, Matthew T. Kaufman, Daniel J. O'Shea, Saurabh Vyas, Xulu Sun, Stephen I. Ryu, Surya Ganguli, and Krishna V. Shenoy. Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron*, 103(2):292–308.e4, July 2019. doi: 10.1016/j.neuron.2019.05.003.

Eric M. Trautmann, Daniel J. O'Shea, Xulu Sun, James H. Marshel, Ailey Crow, Brian Hsueh, Sam Vesuna, Lucas Cofer, Gergő Bohner, Will Allen, Isaac Kauvar, Sean Quirin, Matthew MacDougall, Yuzhi Chen, Matthew P. Whitmire, Charu Ramakrishnan, Maneesh Sahani, Eyal Seidemann, Stephen I. Ryu, Karl Deisseroth, and Krishna V. Shenoy. Dendritic calcium signals in rhesus macaque motor cortex drive an optical brain-computer interface. *Nature Communications*, 12(1):3689, June 2021. doi: 10.1038/s41467-021-23884-5.

Manuel Valero and Liset Menendez de la Prida. The hippocampus in depth: a sublayer-specific perspective of entorhinal–hippocampal function. *Current Opinion in Neurobiology*, 52:107–114, October 2018. doi: 10.1016/j.conb.2018.04.013.

Raphael Vallat. Pingouin: statistics in python. *Journal of Open Source Software*, 3(31):1026, November 2018. doi: 10.21105/joss.01026.

Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, June 2014. doi: 10.7717/peerj.453.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Lupeng Wang and Richard J. Krauzlis. Visual Selective Attention in Mice. *Current Biology*, 28 (5):676–685.e4, March 2018. doi: 10.1016/j.cub.2018.01.038.

Leonhard Waschke, Niels A. Kloosterman, Jonas Obleser, and Douglas D. Garrett. Behavior needs neural variability. *Neuron*, 109(5):751–766, March 2021. doi: 10.1016/j.neuron.2021.01.023.

Andrew J. Watrous, Darrin J. Lee, Ali Izadi, Gene G Gurkoff, Kiarash Shahlaie, and Arne D. Ekstrom. A comparative study of human and rat hippocampal low frequency oscillations during spatial navigation. *Hippocampus*, 23(8):656–661, August 2013. doi: 10.1002/hipo.22124.

James C. R. Whittington, Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E. J. Behrens. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, 183 (5):1249–1263.e23, November 2020. doi: 10.1016/j.cell.2020.10.024.

Francis R. Willett, Daniel R. Young, Brian A. Murphy, William D. Memberg, Christine H. Blabe, Chethan Pandarinath, Sergey D. Stavisky, Paymon Rezaii, Jad Saab, Benjamin L. Walter, Jennifer A. Sweet, Jonathan P. Miller, Jaimie M. Henderson, Krishna V. Shenoy,

John D. Simeral, Beata Jarosiewicz, Leigh R. Hochberg, Robert F. Kirsch, and A. Bolu Ajiboye. Principled BCI Decoder Design and Parameter Selection Using a Feedback Control Model. *Scientific Reports*, 9(1):8881, June 2019. doi: 10.1038/s41598-019-44166-7.

Francis R. Willett, Donald T. Avansino, Leigh R. Hochberg, Jaimie M. Henderson, and Krishna V. Shenoy. High-performance brain-to-text communication via handwriting. *Nature*, 593 (7858):249–254, May 2021. doi: 10.1038/s41586-021-03506-2.

Tom J. Wills, Colin Lever, Francesca Cacucci, Neil Burgess, and John O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science (New York, N.Y.)*, 308(5723):873–876, May 2005. doi: 10.1126/science.1108905.

Guy H. Wilson, Francis R. Willett, Elias A. Stein, Foram Kamdar, Donald T. Avansino, Leigh R. Hochberg, Krishna V. Shenoy, Shaul Druckmann, and Jaimie M. Henderson. Long-term unsupervised recalibration of cursor BCIs. *bioRxiv*, February 2023. doi: 10.1101/2023.02.03.527022.

Shawn S. Winter, Max L. Mehlman, Benjamin J. Clark, and Jeffrey S. Taube. Passive Transport Disrupts Grid Signals in the Parahippocampal Cortex. *Current Biology*, 25(19):2493–2502, October 2015. doi: 10.1016/j.cub.2015.08.034.

