NEURAL CORRELATES OF PARALLEL AND DISTRIBUTED ENGRAMS

HAORAN CHANG
B.Sc. Neuroscience, McGill University, 2017

A thesis submitted
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

NEUROSCIENCE

Department of Neuroscience
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

NEURAL CORRELATES OF PARALLEL AND DISTRIBUTED ENGRAMS

HAORAN CHANG

Date of Defence: November 20, 2023

Bruce L. McNaughton	Professor	Ph.D.
Majid H. Mohajerani Thesis Co-Supervisors	Adjunct Associate Professor	Ph.D.
Masami Tatsuno Thesis Examination Committee Member	Professor	Ph.D.
David R. Euston Thesis Examination Committee Member	Associate Professor	Ph.D.
James Knierim External Examiner, Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University	Professor	Ph.D.
Ian Q. Whishaw Chair, Thesis Examination Committee	Professor	Ph.D.

Dedication

To my mother and father, who have always believed in me.

Abstract

From the waking hours into the depths of sleep, the hippocampus and the neocortex engage in an enigmatic dialogue on that which occurs, that which had occurred and that which could occur eventually. Together, this information weaves the memories of our past and the knowledge of our world. Yet, if one attempts to eavesdrop on this conversation, they would be perplexed to discover that the details that would give rise to such sophisticated structures could mostly be diluted into the mapping of space. What invisible bonds may tie space and memory in the brain, the present thesis offers no answers for. What it does offer, instead, are more enigmas to be fancied over and more confusions to be resolved: (1) The neural representation of space is found across multiple regions of the dorsal cortex and necessitates an intact hippocampus to form; (2) During offline periods, the retrosplenial cortex spontaneously reinstates patterns of activity specifically related to the locations of spatial landmarks; (3) The secondary motor cortex, in contrast, reinstates a conjunction of spatial and non-spatial information, in the forms of recent trajectories undertaken in an environment and the locations of visuo-tactile landmarks, respectively. Together, these results elaborate a spatial code that is heavily redundant and dispersed, with a link to mnemonic processing — a neural correlate of parallel and distributed engrams.

Contributions of Authors

All main chapters of this thesis have been published. Sufficient permissions have been obtained for their reproduction in this thesis. The authors and their contributions will be listed below.

Chapter 2

Published in Esteves et al. (2021).

Authors Ingrid M. Esteves, HaoRan Chang, Adam R. Neumann, JianJun Sun, Majid H. Mohajerani and Bruce L. McNaughton.

Author contributions I.M.E., A.R.N., B.L.M., and M.H.M., designed research; I.M.E., A.R.N., and J.S. performed research; I.M.E., H.C., and B.L.M. analyzed data; I.M.E., H.C., B.L.M. wrote the paper.

Chapter 3

Published in Chang et al. (2020).

Authors HaoRan Chang, Ingrid M. Esteves, Adam R. Neumann, JianJun Sun, Majid H. Mohajerani and Bruce L. McNaughton.

Author contributions HR.C., A.R.N. and I.M.E. conceptualized and developed the experiments. I.M.E. conducted the behavioural recordings. J.S. performed all animal surgeries. HR.C. conducted all data analyses. HR.C. and B.L.M. wrote the article, which all authors commented on and edited. B.L.M. and M.H.M. supervised the study.

Chapter 4

Published in Chang et al. (2023).

Authors HaoRan Chang, Ingrid M. Esteves, Adam R. Neumann, Majid H. Mohajerani and Bruce L. McNaughton.

Author contributions H.C., I.M.E., M.H.M. and B.L.M. designed the experiment. H.C., I.M.E. and A.R.N. conducted the experiments and collected the data. H.C. and I.M.E. performed the surgeries. H.C. analysed the data and wrote the manuscript, which all authors helped to revised.

Acknowledgments

Accumulating knowledge is a good thing. But it only takes a small excess of this consumption for the depraved lunatic to soon find themself drowning under the murky waters of conceptual dead-ends and intellectual stalemates. Every time I sank into this nihilistic abyss, my friends have always been there to drag me back to shore, to remind me that, however decadent the mind must become, a place shall always be reserved for Dionysus on our humble Apollonian stage. Without such good comrades, I would have succumbed to the sins of my own inadequacies and to the absurdity of my existence a long time ago. To them, I owe a great debt of gratitude. *Elnaz Alikarami, Ritwik Das, Aubrey M. Demchuk, Ingrid M. Esteves, Samsoon Inayat, Yagika Kaushik, Soroush Malek, Dun Mao, Milad Naghizadeh, Adam R. Neumann, Rui C. Pais, Rebeca León Parada, Surjeet Singh, David Pere Tomàs-Cuesta and Hyun Choong Yong*. To my supervisors. *Bruce L. McNaughton and Majid H. Mohajerani*. And to the person who gave me the opportunity to pursue my life's passion. *Mark P. Brandon*. Thank you for tolerating all my flaws. Thank you for discovering hope where all hope is lost.

P.S. Rui, my dear friend, I ask you: *How do two individuals, who have committed philosophical suicide, carry on the burden of existence?* Answer: *At dawn, they push the boulder uphill. At midday, they reach the summit of Tartarus. At sunset, they watch the boulder roll back down and, themselves, follow suit. And at dusk, they rejoice in the glory of their shared accomplishment.* In other words, I am glad that we are friends.

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

Simonides of Ceos, a renowned poet of Ancient Greece, was commissioned by the powerful aristocrat Scopas of Thessaly for a recital at his palace. In between his performance, the orator was summoned out to the gate to meet with two of his admirers. By a cruel twist of fate, the ceiling of the recital hall collapsed shortly after he had taken leave, entombing beneath the rubble the remains of his patrons, crushed beyond recognition. To the artist's surprise, he was able to accurately identify the disfigured corpses by name based on his recollections of the seating arrangements, and in doing so, rendered some degree of solace to those in mourning. It was then, under the macabre circumstances of his employment, that Simonides discovered a most powerful mnemonic device: that by committing items to memory in association with places, the orderly arrangement of the places serves to preserve the mental integrity of the items themselves. (Cicero, *De Oratore*, Book II, Lines 350-354)

Over two millennia have passed since Cicero had conceived his fictional tale on the origins of the *method of loci* (Yates, 1966, p. 1-2,17). Yet, this ancient mnemonic technique still remains a highly popular and effective tool for memorisation. Each year, hundreds of "mind atheletes" convene at the *World Memory Championship* to compete for the greatest number of arbitrary items one is able to commit to memory (He, 2023). In the majority of cases, the system used to confer on the participants seemingly superhuman mnemonic capabilities is the *method of loci* (Maguire et al., 2003). To practise the technique, the memoriser constructs in their mind a *memory palace* that is modelled after either a real or a fictional architecture (O'Keefe and Nadel, 1978, p. 389-390). Each item that is to be encoded is then placed at discrete locations within this palace. As the orator mentally traverses the path linking the objects in their memory, all items are thence retrieved following the sequence of the journey.

The effectiveness of the method beckons us to wonder what common grounds may *space* and

memory occupy in the mind. Yet, a reasoning was not to come as easily as through an exercise of the intuition[1]. Nor would the search for this link be confined solely to the philosophies, as a series of neuroscientific discoveries in the 1950-70s brought this enigma to the centre of investigations on the brain's ability to remember. Namely, researchers had uncovered space and memory as two distinct functions attributable to a single structure, the *hippocampus*. From this point on, unravelling the intimate connections between these seemingly disparate axes of our cognition holds great promise towards a scientific understanding of memory, and indeed makes for the object of the present thesis.

In this introductory chapter, I will summarise some key advances from parallel investigations on *learning and memory* and *spatial navigation*, and in doing so, lay a unified foundation for a discourse on space and memory. I will then outline the three studies I have conducted in my doctoral training in an earnest attempt to contribute to the efforts in reconciling the two sides of the mnemonic coin.

1.1 Neuroscience of learning and memory

> The new psychology had eliminated the soul; Dewey and Münsterberg eliminated the willing self, but a self, and ego, remained. What remained for psychological study were consciousness and behavior. Soon, the existence of consciousness, too, would become problematic. (Leahey, 1997, p. 295)

The objective study of *memory* emerged in concert with the Behaviourist school of psychology, which sought to apply scientific principles to the understanding of the mind. Such a goal was believed attainable by adopting of a reductionist view of the psyche, whereby complex behaviours can be decomposed into elementary *reflexes* (Pavlov, 1927, Lecture I; Chirimuuta, 2021). By associating (i.e., *conditioning*) these inherent programmes with arbitrary stimuli found in the surroundings, the agent *learns* to generate the appropriate responses given their external circumstances. The memory component of the nervous system can, therefore, be understood and measured as the adaption of behaviours.

Lending support to this notion were the works of Sir Charles Sherrington, who demonstrated that certain reflexes are entirely supported by circuits of the spinal cord. In particular, nervous pathways organise into *sensory-response reflex arcs*, whose hardwired nature is manifested in the stereotyped contractions evoked by sensory afferences (Sherrington, 1906, Chap. I, III). As we ascend the "nervous echelon", more sophisticated, albeit still predictable, reflexes are provoked by

[1] See Aristotle, *De Memoria et Reminiscentia*, Part II, 452a5-30, for a baffling discourse.

electrical stimulation of the motor cortex (Hughlings Jackson, 1873; Ferrier, 1875). In addition to movement routines, discrete functions, including visual, auditory, somatosensory and speech, occupied well-defined locations over the cerebral mantle (cf. Catani et al., 2012, for review). A problem arose, however, in the form of "uncharted" territories that filled the space between otherwise localised functional centres. Stimulation of these areas did not illuminate their functions, but their damage led to appreciable symptoms (e.g., apraxia, agnosia, ataxia, dyslexia) that could *prima facie* be attributed to a loss of memory (cf. Flaherty et al., 1977, Chap. 15; Botez, 1987, Part 1, 4, 5; Pavlov, 1927). Ambitiously, theorists of the time posited that this interstitial tissue had the role of integrating inputs and linking them to outputs, in a manner that permits complex behaviours to be acquired through the process *association*[2] (Pavlov, 1927). The seat of memory, therefore, would reside in these *associational cortices*, while memories themselves would be functionally localised by the role(s) of the neighbouring primary areas that they bridge.

There to put the theory to the test was Karl Lashley, who, over the course of 30+ years, systematically surveyed of the cortex for the locus of memory (which he referred to as the *engram*) (reviewed in Lashley, 1950). Following the behaviourist tradition, Lashley and his students trained rats to solve mazes, in which, successful navigation from entry to exit required memorisation of a sequence of turns. Retention of the memory could then be assessed by counting the number of errors committed after specific lesions to the cerebral cortex. Such a task was deemed particularly suitable for probing *associative memory*, as the solution presumably required the integration of various sensory modalities. In one experiment, several incisions were made over the cortex to transect the fibres connecting various functional areas, hence disrupting the putative reflex arcs. Surprisingly, the rats did not suffer any measurable loss of memory following this procedure. Deficits in retention were, however, obtained by lesions to surface areas of the cortex, but were not manifested abruptly. Instead, performance deteriorated progressively as a function of the percentage of area ablated, while this trend was precipitously aggravated by the increasing complexity of the maze (Fig. 1.1).

Crucially, the degree of impairments was not explained by the locus of damage, but seemed to depend solely on the extent of the lesions; in rats blinded from birth, damage to the visual cortex

[2]Pavlov believed that two stimuli, received simultaneously or in close succession, excite two discrete "points" over the cortex (1927, Lecture XXII, p. 384-385). A connection is, therefore, formed between these two centres, and this connection is strengthened with repeated pairings of the same two stimuli — *reinforcement learning*. This hypothesis clearly resonates with the concept of Hebbian plasticity, and indeed would explain what was meant by Hebb, when he stated that "[t]he general idea is an old one, that [neurons that fire together wire together]" (1949, p. 70).

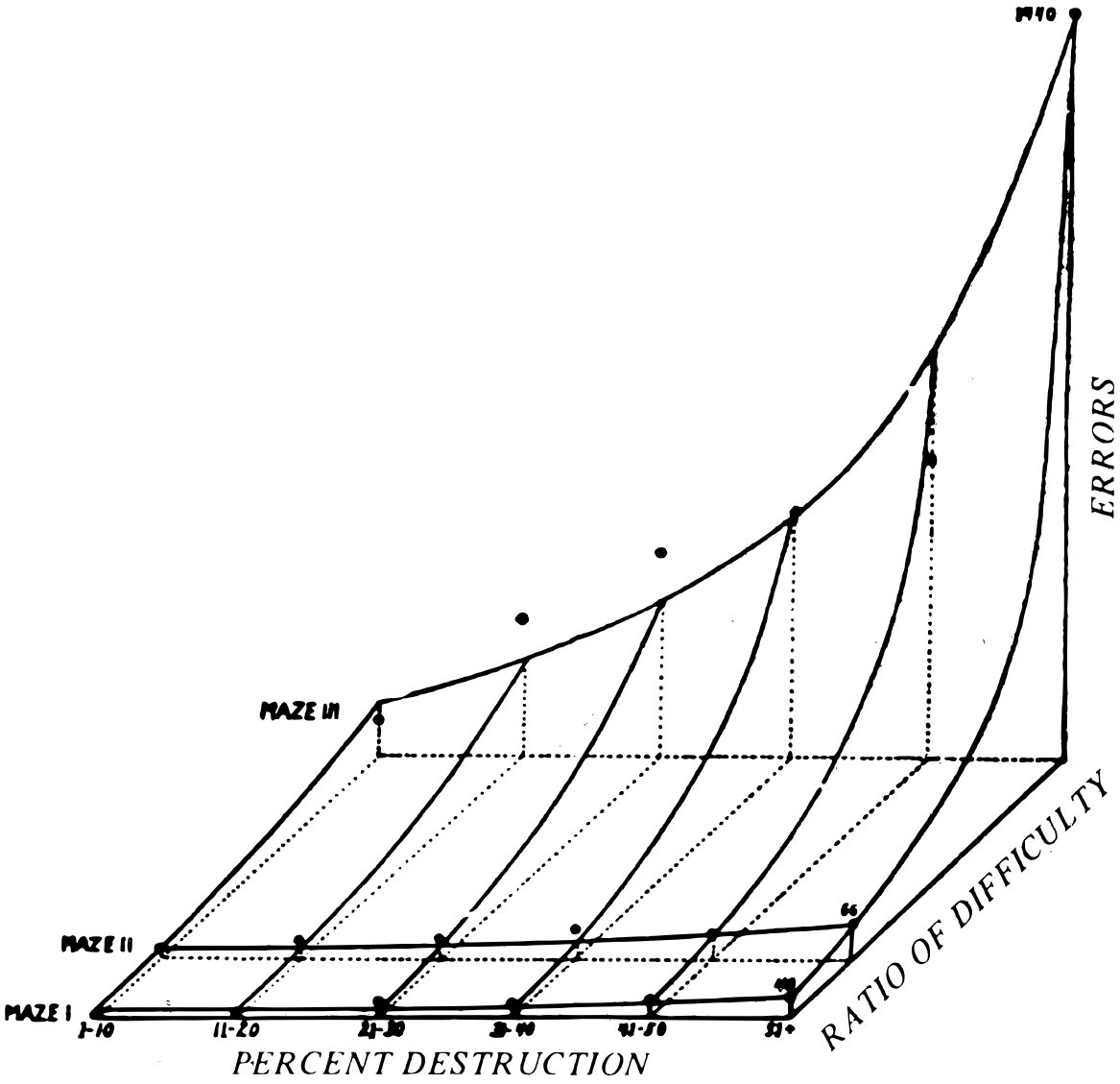

Figure 1.1: The number of errors committed in maze task (z-axis) as a function of the extent of cortical lesions (x-axis) and of the difficulty of the maze (y-axis). Reproduced from Lashley (1929).

equally produced failures in retention. More astonishingly, these maze habits were even preserved following extensive lesions to the cerebellum and spinal tracts, which otherwise resulted in severe motor impairments. Despite rolling and tumbling their way across the maze, the animals made few errors at critical junctions. To explain these extraordinary findings, Lashley formulated two seminal hypotheses on the structure of memory. First, the neural substrate for memory appears to be largely diffuse as opposed to localised, the redundancy of which imparts memory traces a great level of resilience towards injury — a principle he referred as *equipotentiality*. Second, the loss of learned habits is gradual and proportional to the extent of cortical lesions. This, coupled with the flexibility and adaptability of behaviours in solving a familiar task following loss of function, underscores a cooperative engagement of these distributed memory traces during retrieval — principle of *mass action*[3].

It can be argued that Karl Lashley's reputation was superseded by that of his own prodigious student, Donald Hebb, a fact that does little in way of justifying my dedication of a large portion of text, here, to his research. Yet, in my own search for a comprehensive interpretation of my results, I have found little recourse outside of Lashley's two principles. I will elaborate over this issue in the General Discussion. For now, the search for the engram must continue.

1.1.1 The eternal search for the engram

> I sometimes feel, in reviewing the evidence on the localization of the memory trace, that the necessary conclusion is that learning just is not possible. [...] Nevertheless, in spite of such evidence against it, learning does sometimes occur. (A frustrated Lashley, 1950, p. 477-478)

In his failure to isolate a single *locus* per learned habit in the cortex[4], Lashley discovered the parallel and distributed nature of memories. His dedication to the cortex, as the seat of memory, is a reflection of the dominant views of his time, a dedication that the modern scholar may easily label

[3]The concepts of *equipotentiality* and *mass action* trace their origins back to the works of Pierre Flourens, who is the first physiologist to use experimentation to mount an opposition against the theory of *cerebral localisation* (see Tizard, 1959, for review). Flourens noted that behaviour can sustain damage to large areas of the cortex before deficits were expressed, while these deficits mounted gradually as more of the cortex was ablated. His studies, conducted a century before Lashley's, echoed the same messages, which, despite their appreciated importance by the modern scholar, scarcely attracted popularity during the pioneering days of neurophysiology (cf. the Goltz-Ferrier debate, reviewed in Phillips et al., 1984).

[4]Yet, the question remains still as to what neurobiological substrate is responsible for persistent changes in the brain. The search for the *localised* engram, i.e. the elementary unit of storage, was to be succeeded by molecular biologists and cellular electrophysiologists, with discoveries such as *long-term potentiation* and *depression*, *spike-timing-dependent plasticity*, NMDA receptors, per-/post-synaptic signalling cascades (cf. Feldman, 2012, for review)

as misplaced. Because, soon after the publication of Lashley's 1950 *magnum opus*, a functionally differentiated region for memory was indeed to be identified by serendipity. Only, this region resides not in the cortex, as behaviourists and neurologists had initially envisaged.

For decades, Wilder Penfield practised neurosurgery to treat patients suffering from focal epilepsy (see Winter, 2012, for review). In many such cases, the period leading to seizures was marked by an *aura*[5], the contents of which was highly indicative of the focus of epileptogenesis. For instance, an aura characterised by flashing or moving patterns of lights would warn of imminent seizures localised in the occipital lobe visual areas. This distinguishing property of the disorder would be leveraged fully by Penfield to devise a procedure for demarcating the afflicted regions. In particular, by applying a small electrical charge at discrete points over the brain, the neurosurgeon sought to re-elicit the aura reported by individual patients prior to seizure onset (Penfield, 1954). Upon "re-awakening" this aura, the electrode tip would pinpoint the tissue that was to be surgically excised, in order to relieve the patients from chronic seizures.

While focal epilepsy can affect many areas of the brain, the temporal lobe was by far the most commonly afflicted region (Penfield, 1954). Unlike seizures focused elsewhere, for which the auras were predominantly linked to sensori-motor functions, the auras associated with temporal lobe epilepsy displayed markedly richer contents, which Penfield described as *psychical phenomena* (Penfield, 1954, 1955). Of particular interest were the reported "dreamy states", in which the patients sometimes experienced events and people that they had encountered in the past, accompanied by a feeling of familiarity or déjà-vu. When the team investigated the affected areas by electrophysiological means, they noticed that, occasionally, stimulating the temporal lobe would induce the re-experiencing of a distant episode, one that the patients frequently reported as forgotten or inaccessible otherwise, if not evoked artificially. Contrasted to voluntary recall, these experiences portrayed themselves more vividly and realistically[6] — as through *consciousness was doubled* — despite the patients being able to clearly identify them as past happenings. This memory record would unravel in a forward motion for as long as the stimulation was held. In some cases, repositioning the electrode back to the same location and stimulating would cause the same memory to be

[5]Seizures confined to a small region that can serve as a precursor to larger seizures spreading to other areas (Lowenstein, 2018). During auras, patients retain consciousness.

[6]Interestingly, only auditory and visual senses, and occasionally the **sense of position**, composed these experiences. Somatosensation, gustation and olfaction were missing from these pictures (Penfield, 1955, p. 65).

reenacted[7]. Enthusiastically, Penfield posited that somewhere in the brain (likely in the brainstem, he reasoned) exists a mechanism that keeps a permanent "ganglionic record" of all the conscious experiences in a person's lifetime. By stimulating the epileptogenic tissue, for which years of seizures had strengthened its connections with the fibres leading into this record, he was able to "tap into" a fragment of this permanent memory store.

Despite these strongly suggestive observations, Penfield did not consider, at this time, the possibility that he had located Lashley's engram (he believed that temporal lobe held at most a duplicate of the record, but not the record itself) (Penfield, 1955). Nor did he have any reason to believe memory to be localised there, as only mild deficits were ever produced by unilateral excision of medial temporal structures, as substantiated in over ninety epileptic patients he had treated (Penfield and Milner, 1958). That is, all except two patients, who suffered profound generalised amnesia post-operatively, cases that Penfield would document with the aid of a then young PhD student, Brenda Milner. In one patient, the amnesia had retroactively erased four years of memories leading to the surgery. In the other patient, three months had been lost. These patients subsequently became severely impaired in their ability to remember new events, information and people they encountered from day-to-day, while experiencing severe difficulties in navigating familiar surroundings.

These drastic side effects were presented at a conference, and immediately drew the attention of William Scoville (see Squire, 2009, for historical review). A neurosurgeon himself, Scoville had eight patients also suffering from similar amnesic symptoms, following partial bilateral lobectomies in the medial temporal lobe[8], encompassing the uncus, the amygdala, the hippocampus and the parahippocampal gyrus (i.e., the perirhinal, entorhinal and parahippocampal cortices)[9]. In Penfield's two amnesic patients, there was strong evidence as well to suspect that the corresponding structures on the contralateral side were also damaged by seizures, rendering the unilateral treat-

[7]While some writers have interpreted this observation as indicative of a form of localisation of memories/engrams, Penfield's original report provides a more diligent account over the issue: "It was obvious that years of epileptic discharge arising in this zone had sensitized the temporal cortex so that stimulation could produce psychical responses. Stimulations elsewhere in the lobe were without positive effect even when a small increase was made in voltage" (Penfield, 1955, p. 56). It is therefore possible that the recurring epileptic discharges had overtime "hijacked" portions of memories such that evoking the aura would consistently replay these same memories — a grand-scale pathological engram, cell assembly or attractor if you will (cf. Bower et al., 2015).

[8]Out of these patients, only H.M. suffered from epilepsy; the others were treated for schizophrenia and psychosis (see Mauguiere and Corkin, 2015, for review). H.M., being the only "unbiased sample" in this group, became a standard subject for case studies on medial temporal lobe amnesia.

[9]Out of these regions, the hippocampus and the entorhinal, perirhinal and parahippocampal cortices came to be the structures most implicated in declarative memory, based on cross-validation with the pathoanatomy of other patients and the known functions of the other damaged regions (Squire and Zola-Morgan, 1991).

ment into a bilateral lesion in effect (Penfield and Milner, 1958). On account of the similarities between the cases, Milner was invited to study Scoville's patients. There, she noticed that, not unlike Penfield's patients, these subjects endured severe retrograde amnesia, dating as far back as three years of the premorbid period in the most severe case (Scoville and Milner, 1957). Their abilities to learn new information and to keep any record of their daily happenings were virtually non-existent at the extreme, while their conversations mostly surrounded earlier life experiences, which they could seemingly recount with a great degree of detail. Moreover, their attention was unaffected, but diverting this attention away, even for a brief moment, would cause them to forget the object of the task or conversation that they were carrying out. Importantly, the severity of anterograde amnesia roughly corresponded with the posterior extent of the medial temporal lesions.

Based on these patterns of deficits, Milner drew an important set of conclusions that, still to this day, forms the basis of the most prominent theories on memory (1962, p. 270-271). First, she noted that extensive bilateral damage of the medial temporal lobe disproportionately affected recent memories, while more remote memories were seemingly spared (i.e., there is a temporal gradient to retrograde amnesia)[10]. This fact reinforces the notion that persistent engrams are situated in the cortex, as Lashley[11] and early investigators envisaged, and that the hippocampus and parahippocampal gyrus are not the permanent seats of memories. Second, for as long as attention is not diverted, patients are able to retain a great number of new facts (e.g., holding a string of numbers or word associations). This suggests that attention and the neural processes necessary to temporarily keep arbitrary associations[12] are independent of the medial temporal lobe structures as well. Third, taking these two points in consideration together, the arbitrary facts that are held temporarily for some time, whereas they sometimes give way to stable changes and retention, never reach the state of permanence in patients with bilateral hippocampal damage. The role of the hippocampus and peripheral

[10]The existence of a retrograde gradient in amnesia is heavily contested, as results vary inconsistently from patients to patients and from the different methodologies employed for psychological assessments. In general, it appears that semantic memories undergo consolidation, while episodic memories may or may not ever become independent of the hippocampus (see Nadel et al., 2007, for review).

[11]Peter Milner, in examining Lashley's work decades later, proposed an interesting hypothesis: the limitations of Lashley's ablation methods may have caused non-specific damages to the hippocampus (1999). This possibility would certainly explain the lost of spatial memories following these lesions, as shown decades later under the "Morris Water Maze" (Morris et al., 1982; Sutherland et al., 1982).

[12]Known as *working memory* in current jargon, which is believed to be supported by Lorente de Nó's *reverberatory circuits* or Hebb's *cell assemblies* (cf. Durstewitz et al., 2000). This view engenders an interesting hypothesis: though the hippocampus, by virtue of its expansive connectivity with distributed regions of the cortex, may be necessary to the permanent recording of arbitrary associations, cortical reverberations could act as a temporary holding dock for concurrent stimuli, despite these inputs having no prior associations (cf. Hebb, 1949, Chap. IV).

structures is, therefore, to mediate the *consolidation* of new memories into stable traces.

Another important conclusion that Milner derived is that there are more than one type of memory (Milner, 1962). In particular, tasking one patient (H.M.) to trace over a star shape, while only having visibility of the pattern and of his drawing-hand via a mirror, revealed that novel visuo-motor skills could be acquired steadily and persistently. This patient, however, would have no recollection of any prior training and would engage in the task, each day, believing that it was his first exposure. Later studies showed that a myriad of other memory types can still be acquired following bilateral medial temporal damage (Brooks and Baddeley, 1976; Cohen and Squire, 1980); it would appear that damaging these structures specifically impaired the acquisition of *facts* and *autobiographical events*, which can be consciously recalled and verbally declared by the individual (see Squire, 1992, for review). A distinction was, therefore, made between *declarative (explicit) memories*, which are dependent on the medial temporal lobe, and *non-declarative (implicit) memories*[13], which are spared following bilateral damage.

Under the umbrella of declarative memory, a further distinction was proposed by Tulving, that is *episodic versus semantic memories* (Tulving et al., 1972). Interestingly, the motivation behind this proposal stemmed from the works of a group of psychologists who counted amongst the early pioneers of modern *Artificial Intelligence*. In particular, Tulving accredits Ross Quillian, the person who coined the term semantic memory, as well as David Rumelhart, Walter Kintsch and Don Norman, who were contemporaries of certain key pioneers of modern Artificial Intelligence, such as Geoffrey Hinton, Ronald Williams and Terrence Sejnowski (cf. Rumelhart et al., 1986). It was, therefore, by no accident that their views on memory was that of a data structure, which comprise a definite organisation and computational goal[14](I will elaborate on these details later). On the one hand, episodic memory is organised by the temporal-spatial relationships that exist between the different events in an agent's personal experiences (i.e., *what* happened *when* and *where*, and how does

[13]Procedural/habit learning is primarily supported by the basal ganglia (see Packard and Knowlton, 2002, for review). It is the "switch-board" type of memory that behaviourists had initially posited to be supported by the associational cortices. Lashley's maze tasks are, however, hippocampal-dependent as they constitute spatial memories, despite him calling them "maze habits".