Jonathan R. Wolpaw. Brain–computer interfaces as new brain output pathways. *The Journal of Physiology*, 579(3):613–619, 2007. doi: 10.1113/jphysiol.2006.125948.

Jonathan R. Wolpaw and Dennis J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences*, 101(51):17849–17854, December 2004. doi: 10.1073/pnas.0403504101.

Ruth A. Wood, Marius Bauza, Julija Krupic, Stephen Burton, Andrea Delekate, Dennis Chan, and John O'Keefe. The honeycomb maze provides a novel test to study hippocampal-dependent spatial navigation. *Nature*, 554(7690):102–105, February 2018. doi: 10.1038/nature25433.

Wei Wu, Ammar Shaikhouni, John P. Donoghue, and Michael J. Black. Closed-loop neural control of cursor motion using a Kalman filter. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 4126–4129, September 2004. doi: 10.1109/IEMBS.2004.1404151.

Seung-Schik Yoo, Ty Fairneny, Nan-Kuei Chen, Seh-Eun Choo, Lawrence P. Panych, Hyun-Wook Park, Soo-Young Lee, and Ferenc A. Jolesz. Brain–computer interface using fMRI: spatial navigation by thoughts. *NeuroReport*, 15(10):1591, July 2004. doi: 10.1097/01.wnr.0000133296.39160.fe.

Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual Learning Through Synaptic Intelligence. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3987–3995. PMLR, July 2017.

Kechen Zhang, Iris Ginzburg, Bruce L. McNaughton, and Terrence J. Sejnowski. Interpreting Neuronal Population Activity by Reconstruction: Unified Framework With Application to Hippocampal Place Cells. *Journal of Neurophysiology*, 79(2):1017–1044, February 1998. doi: 10.1152/jn.1998.79.2.1017.

Yan Zhang, Márton Rózsa, Daniel Bushey, Jihong Zheng, Daniel Reep, Yajie Liang, Gerard Joey Broussard, Arthur Tsang, Getahun Tsegaye, Ronak Patel, Sujatha Narayan, Jing Xuan Lim, Rongwei Zhang, Misha B. Ahrens, Glenn C. Turner, Samuel S.-H. Wang, Karel Svoboda, Wyatt Korff, Eric R. Schreiter, Jeremy P. Hasseman, Ilya Kolb, and Loren L Looger. jGCaMP8 Fast Genetically Encoded Calcium Indicators. *Janelia Online Resources*, 2020. doi: 10.25378/JANELIA.13148243.

Zheyang Zheng, Roman Huszár, Thomas Hainmueller, Marlene Bartos, Alex H. Williams, and György Buzsáki. Perpetual step-like restructuring of hippocampal circuit dynamics. *Cell Reports*, 43(9), September 2024. doi: 10.1016/j.celrep.2024.114702.

Noam E. Ziv and Naama Brenner. Synaptic Tenacity or Lack Thereof: Spontaneous Remodeling of Synapses. *Trends in Neurosciences*, 41(2):89–99, February 2018. doi: 10.1016/j.tins.2017.12.003.

Yaniv Ziv, Laurie D. Burns, Eric D. Cocker, Elizabeth O. Hamel, Kunal K. Ghosh, Lacey J. Kitch, Abbas El Gamal, and Mark J. Schnitzer. Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience*, 16(3):264–266, March 2013. doi: 10.1038/nn.3329.

Weijian Zong, Horst A. Obenhaus, Emilie R. Skytøen, Hanna Eneqvist, Nienke L. de Jong, Ruben Vale, Marina R. Jorge, May-Britt Moser, and Edvard I. Moser. Large-scale two-photon calcium imaging in freely moving mice. *Cell*, 185(7):1240–1256.e30, March 2022. doi: 10.1016/j.cell.2022.02.017.

Ipshita Zutshi, Athina Apostolelli, Wannan Yang, Zheyang Sam Zheng, Tora Dohi, Edoardo Balzani, Alex H. Williams, Cristina Savin, and György Buzsáki. Hippocampal neuronal activity is aligned with action plans. *Nature*, pages 1–9, January 2025. doi: 10.1038/s41586-024-08397-7.