[14]One way to appreciate the idea of semantic memory is by analogy of a Turing Machine. In this machine, a single stripe of tape holds symbols drawn from some alphabet. When the machine reads a symbol, depending on the identity of the symbol and the "current state" of the machine, it would either write over this data point, move on to a different symbol or change its state. It follows, then, that in such a machine, the data (e.g., 1, 2, 3, 69, apple, bee's) and the operators (e.g., $+$, $-$, $\times$, $\nleq$, eat) share the same lexicon and occupy the same stripe of tape. Data that are stored and operated on also serve to instruct how other inputs or stored information are to be processed. In other words, stored knowledge, in itself, forms the unit of computation (cf. Hofstadter, 1979, Chap. X, XVII).

this event relate to others in these regards?). On the other hand, semantic memory is the knowledge of facts, meanings, algorithms and the rules of the relationships that bind them. This memory is devoid of an autobiographical component, but trading its place are the abilities for *inferential reasoning*, *logical deduction* and *generalisation*. In particular, for a new item to enter the semantic memory base, this input must be compatible with an existing *cognitive structure*. That is to say, new information is always referenced with established priors — *cognitive referents* — such that, the act of assimilating a *percept* situates that percept within the larger structure of knowledge and of relations.

Tulving's crucial prediction is that these two memory systems could act independently from each other. From the outset, empirical support for such a distinction did not seem forthcoming, as it was clear that the amnesia afflicting Penfield's and Scoville's patients impaired new acquisitions of both facts and personal experiences. Early studies that directly assessed this independence were not encouraging either; when Gabrieli et al. attempted to teach patient H.M. new semantic information, in the form of eight infrequent words, the subject was unable to retain any new definitions, let alone to substitute for their synonyms or to place them into a sentence (1988). This dissuasion notwithstanding, a landmark case study that appeared decades after Tulving's initial proposal did provide seemingly unequivocal support for an intact semantic learning system in the absence of episodic encoding. In particular, Vargha-Khadem et al. documented three patients who, despite enduring bilateral hippocampal damage at a young age (one at birth, one at 4 years of age and one at 9), attended "mainstream" schools and attained scores, in assessments of academic competencies, that were within normal margins (1997). In contrast, their ability to record episodic events was grossly impaired. In H.M.'s case, outside of the laboratory setting, the patient did seem to acquire a few vocabularies that were newly introduced into the English dictionary and identify certain prominent figures who rose to fame during the patient's postmorbid period (Gabrieli et al., 1988). Tulving, himself, had previously found that, by employing *errorless* training protocols, profoundly amnesic patients were able to acquire a considerable number of new factual knowledge, albeit at a much slower rate than normal subjects (Glisky et al., 1986b,a; Tulving et al., 1991). Specifically, the acquisition phase of these tasks was designed in such a way as to coax the subjects into always producing the correct responses to prompts, hence reducing *interference* and the compounding of errors during learning.

The interference, in question, is presumed to arise as a result of old cognitive priors forming competing hypotheses against new concepts (or new concepts interfering amongst themselves)[15] (Tulving et al., 1991). Eliminating this interference during learning, therefore, would allow new items to be integrated into the existing knowledge base, in the absence of the disambiguation normally provided by the *when* and the *where* of the episodic memory system. Further support to this hypothesis was lent by a study that examined amnesics' ability to acquire new semantic representations under a *collaborative learning* setting (Duff et al., 2006). The task involved pairing hippocampal patients with a close friend, relative or spouse, where the patient would instruct their partner to arrange a set of abstract pictogrammes in a given order. The key feature of this test is that participants had command over their own interpretations of the abstract geometries, which they would cross-validate on-the-fly with their partner. Hence, a *common ground* would be established out of the patients' existing semantic priors, which their partners are presumed to have a decent access to, owing to their close relationship. Under these conditions, the subjects gradually shortened their initially elaborate descriptions of individual pictogrammes, to few-word labels, at a similar rate as healthy controls, while these acquired short labels persisted across sessions and days. Taken together, these results are commensurate with the notions that, in the case of developmental amnesia, a semantic memory systems can operate independently of an episodic system to establish a semantic/cognitive scaffold, while similarly, in the case of adult-onset amnesia, the existing cognitive structure can accommodate new semantic items when cognitive interference is minimised (see Duff et al., 2020, for review).

A crucial weakness of errorless learning is that such a paradigm may conduce to a lack of generalisability in the learned items. In fact, a core appeal of the semantic memory idea is that such memory can encompass a wide range of cognitive objects that are all interrelated by degrees of similarity, disparity or independence. Within this dense relational structure, the ability to negotiate interference must, therefore, rest at the core of the semantic learning system. In contrast, if learning is only possible by minimising interference and maximising conformity with existing semantic priors, then did new semantic learning actually occur? Under errorless learning paradigms, it has been reported that the new information acquired by amnesics was *hyperspecific* (Glisky et al., 1986b,a;

[15]For instance, when encountering the word 'punctual' for the first time, one might initially find it to resemble the already known word 'puncture' and, consequently, confuse or conflate their definitions.

Bayley and Squire, 2002). In particular, substituting a retrieval cue with a synonym frequently caused failure of retrieval (e.g., "Venom caused seizure" versus "Venom induced ???"), while the subjects were heavily reliant on specific cues, such as the first letter of the target word. These problems are further compounded by the fact that a patient was unable to confidently rate the accuracy of their own answers, which raises doubts about whether the learning demonstrated by amnesics indeed reflects declarative memory, rather than a form of "verbalisable" non-declarative memory[16] (Bayley and Squire, 2002). To address this issue, Stark et al. (2005) modified the errorless learning protocol, where synonyms were substituted in the retrieval cues across learning trials. Surprisingly, this procedure not only rescued the generalisability of the learned words, but also tended to improve retention, a result that seemingly contradicts with the notion of interference. Despite this marked amelioration, the subject was still none the wiser when it came to rating the accuracy of their own answers, which puts into question the usefulness of such learning in guiding reasoning and deductions.

The culmination of such evidence speaks unfavourably of the position of semantic and episodic memories as two separate learning systems capable of acting independently. In fact, if some new semantic learning does occur under episodic amnesia, such learning, as well as the items that are learned, certainly do not appear normal. To Squire and Zola (1998), these results clarify episodic and semantic memories as equally dependent upon medial temporal lobe structures, while episodic memories are only distinguished by an added spatial and temporal context[17]. Though interpretations differ over the exact relationships shared by these two memories, a rational consensus would state that these systems act in an interdependent manner (cf. Ranganath, 2022, for review). This holistic view notwithstanding, the two terminologies remain firmly established in current literature. The reason for it likely recaptures the same appeal that propelled Tulving to initially propose this distinction, that it is a useful concept to form the basis of theoretical proposals. Indeed the existence of two parallel memory systems finds itself especially useful when considering what cog-

[16]Generalisation is not a feature exclusive to semantic memories; certain non-declarative memories also have the ability to generalise (cf. Zeithamova and Bowman, 2020; Poldrack and Foerde, 2008, for review). A popular anecdote speaks of an amnesic with Korsakoff's syndrome. One day, her attending physician decided to offer her a handshake with a pin camouflaged between his fingers. After falling victim to this prank, the patient would refuse all subsequent offers of a handshake from this physician, though she had no recollections of the specific incident that instilled this fear reflex into her. When repeatedly asked about her reasoning, she would sometimes respond: "There are sometimes pings hidden inside hands" (Claparède, 1911, p. 85, translated by this author).

[17]The authors postulated that these spatio-temporal components are provided by frontal lobe structures. Other authors attribute this role mainly to the hippocampus (e.g. O'Keefe and Nadel, 1978; Mishkin et al., 1997).

nitive, neuroanatomical, physiological or functional processes must all take part in supporting the *consolidation* of memories.

1.1.2 Systems consolidation

> Memory is assumed to be distributed and active; there is no homunculus to compare the features of the input with static memory lists. The term *knowledge* is used to encompass innate and developed capacities (neural connections) in addition to *memory*. *Learning* refers to any change in knowledge. According to the [connectionist] model, you do not *have* a store of knowledge; you *are* your knowledge, among other things. (Feldman, 1981, p. 50-51)

In the previous section, we explored how the clinical description of medial temporal lobe amnesia gave rise to the distinction between *explicit* and *implicit* memory systems. In addition, the temporal gradient of memories observed in amnesics cemented the idea of a gradual *consolidation* process taking place in the background with each episodic experience, for which the hippocampus makes a time-limited contribution. Furthermore, the early works performed by Lashley and others on the role of the cortex in memory, in combination with Tulving's putative *semantic* and *episodic* memory systems, suggest that different memory systems may occupy different brain structures, and therefore, be dissociated neuroanatonomically. The links across all these concepts are apparent, yet clearly stating how they combine into a coherent narrative is no mundane task. For this task, we refer to computational models in order to cast these constituent pieces of memory under a holistic view, starting with a consideration of the hippocampo-neocortical memory systems.

Amongst the first researchers to consider the possible synergies between the hippocampus and the neocortex in *learning* was David Marr. A mathematician by training, Marr formulated a series of neural network models, based on the hitherto known anatomy and physiology of these structures. With regards to the cortex, Marr noted that, despite the diverse functions supported by it, the circuit architecture that comprises this mantle is remarkably homogeneous throughout (1970). This implies that certain elementary modes of operation govern its overall information processing. If we are able to uncover a "fundamental hypothesis" from the design of these micro-circuits, then this hypothesis could help define the core working principles and computational objectives of the cortex.

To elucidate this hypothesis, Marr first drew attention to the *redundancy* inherent to the natural world and, by extension, the input patterns that the brain must receive. This point was exemplified by the visual world, which is primarily composed of oriented edges, demarcating continuous regions

in space. There to capture these regularities, at first, are the 'simple cells' and 'complex cells' of the primary visual cortex, which, themselves retinotopically organised, function as *feature detectors* to extract details that tend to be universally relevant under most viewing conditions (cf. Hubel and Wiesel, 1962). Later stages of visual processing are also concerned with redundancy, in that features that are commonly shared by objects of the same category tend to arise in concert. From there, the process of identifying (i.e., to call from memory) the items of a certain visual scene is the same process that reduces redundancy into categories. Therefore, from a visual processing standpoint, an argument can be made that the role of the cortex is to learn regular features in the world and, through this, formulate distinct categories or concepts.

Marr postulated that the laminar organisation of the cortex encapsulates the necessary micro-circuits to acquire and perform this classification task. Put simply, a three-layered *feed-forward* network was proposed (cf. Willshaw et al., 1997; Fig. 1.2**a-b**). The first layer contains *input neurons* that carry sensory afferent signals. These input neurons then project to a layer of *codon* cells[18], which are analogous to the aforementioned simple and complex cells, in that sensory inputs undergo *re-coding* to allow the first-order of regularities — features — to be extracted. In turn, codon cells project onto an *output layer*, which was believed to be comprised of pyramidal cells, each supporting an individuated category (a grandmother-esque layer of neurons, so to speak). The connections across all three layers are modifiable (a Hebbian rule was suggested) to permit the learning of new categories/redundancies and the refinement of existing concepts. In particular, a codon cell that consistently and repeatedly activates in conjunction with a pyramidal cell has their shared synapse strengthened (cf. Hebb, 1949). Such *reinforcements* would, in return, be provided by the redundancies found in the environment, such that the network would learn by itself over time to discover regular feature sets that discern separate categories — *self-organisation*. The end result is that the strength of the connection between a codon cell and an output cell is proportional to the probability that the feature supported by the codon cell belongs to the category represented by the output cell. The pyramidal cell that is most strongly activated by the collection of codon cells that are currently

[18]Marr posited that the Martinotti cells in the deep layers of the cortex could fulfil the role of codon cells. Martinotti cells can be understood as "inverted" pyramidal cells, with axons projecting superficially to the molecular layer, where they would synapse with pyramidal cell apical dendrites to close the loop (Fairen et al., 1984). Stellate cells of layer IV would constitute another population of codon cells, which receive afferent signals faithful to primary sensory inputs. These features led Marr to conjecture that the afferent fibres to Martinotti cells to be plastic, while the fibres to stellate cells to be un-modifiable.

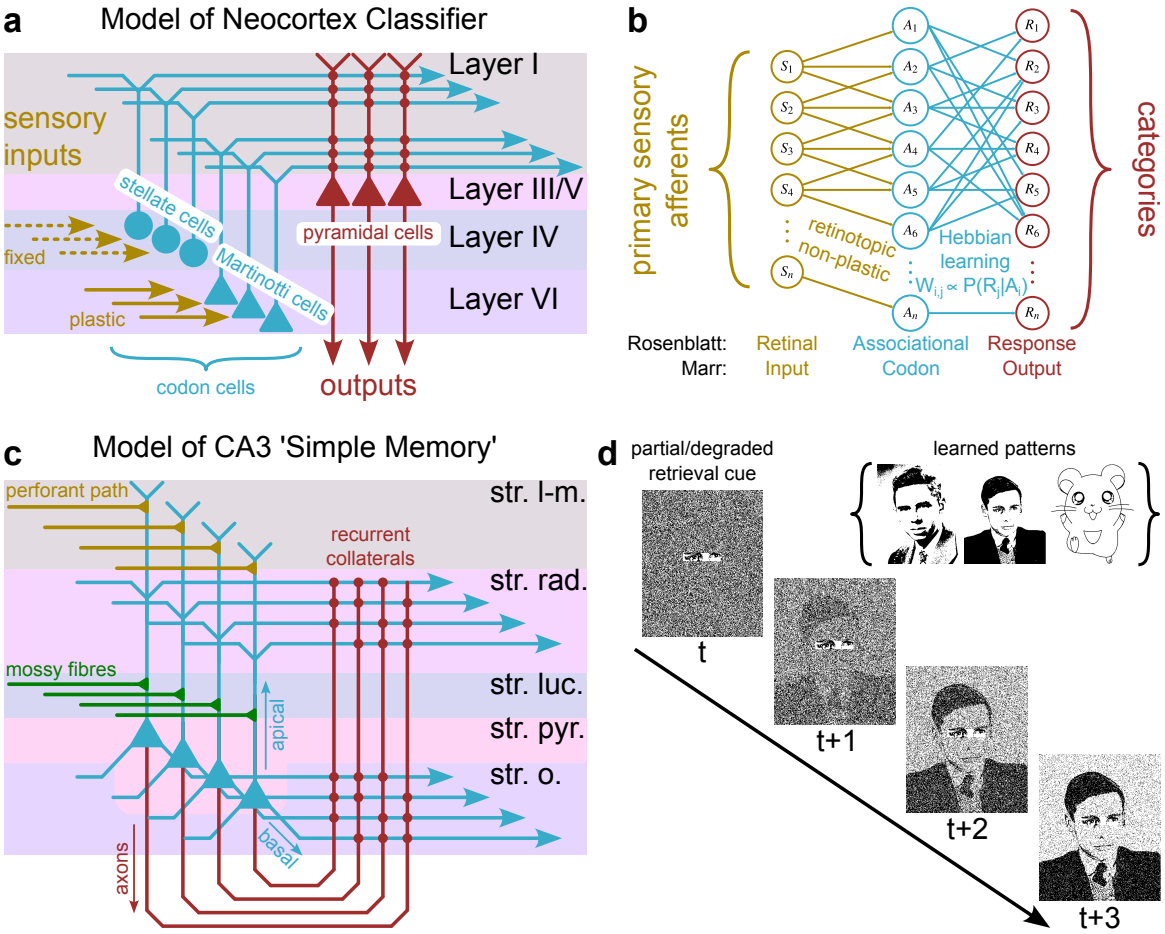

Figure 1.2: Computational models of the hippocampal-neocortical learning systems. **a)** Marr's model of the neocortex as a classifier. **b)** Comparison between Marr's network and Rosenblatt's perceptron. **c)** Marr's model of the hippocampus as an (auto-)associative memory system. Connectivity based on the description provided by Witter (2018). **d)** A stochastic Hopfield network that, after having learned three binary images, gradually retrieves the correct image from a partial cue.

active represents, therefore, the likely category of an input[19].

Marr's model of the cerebral cortex shares a strong resemblance with an earlier model elaborated by Frank Rosenblatt, known as the *perceptron* (Rosenblatt, 1958; Fig. 1.2**b**). The main difference, to the extent that the present review is concerned, is that Marr's network grew vertically (i.e., across cortical laminae), whereas Rosenblatt's model extends horizontally (i.e., from primary regions to associational areas). Over years of advances in knowledge and technology, the perceptron gradually evolved into the *convolutional neural network* (CNN), among other things, which is a special

[19]The appeal of this model can be appreciated through its ability to explain certain "Gestaltian phenomena". For instance, the *law of closure*, whereby objects with incomplete edges are perceived as complete, can be explained by a set of codon cells strongly activating the pyramidal cell corresponding to the category that the object belongs to. The missing codon cells and any noisy codon cells are "hammered away" by the evocation of this true category.

type of *multi-layered perceptron*, where neurons of each subsequent layer receive inputs from a retinotopically constrained *receptive field* of the previous layer (Fukushima, 1980). Hence, with each successive layer, neurons acquire more *abstract* features by integrating features of the previous layer (e.g., a complex cell is formed by combining multiple simple cells with adjacent receptive fields) and are, therefore, better able to distinguish complex nuances across categories. Imperatively, such a network organisation appears to faithfully model the behaviour of neurons and cortical activities along the ventral visual stream (V1 $\rightarrow$ V2 $\rightarrow$ V4 $\rightarrow$ IT). For example, following learning, CNN deep layers support representations similar to those found in the *inferior temporal cortex* (see Lindsay, 2021, for review). This, coupled with its wide success in practical image classification problems, makes convolutional networks the currently most well accepted model of the mammalian visual system (cf. Serre, 2014). Therefore, despite missing some marks on the exact organisation of the circuitry (cf. Willshaw et al., 2015), Marr's general proposal for the neocortex, as a universal classifier that acquires *memories* of categories, certainly captures an essence of its role[20].

An equally important deduction, made by Marr, argued that some of the synapses within the neocortex would only become modifiable during sleep[21] (Marr, 1970, p. 215). Curiously, the reasoning that was offered recaptures many of the same ideas concerning the issue of *interference* in semantic learning. In fact, it was inferred that, for new information to be incorporated into existing associations, the incoming patterns need first be verified against the entirety of the already estab-

[20]Marr did not specify whether the cortical mantle would act as a single (massive) *hidden*/codon layer, though the manner in which he developed his ideas certainly implied so (1970). This peculiarity notwithstanding, he did suggest, also, that Martinotti cells, with modifiable afferent synapses, would receive intercortical projections (p. 231). Therefore, it is possible that he envisaged, eventually, a multi-layered feed-forward architecture, akin to the one found in CNNs (see Kohonen, 1988, for what a generalised framework of Marr's models could look like). By *Universal Approximation Theorem*, a neural network with just a single hidden layer should be able to accurately solve any arbitrary classification task, provided that there is a *very* large number of hidden neurons (Hornik et al., 1989). Such networks, however, pale in efficiency when faced against deep networks (cf. Minsky and Papert, 1969; Bengio et al., 2009).

[21]To be precise, it was suggested that the synapses between the codon and pyramidal layers would be modifiable at all times, whereas the weights between the input and codon layers would be frozen until sleep (Marr, 1970, p. 231-232). The full reasoning offered by Marr was somewhat esoteric, but can be related to the modern-day concept of *vanishing gradient* (cf. Balduzzi et al., 2017). In multi-layered feed-forward networks, many learning rules, such as the popular *backpropagation* algorithm, cause the changes in synaptic weights to decrease exponentially in each preceding layer. Consequently, the shallow layers may exhibit very little learning, causing the entire network to converge onto a sub-optimal *local minimum*. Yet, combating the vanishing gradient by increasing the rate of synaptic change in the lower layers may interfere with the consistency of the inputs (e.g., imagine if the simple cells in V1 were to constantly change their receptive fields!), which, conversely, leads to unstable learning and the *exploding gradient problem*. One solution, close in spirit to Marr's, is the adaption of *deep belief networks*, along with the *wake-sleep* algorithm that is sometimes used to train them (Hinton et al., 2006). Such networks make use of a "sleep phase" to tune the weights in the preceding layers, so that they are more conforming to the representations in the deeper layers. During this sleep phase, no input is passed to the network. Instead, *pseudo*-inputs are generated by the network itself from a deep *associative memory* store, which reflect the network's internal model of external sensory patterns (see also Friston et al., 2021).

lished *probability structure* (cf. semantic structure), as well as all other input events. For this to be possible, all the weights within the network must become simultaneously plastic, so that the relationships between novel and familiar concepts are considered as a whole. In exchange, because the network is unable to perform classification on sensory inputs during this "training period", the most opportune time for the latter to take place would be during the sensory-deprived moments of sleep.

Given that the cortical network cannot perform classification and learning concurrently, a temporary memory store, therefore, becomes necessary for keeping and, later on, reinstating all the input events an agent experiences over the course of a day. To adequately perform this role, the putative memory system must satisfy two requirements. First, the system must be able to keep track of the set of cortical pyramidal cells that were concurrently active during individual experiences — i.e., to *associate* the simultaneous elements that, collectively, comprise an experience. Second, because no actual associations were formed in the cortex as these experiences occurred, complete events kept in this memory must be addressable and retrieved through the presentation of partial information that is constructed from the existing associations in the cortex. Marr (1971) proposed that the *hippocampus*, by virtue of its densely *recurrent* collateral structure, would serve the role of this *simple memory*[22] (Fig. 1.2**c**). Specifically, the CA3 subfield of the hippocampus is largely populated by pyramidal neurons, for which as many as two-third of the inputs are derived from other pyramidal cells within CA3[23] (Li et al., 1994). If parallel inputs, arising from distributed neocortical locations, were to converge onto CA3 and excite an arbitrary population of pyramidal cells, then the large availability of interconnected Hebbian synapses would guarantee these inputs to be associated amongst themselves — *autoassociated*. A subsequent activation of a subset of this original population would cause collateral excitation of all other pyramidal cells that were active during the initial experience. This self-sustained and propagated wave of excitation would eventually lead to

[22] In actuality, Marr formulated two models of associative memory, but quickly abandoned the first model, quoting a constraint on storage capacity (Marr, 1971, p. 31). In effect, the first model is closely related to the modern-day *Restricted Boltzmann Machine* (Hinton, 2012), while the second model finds its equivalent in the *Hopfield network* (Hopfield, 1982). Furthermore, the *energy landscape* between these two networks can be expressed by the same mathematical formulation, signifying that they can effectively perform the same functions (cf. Willshaw and Buckingham, 1990; Hinton, 2012).

[23] Recurrent collateralisation is present at both the apical (*str. radiatum*) and the basal dendrites (*str. oriens*) of CA3 pyramidal cells (see Amaral and Witter, 1989; Witter, 2018, for review). In particular, CA3 pyramidal cell axons ramify inside both *str. radiatum* and *str. oriens*, while *perforant path* afferents project onto the distal apical dendrites in *str. lacunosum-moleculare* (cf. Fig. 1.2**c**). Furthermore, studies have shown that these basal and apical dendrites express different types of plasticity and synaptic integration (Brzdak et al., 2019; for review, see Spruston, 2008). Therefore, clarifying the putative differences in roles between these two recurrent networks may shed new lights onto the computational basis of the CA3 associative memory (cf. Knierim and Neunuebel, 2016).

the retrieval of the full memory pattern — *pattern completion* (cf. Fig. 1.2**d**).

In addition to being able to act as a *content-addressable* memory (cf. Hopfield, 1982), recurrent network models display extra properties to help further illuminate the functions of the hippocampus. In fact, researchers soon figured out that, by feeding a steam of input patterns, as opposed to stationary images (e.g., Fig. 1.2**d**), the network learns to extract regularities in the unfolding of events over time (Kohonen, 1988, p. 16-20). This occurs due to the fact that any output produced by the network at one time point, will, at the next time point, be *fed back* as input onto itself, via the recurrent collaterals, in conjunction with the new external inputs. In this sense, the network has a short-term memory of the previous patterns that it had received, which are kept actively in *reverberation* within the circuits. Therefore, when learning occurs at each iteration, the current inputs are not only auto-associated with themselves, but also with the state of the network from the previous time step — *temporal associations*. It follows that the recurrent architecture, when employed in this manner, is able to learn and, later on, recall sequential patterns of activities[24]. In other words, such a memory system is predisposed to confer a sense of time and space[25] (i.e., contextual information) to individual experiences, consistent with the involvement of the hippocampus in episodic memories. Furthermore, such networks display generative properties[26], implying a role of the hippocampus in forming predictions and in imagining hypothetical scenarios. Indeed, Hassabis and colleagues have shown that hippocampal patients were severely impaired in their ability to imagine fictitious experiences that take place within a environmental context that they have been cued with (Hassabis et al., 2007; see also Pfeiffer and Foster, 2013). In particular, the forged details came as fragmented pieces that lacked spatial coherence, as though the disparate elements of an experience failed to be chained into a narrative taking place within a unified context (cf. O'Keefe and Nadel, 1978, cognitive mapping theory).

[24]For some time, variants of recurrent neural networks, such as the *long short-term memory* network, have been employed extensively in solving tasks involving sequential inputs (see Yu et al., 2019, for review). Recently, however, its popularity had been displaced by the advent of *transformers*, which form the backbone of the widely successful language models ChatGPT and GPT-4. Paradoxically, transformers do not have a recurrent architecture; it was demonstrated that the feedback mechanism holds no particular advantage in extracting dependencies in sequential structures (Vaswani et al., 2017). A severe limitation recently uncovered in the GPT-4 model relates to its inability to predict many steps into the future (Bubeck et al., 2023). In particular, such capacity appears to rely on a dynamic internal model that plays out multiple scenarios of future events within the constraints of past occurrences, a capacity that the hippocampus appears to fulfil (cf. Hassabis et al., 2007).

[25]With some tweaks, recurrent networks can also be made to model a spatial cognitive map. These models, known as *continuous attractor networks*, will be explored in the section on the *relational interpretation*.

[26]For example, a recurrent net, trained on various text documents, could both complete and generate long strings of text (Hinton, 2012). It does so while displaying an impressive level of adherence to grammar, syntax, style and comprehension.

In interim summary, Marr's computational models laid the theoretical groundwork for how the two parallel and interdependent systems of the neocortex and the hippocampus may work in synergy to achieve complementary roles in the encoding, processing and retrieval of (declarative) memories (cf. McClelland et al., 1995). On the one hand, the cortex performs the "sophisticated" role of learning, at times, abstract categories from input patterns that reflect the orderly redundancies found in the natural world. Yet, it is unable to perform this learning "on-the-spot" as doing so risks interference with the established memory structure. In response to this limitation, on the other hand, the hippocampus plays a time-limited role in rapidly storing cortical patterns of activation elicited by individual episodes of experience. Further modelling works have suggested that this 'simple memory' may, in fact, be very sophisticated still, as evident in its ability to encode spatio-temporal sequences and to formulate predictions as well as fictional scenarios. Marr predicted that, during periods of sleep, these provisional memories would be addressed and retrieved by the cortex by presenting partial pieces of the whole pattern, which can be produced from the existing representations in the cortex[27]. This interplay between the cortex and the hippocampus, between the classifier and the auto-associative memory (cf. Hinton et al., 2006) and between the slow and the fast learners (cf. McClelland et al., 1995), hence, elaborates a systems-level process through which memories of individual experiences, first held (temporarily) by the hippocampus, is gradually assimilated into cortical associations — *systems consolidation*.

On the basis of these computational perspectives, an important inference can be drawn for a link between the hippocampo-neocortical memory systems and Tulving's episodic/semantic dichotomy. In fact, the categorical learning performed by the cortex alludes to the establishment of a semantic structure, while the sequential associative memory of the hippocampus is reminiscent of episodic memories, which are distinguished by spatio-temporal contexts. It is, therefore, conceivable that the two independently operating memory systems envisaged by Tulving have an anatomical basis in the hippocampal and cortical networks. In the preceding section, we reviewed strong evidence against the independence between semantic and episodic memories. Yet, the possibility still remains that these two memories may interact in a synergistic manner as delineated by the computational models (i.e., interdependent rather than independent). A question arises, therefore, about the nature of the information stored in the hippocampus and the manner in which this information is transformed as

[27]This phenomenon, termed *reactivation*, will be explored in the section on the *mnemonic dialogue on space*.

it exits the hippocampus to be incorporated into cortical associations.

Current conceptualisations of systems consolidation is divided between two major theories. The core of their discrepancy lies precisely in the differential treatment of semantic and episodic memories in systems consolidation. The first theory, known as the *standard memory consolidation* theory (and its sister theory, the *indexing theory* (cf. Nadel et al., 2007)), asserts that both semantic and episodic memories rely on the hippocampus for a limited time, until they are consolidated into cortical networks (Teyler and DiScenna, 1986a; Frankland and Bontempi, 2005). This is the view that Marr held, citing that, due to a limited storage capacity (cf. Hopfield, 1982), memories need to be periodically purged from the hippocampus, after they had undergone consolidation. As such, the information stored in the hippocampus must either be in a raw form or in the form of 'indices' pointing to the original patterns of activity elicited over distributed cortical sites during behavioural exposure. Modifications on the existing weights in neocortical networks, guided by the information held in the hippocampus, which are reinstated during offline periods, therefore allows for this new information to be gradually incorporated into permanent cortical representations independent of the hippocampus. In the clinical studies highlighted earlier, hippocampal patients suffer from a *temporally graded* form of retrograde amnesia, in that remote memories are more likely to be spared than recent memories. Such appeared to be true of both semantic and episodic memories. Therefore, from the point of view of the standard model, semantic memories and episodic memories are treated in the same way during the process of consolidation.

This perspective soon came to be challenged, however, as new data began to portray an ambiguous narrative over the temporal susceptibility of episodic and semantic memories in medial temporal amnesia (reviewed in Nadel and Moscovitch, 1997; Nadel et al., 2007). In particular, the reach of retrograde amnesia in relation to episodic memories appears to be extensive, covering anywhere from decades of the premorbid period to a lifetime in patients with substantial bilateral medial temporal damages. This renders the notion of the hippocampus, as a "temporary" memory store, rather dubious with respect to episodic memories. Furthermore, over the span of this retrograde period, there is little suggestion for a temporal gradient[28], that is the propensity for details to be lost is uncorrelated with the age of the memory episode. This sharp, yet, non-parametric decline in episodic

[28]More precisely, patients with temporal lobe amnesia tend to lose details in older memories at a similar rate as intact individuals (Steinvorth et al., 2005).

recall, therefore, cannot attest for a systematic consolidation process that operates gradually and continuously. In contrast, semantic memories do appear to follow certain trends delineated by the standard model. Particularly, semantic knowledge acquired during the premorbid period generally shows preservation, i.e., no retrograde amnesia. In this respect, consolidation of semantic memories seems to operate over a relatively and reasonably short timescale. Though some form of new semantic learning does occur, nevertheless, as previously addressed, the learning displayed is a far cry from that expected of healthy semantic acquisition.

Analogous reports have also surfaced from animal studies, where different memory tasks express different levels of impairments from the time the hippocampus was lesioned (reviewed in Sutherland et al., 2010). In particular, *spatial memories* consistently show a "flat gradient", that is they are never consolidated to the extent that they can survive complete hippocampal ablations. In contrast, other types of memories, such as contextual fear memories and flavour-place paired-associates, have been shown to persist, if enough time was allowed for their putative consolidation.

These contrasting reports between episodic and semantic memories have motivated an alternative theory to be proposed, known as the *multiple trace theory* (Nadel and Moscovitch, 1997). Under this perspective, episodic memories, however old, always rely upon the hippocampus. This is because the indices stored by the hippocampus are, in and of themselves, not the raw information that becomes incorporated into neocortical representations. Instead, the cortex only learns to extract the statistical regularities found across multiple episodes. Such regularities are hypothesised to emerge from the accumulation of multiple copies of commonly accessed information; each time an episodic trace is retrieved, a copy of that trace is created — *multiple traces*. The indices themselves, which record all the arbitrary associations that are unique to a given episode and augmented by a spatial and temporal context provided by the hippocampus, never fully exit the hippocampus. Rather, only the redundancies, the gists or the information that could subserve the basis of categorical knowledge — *semantic information* — become consolidated.

This interpretation is certainly commensurate with the computational models previously elaborated. As Marr held, the architecture of the neocortex is suited for the acquisition of categorical information structured around redundancies, regularities and the relationships across these regularities. In contrast, the hippocampal network is poised to form arbitrary associations amongst cortical patterns, without consideration for any inherent semantic structures found within these episodes,

other than the structures that are already present in the cortex at the time the information is perceived and encoded. From the point of view of the *multiple trace theory*, the neuroanatomical location(s) where a type of memory resides, and the temporal susceptibility of said memory to loss following hippocampal injuries, are defined by the distinct computational objectives of different memory networks.

In closing, the clinical case study of amnesics, alongside animal experimentation, led to the defining and distinguishing of *explicit* and *implicit* memories, hippocampal and neocortical learning systems, the putative distinction between *episodic* and *semantic* memories, and the concept of memory consolidation. By arguing from computational perspectives, we unified these concepts and brought forth the views of a "hippocampal arbitrary episodic associator" and a "neocortical semantic features extractor". Unresolved nuances in how these networks interact led to different interpretations on the nature of information storage, transformations and consolidation. Lesser emphasis, however, was given to spatial/contextual information, which, as briefly discussed, constitute defining features for episodic memories and for the computational properties of the hippocampal network. Having taken care of one side of the mnemonic coin, we will now explore the other side.

1.2 Hippocampal encoding of "space"

The pioneering studies on patient H. M. (Scoville and Milner, 1957) firmly placed the hippocampus at the centre of discussions over declarative memories. This realisation mounted pressure on the development of animal models in which more invasive procedures can be performed with the hope of elucidating the neural mechanisms underlying this mnemonic process. A breakthrough came with the advent of single-unit recording techniques in awake and freely-behaving rats, which permitted James B. Ranck (1973) to isolate a variety of neuronal responses tied to a diverse range of behaviours in hippocampal and parahippocampal regions. This level of complexity and diversity observed from the "receptive fields" of hippocampal neurons was expected from a structure, for which the putative function concerns the association of distributed attributes of experiences. However, it may be argued that it was a contemporaneous study, conducted by O'Keefe and Dostrovsky (1971), that set the precedence for a great majority of research to come.

While recording single-unit activities from the CA1 subregion, in rats exposed to an "open-field" platform, the experimenters observed the same complex behaviourally-related patterns as Ranck,

but also noted a small population of neurons[29] whose firing was circumscribed to specific locations where the animal was found within the environment. The discovery of these *place cells* eventually led to the adaptation of the now widely acclaimed *Cognitive Mapping* theory[30] (Tolman, 1948) to describe hippocampal functions (O'Keefe and Nadel, 1978). Central to this framework is the idea that the hippocampus constructs and stores an *absolute* coordinate system for unique spatial environments, within which the agent is able to describe the allocentric relations between the individual elements that compose its experiences over space and time. It is perhaps due to the uni-dimensional nature of the framework (i.e., space), which allows concrete experimental designs and research questions to be formulated in a ready manner, that space became the basis on which a large body of modern investigations (this thesis included) have been conducted. Ironically, this strong emphasis on space also served to obscure, to some extent, the role of hippocampal physiology in supporting declarative memory.

In the subsequent chapters of this thesis, mnemonic phenomena in the brain will be elaborated from the perspective of spatial representations, as described through functional recordings. It is therefore of paramount importance, here, that an understanding of the relationship between space and memory be reached, at least on conceptual grounds. To this end, two competing views on the hippocampal encoding of space will be introduced and contrasted: the *geometric* interpretation and the *relational* interpretation. Through reviewing the successes and failures of these disparate propositions, the nature of the functional encodings supported by the hippocampus will be discussed.

1.2.1 The geometric interpretation

The "Euclidean" property of place cells inspired early investigators to propose hypothetical models, which could account for the computations necessary to support place coding, based on the integration of geometric quantities found within an environment. One such early model seemingly applied the mathematical formalisation of space to directly explain the neurophysiological phenomenon. In a Cartesian coordinate system, the intersect between two orthogonal lines is used to determine the location of an element (including the agent) in the absolute space defined over a

[29] In this original report, only eight out of the seventy-six cells recorded displayed spatially-localised responses. However, with better recording techniques and larger environments, it was later determined that a great fraction of CA1 pyramidal cells exhibits this property (O'Keefe, 1976; Wilson and McNaughton, 1993).

[30] Note that Tolman formulated this theory originally as a rebuttal to the sensory-response theory popular during his (and Lashley's) times, majorly based on the experimental observations of flexible solutions to navigational problems, which could not arise as a result of procedural learning or conditioned responses.

navigational plane. It therefore stands to reason that, given a neural code capable of informing the agent of their distance away from the geometric borders of an environment, integration over such signals would yield a representation similar to that of place cells (Fig. 1.3**a**) (O' Keefe and Burgess, 1996; Burgess et al., 2000; Hartley et al., 2000). The neural correlate of such a signal was discovered in the subiculum, in the form of neurons with receptive fields stretching along an edge of an environment — termed *border cells* (Lever et al., 2009). With this discovery came soon the realisation, however, that such a mechanism is unlikely to support place field formation. For one, most of these border cells were limited to encoding the locations immediately adjacent to a wall, therefore failing to provide a uniform metric of scaled distances[31]. Secondly, in an environment with complex geometries or in one that is devoid of borders, the reliance upon such rigid quantities would contribute to a large accumulation of errors unfavourable to supporting a stable representation of space (cf. Geva-Sagiv et al., 2015).

Perhaps the solution lies in an alternative form of coordinates, one which preserves its uniformity over distance. As luck would have it, the discovery of border cells was paralleled with the identification of *grid cells* in the superficial layers of the medial entorhinal cortex (Hafting et al., 2005; see also Krupic et al., 2012). This latter cell type is characterised by multiple place fields distributed periodically at locations defined by a tessellation of equilateral triangles (Fig. 1.3**b**). The resulting hexagonal lattice, therefore, provides an invariant measure of spatial intervals in two-dimensions. Grid cells are distinguished by three properties: spacing (interval between place fields), orientation (angle of the lattice relative to the environment) and phase (displacement of the lattice relative to the environment). When integrating the firings of a small set of grid cells that vary in spacing and orientation, but that share a fixed spatial phase, the constructive and destructive interference would yield a periodic pattern with a central peak, resembling a place cell (Solstad et al., 2006). Hence, within the restricted space of the laboratory environment, the linear integration of different grid cells would yield a single place field. Unfortunately, this promising lead was yet again met with disappointment when bilateral lesions to the medial Entorhinal cortex caused only partial deficits in place coding[32]

[31] However, the width of place fields does increase as a function of distance away from borders, while the fraction of place cells decreases, as are the expected behaviours if place cells were to emerge from this biased spatial metric (O' Keefe and Burgess, 1996; Tanni et al., 2022)

[32] The discovery of *grid modules*, later on, hammered one more nail into the coffin (Stensola et al., 2012; Waaga et al., 2022). The rat medial Entorhinal cortex supports a small handful of functional modules, where each module contains a set of grid cells that share similar orientation and spacing, but differing phase. This limited set of inputs would not provide the diversity necessary to produce a well localised interference pattern. However, this finding conforms well with

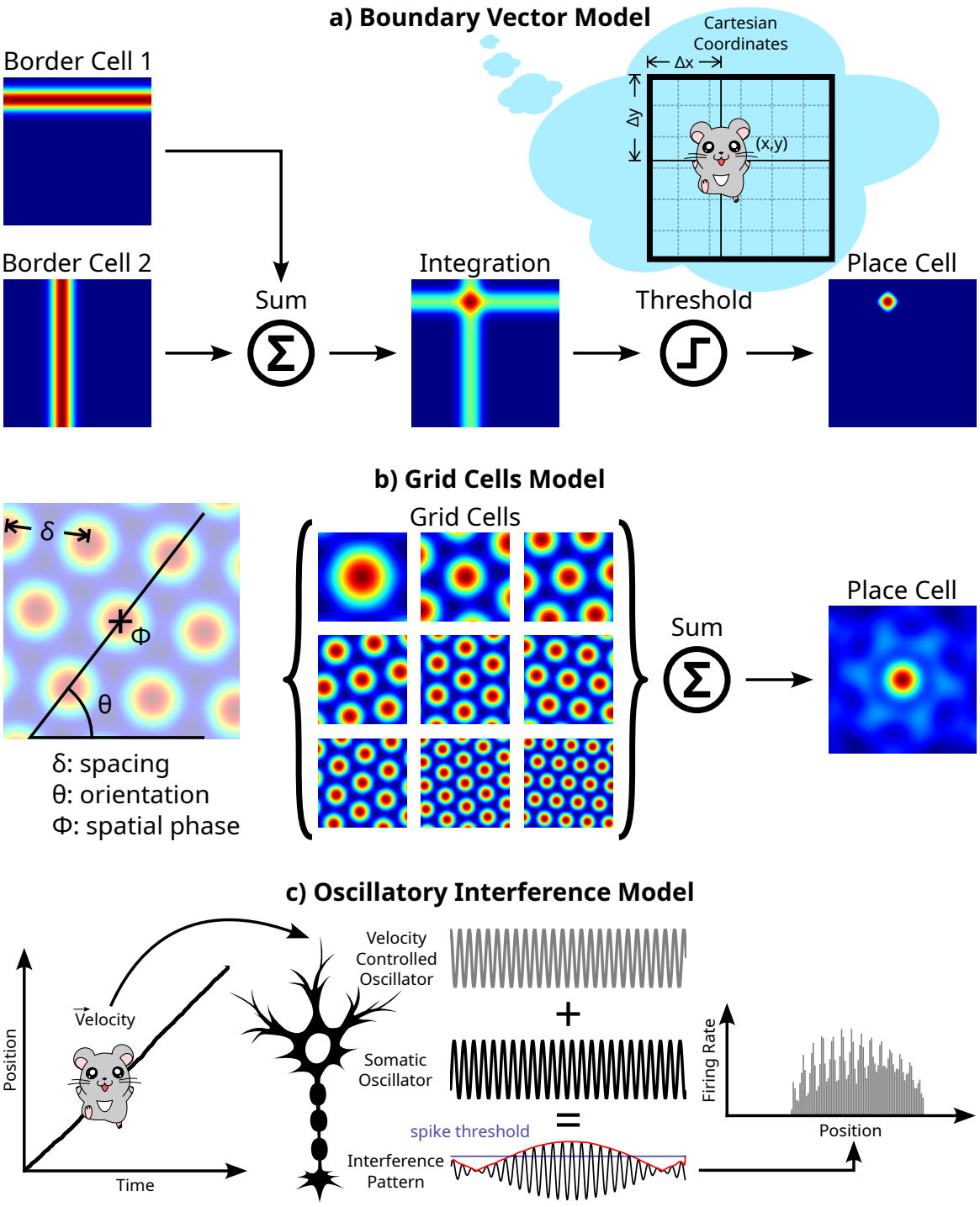

Figure 1.3: Computational models for place cells that follow the *geometric interpretation*: **a** Boundary vector model, based on O' Keefe and Burgess (1996); **b** Model following the integration of grid cells, based on Solstad et al. (2006); **c** Oscillatory interference model, based on Lengyel et al. (2003).

(Miller and Best, 1980; Hales et al., 2014). Therefore, although grid cells may partially contribute to the place code, they are not necessary for sustaining place cells. Furthermore, during hippocampal inactivation, the grid pattern in grid cells is severely disrupted[33] (Bonnevie et al., 2013). Together, these findings implying that possible bidirectional exchanges between place and grid cells would occur asymmetrically, with the former exerting more influence over the latter, rather than the other way around.

Looking at yet another alternative, it is possible that place cells do not arise from the integration of external properties, but rather depend on internally-generated rhythms. In fact, if a mechanism exists to count the number of steps an agent takes in any direction, then the accumulation of these displacements would yield trajectories in absolute space without the need to reference external environmental cues. This ability to keep track of self-motion vectors — known as *path integration* or *dead reckoning* (McNaughton et al., 1996; Whishaw, 1998) — is well supported by behavioural observations (cf. McNaughton et al., 1991, for review); a major inspiration for Tolman's Cognitive Map theory comes from the rat's ability to plan new trajectories in a learned environment to either circumvent obstacles or to increase efficiency, as though as they could mentally construct a new vector in geographical space based on the information they have acquired from previous explorations. Soon after the discovery of place cells, O'Keefe and Nadel suspected that *theta-rhythms* may contribute a key role in supporting this computation (O'Keefe and Nadel, 1978). Theta-rhythms in the hippocampus and the medial entorhinal cortex have their amplitude and, to a lesser extent, frequency modulated by the running speed of the animal (McFarland et al., 1975; Hinman et al., 2011). Likewise, the firing rates of place cells are modulated by running speed and heading direction as well[34] (McNaughton et al., 1983). When these field oscillations and cell firing rates are simultaneously considered, a stereotyped interaction, known as *phase precession*, occurs between the two: upon the animal entering a neuron's place field, the place cell would fire at a precise phase angle in relation to theta, while this phase preference would progressively shift to an earlier angle with each subsequent

the expectation of the Continuous Attractor model (discussed later).

[33] It is worth noting that in preweaning rat pups, place cells emerge earlier than grid cells, although they reach maturity at a slower rate than grid cells (Wills et al., 2010).

[34] Interestingly, disruptions of the head-direction cells network, by lesions of the lateral mammillary nuclei, caused place fields to repeat themselves in identical compartments arranged around a radial arm (i.e., a half circle) (Harland et al., 2017). Previously, this effect of repeated place field maps in identical compartments had been shown in normal animals, as long as the compartments are arranged in parallel orientations (Skaggs and McNaughton, 1998; Spiers et al., 2015). These results provide a strong suggestion that the hippocampus makes use of small differences in internal cues to dissociate spatial contexts.

theta cycle, up to the point of covering nearly a full 2π range at the exit of the place field (O'Keefe and Recce, 1993; see also Geisler et al., 2007). Such orderly relationship could arise as a result of the interference between two oscillators tuned to slightly different frequencies (Fig. 1.3**c**) (O'Keefe and Recce, 1993; Lengyel et al., 2003; Burgess et al., 2007). If one of these oscillators was modulated by the running speed of the animal, then spatial displacements over time would be tracked by the phase difference between the two oscillators, effectively achieving an *integral* operation of velocity over time[35]. This proposition was however brought into doubt when place cells and grid cells were discovered in bats, an animal specie in which theta oscillations occur intermittently as opposed to continuously in rodents (Ulanovsky and Moss, 2007), hence not being able to act as a "tracking device" over extended periods. Furthermore, hippocampal place cells were unimpaired by the disruption of theta-rhythms through pharmacological inactivation of the Medial Septum[36] (Mizumori et al., 1989; Koenig et al., 2011).

These findings alone do not altogether refute the possibility for path integration to be supported by entorhino-hippocampal circuits. In fact, it has been later found that the modulation of firing rates of certain Entorhinal neurons by speed — dubbed *speed cells* — persists in the absence of field theta[37,38] (Hinman et al., 2016). Therefore, parallel and redundant sources of information could converge onto these circuits to maintain spatial localisation by path integration. To test the influence of self-motion cues on place cells, Terrazas et al. (2005) devised a behavioural apparatus that progressively restricted the availability of self-motion information. Three conditions were assessed: 1) free running over a circular track; 2) passive travel around the circular track in a toy train cart that is actuated either by the animal via a lever or by the experimenter; 3) the environment (maze track

[35]The interference pattern produced by two slightly out-of-tune oscillators would by itself be a periodic pattern. Therefore, constructing of a single place field through this method would require further mechanisms to be in place. For this reason, oscillatory interference models are most commonly used to describe grid cells.

[36]Although grid cells are severely disrupted. Interestingly, new place fields are also formed in a novel environment during septal inactivation, which remain stable following washout of the pharmaceutical agent (Brandon et al., 2014). In contrast, place fields also form in a novel environment in the presence of an NMDA-receptor antagonist, but these representations remapped following washout (Kentros et al., 1998). These reports suggest that the spatial and memory aspects of the hippocampal code may in fact be dissociated.

[37]However, optogenetic stimulation of the Medial Septum, which entrains hippocampal field theta with the frequency of the stimulation, can both initiate locomotion and maintain a steady running speed proportional to the frequency of the stimulation (Fuhrmann et al., 2015). Therefore, self-motion cue information likely reside in distributed circuits with varying degrees of involvement in action planning, motor efference copy and feedback control.

[38]It may be tempting to infer that grid cells have a stronger involvement in path integration, based on the fact their spatial tuning patterns necessitate theta. However, grid cells collapse in the dark with the absence of stable visual anchors (Chen et al., 2016), whereas place cells persist. Perhaps path integration can only survive for brief periods before its accuracy succumbs to the accumulation of errors. However, this view does not explain the resilience of place cells, apart from the potential odour trails left by the animals.

and curtains) is rotated around the animal that is sitting inside a stationary cart. Each gradual restriction in self-motion cues (i.e., no restriction, motor efference copy and vestibular inputs) saw an increase in the width of place fields, with an accordingly longer slope over theta phase precession[39]. Similarly, in the dark with restricted optic flow information (but also visual cue landmarks), place field width also saw a modest expansion paralleled with a decrease in stability (Quirk et al., 1990; Markus et al., 1994). From these results, it may be inferred that place cells reflect an active tracking of self-motion cues, the availability of which dictates the precision of the encoding. However, what is important to note in these studies is that, in-between active and passive movement conditions and also light/dark conditions, a sizeable portion of place fields undergo remapping. Other studies showed that these (partial) remappings occurred spontaneously when switching between conditions (Chen et al., 2013; Quirk et al., 1990), without having the animals re-enter the maze, which could serve to induce a new spatial context. Therefore, not all place cells could arise from a simple linear combination of self-motion signals, as is the expected result if the goal is to minimise the error in path integration using the information at hand.

Perhaps, then, the place code is supported by an integration of all of these quantities. In fact, all of the aforementioned manipulations had an appreciable impact on place cells, if not a destructive one. Accordingly, virtually all models of place cells, including most of the previously discussed ones, argue for the necessity of integrating diverse sources of sensory information in order to support a reliable hippocampal encoding of space. This rectification notwithstanding, behavioural observations made in hippocampal-lesioned rats have cast tantalising doubts over the hippocampus's role in spatial learning. An extensive body of experiments conducted by Whishaw and colleagues (reviewed in Whishaw, 1998) have shown that the hippocampus, despite its intricate functional correlates with space, may in fact not be needed for solving certain core Cognitive Mapping problems. In an initial experiment, hippocampal-lesioned or fimbria-fornix transected rats learned to swim to a hidden platform submerged under milky water in a circular pool — termed the Morris Water Maze task. The platform was located at one of four quadrants, delineated by the cross-hair formed from the four cardinal points, with the animal starting from one of these cardinal points. During probe trials,

[39]Computer simulations suggest that the ability to tell the animal's position from a population of place cells is not negatively impacted by the expanding width of place fields in a two-dimensional environment (Zhang et al., 1998). An interesting question that follows is: why are place fields usually well localised and scale proportionally with the availability of sensory inputs?

the platform was removed and behavioural performance was assessed by the time latency to reach the correct quadrant and the time spent in the correct quadrant. With extensive training, the lesion rats approached the hidden platform location with a latency comparable to the control group[40,41], albeit requiring substantially more training time. The crucial finding was reported in a separate experiment, in which the training trials were initiated from only three out of the four cardinal points. During the probe trial, the rats were started from the withheld cardinal location. Surprisingly, both control and lesion groups swam to the learned platform location with comparable performances. Together, these results demonstrate that the ability to formulate a holistic representation of a physical environment (i.e., a cognitive map), such that a new path towards a familiar goal can be devised from any location, is preserved in the absence of the hippocampus[42]. The hippocampus, in turn, seems to play a crucial role in catalysing the learning of allocentric maps, but is not, in and of itself, necessary for supporting navigation using such maps.

In summary, these *prima facie* attempts at establishing the functional patterns of the hippocampus within a Newtonian framework were quickly disillusioned by the complexity and resilience of the place code. Crucially, the mechanism underlying the generation and sustain of this space-centric signal does not adhere to a unimodal parametric space, in a manner that is counter-intuitive to a geometric interpretation of the phenomenon. To be clear, the body of results discussed here should not be taken as evidence for a lack of contribution of geometric quantities to hippocampal function. In fact, each of the aforementioned manipulations has an appreciable impact on the place code, despite the fact that none of them alone is required for place cells. It is precisely this sufficiency-over-necessity nature of place cells that makes its understanding elusive. Compared to other canonical

[40]Interestingly, the lesioned rats spent significantly more time in the correct quadrant compared to controls. This seemingly extraordinary result can be explained by a follow-up experiment, in which the platform was relocated to a different quadrant following learning of a familiar location. Here, control rats acquired the new place association quickly, while lesioned rats consistently swam to the old location, but eventually learned the new location at a similar rate displayed during initial training. Hence, the hippocampus appears to be crucial for cognitive flexibility, transfer learning and/or "forgetting".

[41]Another interesting behavioural feature displayed by lesioned rats is to swim in tight circular loops, all the while maintaining a focused direction of travel towards the learned platform. One possible interpretation is that hippocampal-lesioned rats have a less precise notion of current location and travelled distance, owing to a deficit in path integration, causing them to adopt a navigation strategy that broadens their scope of search. This hypothesis is corroborated by a later demonstration that fimbria-fornix transected rats display major deficits in path integration in the dark (cf. Whishaw et al., 2001), although contradictory results have been reported in hippocampal-lesioned rats (Alyan and McNaughton, 1999).

[42]It is important to note that this behaviour is only rescued under a testing environment that contains stable distal visual landmarks (Whishaw et al., 2001), suggesting a deficit in path integration (see footnote 41). However, the fact that any allocentric navigation strategy can be manifested by hippocampal-lesioned rats casts doubts over the hippocampus's exclusive role in allocentric coding.

spatial cell types[43], place cells seem to express a level of resilience that insinuates a role beyond the coding of space. In the next section, we depart from this space-centric view in order to elaborate a more general perspective on hippocampal coding and functions, that is the mapping of relations.

1.2.2 The relational interpretation

Despite O'Keefe and Nadel's initial emphasis on the mapping of physical space to serve as a foundation for associative memory, some subsequent theorists have attributed a broader role to the hippocampus in encoding generalised features of experiences. This shift in paradigm owes in part to the experimental observations elaborated earlier, but also to certain "anomalies" in the functional patterns of the hippocampus that described orderly relationships outside of the spatial domain. In fact, a multitude of experiments have demonstrated that hippocampal neurons can encode a variety of experiential features not necessarily tied with space. These non-spatial representations can, generally speaking, be filed under the broader phenomenon of *rate remapping*.

Early on, it was observed that individual place cells tend to exhibit different firing rates on the inbound and outbound excursions within the same arm of a radial-arms maze, despite preserving the locations of their place fields (McNaughton et al., 1983). All of a sudden, a novel dimension in the hippocampal code had been made available for describing behavioural/sensory parameters ulterior to space. This discovery is especially meaningful to memory researchers who, until now, encountered difficulty in establishing the mechanistic underpinnings of episodic associations within the framework of space. In this experiment, the non-spatial parameter in question appears to be the directionality of movements. A later study has also observed this direction-based rate modulation in an open field environment in subsets of place cells (Markus et al., 1995), suggesting that rate remapping is a conserved property. However, going from a radial-arms maze to an open field platform saw a drastic decrease in directional modulation, indicating that these differential firings may be a result of environmental geometries and the linearity of motion. Nevertheless, when the demand of the task was shifted from random foraging to a directed searching strategy, an increase in directional modulation was measured. Therefore, rate remapping appears to reflect specific aspects of task demands

[43] Other known canonical spatial cell types all seem to have an Achilles's Heel. Grid cells are disrupted by attenuation of theta-rhythms (Brandon et al., 2011; Koenig et al., 2011) or inactivation of the hippocampus (Bonnevie et al., 2013). The head direction cells network loses directional tuning following lesions to structures important in processing vestibular inputs (Blair et al., 1998; Bassett et al., 2007). The ability to pinpoint a single source of necessity for a particular neural code facilitates the attribution of function to this code.

or movement stereotypes, all of which cannot be directly explained by the geometry of space.

Additional studies have demonstrated similar rate remapping phenomena associated with a variety of non-spatial features (see O'Keefe and Krupic, 2021, for review). In particular, Itskov et al. (2012) recorded CA1 neurons that responded to distinct tonal stimuli, with the magnitude of their firing rates reliably discriminating the identity of the sounds. These same neurons, however, lost their auditory selectivity when the animal was relocated to an adjacent testing platform in the same room. The same conclusion was reached in a separate study testing for texture discrimination (Itskov et al., 2011), and again when testing for object discrimination (Manns and Eichenbaum, 2009) and odour discrimination (Manns et al., 2007; but see Wood et al., 1999). More abstract quantities have also been implicated in rate modulations. For instance, in animals performing a spatial alternation task in a T-maze, subsets of place cells in the central arm showed differential firing rates between left- and right-going trials (Wood et al., 2000). Furthermore, these rate modulations seem to reflect both retrospective coding (location from which animal arrived to the current place) and prospective coding (from the current place, the future goal to be reached)[44] (Ferbinteanu and Shapiro, 2003; see also Bower, 2005). Therefore, rate remapping (or *feature-in-place cells*) appears to capture a broad set of sensory/behavioural features and confine them into distinct spatial locations — *what* happened *where* (cf. O'Keefe and Krupic, 2021, for review).

The study of odour discrimination has led to an interesting proposal for a role of the hippocampus beyond the encoding of space. In one study, rats were tasked with learning a sequence of five odours and, later on, when re-exposed to a pair of odours drawn from the previously series, pick the odour that came earlier in the sequence (Fortin et al., 2002). It was found that, whereas normal animals could perform the task relatively well, hippocampal lesioned rats performed the task near chance levels. This implicated a role of the hippocampus in encoding the sequential order of items. A neural correlate for such code was found in the systematic rate remappings that occurred with successive exposures to different odours in a given location (Manns et al., 2007; see also Allen et al., 2016). These investigations, finally, culminated into the discovery of a purely non-spatial code: *time cells*. Pastalkova et al. (2008) found that, by introducing a running wheel (i.e., a delay period) in a spatial alternation task at the choice-stem of a figure-eight maze, CA1 pyramidal cells spontaneously organised into a temporal sequence tiling the entire duration of the delay period. Just

[44] Prospective and retrospective coding are also found in *awake replay*. See below.

like place cells, these time cells consistently fired at the same delay time across trials and exhibited precession around field theta. Rather intriguingly, these time cells ceased to exist when the alternation (i.e., the working memory) component to the task was removed or during spontaneous running inside the animals' homecages. Not only did these results demonstrate the contingency of this time code upon a working memory demand, but also immediately ruled out locomotion as a possible confound. Later, it was shown that time cells spontaneously organise during delay, still, in the absence of motion (MacDonald et al., 2013).

The existence of spontaneously organised sequences, be it for the encoding of odour sequences or delay time, led to the proposal of the *relational theory* (Cohen and Eichenbaum, 1993; Eichenbaum, 2004) and, analogously, of the view that the hippocampus is a sequence generator (Buzsáki and Llinás, 2017; Buzsáki and Tingley, 2018). The main postulate put forth is that the hippocampus's ability to establish relationships extends far beyond those that exist between locations within an allocentric coordinate space. Instead, all manners of quantities, spatial or non-spatial, can be assimilated into relationship structures, prescribing order, similarity and distance to arbitrary information that constitute holistic episodes of experiences. This is made possible by the sequential firing of cells that is spontaneously self-organised amongst hippocampal neuronal populations (likely owing to the recurrent connectivity). This initially *content-free* sequence becomes coincident upon cortical patterns of activity during experience[45]. The associations that are formed between the distributed cortical patterns and the hippocampal sequences, therefore, binds semantic contents into orderly relationships as experiences unfold. In other words, the hippocampus does not form associations between items and locations over a cognitive map, but between items and the occurrence of said items in an orderly fashion.

The appeal and generalisability of the relational interpretation notwithstanding, there is a fundamental challenge to this view. Whereas the encoding of space is exceptionally resilient, as highlighted in the previous section, rate remapping and the encoding of other non-spatial quantities seem to be manifested only under special experimental conditions and task demands. This aspect, alone,

[45] A common argument for internally organised sequences, as a fundamental organising principle of hippocampal function, is the concept of theta sequences (reviewed in Buzsáki and Tingley, 2018). The firing of both place cells and time cells follow a precise phase relationship according to the cycle of theta — phase precession. Sequential activity patterns in the hippocampus can, therefore, be accurately described by the cycle of internally generate theta oscillations, without the need to refer to external quantities such as displacement or elapsed time. Remember, however, that place cells survive theta inhibition.

makes a strong case, still, for the *cognitive mapping theory*. Overall, neither a purely space-centric view nor an overly generalised perspective can account for the functional properties of the hippocampus. The conservative way to approach the problem of *What does the hippocampus encode?* is to situate oneself somewhere in-between these two extremes, without gravitating too close to either camps.

Throughout this review, we turned to computational models in order to provide an intuitive and holistic understanding of the experimental observations and theories. In closing, I will describe a network model that, though much more complex than the previous formulations, does faithfully account for a large portion of the evidence described so far. Starting with a recurrent network model, as with Marr's 'simple memory', we pre-define the weights of the network in such a way as to embed a metric of distance. In Samsonovich and McNaughton (1997), this is achieved by defining the weights of each neuron to all other neurons according to a Gaussian function of the distance separating them[46]. In doing so, individual neurons in the network are now imparted with "knowledge" of their own distance away from all other neurons, similar to how any two points on a Euclidean map can be related to each other by their separation. The weight matrix between all these neurons, therefore, collectively forms a spatial chart, with coordinates as neurons and a distance metric as a Gaussian function.

Another way to conceptualise this model is by comparing it with Marr's recurrent network. In Marr's network, groups of neurons that are coincidentally activated have their synapses strengthened, in accordance with the putative auto-associations that bind discrete items composing entire episodic events. Here, these associations are formed on the basis of distances across spatial locations, wherein neurons representing adjacent locations share stronger synaptic connections than neurons representing distant points. This is equivalent to associating place cells with neighbouring place fields. It follows that the measure of distance between two neurons does not reflect the actual physical spacing between the neurons[47], but rather the relative distances between points within some abstract environment. In fact, just like the Hopfield network is capable of storing many distinct pat-

[46] More often, the Mexican Hat function is used to provide some lateral inhibition (centre-on, surround-off) (cf. Zilli, 2012).

[47] Doing so would yield a topographic network sheet organised around physical space. In practice, it is convenient to scale the topography of the spatial environment to the arrangements of the neurons, so that when visualising the states and evolution of the network, a one-to-one correspondence can be drawn between the activity of the network and the agent's location in space.

terns, multiple charts can be stored within the network's weight matrix to represent separate spatial environments.

When the parameters of this model are set just right, a single "bump" of activity is observed at any given point in time over the network sheet. This bump reflects a form of self-sustained activity perpetuated by a group of neurons sharing place fields in each others' vicinity, providing with their neighbours enough excitation to overcome some global level of inhibition. This local area of elevated activity, therefore, corresponds to the estimated current location of the agent, i.e. a set of place cells activated by the agent's entry into their place fields. This bump travels smoothly in all directions over the network sheet, tracking continuous changes in animal position and giving rise to sequential activations of place cells chained along trajectories. The location and movement of this bump is influenced by inputs onto the network. There is no limit on the type of information that can serve as input (e.g., velocity, heading direction, visual scene, etc.) (cf. McNaughton et al., 1991; Zilli, 2012). These separate sources of information converge onto the network and excite a set of candidate place cells whose locations match, to varying degrees, with the evidence presented (cf. McNaughton et al., 1991). These inputs, in combination with the current bump of activity sustained by the recurrent excitations, update the network state to formulate the next prediction of the agent's location. In other words, this network performs a form of *maximum a posteriori* estimation on the agent's location given a set of sensory/behavioural parameters:

$$\arg\max_{l} \quad f(I|l)g(l), \tag{1.1}$$

where $f(I|l)$ is the likelihood of observing some combinations of inputs I given the agent's location l and $g(l)$ is the prior probability of the agent's location, given by the self-sustained bump over Euclidean space.

This family of models, known as *continuous attractor neural networks* (CANNs), are extremely flexible and generalisable[48]. For instance, due to the varying influences of the input space on the

[48]CANNs have been exceptionally successful at modelling grid cells. In particular, a network wired in the manner described here would spontaneously give rise to a periodic pattern similar to that observed in grid cells. Because this single hexagonal pattern is supported by an entire network of neurons, if follows that all neurons within this network should share grid fields with the same scale and orientation, but shifted phases. This turned out to be exactly the case. In fact, the medial entorhinal cortex appears to support a small number (four in this case) of *grid modules* each containing groups of grid cells that shared the same scale and orientation and that operated synchronously (Stensola et al., 2012). The modules, themselves, operated independently of each other. Therefore, each grid module appear to reflect a discrete continuous attractor network.

network, the firing rates of the same place cells may vary under different sensory/behavioural conditions, accounting for rate remapping phenomena (cf. Solstad et al., 2014). Moreover, a decrease in the availability of inputs would lead to an accumulation of errors compounded over multiple steps of inaccurate model updates, consistent with experimental observations. An especially interesting prospect reveals itself when considering the metric of relative distances embedded in the network weights. In fact, the relationship between neurons need not satisfy a Cartesian coordinate system. Instead, any abstract geometry or manifold can potentially be learned by the network to describe distances/relationships in non-Euclidean space (cf. Jayasumana et al., 2013), in accordance with the relational theory. Overall, CANNs provides a decent computational framework for conceptualising both the cognitive mapping and the relational memory theories. Begin by defining a Euclidean structure into the network connectivity, then map inputs onto this network, we derive a model of the cognitive mapping theory. Trade the Cartesian system for a manifold defining distances over abstract relations (e.g., time, objects, odours, etc.), we obtain a model for the relational memory theory (see Fig. 1.4 for a working example).

Nevertheless, when a model is overly generalisable, predictions that could help situate the exact functional role of hippocampal activities cannot be formulated. Similarly, a consensus between the cognitive mapping theory and the relational theory, between the geometric and the relational interpretations, is yet to be reached, just as in the case of hippocampus's involvement in memory processing. It follows that any tangible way of reconciling the spatial and mnemonic aspects of hippocampal functions would appear even less likely. In the next section, we explore a physiological phenomenon that provides a strong bridge between space and memory and argue for its use as a promising experimental strategy to uncover the links relating the two side of the mnemonic coin.

1.3 The Mnemonic Dialogue on Space

In the two preceding sections, we independently addressed the involvements of the hippocampus in memory storage and spatial coding. Though a few allusions have been made about possible grounds over which space and memory interact (e.g., hippocampal patients are amnesiac for spatial/-contextual information; place field locations remain persistent across re-exposures to an environment), the exact relationship between these separate axes of our cognition remains to be elucidated. In fact, to date, the relevance of space in memory, or of memory in space, remains subject to much

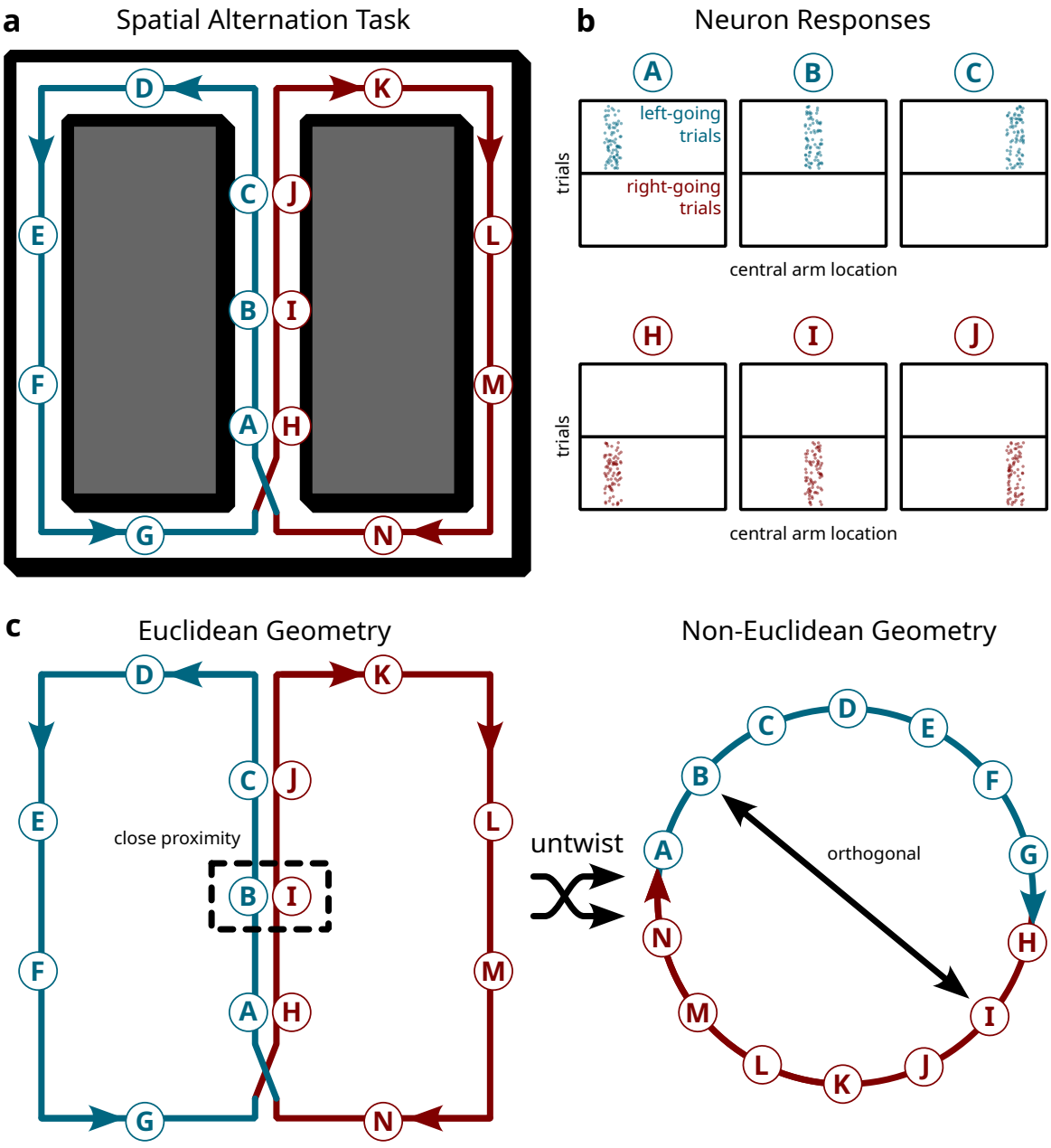

Figure 1.4: A relational theoretic explanation to "splitter cells". **a** Suppose rats perform a spatial alteration task in an eight-maze, as described in Wood et al. (2000). Neurons A-N support place fields at the labeled locations. **b** Along the central arm of the maze, neurons with overlapping place fields exhibit differential firing depending on the trial identity (left- or right-going). These differences can be accounted by rate remapping, but the relational theory offers a potential alternative explanation. **c** Within the Cartesian framework, neurons B and I share overlapping place fields. However, if the stereotypical loop that the animals perform under the task demands is "untwisted", a simple manifold in the form of a circle is obtained. Inside the geometry of this new non-Euclidean space, points that used to be close to each other are now on opposite sides. This accounts for the seemingly independent sequences A-B-C and H-I-J observed along the same central arm of the maze. According to the relational theory, distances are not defined by Euclidean geometry of physical space, but by abstract relationships, e.g., task rules.

speculation (cf. Lisman et al., 2017), with a unified theory yet to be established by an experimental breakthrough. Nevertheless, it may be argued that the discovery that came closest to such a breakthrough was that of the phenomenon of *reactivation*. In this closing section, we review the relevant literature that implicates *reactivation* as a functional correlate in the retrieval of spatial memories and in the exchange of information between the hippocampus and the cortex. This review, by no means, provides a definitive framework for reconciling the two sides of the mnemonic coin, far from it. However, it does make the case for *reactivation* as a most promising experimental strategy that may eventually lead to the much anticipated breakthrough.

Before we describe how *reactivation* was discovered, it is useful to situate this phenomenon within Marr's theoretical frameworks previously elaborated, which indeed predicted its existence. Marr's models proposed that, during awake periods, incoming sensory information or *events* are mapped onto patterns of activity diffused over the neocortex. These arbitrary patterns are associated and stored by the hippocampus, either in raw forms or in the form of 'indices'. During sleep, those patterns corresponding to individual *events* are retrieved from the hippocampus by the presentations of partial information, derived from existing representations in the neocortex, a process made possible by *pattern completion* and the *content-addressable* memory system of the hippocampal network. In this manner, new information that could potentially interfere with the existing semantic structure is gradually consolidated into neocortical representations. Given this mechanism, it should be possible in theory to observe, both in the hippocampus and in the cortex, arbitrary patterns of activity that can be linked to previous sensori-behavioural experiences. More precisely, given a pattern of activity that can statistically explain a behavioural state associated with a unique experience, if such a pattern is indeed successfully encoded, it should in principle be reinstated during offline retrievals following the behavioural experience. The experimental observation of such patterns is what is defined as *reactivation*.

As established, a great percentage of the activities in the hippocampus are functionally related to space. It was, therefore, within the spatial domain that this theorised mnemonic phenomenon was first discovered. While conducting a typical "place cells" study in rats navigating an eight-arms maze, Pavlides and Winson (1989) decided to confine the animals to one arm of the maze in which a place field was found, while completely limiting access to another arm where a simultaneously recorded place cell would respond. Surprisingly, during the sleep periods following this exposure,

the place cell associated with the accessed arm showed elevated firing rates compared to the place cell localised to the unaccessed arm. This finding provided the first experimental hint that hippocampal activities tied with recent experiences are spontaneously reinstated during offline periods. Moreover, the same tendency was observed during awake quiescent periods[49].

This single-unit demonstration of experience-dependent changes in hippocampal dynamics satisfies the prediction of there being reinstatement of activities related to recent behavioural events. However, it fails to substantiate the notion of arbitrary associations being formed around unique episodes. According to Marr's model, arbitrary patterns of activity that coincidentally occur are expected to undergo *auto-association*. Within the context of place cells, the structure of such putative associations can be statistically predicted from the geometry of an environment. In fact, if an agent were to participate in the random exploration of some space, two locations that are in close proximity to one another are more likely to be jointly visited by the agent than two locations that are separated by some distance. It follows that two place cells with adjacent place fields are more likely to activate in close succession to one another in time, hence be more likely to belong to the same episodic experiences. In other words, place cells with neighbouring place fields are more likely to be (auto-)associated, with the probability of associations decreasing smoothly with spatial distance under random exploration. This was precisely the finding reported by Wilson and McNaughton (1994). While simultaneously recording from 50-100 CA1 pyramidal cells in rats exploring open-field environments, the investigators found that pairs of place cells that expressed a stronger degree of temporal correlation during exploration retained these similarities in firing during the post-task sleep periods. Therefore, hippocampal neurons that are more likely to co-activate or to be recruited as part of the same behavioural episodes carry over these *functional connectivities* into the subsequent offline periods.

Another prediction from the recurrent network model is that associations are not only realised over space, but over time as well. Because the network retains in reverberation "memories" of activities from recent time points, arbitrary associations stored by the hippocampus are expected to retain not only instantaneously coincident patterns, but patterns of complete episodes that unravelled over some duration of time. For place cells, this insinuates at the reactivation of entire trajectories previously undertaken by an agent, which are composed of place cells that did not necessarily have any

[49]Known as *awake reactivation*. More on this later.

overlap between their place fields (i.e., show no correlation in their time-series). These sequentially organised patterns of reactivation — *replay* — were indeed observed during sleep[50], initially, in the form of a temporal delay between pairs of neurons reflecting the order of activation of place cells during track running (Skaggs and McNaughton, 1996). Soon following, more complete sequences composed of more than three cells were reported to replay following both a wheel running task (Nádasdy et al., 1999) and after natural travels over long linear tracks (Lee and Wilson, 2002). In the latter study, continuous trajectories that were replayed spanned over a metre and were composed of certain place cells with no overlapping fields and, consequently, no coincident spikes. In addition, the sequence itself was compressed 20-fold, i.e., it was replayed over a much faster timescale, in relation to the sequence observed during behaviour[51]. Therefore, patterns of activity that occur in succession can still be assimilated into the same memory trace, elaborating complete and uninterrupted episodes from previous experiences.

A most interesting feature of recurrent networks is their generative property. On the basis of the stored information, the network could spontaneously formulate fictional plans and predict future outcomes given a present set of constraints. These properties find their neural correlates in *awake replay*. Foster and Wilson (2006) noticed that, following just a single exposure to a linear track, the immediately traversed path would be replayed once the rat reached the end of the alley, still awake but pausing. Interestingly, the order of the sequence during replay was reversed from that which occurred during behaviour. Later on, it was shown that, prior to the initiation of navigation, replay tends to unfold in the forward direction, compatible with the observed behavioural sequence — *forward replay* — whereas the sequence's order would be reversed during replays at the end of the excursion — *reverse replay* (Diba and Buzsáki, 2007). These *prospective* and *retrospective* sequences, that become spontaneously organised during the offline periods surrounding behavioural

[50] When entering a place field, the probability of individual spike occurrences for the associated place cell varies as a function of the phase of theta field potentials. With each cycle of theta, the place cell fires progressively earlier in phase, known as *phase precession* (O'Keefe and Recce, 1993; Skaggs et al., 1996). This phenomenon is believed to provide a mechanism to temporally disambiguate the order in which overlapping place fields are explored by the animal (see Drieu and Zugaro, 2019, for review). Indeed, it has been shown that passive transportation inside an environment degrades theta precession and, consequently, disrupted the temporal order during sequential replay of trajectories (Drieu et al., 2018).

[51] However, the time scale of these replay events is comparable to that of *theta sequences* — place cells with overlapping fields are organised in a temporal chain within a single cycle of theta owing to the phase precession phenomenon (see Poulter et al., 2018; Drieu and Zugaro, 2019, for review). Moreover, perturbation of theta sequencing during behaviour disrupted the temporal structure of subsequent replays (Drieu et al., 2018; see also Liu et al., 2023). Therefore, theta modulation of hippocampal ensembles is a likely mechanism for establishing temporal associations across sequences of events during the encoding of memories.

experiences, are commensurate with the expected behaviours of a generative network model.

Under this framework, it should in theory be possible for awake replays to convey information regarding possible future scenarios, as a mechanism for exploring hypothetical outcomes. Pfeiffer and Foster (2013) discovered that, within an open field arena, spontaneous forward replays of trajectories closely follow the ensuing path animals take to reach a learned goal location (see also Wikenheiser and Redish, 2015). Importantly, the origin of these replay trajectories commonly intersected the current location of the animals, implicating their role in prospection and planning. In an experiment that combined two types of memory tasks (reference memory that relies on remembering a set of goal locations in an environment; and working memory that requires keeping track of all previously visited goals), researchers showed that the nature of the task heavily biased the contents of awake replay (Xu et al., 2019). In particular, the working memory task, which requires tracking of retrospective paths, saw a greater number of reverse replays to previously visited locations. In contrast, the reference memory task was dominated by forward replay trajectories to prospective goal locations. Therefore, it seems highly plausible that the hippocampus supports a generative model capable of engendering prospective and retrospective plans premised on the agents' internal goals. When animals engaged in the spontaneous foraging inside an open field, as opposed to a goal-oriented navigation, the replayed trajectories during awake quiescent periods did not depict any prospective or retrospective paths, but rather resembled Brownian diffusion or a random walk that shared no similarity with actual behaviours (Stella et al., 2019). The spontaneous organisation of hypothetical paths can therefore be goal-directed or fictitious, as expected of generative models (cf. Hinton et al., 2006) and can, indeed, contribute a mechanism for *path integration* (see previous section).

An extreme exemplar of the generative property of the hippocampus is found in *preplay*. Dragoi and Tonegawa (2011, 2013); Ólafsdóttir et al. (2015) observed that, even before mice gained any exposure to an environment or to a portion of the environment, trajectory sequences would be spontaneously replayed (preplayed) to elaborate a novel set of place cells tiling the to-be-discovered locations (see also Mizuseki and Buzsáki, 2013). These results would suggest that hippocampal charts are determined *a priori* to encompass any arbitrary spatial contexts (cf. Samsonovich and McNaughton, 1997). The existence of such *pre-configured maps* does, however, seem implausible; arguing from the perspective of continuous attractor networks, for the network to start tracking

animal locations using motion cues without the need for any *de novo* learning, some unknown sophisticated mechanism must be in place to preset the network weights to their optimal values (cf. Zilli, 2012, for review). Extending this thought further, the idea that the hippocampus, as a memory system, is capable of synthesising contents that extrapolate beyond the rudimentary pieces of information that it has in store, does not seem likely either. In fact, an independent research group has failed to replicate the previous results and proposed that demonstrations of preplay may be a result of statistical artefacts or biases introduced by experimental designs (Silva et al., 2015). The extent to which the generative properties of the hippocampal network lends itself to the composition of fictitious materials and to the *pattern completion* of discontiguous events over space remains to be further explored.

The theorised goal of hippocampal reactivation/replay, as delineated by Marr's framework, is to enable the reinstatement of a set of cortical patterns conjointly expressed during previous episodic experiences. Conversely, the initiation of hippocampal reactivation, by pattern completion, is hypothesised to rely on the spontaneous conveyance of partial cues by the cortex, derived from existing semantic representations. It therefore stands to reason that reactivation would be similarly observed in the cortex. Indeed, a series of studies have demonstrated the reactivation of recent patterns in a diversity of cortical regions. Usually, the information conveyed during these offline events could be tied to the known functions and the neuro-behavioural correlates of the respective areas. For instance, the medial prefrontal cortex reactivated for task rules (Peyrache et al., 2009), the posterior parietal cortex for egocentric parameters (Wilber et al., 2017), the auditory cortex for sound tones (Rothschild et al., 2017) and the visual association cortex for grating orientations (Sugden et al., 2020). This conformity between online functions and offline mnemonic contents is of great theoretical relevance. For offline cortical patterns to serve as partial retrieval cues, these patterns should in theory be supported by existing (semantic) representations in the cortex. These region-specific patterns of reactivation could, therefore, be composed of a mixture of spontaneous activations of locally-encoded semantic features[52] and individual patterns elicited by the hippocam-

[52]There is however a huge problem with this conjecture. In all aforementioned studies, cortical reactivation typically follows from hippocampal *sharp-wave ripple* events. If these cortical reactivations were to subserve partial retrieval cues, this temporal relationship should be inverted instead. One possibility holds that because, at any given moment, a local region of the cortex is more likely to be recruited in the global retrieval of some experience, rather than to be the initiator of a retrieval event, a positive delay would be observed on average between hippocampal and cortical activations (cf. Rothschild et al., 2017).

pus during globally coherent retrievals of distributed cortical representations that collectively define complete episodic experiences. In fact, a study that simultaneously recorded ensembles of neurons from four distinct cortical areas in a macaque found that reactivations were coordinated across all regions, suggesting that offline retrieval involves distributed cortical sites (Hoffman and McNaughton, 2002).

The contents of cortical reactivations, however, do not always follow from the functions of the region. In some studies, cortical reactivations were linked to previously visited spatial locations (Qin et al., 1997; Ji and Wilson, 2007; Euston et al., 2007; Jadhav et al., 2016). These studies not only hinted strongly at the possible existence of spatial coding in extra-hippocampal regions, but elucidated a universal functional code for the mnemonic exchange between the hippocampus and the cortex. In particular, Ji and Wilson (2007) described sequential replay of spatial trajectories in the visual cortex, similar to those observed in the hippocampus. Importantly, these events were coordinated with hippocampal replays, portraying the same behavioural sequences (see also Qin et al., 1997). It therefore appears that the hippocampus and cortex engage in coherent reinstatements of the same behavioural experiences, as postulated by Marr and the indexing theory of consolidation.

In summary, the body of works reviewed here underline the study of reactivation as a promising research direction for reconciling the dichotomy of space and memory. In the hippocampus, sets of place cells spontaneously organise into firing chains, reinstating sequences of previously explored trajectories. In the cortex, we equally find the reactivation of previous spatial experiences, the occurrence of which is synchronised with hippocampal replay of similar behavioural episodes. Overall, replay phenomena confirms many predictions from the computational models on memory networks, including the generative properties of the hippocampal recurrent network and the linking of distributed patterns in the cortex into holistic episodic traces stored by the hippocampus. It is, therefore, within the context of this "mnemonic dialogue on space" that we situate the purpose of the present thesis.

1.4 Thesis outline

There are three key pieces of information to be extracted from this general review:

1. The hippocampus and the neocortex constitute two parallel and complementary learning systems;

2. The hippocampal functional code, however counter-intuitive, is one that is predominantly explained by the mapping of physical space; however, a significant portion of the variance in spatial firing is accounted for by rate remapping — i.e., by what external and internal events occur at a given place;

3. During offline periods, patterns of activity that are linked with recent experiences, which may or may not convey spatial attributes, are spontaneously reinstated in both the cortex and the hippocampus.

Within this framework, two exceptional studies have provided an experimental demonstration for how all three of these theoretical concepts come together synergistically. In Qin et al. (1997) and Ji and Wilson (2007), it was shown that certain cortical regions (in the posterior parietal and visual cortices, respectively) also expressed place coding in certain capacities. In the posterior parietal cortex, cells fired at discrete but identical locations across arms of a maze (cf. Nitz, 2012; Wilber et al., 2017), while the visual cortex neurons exhibited less specific place fields. Moreover, in both studies, cortical ensembles reactivated cohesively with hippocampal neurons for the same spatial locations/trajectories. These studies offered a preliminary insight into how the hippocampal place code may be communicated with the cortex, presumably for the purpose of memory consolidation.

The present thesis aims to expand upon these promising discoveries by testing for the generalisability of these results across different cortical areas and by further elaborating upon the nature of the information reactivated by the cortex. I must confide that the findings reported in Chapter 3 and 4 were completely serendipitous. Therefore, the hypotheses associated with them will be stated here *a posteriori*. Nevertheless, consider the following propositions:

Chapter 2: *There is evidence to believe that spatial coding may be found across a diversity of cortical areas. There are also reasons to believe that the network architecture of the hippocampus affords it a unique capacity to encode space, as elaborated in this General Introduction. Therefore, we hypothesize that many areas of the dorsal cortex, including primary/secondary motor, somatosensory, posterior parietal and retrosplenial cortices would exhibit spatially correlated responses during online navigation. Furthermore, extensive destruction of the hippocampus would lead to an impairment in these representations, owing to the possibility that they are modulated by the hippocampal place code.*

Chapter 3: *Studies suggest that the cortex reactivates in conjunction with the hippocampus to reinstate spatial contents from recent experiences. However, extra-spatial features related to the distinct functions of specific cortical regions have also been found to reactivate. Given that the retrosplenial cortex receives direct projections from the dorsal hippocampus (Monaghan and Cotman, 1985), it is possible that this area may be exceptionally prone to reactivating spatial information in a similar way as the hippocampus.*

Chapter 4: *In contrast to the retrosplenial cortex, the secondary motor cortex is situated farther away from the hippocampus synaptically. Therefore, we expect the reinstatement of patterns related to motor sequences (cf. Eckert et al., 2020) and less so to spatial experiences.*

In the following chapters, I will describe the experiments that were conducted to answer each hypothesis and discuss their findings in isolation. I will argue that hypothesis 1 has been confirmed, while hypotheses 2 and 3 held, curiously, not for themselves, but for each other (i.e., invert the regions and the hypotheses hold). In the General Discussion, I offer one possible, somewhat speculative, interpretation for what these findings may represent for the broader framework of *learning and memory*.

Chapter 2

Spatial Information Encoding across Multiple Neocortical Regions Depends on an Intact Hippocampus[53,54]

Abstract

There has been considerable research showing populations of neurons encoding for different aspects of space in the brain. Recently, several studies using two-photon calcium imaging and virtual navigation have identified "spatially" modulated neurons in the posterior cortex. We enquire here whether the presence of such spatial representations may be a cortex-wide phenomenon and, if so, whether these representations can be organized in the absence of the hippocampus. To this end, we imaged the dorsal cortex of mice running on a treadmill populated with tactile cues. A high percentage (40–80%) of the detected neurons exhibited sparse, spatially localized activity, with activity fields uniformly localized over the track. The development of this location specificity was impaired by hippocampal damage. Thus, there is a substantial population of neurons distributed widely over the cortex that collectively form a continuous representation of the explored environment, and hippocampal outflow is necessary to organize this phenomenon.

2.1 Introduction

The neural mechanisms underlying the emergence of place coding continue to elude complete understanding. In the hippocampal and parahippocampal regions reside large populations of neurons displaying specialized spatial firing patterns (Moser et al., 2015). Of particular interest are hippocampal "place cells," which share the same anatomic substrate as episodic memory functions

[53] Reproduced from Esteves et al. (2021) with authors' permission.

[54] **Author contributions** H.C., I.M.E., M.H.M. and B.L.M. designed the experiment. H.C., I.M.E. and A.R.N. conducted the experiments and collected the data. H.C. and I.M.E. performed the surgeries. H.C. analysed the data and wrote the manuscript, which all authors helped to revised.

(O'Keefe and Dostrovsky, 1971). Hippocampal place cells fire at one or a few specific locations in an environment and collectively are believed to form the neural basis of the cognitive representation of space (O'Keefe and Conway, 1978). It is now recognized that once "allocated" to a place, mainly through self-motion-related processes (McNaughton et al., 2006), these neurons become modulated by other attributes of the current state of the brain (e.g., sensory input, motor actions, plans, affective state), a phenomenon dubbed "rate remapping" Leutgeb et al. (2005); Sheintuch et al. (2020).

The response characteristics of place cells are indicative of complex integrations of information spanning multiple sensory modalities and levels of cortical processing. Studies have shown that place cell activity is modulated by a broad range of environmental features such as visual (O'Keefe and Conway, 1978; Muller and Kubie, 1987; O'Keefe and Speakman, 1987; Quirk et al., 1990), olfactory (Save et al., 2000), auditory (Moita et al., 2003), and tactile information (Gener et al., 2013). Hippocampal place cells can be driven by spatial orienting cues in a novel environment and still fire in the correct locations in the dark or in the absence of cues once the environment becomes familiar, supporting the notion that cortical–hippocampal interactions are involved in the formation of associations among places, movements, and cues established during the learning of an environment (O'Keefe and Speakman, 1987; McNaughton et al., 1989).

Although place cells have predominantly been described in the hippocampal formation, a number of recent studies have reported neurons with similar firing patterns as place cells in multiple neocortical regions (Nitz, 2009; Harvey et al., 2012; Fiser et al., 2016; Mao et al., 2017; Pakan et al., 2018; Saleem et al., 2018; Minderer et al., 2019). The parietal cortex presents cells with spatially modulated activity patterns important to solving navigational tasks and also presents population activity with specific choice trajectories (Nitz, 2006; Harvey et al., 2012; Wilber et al., 2014). Visual responses have been shown to be modulated by spatial location and controlled by navigational signals (Saleem et al., 2018). Additionally, retrosplenial cortex neurons have been shown to encode spatial activity sequences resembling the activity of hippocampal CA1 place cells, and the formation of these spatial sequences relies on an intact hippocampus (Mao et al., 2017, 2018; Chang et al., 2020).

Given the potential interplay between the hippocampus and neocortex at large, we surveyed and compared multiple dorsal neocortical regions (including primary, secondary, and associational areas) to investigate the extent to which other neocortical regions also express continuous spatial

representations. With targeted bilateral excitotoxic lesions to the dorsal hippocampus, we examined the degree to which these representations relied on an intact hippocampus.

2.2 Materials and Methods

2.2.1 Experimental design

Thy1-GCaMP mice were implanted with a 5 mm cranial window above the dorsal cortex. Animals were head restrained and trained to run for reward over a 150-cm-long treadmill belt on which were mounted several tactile cues. Two-photon calcium imaging was conducted across different neocortical regions in animals with and without excitotoxic hippocampal lesions. We tracked the spatial activity across different cortical areas and evaluated how the spatial representation found in the neocortex was overall influenced by the hippocampus.

2.2.2 Animals

A total of eight adult transgenic Thy1-GCaMP6s mice (weight, 22–30 g; age, 2–4 months), specifically expressing the calcium indicator in excitatory neurons, were used for this study. Mice were housed in standard rodent cages, maintained at 24°C room temperature and under a 12 h light/dark cycle (lights on at 7:30 A.M.), with free access to food and water before the beginning of training. All experiments were performed between 7:30 A.M. and 7:30 P.M. Procedures were conducted in accordance with the guidelines established by the Canadian Council on Animal Care, and using protocols approved by the Animal Welfare Committee of the University of Lethbridge.

2.2.3 Surgery

Before surgery, mice were injected with 0.5 ml of a 5% dextrose and 0.9% saline solution mixed with atropine (3 µg/ml, s.c.), buprenorphine (0.05 mg/kg, s.c.), and dexamethasone (0.2 mg/kg, i.m.). Animals were then anesthetized with isoflurane (1–1.5%, O2; 0.5–1 L/min) and placed in a stereotactic frame with body temperature maintained at 37.0 ± 0.5°C through a heating pad with a closed-loop control system. Once the skull was exposed, a custom-made titanium headplate was fixed to the skull using adhesive cement (C&B Metabond, Parkell) and dental acrylic. Bregma was identified and marked in the cement. A 5-mm-diameter craniotomy was made following bregma-referenced coordinates: 1.5 mm anterior to -3.5 mm posterior; ±2.5 mm medial–lateral (Fig. 2.1A).

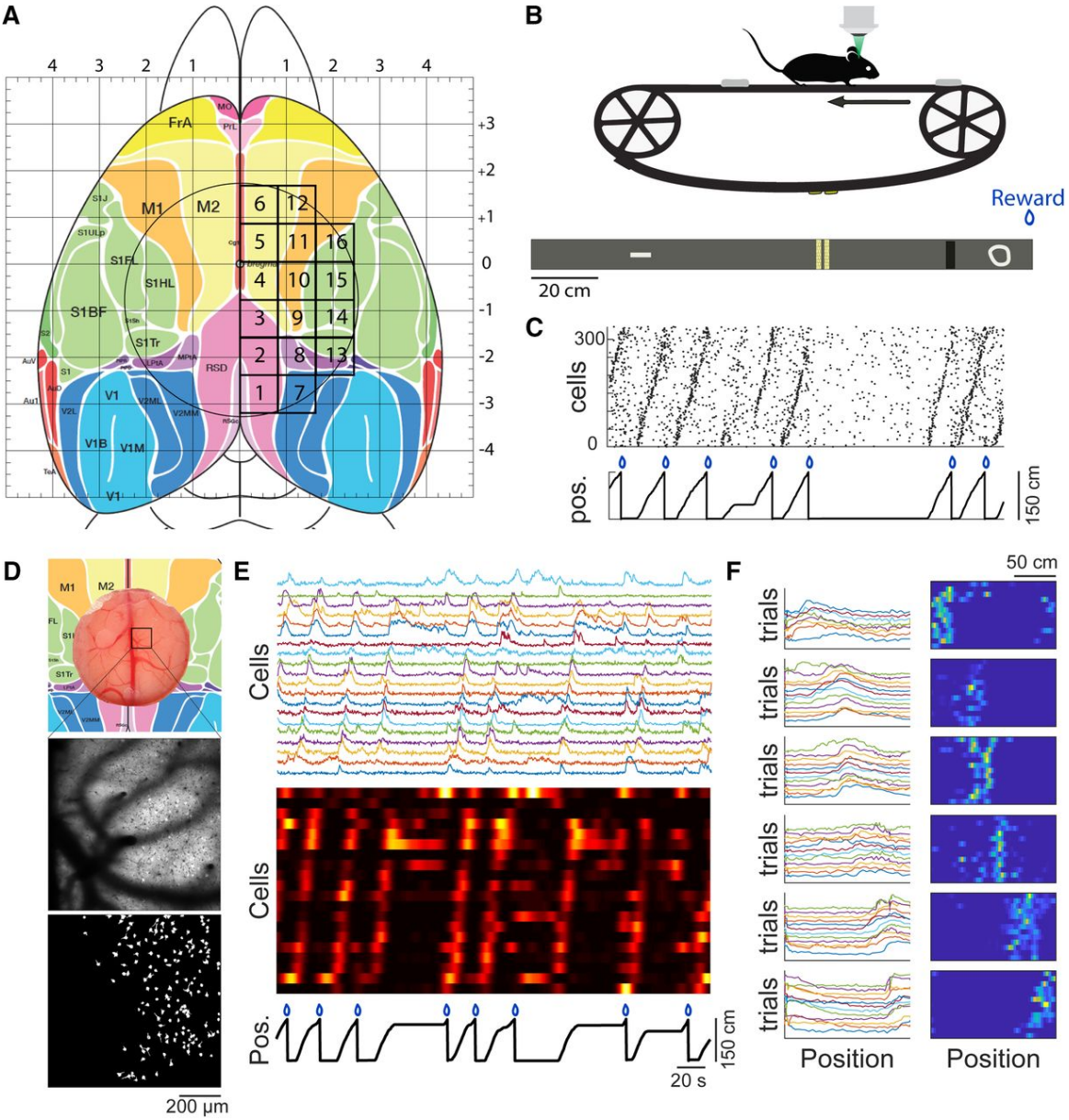

Figure 2.1: Cranial window implant, treadmill track, and two-photon Ca^{2+} imaging. A, Cranial window implant position (circle) with all the regions of neocortex imaged (black boxes; adapted from Kirkcaldie et al., 2012). B, Top, Basic setup for recording on treadmill track using two-photon Ca^{2+} imaging. Bottom, Diagram of belt lined with the tactile cues. C, Sequential activity of neocortical (RSC) spatially selective cells during running on the tactile belt. Neurons were ordered by their peak response positions. Black trace shows the position, and blue drops indicate reward. D, Cranial window of one animal (top), field of view of one imaging session (middle), and identified neurons (ROIs) from one session (bottom). E, Plot of the normalized, raw calcium $\Delta F/F$ time courses (top) and deconvolved trace smoothed using a $\sigma = 8$ s. Gaussian kernel (bottom) of 20 simultaneously imaged neurons in M2. Animal position and reward are shown below. F, Example of normalized activity of the six spatially selective cells as a function of location for multiple laps (left, raw trace; right, deconvolved). Pos., Position.

Three layers of coverslips affixed with optical adhesive (NOA71, Norland) were implanted over the craniotomy, attached to the skull with Vetbond (3M). A rubber ring was fixed over the headplate to form a well to hold water between the imaging region and the immersion objective. Surgeries were concluded with injections of meloxicam (Metacam; 1.0 mg/kg, s.c.) and enrofloxacin (Baytril; 10 mg/kg, s.c.). The same drugs were administered once daily for 3 d. Mice were then allowed to recover for 1 week. For bilateral hippocampal excitotoxic lesions, NMDA (15 mg/ml, in 1× PBS) was injected before implanting the coverslip. Injections were made with a micropipette, with tip diameter between 20 and 30 µm, loaded on a nanoliter injector (Nanoject II, Drummond Scientific), at a speed of 9.2 nl/pulse over eight pulses (in total: 9.2 nl/pulse × 8 pulses = 73.6 nl; interpulse interval, 15 s) at each site. In total, the following four injection sites were used (two in each hemisphere): anteroposterior (AP), -2.3 mm; mediolateral (ML), 1.5 mm; dorsoventral (DV), 1.8 mm; and, AP, -3.2 mm; ML, 2.5 mm; DV, 2.0 mm. Diazepam (5 mg/kg, i.p.) and phenobarbital (30 mg/kg, s.c.) were administered right after surgery to suppress potential seizures.

2.2.4 Treadmill apparatus

Once recovered, mice were water restricted throughout training and imaging sessions. During the water restriction period, mice were given ad libitum access to water for a maximum of 30 min/d in their home cages after training/imaging sessions, and their weight was monitored to be maintained to at least 85% of their baseline weight (average weight 3 d before water restriction). Animals were gradually accustomed to run on the treadmill while head fixed. The treadmill consisted of a 150 cm Velcro (Country Brook) belt with tactile cues (made of hot-glue stripes, reflective tape, and Velcro) inserted at several locations (Fig. 2.1B). Polyamide wheels (diameter, 10 cm) were attached to both ends of the treadmill to guide the belt, and an optical encoder (Avago Tech) attached to the wheel shaft was used to monitor belt movement. After each lap, a drop of sucrose water (10% concentration; ∼2.5 µl volume) was delivered to the animal. The reward was dispensed by a solenoid pinch valve (Bio-Chem Laboratories) whenever a photoelectric sensor (Omron) was triggered by the reflective tape attached to the opposite side of the belt. A custom-designed circuit with a microcontroller (Arduino UNO, Farnell) was used to control the reward delivery and to monitor the encoder.

2.2.5 Two-photon imaging

After training, imaging experiments were conducted using a Bergamo II Multiphoton Microscope (THORLABS). Samples were excited by a Ti:Sapphire pulsed laser (Coherent) tuned to a wavelength of 920 nm (∼80 mW output power measured at the sample). Focusing of the beam and light collection were achieved by a 16× water-immersion objective lens (numerical aperture, 0.8; Nikon). Scanning was conducted by Galvo-Resonant X-Y mirrors. Emitted fluorescent lights were detected by a GaAsP photomultiplier tube (Hamamatsu). Samples were acquired from a 835 × 835 μm field of view (FOV) at depths between 130 and 190 μm (layers II/III). Images were digitized at a sampling rate of 19 Hz, and at a resolution of 800 × 800 pixels. Photosensor and encoder signals from the treadmill were acquired and synchronized with imaging using Clampex software (Molecular Devices).

Imaging data from all animals were acquired from the right hemisphere. The hemisphere was divided into 16 regions partitioned by grid (Fig. 2.1A). Each region spanned the same dimensions as the imaging FOV (Fig. 2.1D). In total, 16 recording sessions were conducted per animal, one for each region, and each recording was 5–10 min in duration. The imaging experiments were performed over the course of 1 week, after at least 1 month of familiarization on the track/belt. The cranial window allowed us to image five different neocortical regions. For data analysis, we considered grids 1–3 as restrosplenial cortex (RSC); grids 4–6 as secondary motor cortex (M2); grids 10 and 11 as primary motor cortex (M1); grids 8 and 13 as posterior parietal cortex (PPC); and grids 14 and 15 as primary somatosensory cortex (SS1; Fig. 2.1A).

2.2.6 Data analysis

Image preprocessing was conducted automatically using the Suite-2P software suite (Pachitariu et al., 2016), as previously described (Mao et al., 2017, 2018). Regions of interest (ROIs) detected as candidate neurons were visually inspected. Neuropil contamination was estimated from the surround of ROIs and was subtracted (Bonin et al., 2011). To infer relative underlying firing rates for each ROI, the ΔF/F time courses were deconvolved using constrained non-negative matrix factorization (Pnevmatikakis et al., 2016). All analyses were conducted using MATLAB (R2017a, MathWorks). All subsequent analyses were conducted on the deconvolved time courses using data normalized between 0 and 1, unless stated otherwise (Fig. 2.1E).

Spatially selective cells were identified based on the following criteria: the length of the belt was divided into 50 equally spaced bins (3 cm each). The activity of each neuron was mapped to corresponding position bins for each lap. This position-mapped activity was normalized by the time spent in each bin (occupancy) and then filtered with a Gaussian window (4.5 cm SD). Spatial information (SI) was calculated for each cell using the following formula (Skaggs et al., 1993):

$$\mathrm{SI} = \sum_{i=1}^{N} p_i \frac{f_i}{f} \log_2(\frac{f_i}{f}), \tag{2.1}$$

where N is the total number of bins, p_i is the occupancy probability in the i^{th} bin, f_i is the activity in the i^{th} bin, and f is the overall activity (averaged f_i across all bins). We generated a shuffled distribution of spatial information by circularly shifting the time courses 1000 times. To be classified as significantly spatially tuned, the original SI was required to be higher than the 95^{th} percentile of the shuffled distribution, the mean activity within a place field (PF) was required to be 2.5 times higher than the mean activity outside of field, and the peak activity was required to occur inside the place field in at least a third of the laps (Fig. 2.1F). Note that SI is a conservative measure. For example, many hippocampal cells have several clear place fields and hence show less spatial information overall. The spatial tuning characteristics of place fields were derived by obtaining a continuous wavelet transform, W, over the spatial response curve f_i of each neuron with a Mexican Hat mother wavelet, Ψ, as follows:

$$W(\sigma,\tau) = \frac{1}{\sqrt{\sigma}} \sum_{i=1}^{N} f_i \Psi(\frac{1-\tau}{\sigma}), \tag{2.2}$$

where

$$\Psi(t) = \frac{12}{\pi^{\frac{1}{4}}\sqrt{3}}(1-t^2)\exp\frac{-t^2}{2}. \tag{2.3}$$

The local maxima in the resulting spectrum indicate the location (given by τ) and the width (given by σ) of place fields. Peaks with values lower than three median absolute deviations from the median of the wavelet coefficients at the lowest scale ($\sigma = 1$) were rejected. Peaks that fell within the receptive field of a local maximum at a higher scale σ were rejected. The mean activity within a place field was required to be 2.5 times higher than the mean activity outside of the field. The peak activity was required to occur inside the place field in at least a third of the laps (Fig. 2.1F). Place fields must be

wider than 5% of the length of the environment, but narrower than 80%. Sparsity, a measure of how dispersed the firing profile of a neuron is in relation to the environment, was calculated as follows:

$$\text{sparsity} = \frac{(\sum_{i=1}^{N} p_i f_i)^2}{\sum_{i=1}^{N} p_i f_i^2}, \tag{2.4}$$

where f_i is the mean activity in the i^{th} spatial bin and p_i is the probability of occupancy in the i^{th} spatial bin over a total of $N = 50$ spatial bins (Jung et al., 1994). Sparsity ranges between 0 and 1, where smaller values indicate finer spatial tuning profiles.

Population activity was decoded using an independent Bayesian decoder (Zhang et al., 1998; Mao et al., 2018). In brief, for every time bin, we estimated the probability of the animal being at a position x given the population response of all imaged neurons as follows:

$$\mathrm{P}(x|n) = C \prod_{i=1}^{N} f_i(x)^{n_i} \exp -\tau \sum_{i=1}^{N} f_i(x), \tag{2.5}$$

where $f_i(x)$ is the mean deconvolved fluorescence trace over position x and n_i is the time course vector of the i^{th} neuron within a time bin of length τ, N is the number of neurons, and C is a normalizing constant that makes the probability distribution across all positions sum to 1. Running periods of odd trials were used for training the model, and even trials were used to evaluate the error of decoding. A decoded position was defined as the position with the highest probability for any given time bin, and the absolute value of the difference between true position and decoded position was defined as the Bayesian decoding error. The direct implementation of a Bayesian decoder based on a Poisson likelihood distribution, for use with deconvolved calcium traces, poses a limitation. Indeed, we cannot assume that the underlying distribution of deconvolved firing rates is drawn from Poisson point processes. However, considering that there is a marked linear relationship between calcium-inferred firing rates and true firing rates Pnevmatikakis et al. (2016), the estimator used in the present study is still valid as a close approximation. We expect improved and standardized methodologies to become available in future studies with the maturation of calcium imaging techniques.

The anatomic density distribution map was obtained by discretizing the entire window into 120 × 60 grids of 20 × 20 pixels (435.76 μm^2) and then calculating the number of cells in each grid. The density distribution map of spatially selective cells was obtained by dividing the number of spatially

tuned cells by the total number of cells detected in each grid. Data were smoothed using a $\sigma = 2$ pixel Gaussian kernel. Maps were aligned by locating bregma (marked outside the window during surgery) in the upper part of grid 4 (see Fig. 4A).

Table 2.1: Summary of all statistics

Figure	Variable	*n*	Test	Results
3C, right	Fraction of spatially selective cells per group	Value per animal; n = 4 each group	t test	All values are mean ± SEM Control, 0.60 ± 0.02; lesion, 0.26 ± 0.03; p <0.001

3C, left	Fraction of spatially selective cells per region	Value per session recorded in each region
		
		RSC: control = 11;
		lesion = 9;
		PPC: control = 7;
		lesion = 5;
		M2: control = 10;
		lesion = 8;
		M1: control = 8;
		lesion = 8;
		SS1: control = 8;
		lesion = 8;
			Wilcoxon rank-sum	All values are median (interquartile range)
				
				control, 0.52 (0.63–0.47);
				lesion, 0.16 (0.29–0.14); p = 0.0008
				control, 0.42 (0.45–0.36);
				lesion, 0.23 (0.26–0.19); p = 0.048
				control, 0.73 (0.83–0.70);
				lesion, 0.25 (0.48–0.05); p = 0.0003
				control, 0.72 (0.76–0.61);
				lesion, 0.36 (0.44–0.25); p = 0.0003
				control, 0.56 (0.64–0.47);
				lesion, 0.27 (0.37–0.18); p = 0.0002

3D, left	Running speed	Value per animal	t test	All values are mean ± SEM Control, 18.4 ± 2.3 cm/s; lesion, 14.5 ± 2.1 cm/s; p = 0.249
3D, right	Number of laps	Value per session ecorded in all region control = 44; lesion = 38	Wilcoxon rank-sum	All values are median (interquartile range) Control, 20.5 (17.0–25.0); lesion 18.0 (14.0–24.0); p = 0.2279
4C	Cells detected per region per group	Value per animal; n = 4 each group	Wilcoxon rank-sum	All values are median (interquartile range) RSC: control, 313.0 (407.0–243.5); lesion, 403.5 (448.0–389.5); p = 0.3429 PPC: control, 125.5 (163.0–59.5); lesion, 53.0 (80.0–30.0); p = 0.2000 M2: control, 254.0 (280.5–179.5); lesion, 134.5 (252.5–94.0); p = 0.4857 M1: control, 208.5 (278.5–173.0); lesion, 321.5 (467.0–211.0); p = 0.3429 SS1: control, 154.5 (174.0–140.5); lesion, 203.5 (262.5–156.0); p = 0.3429
5B, right	Place field width, spatial information and sparsity per group	Value per animal n = 4 each group	t test	All values are mean ± SEM: Place field width: control, 43.0 ± 2.3; lesion, 42.38 ± 2.75; p = 0.876 Spatial information: control, 1.69 ± 0.02; lesion, 1.176 ± 0.11; p = 0.004 Sparsity: control, 0.239 ± 0.005; lesion, 0.349 ± 0.025; p = 0.005

5B, left	Place field width, spatial information and sparsity of the spatially selective cells per region	Value per session recorded in each region RSC: control = 11; lesion = 8; PPC: control = 7; lesion = 5; M2: control = 10; lesion = 7; M1: control = 8; lesion = 7; SS1: control = 8; lesion = 8;	Wilcoxon rank-sum	All values are median (interquartile range) Place field width: RSC: control, 48.77 (57.47–44.18); lesion, 43.69 (48.11–42.15); p = 0.07 PPC: control, 46.94 (50.40–43.58); lesion, 44.4 (63.17–40.62); p = 1.00 M2: control, 41.62 (48.43–38.67); lesion, 40.77 (53.87–34.23); p = 0.81 M1: control, 38.09 (41.73–36.33); lesion, 38.88 (44.15–34.90); p = 0.95 SS1: control, 36.64 (39.24–34.40); lesion, 40.53 (45.91–38.31); p = 0.13 Spatial information: RSC: control, 1.34 (1.53–1.25); lesion, 0.80 (1.10–0.73); p = 0.025 PPC: control, 1.49 (1.53–1.36); lesion, 0.90 (1.10–0.82); p = 0.025 M2: control, 1.79 (1.94–1.59); lesion, 1.07 (1.32–0.88); p = 0.0001 M1: control, 1.84, (1.97–1.73); lesion, 1.49 (1.66–1.24); p = 0.0093 SS1: control, 1.76 (1.93–1.71); lesion, 1.15 (1.40–1.05); p = 0.0006 Sparsity: RSC: control, 0.29 (0.31–0.25); lesion, 0.43 (0.47–0.36); p = 0.0025 PPC: control, 0.26 (0.29–0.36); lesion, 0.41 (0.43–0.34); p = 0.0101 M2: control, 0.21 (0.25–0.19); lesion, 0.36 (0.42–0.30); p = 0.0001 M1: control, 0.20 (0.22–0.19); lesion, 0.27 (0.32–0.24); p = 0.0022 SS1: control, 0.22 (0.23–0.19); lesion, 0.34 (0.35–0.29); p = 0.0002
6B, Bottom	Peak activity	Value per animal n = 4 each group	t test	All values mean ± SEM Control, 3.932 ± 0.407; lesion, 2.987 ± 0.276; p = 0.011

8B	Decoding error as function of position	Value per animal n = 4 each group	Two-way ANOVA	Source of variance Group: df = 1; SS = 15,457.119; MS = 15,457.119; F = 2694.449; p <0.001 Position: df = 49; SS = 3876.430; MS = 79.111; F = 13.790; p <0.001 Group × position: df = 49; SS = 1725.192; MS = 35.208; F = 6.137; p <0.001 Residual: df = 300; SS = 1720.996; F = 5.737 Total: df = 399; SS = 22,779.738; F = 57.092
8C	Decoding error for each position	Value per group control = 2200; lesion = 1850	KS test	Control vs lesion, p <0.0001
8D	Decoding error per region	Value per animal n = 4 each group	t test	All values mean ± SEM RSC: control, 10.01 ± 1.33; lesion, 23.1 ± 2.17; p = 0.0022 PPC: control, 16.43 ± 0.64; lesion, 22.96 ± 0.63; p = 0.0004 M2: control, 6.48 ± 1.05; lesion, 19.78 ± 3.25; p = 0.0081 M1: control, 6.44 ± 0.98; lesion, 18.38 ± 2.61; p = 0.0053 SS1: control, 7.61 ± 0.93, lesion 20.19 ± 1.40; p = 0.0003

2.2.7 Histology

Mice were perfused with PBS, and brains were postfixed with 4% paraformaldehyde for 24 h after the end of the experiments. Brains were then cryoprotected in 30% sucrose solution (with 0.02% sodium azide) and sectioned in the coronal plane at 40 µm using a blockface imaging system composed of a sliding microtome and an Olympus MVX10 microscope. We used a NanoZoomer scanning microscope (Hamamatsu Photonics) to acquire images of sections. Images of coronal sections from -0.8 to -3.8 mm AP (Paxinos and Franklin, 2003) were used to evaluate the extent of hippocampal lesions.

2.2.8 Statistical analysis

All statistical tests performed in this work were conducted using MATLAB functions (catalog #R2017a, MathWorks). Further details of all statistical tests implemented in this study are provided in Table 2.1.

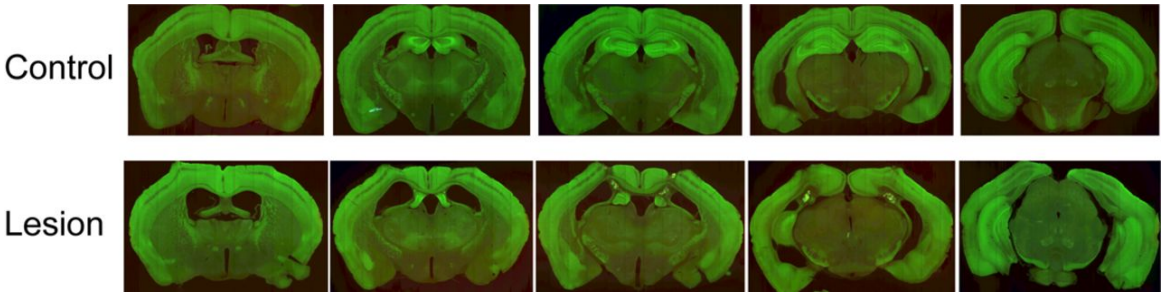

Figure 2.2: Histology. Example histology of one representative mouse from each group. Note the tissue loss in the dorsal hippocampal region in the lesion group.

2.3 Results

Using two-photon Ca^{2+} imaging, we systematically recorded neurons from the superficial cortical surface (layers II/III) of the right hemisphere exposed under a 5 mm craniotomy in four mice in a head-fixed, treadmill assay (Fig. 2.1A,B). We found neurons highly tuned to the positions of animals on the belt, as illustrated in Figure 2.1C. Ca^{2+} imaging was also conducted on a second group of four mice with dorsal hippocampal lesions (bilateral) to evaluate whether the spatial representations were dependent on an intact hippocampus. The animals in the lesion group sustained extensive neuron/tissue loss in the dorsal hippocampal formation (Fig. 2.2). Running speed was similar across groups (Fig. 2.3D, left), and no significant difference was found between the number of laps performed by each group (Fig. 2.3D, right).

2.3.1 Spatially selective cells are widespread in the neocortex and are dependent on an intact hippocampus

We found neurons with spatial selectivity in all neocortical areas examined, as shown in Figure 2.3, A and B, left. However, in the lesion group, the fraction of cells expressing high spatial information content relative to position on the belt was greatly reduced (Fig. 2.3A,B, right). Averaging the cell fractions over all ROIs, the control animals presented a fraction of spatially selective cells that was 2.3 times higher than in the lesion group (Fig. 2.3C, right). By comparing this fraction per region, the median from the control group was significantly higher than that of the lesion group in all neocortical regions (Fig. 2.3C, left). These results show that cells with spatial selectivity were present in all neocortical regions examined, and that the emergence of those properties was disrupted by hippocampal lesions. Cell density maps for each animal are shown in Figure 2.4A. Among the

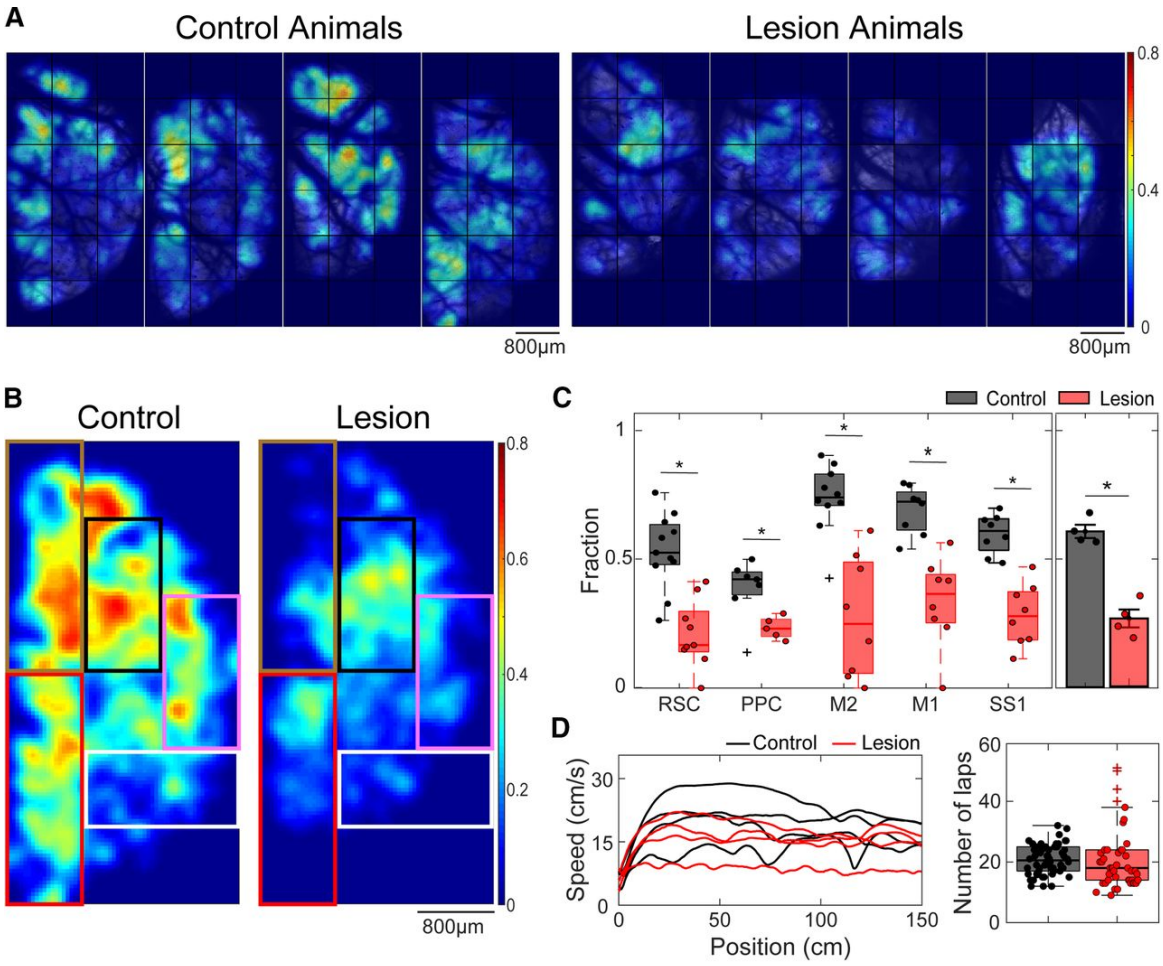

Figure 2.3: Distribution of spatially selective cells in the neocortex and its hippocampal dependence. A, B, Density distribution map with all the spatially modulated neurons of each animal and group, showing that cells expressing spatial coding characteristics were found in multiple areas of the neocortex and that hippocampal lesions disrupted the spatial coding in those regions. Color bar shows the fraction of cells that passed the criteria for being spatially selective cells inside an area of 20 × 20 pixels (435.76 μm2); Colored boxes refer to the different neocortical regions imaged (RSC, red; M2, brown; M1, black; PPC, white; SS1, magenta). C, Fraction of neurons with spatial selectivity for each area (left: line, median; box, 25th and 75th percentiles; dots, cell fraction for an individual session recorded in each region from all animals; whiskers, minimum and maximum values; $*p < 0.05$) and per group (right: mean ± SEM over animals; dots, cell fraction for individual animals; n = 4 in each group; $*p < 0.05$). There was a significant reduction of spatially selective cells after lesioning for all regions as well as for the regions pooled. D, Left, Speed profiles as a function of position from all trials. D, Right, Number of laps performed by each group during all sessions. Line, median; box, 25th and 75th percentiles; dots, values for an individual session recorded for all animals; whiskers, minimum and maximum values; + signs, outliers. No significant difference was found between the speed and the number of laps performed by each group. For exact n and p values, see Table 2.1.

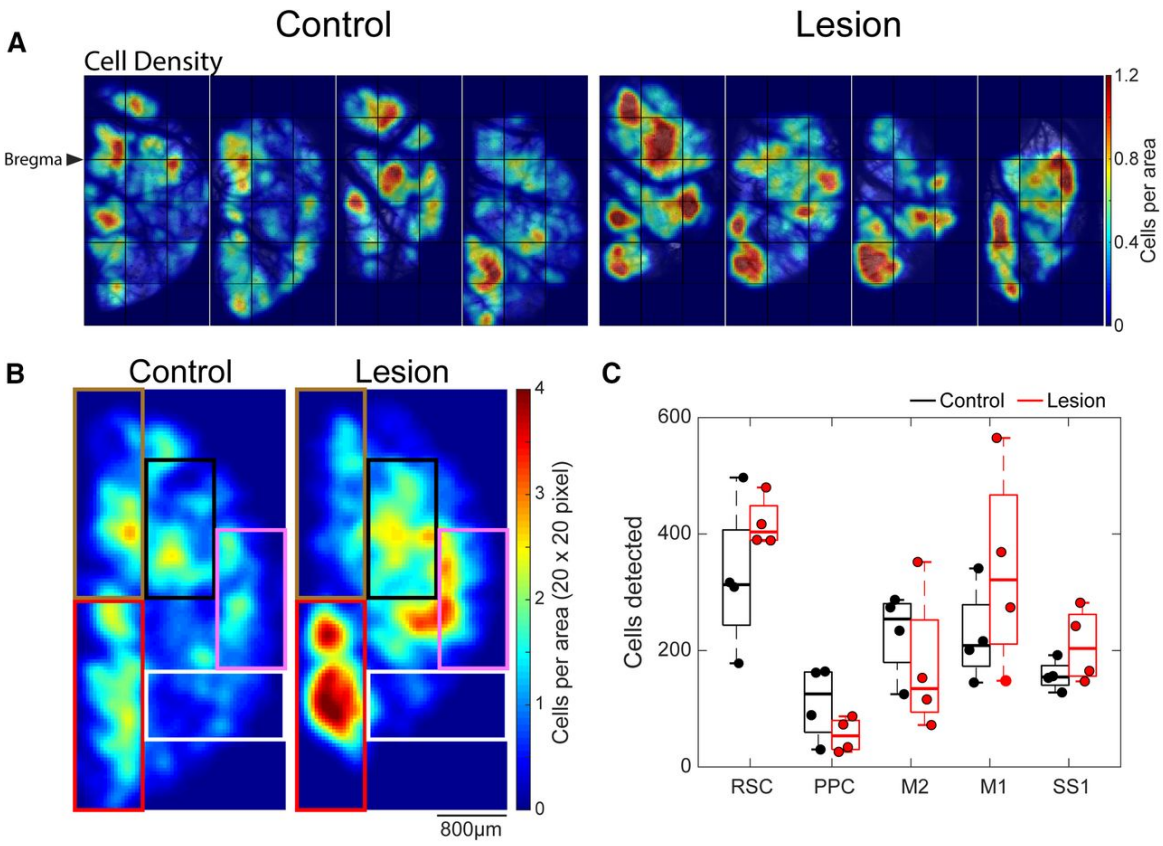

Figure 2.4: Distribution of cells. A, Density distribution map of all the cells detected across the window for the four control animals (left) and the four animals with bilateral hippocampal lesions (right). B, Density distribution map of all cells detected of each group on the same map. Color bar shows the number of cells inside an area of 20 × 20 pixels (435.76 µm2). Colored boxes refer to the different neocortical regions imaged (RSC, red; M2, brown; M1, black; PPC, white; SS1, magenta). C, Box plot of all the cells detected separated per region. Line, median; box, 25th and 75th percentile; dots, values for an individual animal; whiskers, minimum and maximum values. No significant differences were found in the total number of cells detected in any neocortical region across groups. For exact p values, see Table 2.1.

cells detected during linear treadmill running, there was no difference in the total number of cells across groups in any neocortical region (Fig. 2.4B,C).

2.3.2 Spatially selective neurons in the hippocampal lesion group convey less spatial information and lower sparse coding characteristics

We next evaluated whether the neurons that passed the criteria for being spatially selective exhibit specific differences between groups. For cells passing the criteria, the distribution of PF widths in both groups was very similar. However, the SI and the sparsity distribution showed that the lesion group cells exhibited lower spatial information and higher sparsity index (i.e., were less selective)

when compared with the control group (Fig. 2.5A). On average, there were no differences in the PF width, but control animals' cells had higher spatial information, and sparser coding than the lesion group (Fig. 2.5B, right). These results were confirmed when the spatially selective cells were evaluated separately by region (Fig. 2.5B, left).

2.3.3 Hippocampal lesion impairs the ability of neocortex to create a uniform representation of space

Next, we checked for differences between the encoding patterns of both groups. We first evaluated the position-mapped activity for each group (Fig. 2.6A, top) and the aligned trial-averaged profile (Fig. 2.6A, bottom) for all imaged neurons regardless of spatial specificity. This analysis showed that the control group cells had more spatially compact firing patterns with sparser coding. When we averaged the unnormalized position maps and the aligned activity profiles, we observed no difference in the overall activity between groups (Fig. 2.6B, top), but the aligned trial-averaged activity showed that the lesion group cells had significantly reduced peak activity (Fig. 2.6B, bottom).

Next, we evaluated only the neurons that met our criteria for spatial selectivity (see Materials and Methods). We first cross-validated the position map by dividing the trials into two sets and by using the odd-numbered trials to find the position at which each neuron fired maximally. Then we quantified the position map of the odd and even trials sorted by the peak of the odd trials and found that the preference for position was consistent, with neurons presenting sequential firing during movement covering the entire belt for both groups. However, the lesion group presented a less uniform representation over position, with more cells having peak activity by the cues and the reward site (Fig. 2.6C). With the correlation map of those cells, we quantified the similarity between cells tuned to distinct locations. The control group showed cells with a high correlation between nearby locations (near the diagonal) and a steep drop-off in correlation with the distance. The lesion group cells also had a high correlation near the diagonal. Still, there appears to be a tendency in the lesion group that the correlation with the distance only drops steeply with distance (as with the control group) at the locations of cues (Fig. 2.6D). Analysis of the averaged position map and the trial-averaged activity aligned for each neocortical region are shown in Figure 2.7.

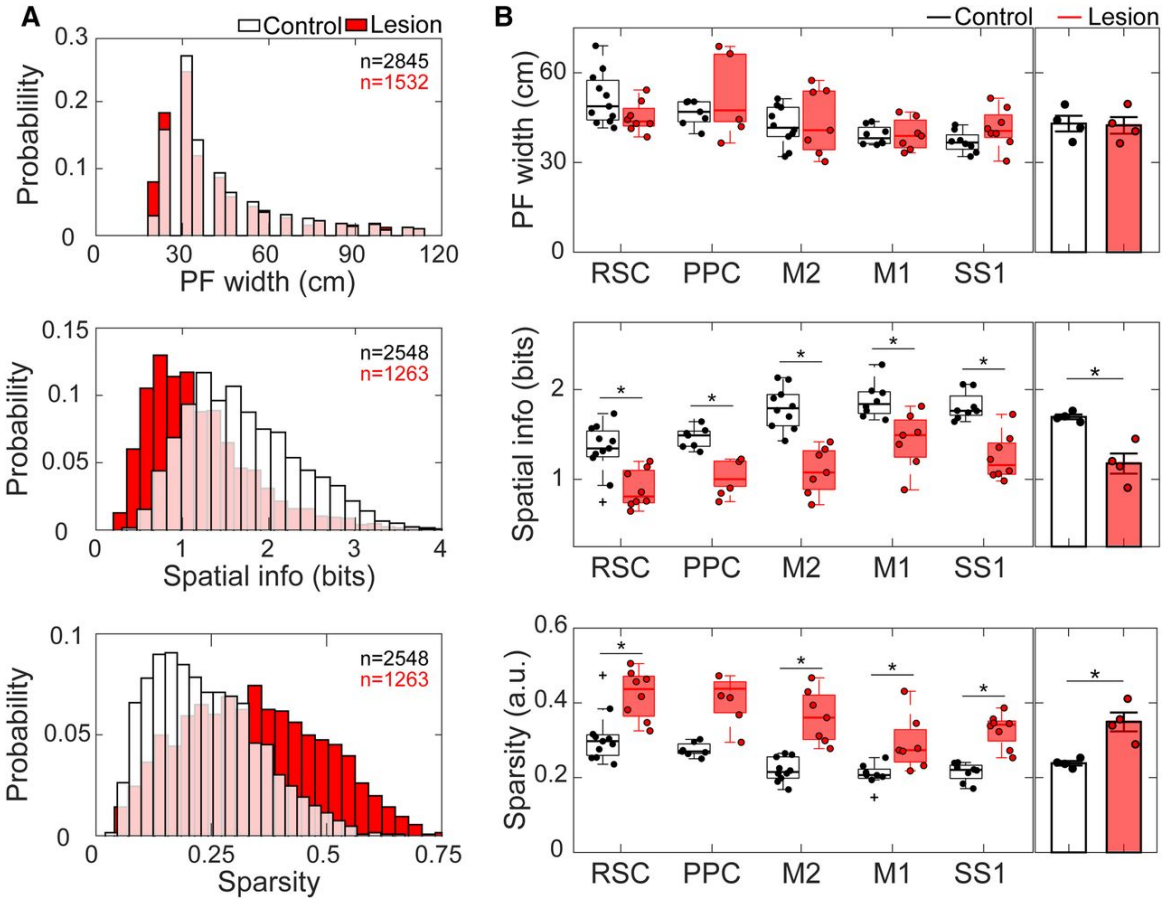

Figure 2.5: Spatially selective cell characteristics with and without bilateral hippocampal lesions. Data shown were for cells that passed the spatial selectivity criteria (see Materials and Methods). A, Probability distributions of place field widths, SI, and sparsity for control group (white) and lesion group (red). B, Left, Box plot of place field widths, SI, and sparsity separated by region. Line, median; box, 25th and 75th percentiles; dots, values for an individual sessions recorded in each region from all animals; whiskers, minimum and maximum values; + signs, outliers; $*p < 0.05$. B, Right, Bar plots of the average place field width, SI, and sparsity for neurons per group. Error bars denote SEM over animals (control, n = 4; lesion, n = 4; $*p < 0.05$). The average place width did not change with the hippocampal lesion. However, lesioned animals presented cells with reduced spatial information content, along with reduced sparsity. For exact n and p values, see Table 2.1.

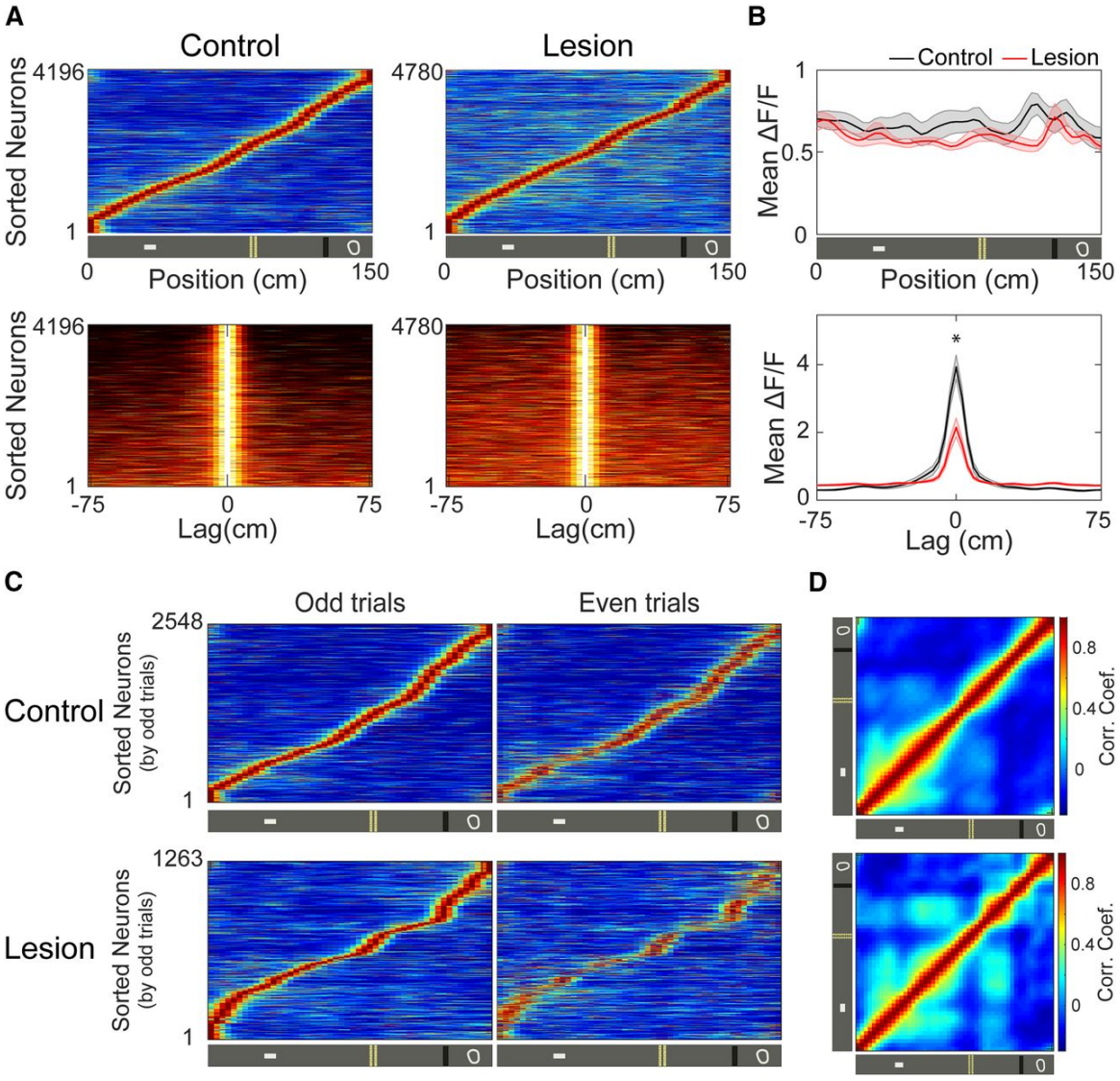

Figure 2.6: Cell activity during the task with and without bilateral hippocampal lesions. A, Pooled data from all mice showing activity profiles for all imaged neurons regardless of spatial specificity. Trial-averaged activity map for all cells in both groups as a function of position (position map). Top, Activity profiles were normalized and ordered by their peak response positions. Bottom, Aligned position map of all cells (circularly shifted to be aligned in the middle) sorted by their SI values. B, Grand averages for both groups of the trial-averaged activity (non-normalized) as a function of position (top) and with their peak aligned in the middle (bottom). Shaded area represents SEM over animals (control, n = 4; lesion, n = 4; $*p < 0.05$). The lesion group cells had less spatially compact firing patterns, with lower peak activity of the distributions. C, Cell activity of the neurons that passed the criteria for being spatially selective cells. Left, Position map obtained from odd trials, for only the spatially selective cells of each group. Right, Same as in left but for even trials and ordered by the position of the maximum response of odd trials. D, Pearson's correlation matrices of population vectors between pairs of positional bins for both groups with only the spatially selective cells. For exact n and p values, see Table 2.1.

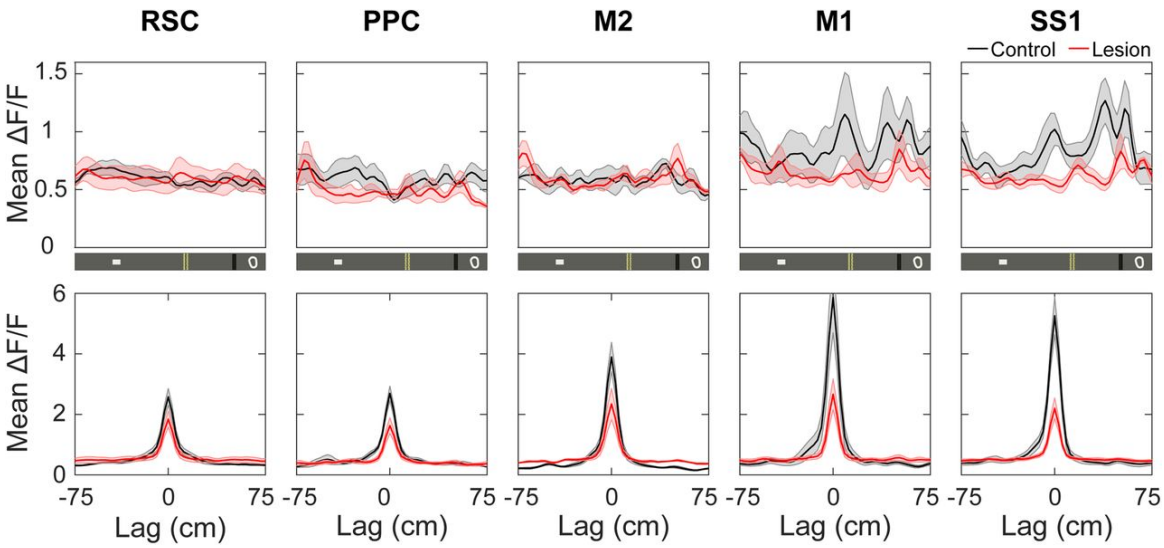

Figure 2.7: Trial-averaged activity (non-normalized) as a function of position for each region. Grand averages for both groups for all detected neurons, evaluated separately by neocortical region, of the trial-averaged, not normalized, activity as a function of position (top), and with their peaks aligned in the middle (bottom). Shaded area represents SEM over animals (control, n = 4, lesion, n = 4). Generally, there is no difference in the overall activity and the peak of the aligned atrial-averaged activity is higher in the control group.

2.3.4 Hippocampal lesion increases decoding error of position

Finally, we applied Bayesian decoding to estimate the position with the highest probability given the activity of all imaged neurons (Fig. 2.8A). The decoding error as a function of position for the lesion animals was significantly higher at all positions on the belt (Fig. 2.8B). The probability distribution of the decoding error showed that the median error from the control group was 7.03 cm. In contrast, the median decoding error in the lesion group was 20.35 cm (Fig. 2.8C). By computing the error by region, we also observed a significant difference between the two groups, the average decoding error from the lesion group was at least 40% higher in all the regions evaluated (Fig. 2.8D). The distribution map—position decoded as a function of the actual position of the animal—also showed a substantial between-group difference in the decoding performance (Fig. 2.8E). The red diagonal stripe in the control group and the higher dispersion around the diagonal in the lesion group show that the representations of position corresponded more with the actual location of the animal in the control group. Similar results were obtained when the decoding error was evaluated for separate neocortical areas (Fig. 2.9).

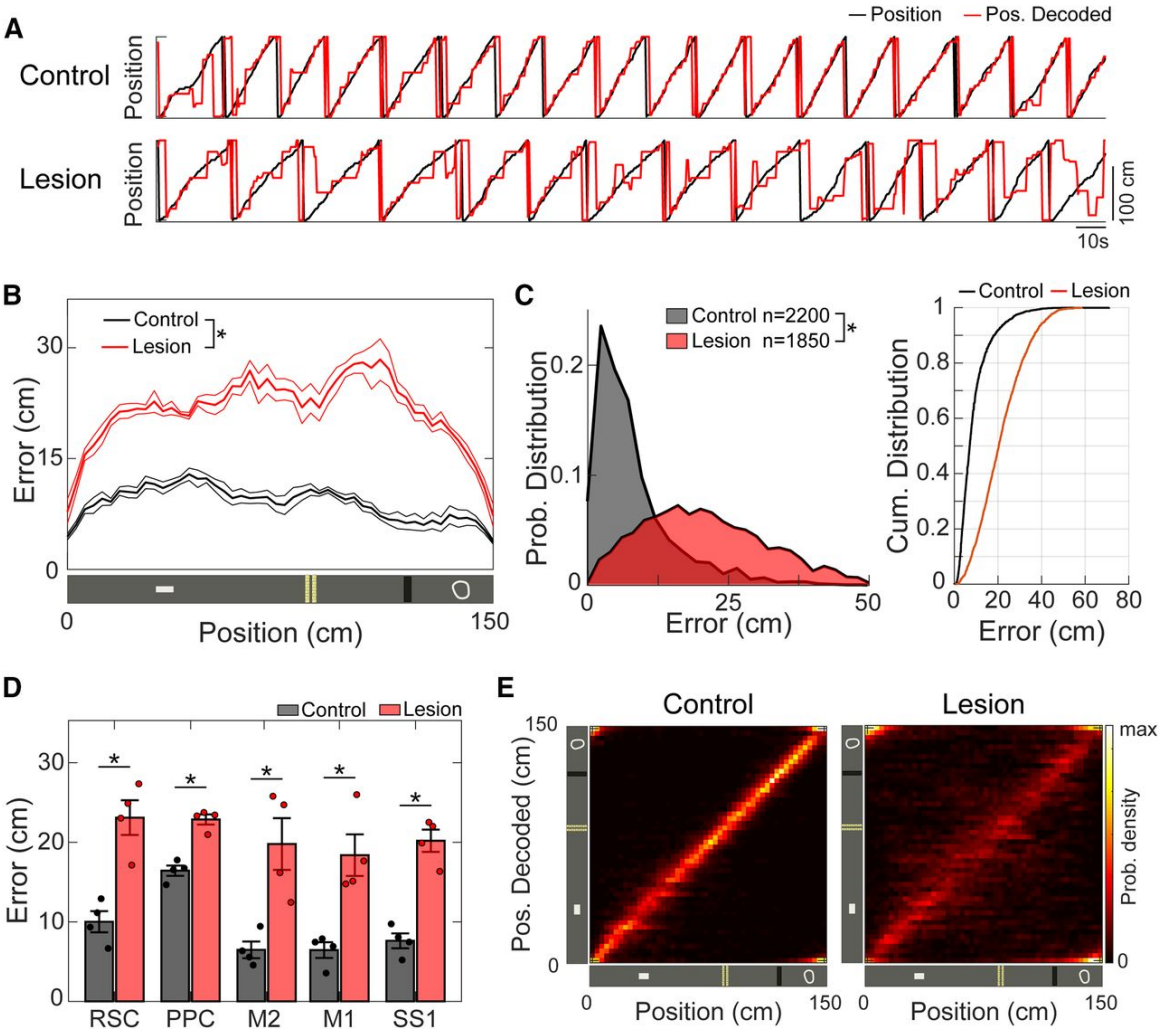

Figure 2.8: Decoding population activity between the control and lesion groups. A, Position (black) and position-decoded (red) traces of a session for a control and a lesioned animal. B, Decoding error as a function of position for the control group (black) and the lesion group (red) averaged across animals and regions ($*p < 0.05$; n = 4 per group). C, Left, Distribution of the decoding error for the control (black) and the lesion group (red); $*p < 0.05$. C, Right, Cumulative distributions of left. D, Mean decoding error for the five regions in the control and the lesion groups (error bars denote SEM over animals; $*p < 0.05$). Dots represent average decoding errors for each animal. E, Probability density map of all animals showing the distribution of position decoded from the activity of all neurons (y-axis) as a function of the location of the animal (x-axis). The red diagonal stripe indicates that the decoder from the control group gives a more accurate estimation of position. For exact n and p values, see Table 2.1.

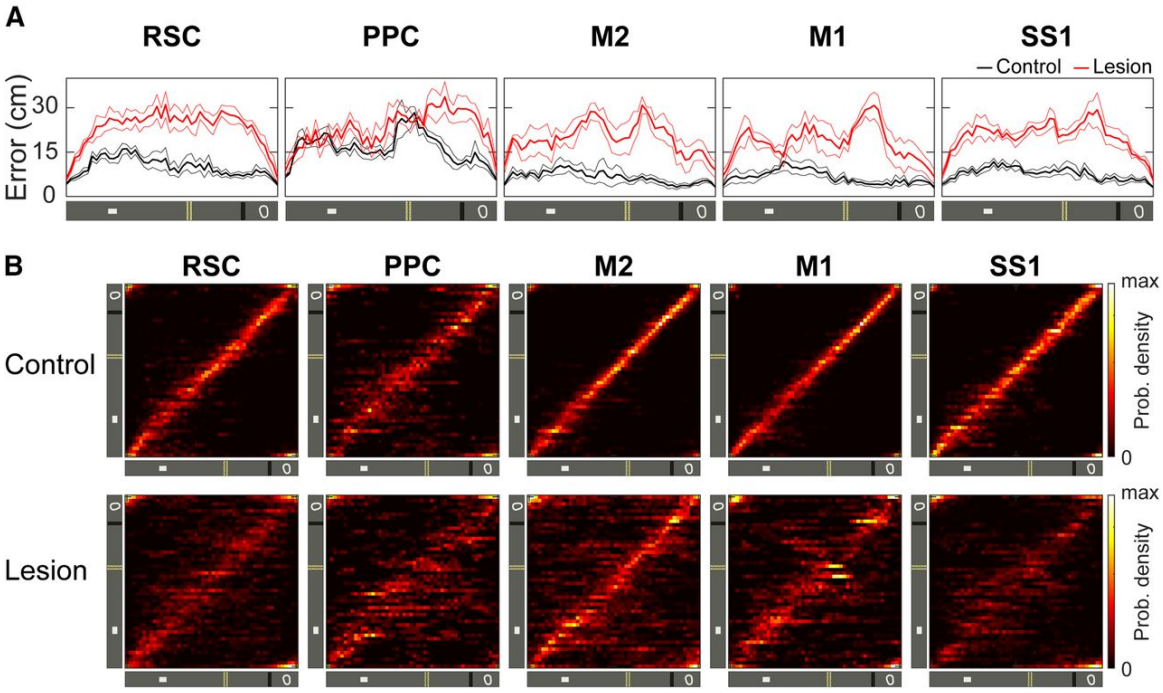

Figure 2.9: Decoding error for the five neocortical regions. A, Decoding error as a function of position for the control group (black) and the lesion group (red) averaged across animals for the five neocortical regions (n = 4/group). B, Probability density map showing the distribution of position decoded from the activity of all neurons (y-axis) as a function of the location of the animal (x-axis) for the five neocortical regions.

2.4 Discussion

By combining a 5 mm craniotomy with a treadmill assay, we found a large fraction of spatially selective neurons in the superficial layers of all neocortical regions surveyed. Bilateral lesions of the dorsal hippocampus significantly reduced the number of neurons exhibiting location-specific firing along the treadmill track. Furthermore, the place fields of the residual spatial neurons following lesion were not wider, but showed reduced sparsity and conveyed less spatial information. Together, these results suggest that, although a hippocampal lesion does not abolish spatially selective firing in neocortical areas, hippocampal inputs are necessary to maintain a precise cortical representation of space. We further supported these conclusions by showing that the positions of the animals on the track can be accurately estimated with the population activity of neocortical ensembles using a Bayesian approach, while a hippocampal lesion resulted in a significant decrease in decoding accuracy.

In our study, there was a general trend for anterior regions to express higher fractions of spatially selective neurons, with the posterior parietal cortex exhibiting the lowest numbers. Although PPC is the region less affected by the lesion, it is also the region with a lower number of cells detected (expressing spatial coding characteristics or not), and a higher decoding error in the control group. Such a discrepancy may be because of the presence of dense vascular structures over the parietal region that can impact the signal-to-noise ratio of the recordings using in vivo two-photon Ca2+ imaging. Several studies have shown the posterior parietal cortex to be a critical component of the navigational system (Kolb et al., 1983; Parron and Save, 2004; Nitz, 2006, 2012; Whitlock et al., 2012). The PPC has been shown to integrate input from multiple sensory modalities, and to encode spatial information in both egocentric (body-centered) and allocentric (world-centered) frames of reference (McNaughton et al., 1994; Nitz, 2006; Wilber et al., 2014). The construction of spatial cognitive maps relies on the ability to translate navigational signals, which arise from a body-centered frame of reference in primary cortical areas, to a world-centered frame of reference (McNaughton et al., 2006). The PPC has been postulated as the region where the transition from an egocentric to an allocentric coding scheme occurs (Burgess et al., 1999).

Although single-unit recordings in behaving animals have revealed many forms of spatial cell types in extrahippocampal regions, the precise nature of these signals remains poorly understood (Knierim and Neunuebel, 2016; Grieves and Jeffery, 2017). Recently, a growing body of studies

that used two-photon calcium imaging and treadmill assays have described spatially tuned neurons in many neocortical areas, including the posterior parietal cortex (Nitz, 2006; Harvey et al., 2012), the retrosplenial cortex (Mao et al., 2017, 2018), and the visual cortex (Fiser et al., 2016; Pakan et al., 2018; Saleem et al., 2018; Minderer et al., 2019). In these studies, special emphasis was given to occipitoparietal areas involved in visual processing. Interest in those regions may be justified by the anatomic connectivity these areas receive from the dorsal hippocampus (Strange et al., 2014; Skelin et al., 2019). Nevertheless, in the present work, we indiscriminately surveyed multiple neocortical regions (including primary, secondary, and association areas), and demonstrated that, in general, these spatial representations do not differ in fundamental features across regions, although it may be expected that differences may emerge when cues and behavioral context are manipulated, which was not done in this survey. Furthermore, we directly investigated the link between the neocortex and the hippocampus with respect to the coding of location and characterized the extent to which cortical representations relied on hippocampal inputs.

The current results fittingly complement the previous reports by highlighting the importance of hippocampal feedback to cortical spatial coding. In fact, a good example of this has been given when a lesion to the hippocampus caused medial entorhinal grid cells to lose their spatial periodicity, with some adopting instead the response schemes of head direction cells (Bonnevie et al., 2013). However, following lesions of the medial entorhinal cortex, place cells could still be observed (Miller and Best, 1980; Brun et al., 2008; Schlesiger et al., 2015). Similarly, specific lesions of the perirhinal cortex, of the postrhinal cortex, and of the presubiculum, three other important hubs for corticohippocampal inputs, did not lead to complete disruption of the spatial specificity of place cells either (Muir and Bilkey, 2001; Calton et al., 2003; Nerad et al., 2009). These data suggest that spatial representations in the hippocampus may emerge from highly parallelized streams of information and can adapt to a substantial amount of signal loss. In contrast, spatial signals found in regions outside of the hippocampus appear to rely more heavily on hippocampal inputs.

Overall, much of what is known about the general behavior of rodents that have undergone hippocampectomy has been described by a large body of research in rats (Nadel, 1968; Mumby et al., 2002; Clark et al., 2005; Faraji et al., 2008; Ocampo et al., 2018), but not all results from rats can necessarily be generalized to mice. As expected, it has been shown that hippocampal lesions in mice also impaired spatial working memory and spatial reference memory, and disrupt some

species-typical behavior (e.g., rearing, exploring, and hoarding). However, anxiety in lesioned mice was not uniformly diminished, and lesioned mice did not present an ill-groomed appearance and did not show alteration in locomotor activity (Arns et al., 1999; Deacon et al., 2002; Deacon and Rawlins, 2005). In addition, another study conducted to evaluate the effect of dorsal and ventral hippocampal lesions in mice showed that although hippocampal lesions made mice hyperactive during the habituation period in the open field, dorsal lesions were not able to produce a significant difference in the maze running time (Ammassari-Teule and Passino, 1997). Together, these works suggest that lesions of the hippocampus can produce in mice hyperactivity in a novel environment, can affect responses to external stimuli (spatial novelty), and can reduce directed exploration (rearing and head dipping), but all these effects were not because of impaired motility. At least for the task presented in this study, we could not detect any significant locomotion difference between both groups. Lesion and control groups were overall equally responsive during the entire time that they were tested, and hippocampal lesions did not detectably alter the performance of the task by the mice. The lack of difference may rather reflect the case that both groups were well habituated to the task.

One outstanding question that merits further investigation is how the experience of the animals affects the time course of the formation of spatial representations in the neocortex. Several works have shown that the hippocampal place code can emerge during the first traversal of a novel environment, that hippocampal population activity accurately represents the position of the animal after $\sim$10 min of experience in a new environment, and that place fields can expand with experience (Hill, 1978; Wilson and McNaughton, 1993; Mehta et al., 1997). In the neocortex, the emergence of spatial representations in the RSC has been shown to progressively improve with experience (Mao et al., 2017). But for other cortical regions, the dynamics of spatial representations still remain to be elucidated.

In summary, our results provide new insights into the importance of the hippocampus in shaping neocortical activity during the exploration of a controlled environment. Although this study offered a general perspective on the encoding of space in the neocortex, further research using more complex assays is needed to test various sensory/cognitive features to better characterize those cells and to identify region-specific attributes. We demonstrated that spatial representations are widely observed in the dorsal neocortex and that these representations express a similar degree of reliance of

hippocampal feedback. In the framework of the "index theory" (Teyler and DiScenna, 1986a; Teyler and Rudy, 2007; McNaughton, 2010), our results may suggest that, without the hippocampus, the cortex fails to express a unique code for each experience, and this failure likely explains the failure of the cortex to bind together attributes of an experience into unique episodic memories. This possibility remains to be investigated by placing animals in familiar and novel environments before and after hippocampal lesion.

Chapter 3

Coordinated Activities of Retrosplenial Ensembles during Resting-State Encode Spatial Landmarks[55,56]

Abstract

The brain likely utilizes off-line periods to consolidate recent memories. One hypothesis holds that the hippocampal output provides a unique, global linking or 'index' code for each memory, and that this code is stored in the cortex in association with locally encoded attributes of each memory. Activation of the index code is hypothesized to evoke coordinated memory trace reactivation thus facilitating consolidation. Retrosplenial cortex is a major recipient of hippocampal outflow and we have described populations of neurons there with sparse and orthogonal coding characteristics that resemble hippocampal 'place' cells, and whose expression depends on an intact hippocampus. Using two-photon Ca^{2+} imaging, we recorded ensembles of neurons in the retrosplenial cortex during periods of immobility before and after active running on a familiar linear treadmill track. Synchronous bursting of distinct groups of neurons occurred during rest both prior to and after running. In the second rest epoch, these patterns were associated with the locations of tactile landmarks and reward. Complementing established views on the functions of the retrosplenial cortex, our findings indicate that the structure is involved with processing landmark information during rest.

3.1 Introduction

The Retrosplenial Cortex (RSC) performs key functions in the processing and long-term storage of visuospatial information. The structure receives strong afferent projections from the dorsal hip-

[55] Reproduced from Chang et al. (2020) with authors' permission.

[56] **Author contributions** HR.C., A.R.N. and I.M.E. conceptualized and developed the experiments. I.M.E. conducted the behavioural recordings. J.S. performed all animal surgeries. HR.C. conducted all data analyses. HR.C. and B.L.M. wrote the article, which all authors commented on and edited. B.L.M. and M.H.M. supervised the study.

pocampus, both directly and indirectly through the subiculum (Insausti et al., 1997; Miyashita and Rockland, 2007; Vogt and Miller, 1983; Wyss and Van Groen, 1992; Van Groen and Wyss, 1992). In both humans and other animals, lesioning the RSC produces deficits in learning and retrieval of spatial information (Maguire, 2001; Sutherland et al., 1988; Vann and Aggleton, 2004; Cooper and Mizumori, 1999; Whishaw et al., 2001). In particular, RSC has been shown to encode environmental landmarks, directional heading and conjunctive features of space in both egocentric and allocentric frames of reference (Chen et al., 1994; Jacob et al., 2017; Smith et al., 2012; Alexander and Nitz, 2015, 2017; Vedder et al., 2017; Mao et al., 2017). The formation of unique spatial sequences in the RSC relies on an intact hippocampus (Mao et al., 2018). These attributes make RSC a prime candidate for investigating the neocortical patterns of neural activity that support the encoding and consolidation of spatial memories.

Reactivation of behavioural memory patterns has been reported in multiple neocortical areas, including both associational and primary/secondary regions (Qin et al., 1997; Hoffman and McNaughton, 2002; Ji and Wilson, 2007; Euston et al., 2007; Peyrache et al., 2009; Wilber et al., 2017; Gardner et al., 2019; Trettel et al., 2019). Similarly, direct optogenetic stimulation of *c-fos* tagged retrosplenial ensembles was successful in reproducing contextual fear-conditioned responses (Cowansage et al., 2014). Yet, despite the overwhelming signs of its implication in memory functions, spontaneous reinstatement of task-related memory traces in the RSC during off-line periods has so far not been reported. To investigate this possibility, we adapted standard experimental procedures used for studying reactivation in spatial tasks to two-photon calcium imaging. A secondary motivation for this study was to determine whether memory trace reactivation can be observed in neocortex using two-photon imaging which has much lower temporal resolution than electrophysiological recording. Such a demonstration would open the door to future studies that could exploit the ability to record simultaneously from very large populations distributed widely over the cortex, a capability not currently available with electrophysiology. Accordingly, we recorded from large ensembles of neurons in the dysgranular RSC (dRSC) before, during and after virtual navigation over a linear treadmill populated with tactile landmarks.

3.2 Materials and Methods

3.2.1 Subjects and surgical procedures

All animal procedures were conducted in compliance with the guidelines established by the Canadian Council for Animal Care and were approved by the Animal Welfare Committee of the University of Lethbridge. Three Thy1-GCaMP6s transgenic mice, aged between 2 and 8 months, were used. Following surgery, animals were single-housed in clear plastic cages under a 12-hours light/dark cycle. Mice were administered dexamethasone (0.2 mg/kg, intramuscular) and 0.5 ml of a mixture of 5% dextrose and atropine (3 μg/ml, subcutaneous), before being anaesthetized with 1.5% isoflurane. Body temperature was maintained at 37°C using a regulated infra-red heating pad. Lidocaine (7 mg/kg) was injected subcutaneously under the incision site. A 5 mm diameter craniotomy was performed over the dorsal cortex (+2 to -3 mm AP; $\pm$2.5 mm ML) (Fig. 3.1B). A compound glass window, composed of a 7 mm diameter round cover-slip stacked over two 5 mm diameter cover-slips (held together with optical adhesive NOA71, Norland), was implanted over the craniotomy and retained using tissue adhesive (Vetbond, 3M). A custom titanium head-plate was fixed to the skull with Metabond (Parkell) and dental adhesive. A rubber ring was attached along the perimeter of the head-plate to retain distilled water during imaging and to insulate the recording site from light contamination. Mice recovered for a minimum of one week before the start of experiments.

3.2.2 Behavioural task

Water-restricted mice were habituated to head-fixation and underwent daily training to run over a linear treadmill track for over 2 weeks, as previously described (Mao et al., 2017, 2018). The duration of each training session began with 15 minutes and was gradually extended to one hour over the course of five days. The treadmill consisted of two 3D-printed polyamide wheels with radii of 5 cm (Fig. 3.1A). The wheel centres were separated by 40 cm. Position was decoded from an optical encoder (Broadcom) attached to the front wheel. A 150 cm long belt made from the soft fabric of a Velcro strip (Country Brook) was looped around the wheels. A photo-reflective tape was applied to one spot underneath the belt. The tape triggered a photoelectric sensor (Omron) that opened an electromagnetic pinch valve (Bio-Chem Valve) to dispense sucrose water reward.

Once performance reached over 50 laps per hour, the duration of running was reduced to 20

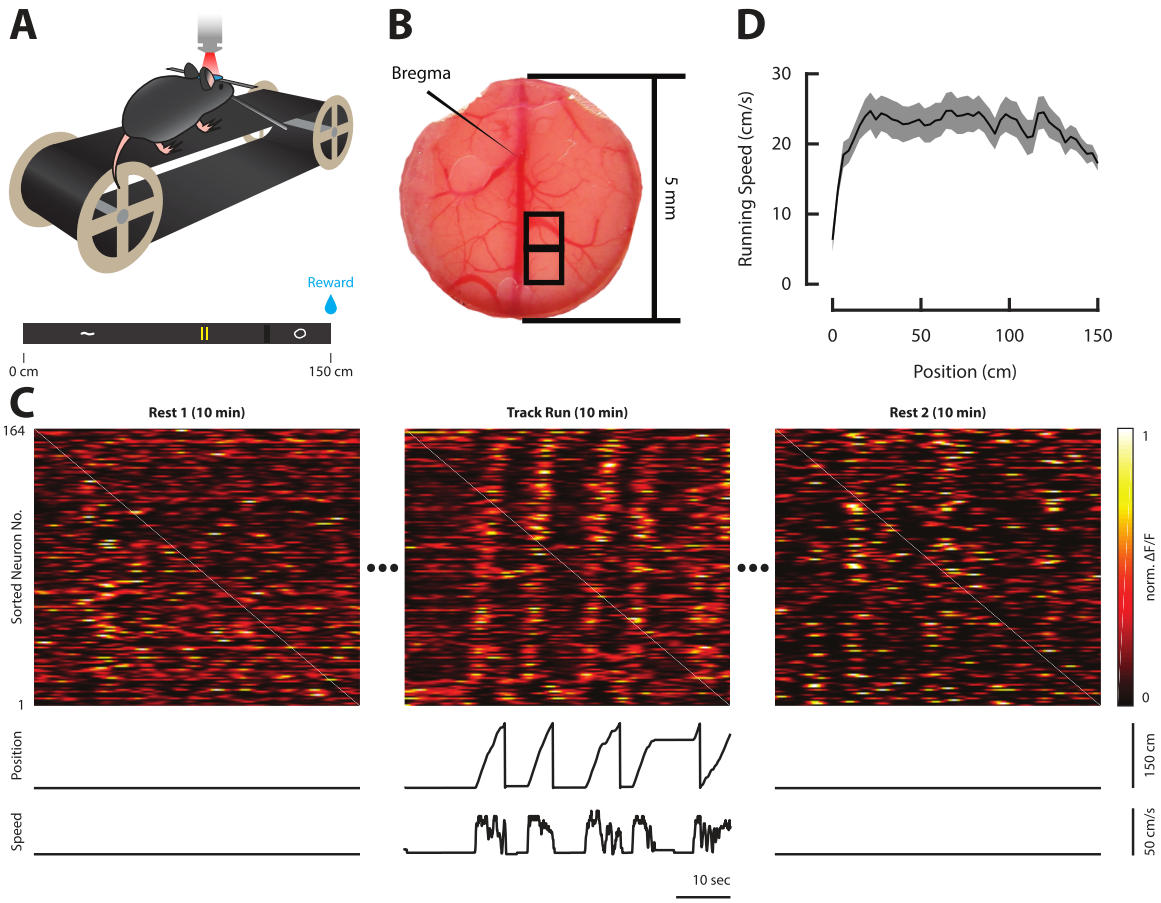

Figure 3.1: Imaging apparatus and experimental design. (A) Water-deprived mice were head-restrained over a 150 cm long treadmill belt on which were mounted several tactile cues. An LED light illuminated the portion of the belt in front of the mice, who were thus able to see the incoming landmarks. Cue positions and sizes are illustrated to scale with the belt over the x-axis. (B) Example of cranial window implant for one mouse. The windows inside of which imaging was conducted for this study are delineated by black boxes. (C) Each imaging session was divided into three 10 minute blocks. Before and after running, animals rested quietly over the belt, while surrounded by a cardboard enclosure. The belt was clamped during rest and released during run. A 1-minute segment from each imaging block is illustrated from one example session. The time courses of deconvolved $\Delta F/F$ for all simultaneously imaged neurons were z-scored and temporally smoothed with a $\sigma = 1$s Gaussian kernel. Traces for each neuron were normalized by the peak (range between 0 to 1). Neurons were sorted by the location with the highest average response during running. Animal position and linear velocity are shown below. (D) The average running speed as a function of location (n=3 mice; n=13 sessions). Shaded region denote S.E.M.

minutes and animals were made to rest for 20 minutes before and after running by clamping the belt. The belt was lined with four distinct tactile cues (Fig. 3.1A), made up of hot glue (first and last cues), reflective tape (middle cue), and a strip of Velcro (second cue before reward). During navigation, a dim LED light illuminated the area in front of the animals, so that they could see approaching landmarks. During rest, a cardboard enclosure was placed around the mice to substitute for a flower pot used in previous paradigms (Lansink et al., 2009). Mice were trained under this new protocol for another 2 weeks or until movement was detected for less than 30% of the duration of rest. For imaging experiments, the duration of REST1, RUN and REST2 epochs were shortened to 10 minutes each (Fig. 3.1C). During rest, the belt was clamped to discourage movements. With the belt clamped, the treadmill encoder was still sensitive enough to detect any small movement made by the animals.

3.2.3 Two-photon imaging

All imaging data were acquired using a Thorlabs Bergamo II multi-photon microscope. A Ti:Sapphire laser (Coherent) tuned to an excitatory wavelength of 920 nm was passed to the tissue through a 16X water immersion objective (Nikon, NA 0.8, 80-120 mW output power measured at the sample). Rasterization was conducted by Galvo-Resonant scanners. The emitted GCaMP6 signals were amplified using a GaAsP photomultiplier tube (Hamamatsu) and digitized to a resolution of 800 x 800 pixels at a sampling rate of 19 Hz. We imaged an 835 x 835 μm window over layers II-III of the agranular retrosplenial cortex at depths between 100-200 μm (imaging windows centred at -1.8 to -2.5 mm AP, 0.5 mm ML) (Fig. 3.1B). A strip of Velcro wrapped around the body of the objective was lowered to the level of the rubber ring to block ambient light.

3.2.4 Image preprocessing and place field analysis

Image preprocessing was performed as previously described (Mao et al., 2017, 2018). Registration and identification of regions of interest (ROIs) were conducted automatically using Suite2p (Pachitariu et al., 2016). Neurons were manually selected based on the morphology of the ROIs and the presence of distinct calcium deflections in the fluorescence trace. The raw fluorescent time-course for each neuron was extracted from the corresponding ROIs. Neuropil contamination, estimated from the surround of ROIs, was subtracted (Bonin et al., 2011). Baseline fluorescence was

approximated and the ratio $\Delta F/F_0$ was obtained. Firing rates were inferred by deconvolving the ratio time-courses using constrained non-negative matrix factorization (Pnevmatikakis et al., 2016). All subsequent analyses were conducted using the deconvolved time-courses in MATLAB R2017a (MathWorks).

Two criteria were used to classify neurons that express spatial-selectivity (henceforth termed 'place' cells). Spatial information (I; bits) was given by (Skaggs et al., 1993):

$$I = \sum_{i=1}^{N} p_i \frac{f_i}{f} \log_2 \frac{f_i}{f} \tag{3.1}$$

where for N spatial bins, p_i is the occupancy probability in bin i, f_i is the mean deconvolved fluorescence in bin i and f is the overall mean fluorescence. We generated a null distribution of spatial information by circularly shuffling the time courses 1,000 times. To be considered spatially-receptive, neurons must contain spatial information higher than the 95^{th} percentile of the null distribution. We obtained a continuous wavelet transform W over the spatial tuning curve of each neuron using a Mexican Hat mother wavelet Ψ:

$$W(\sigma,\tau) = \frac{1}{\sqrt{\sigma}} \sum_{i=1}^{N} f_i \Psi\left(\frac{i-\tau}{\sigma}\right) \tag{3.2}$$

$$\Psi(t) = \frac{2}{\pi^{\frac{1}{4}}\sqrt{3}}(1-t^2)\exp\frac{-t^2}{2} \tag{3.3}$$

where σ and τ are the scale and translation parameters respectively. For initial detection, the local maxima in the transform indicated the locations (given by τ) and the widths (given by σ) of potential place fields (Du et al., 2006; Tary et al., 2018). To remove spurious peaks caused by noise, we obtained a threshold λ of 3 median absolute deviations from the median from the wavelet coefficients at the lowest scale ($\sigma = 1$). Only local maxima with values higher than λ in the transform were selected as potential place fields. Local maxima that were contained within the receptive field of a local maximum at a higher scale σ were removed. A place field must be wider than 5% of the length of the environment, but narrower than 80%. The mean activity within a place field must be 2.5 times higher than the mean activity outside of place fields. When comparing the activity inside a place field and outside any place field, the peak activity must occur within the place field in a third of the laps. Cells that, after passing all criteria, still contained at least one place field were classified

as spatially-receptive neurons.

3.2.5 Identification of Neuronal Ensembles

To identify and group neurons that expressed spontaneous co-activity during rest, we applied agglomerative clustering to the correlation matrices of rate vectors (Fig. 3.2B) (for comprehensive review on hierarchical clustering, see Murtagh and Contreras 2012). First, we removed movement epochs by identifying any region of time in which the instantaneous velocity deviated from 0. Rate vectors were smoothed using a $\sigma = 1s$ Gaussian kernel to reduce temporal jitter and to increase temporal correlation. Then, we computed the Pearson correlation matrix of rate vectors between each neuron pair. Correlation coefficients r were converted to a distance metric $d = 1 - r$. Finally, we performed agglomerative clustering on the upper triangle of the distance matrix using unweighted average distance linkage criterion. We set a cut-off threshold of $d < 0.75$, which corresponds to an average correlation coefficient of $r > 0.25$ within a cluster. Clusters containing less than 10 members were rejected.

For each ensemble, Synchronous Calcium Events (SCEs) were identified by calculating the multi-unit activity (MUA) from the z-scored deconvolved ΔF/F of all neurons that were part of the ensemble. Any continuous segment of time in which the MUA exceeded 3 standard deviations above the mean were classified as SCEs (Fig. A.1). SCEs that occurred less than 250 milliseconds apart were identified as part of the same SCE. Inter-event intervals were measured between the onset of each SCE, within the same ensemble.

3.2.6 Bayesian Reconstruction

Methods for Bayesian reconstruction were as previously described (Mao et al., 2018; Zhang et al., 1998). Briefly, we obtained the probability for the animal to be at a position x given the population firing vector n:

$$\Pr(x|n) = \frac{\Pr(n|x)\Pr(x)}{\Pr(n)} = C(\prod_{i=1}^{N} f_i(x)^{n_i}) \exp -\tau \sum_{i=1}^{N} f_i(x) \tag{3.4}$$

where $f_i(x)$ is the mean deconvolved fluorescence of neuron i as a function of position derived from training data, and n_i is the time-course vector a mean activity within time bins of length τ obtained from testing data. The resulting matrix $\Pr(x|n)$ contains the probability for the animal to be at any

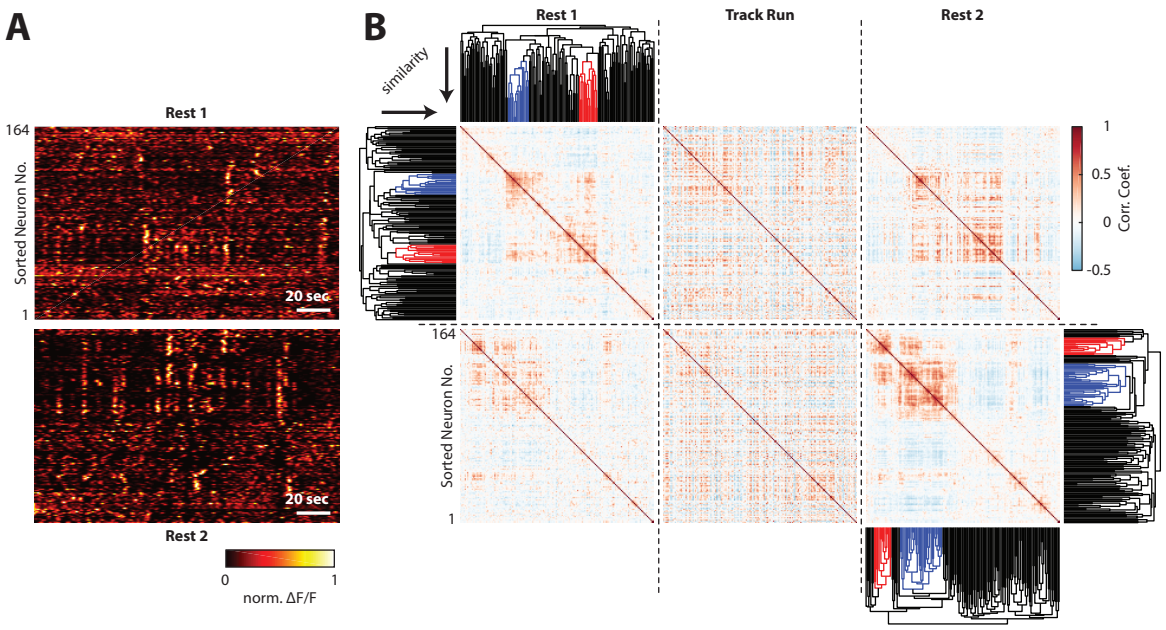

Figure 3.2: Sub-populations of retrosplenial neurons repeatedly co-activate during rest. (A) 200-sec segments of the deconvolved time courses of REST1 and REST2 were taken from the same session as in Fig. 3.1. Sorting the neurons by firing rate vector similarity (see B) reveals groups of neurons that were spontaneously and repeatedly co-active (in this article, we term these synchronous groups 'ensembles'). (B) Pearson correlation matrices were computed between the rate vectors of all cell-pairs for each imaging block. Hierarchical clustering was conducted independently on the correlation matrices of REST1 and REST2 to group neurons expressing synchronous patterns of activity (clustered neurons are grouped by colour in the dendrogram; see Methods). From the root of the dendrogram, clusters are sorted by ascending order of similarity. Neurons are sorted as to maximize the similarities between adjacent neurons. Neurons in the top row (including neurons in A) are sorted according to the REST1 clustering results, while neurons in the bottom row are sorted by REST2 cluster linkages. In other words, direct comparisons by eye is possible between plots over the same row. In this example recording, two ensembles were detected in both REST1 and REST2. Note that, in general, REST1 and REST2 may not necessarily contain the same number of ensembles. The red and blue colours identify ensembles that were independently detected in REST1 and REST2. The colours were randomly chosen and do not imply any relationships between the two epochs.

position x for every time point in n_i (Fig. 3.5A, Fig. A.1). The decoded position over time is given by $\arg\max_x \Pr(x|n)$. Sessions in which the animal covered less than 10 laps were rejected from analysis (Table A.1).

To validate the quality of decoding, fluorescent time courses from odd laps were used as training sets and decoding errors were obtained as the absolute value of the difference between decoded position and real position on even laps (Fig. 3.5C). For rest epochs, the model was trained on all trials during running periods, and decoding was conducted on the resting-state population vectors that corresponded with SCEs. To evaluate the likelihood that positions were decoded by chance, we shuffled the identities of neurons and performed the decoding procedure 1,000 times (Fig. 3.5B, Fig. A.1). This method ensures that the temporal structure of the time courses is preserved. A p-value is obtained at each position for each SCE-associated time frame by counting the instances where $Pr(x|n)$, decoded from rest, was higher than the randomly permuted data.

As a secondary shuffling criterium, we used a 'time bins shuffle'. We randomly permuted the time bins in the posterior probability matrix $Pr(x|n)$ 1,000 times and obtained p-values by counting the number of instances in which the probabilities during SCE-associated time frames were higher than the shuffled probabilities. This method was included to account for the mismatch created between the place fields' tuning profiles $f_i(x)$ and the resting state population firing vectors n_i from 'cell identity shuffling' (Foster, 2017). However, this method is susceptible to a different source of noise by considering time frames that were outside of SCEs, hence creating a mismatch between temporal structures. We included the results obtained from this shuffling method in Fig. A.3 and Appendix A.1B.

3.3 Results

3.3.1 Sub-populations of retrosplenial neurons are spontaneously and recurrently co-activated during rest

We began by investigating the population dynamics in the retrosplenial network during quiet-wakeful periods. We only considered regions of the recording in which animals were immobile. We observed spontaneous and concerted activity among groups of neurons, during resting periods both before (REST1) and after (REST2) exposure to task (Fig. 3.2A). For the remainder of this article, we shall refer to these groups of synchronous neurons as 'ensembles'. Ensembles were identified based

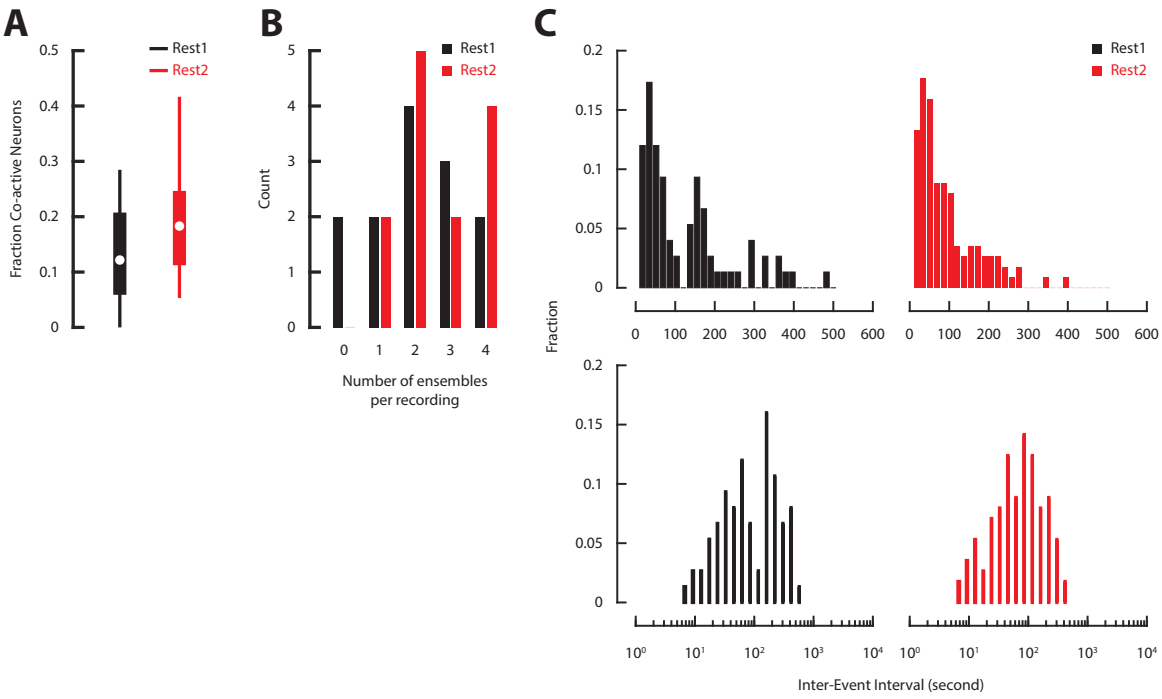

Figure 3.3: During the resting state, sparse subsets of neurons engage in synchronous activation. (A) The fraction of neurons that were co-active during REST1 and REST2 (white dot, median; box, first and last quartiles; whiskers, minimum and maximum values; solid dots, outliers; Wilcoxon signed-rank test n.s.). These fractions include all neurons that were classified as part of an ensemble in a recording, for all 13 imaging sessions conducted. (B) Distribution of the number of ensembles simultaneously detected during rest. For each of the 13 imaging sessions, we count the number of ensembles that co-existed in REST1 and in REST2. (C) Distributions of the time interval between successive SCEs over linear (top) and logarithmic scales (bottom) (n=102 REST1 SCEs; n=147 REST2 SCEs). We took the difference between the onset time-stamp of each adjacent SCE that was part of the same ensemble.

on the similarity between the rate vectors across pairs of neurons (Fig. 3.2B; see Methods). Ensemble neurons formed a small subset of the population. On average, 13.5% of neurons expressed co-activity during REST1 ($\pm$2.6% s.e.m.), while 20.2% ($\pm$3% s.e.m.; n=13 sessions; n=3 mice) were synchronous during REST2 (Fig. 3.3A). Per imaging session, an average of 2.1 ensembles were simultaneously active in REST1 ($\pm$0.4 s.e.m., n=13 sessions; n=3 mice), and 2.6 ensembles in REST2 ($\pm$0.3 s.e.m., n=13 sessions; n=3 mice) (Fig. 3.3B). Though there was an overall tendency for REST2 neuronal populations to express a higher degree of synchrony, the effect was not significant with the current sample size.

Neurons grouped into ensembles exhibited synchronous activity, termed Synchronous Calcium Events (SCEs) (Malvache et al., 2016). These events occurred at a mean rate of 0.66 events/minute

($\pm$0.06 s.e.m.; n=27 ensembles) during REST1 and 0.53 events/minute ($\pm$0.05 s.e.m.; n=34 ensembles) during REST2. The time interval between events appeared to be log-normally distributed (Fig. 3.3C), with a mean interval of 70 seconds during REST1 (+1.13/-0.88 seconds s.e.m.) and 55 seconds in REST2 (+1.1/-0.91 seconds s.e.m.). Together, the results suggest that SCEs occur in close succession to each other (at a rate of $\sim$ 1 SCE per minute) over sparse regions of time.

3.3.2 Resting-state ensembles are stable following active virtual navigation

Next, we examined whether neuronal ensembles remained stable following virtual navigation. We began by comparing the correlation structure of the entire neuronal population across the three behavioural epochs. A Pearson correlation matrix was calculated for the rate vectors across each epoch (Fig. 3.2B). The correlation coefficients between cell pairs in the upper matrix were in turn correlated across all three epochs. The population activity between resting periods exhibited a greater degree of similarity than between rest and run epochs (Fig. 3.4A). Explained variance (EV) was used to quantify the percentage of the variance in the pairwise correlations of the neuronal population in RUN that could be explained by the REST2 correlations, given the correlations that existed between REST1 and RUN (Fig. 3.4B) (Kudrimoti et al., 1999; Tatsuno et al., 2006). A higher fraction of the variance in the RUN correlations was predicted by the REST2 correlations, compared to REST1. Therefore, although the REST epochs were more similar in correlation structures, REST2 correlations exhibited a higher degree of resemblance with the RUN structures than REST1. This result provided the first evidence for reactivation. Looking into ensembles, the likelihood of obtaining a given degree of overlap between REST1 and REST2 co-active neurons was modelled after a hypergeometric distribution. The latter is a probability density function which, in the present scenario, can be used to describe the probability of redrawing k neurons out of n draws from a population of N neurons, given that K neurons were already drawn previously. The degree of overlap between resting-state synchronous neurons was significantly above chance, with an average p-value$<$0.01 (Fig. 3.4C). However, a sizeable fraction of neurons was also substituted following exposure to task; the stability index averaged at 0.21 ($\pm$0.03 s.e.m.). In the next section, we explore the contributions of these reorganisations to the encoding of space.

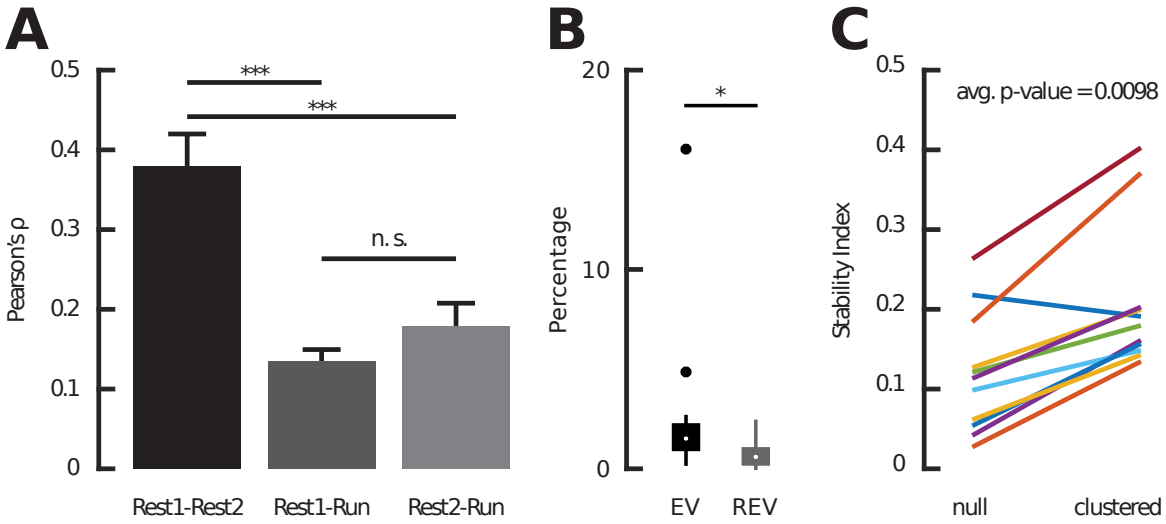

Figure 3.4: Population and ensemble activities are stable following task exposure. (A) For each session, we take the upper triangle of the correlation matrices for the firing rate vectors of all cells for each individual imaging block (n = 3 mice; n = 13 sessions). We compute the Pearson correlation between the correlation coefficients of the cell-pairs as a measure of similarity between the population activity structure across the three imaging blocks. The mean and standard errors of the resulting Pearson correlation coefficients are shown (one-way ANOVA; *** $p < 0.001$). The lack of significant difference between REST1-RUN and REST2-RUN possibly reflects the fact that the animals were already highly experienced on the task. (B) Explained variance was calculated as $EV = (\frac{R_{RUN,REST2} - R_{RUN,REST2}R_{REST1,REST2}}{\sqrt{(1-R^2_{RUN,REST1})(1-R^2_{REST1,REST2})}})^2$, where the R values correspond to Pearson correlation coefficients in A (Kudrimoti et al., 1999; Tatsuno et al., 2006). These values were controlled against the reverse explained variance (REV). REV was obtained using the same equation, but by substituting REST2 with REST1 epochs and vice versa. Explained variance was significantly higher than reverse explained variance, suggestion that reactivation had occurred ($p = 0.0183$; Wilcoxon rank sum test; n = 13 sessions; n = 3 mice; white dot, median; box, first and last quartiles; whiskers, minimum and maximum values; solid dots, outliers). (C) To assess the fraction of neurons that remained co-active after Run, we calculated a stability index $\frac{O}{N_1+N_2-O}$ where N_1 is the number of co-active neurons detected in REST1, N_2 is the number of co-active neurons detected in REST2 and O is the number of neurons that were co-active during both resting blocks. To determine how likely the resulting values were obtained by chance, we estimate the expected number of overlapping neurons between the two blocks as the expected value of the corresponding hypergeometric distribution (shown as 'null'). The average p-value from the hypergeometric distribution, indicating how likely the proportion of overlap O was due to chance, is reported.

3.3.3 Resting-state co-activity encode for the location of landmarks following navigation

We employed a Bayesian approach to determine whether SCEs contained information relevant to previously explored space. We first confirmed that a Bayesian model can accurately decode animal position from neural activity acquired during running. We estimated the model parameters using the activity during even lap runs and measured the difference between decoded position and actual position on odd trials (Fig. 3.5C). The mean decoding error was 11.8 cm ($\pm$0.3 s.e.m.), confirming the accuracy of the model. For resting data, we derived the model parameters from all running laps, and decoded position over the time frames during which SCEs occurred (Fig. 3.5A). The generated output gave the probability for the animal to be at any given position over the linear track at any given frame of time. To control for spurious results, the same analysis was conducted 1,000 times with the identity of the neurons randomly shuffled (Fig. 3.5B). A p-value could therefore be obtained by accumulating the number of instances in which a position contained a higher probability of being decoded than the shuffled distribution (Fig. 3.6A).

For REST2 SCEs, the fraction of significantly decoded positions was anti-correlated with the distance from local landmarks (Fig. 3.6C, Appendix A.1). This trend indicates that synchronous activity patterns during post-task rest decode more frequently locations in proximity to landmarks. In contrast, pre-exposure activities were not significantly related to landmark locations. The decoding fractions were anti-correlated with the distance from the reward in REST2, suggesting that higher precedence was gradually assigned to locations leading up to the reward. These results were consistently observed in all three animals tested (Fig. A.4-A.5).

3.3.4 Reorganisation of ensemble members is tied to environmental features

Lastly, we looked for the presence of any discernable features in the spatial coding characteristics of synchronous neurons. Both prior and subsequent to task exposure, ensembles did not selectively recruit highly spatially-receptive neurons. In REST1, 59% ($\pm$7% s.e.m.) of co-active neurons were classified as 'place cells', while 58% ($\pm$4% s.e.m.) was the average for REST2. Over the entire population, 57% ($\pm$4% s.e.m.) of neurons were spatially-receptive. Surprisingly, the spatial information content of the portion of co-active neurons that were lost subsequent to task was significantly higher than that of neurons that were gained (Fig. 3.7A). To explain this intriguing result, we eliminated the summation step in Eq. 3.1, which computes the spatial information content

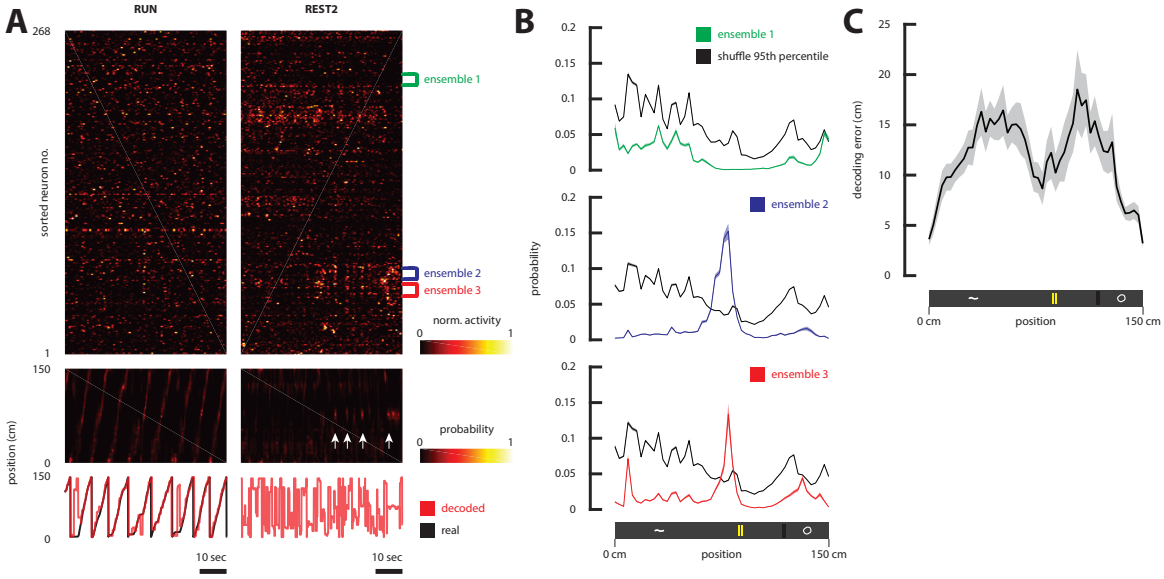

Figure 3.5: Bayesian reconstruction of location from resting neural activity. (A) One-minute segments of the time courses of neural activity during RUN and REST2 were taken from an example recording session (top). Neurons were sorted by similarity between the rate vectors of cell-pairs during REST2. Three ensembles were simultaneously identified in this data and are highlighted in this plot. The outputs from Bayesian reconstruction of position are represented as a matrix of probability that the animal is at a given location for any given position in time (middle). The parameters of the Bayesian model were estimated using all running laps and were tested on the resting state activity. White arrows denote the locations in time where SCEs were detected from ensemble #2. The decoded positions, defined as the location of highest probability, are plotted in conjunction with real positions (bottom). (B) The mean decoded probability as a function of location, during all frames in which a SCE occurred, was plotted for all three ensembles (shaded regions: s.e.m.). Dark plots represent the 95^{th} percentiles obtained from shuffled data. For ensembles 2 and 3, the decoded probability at the location right in front of the middle landmark was higher than the probability obtained from shuffled data. (C) The mean decoding error as a function of position (shaded regions: s.e.m.; n=13 sessions). The Bayesian model was trained on even trials, and errors were evaluated with odd trials.

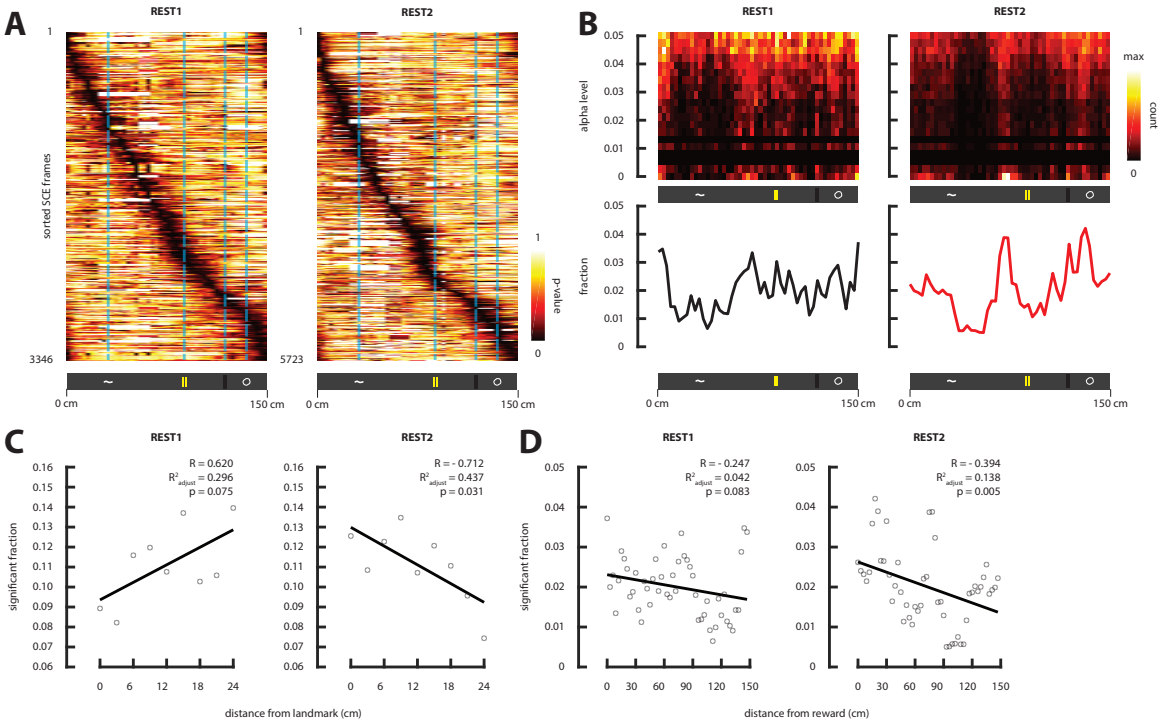

Figure 3.6: REST2 ensemble co-activation decodes positions near landmarks and reward. (A) For each time frame associated with a SCE, p-values were obtained as the fraction of probabilities obtained from shuffling that were greater than the probabilities obtained from the true data, at every given position. In other words, we evaluated the degree to which a decoded position during a SCE is likely due to chance (the lower the p-value, the less likely the decoding was obtained by chance). These time frames were sorted according to the location of most significant decoding. Results from all SCE-associated time frames (obtained from all three animals, all 13 sessions) are shown for REST1 and REST2. (B) Joint density distributions of the fraction of significantly decoded positions at different α levels (p-values). α levels were binned over a logarithmic scale. The fractions of significant decoding ($p<0.05$; analogous to summing over the columns) as a function of position is shown below. Notice in REST1, the fractions are more uniformly distributed, while in REST2, the fractions are more concentrated near the locations of landmarks. (C) The fraction of frames that significantly decoded position ($p<0.05$) is compared with the distance from the closest landmark over a scatter plot. Since the landmarks were not separated uniformly over the same distance, the fractions were normalized by the occupancy over each spatial bin. Linear regression fit line depicted in black. (D) Same as (C), but evaluating the distance needed to be travelled before receiving reward.

of neurons. This in turn yielded a vector of the relative contribution of each location to the total spatial information of a neuron (Fig. 3.7B). We found that a large fraction of spatial information in REST1 ensembles was contributed by the very beginning of the track, right after animals received a reward. In comparison, REST2 ensembles expressed higher spatial information over the first 30 cm and last 10 cm, corresponding with the first and last landmarks. This observation concurs with the results obtained from Bayesian analysis, where a large fraction of decoded positions in REST1 corresponded with the beginning of the track (Fig. 3.6B). The spatial distribution of the locations of place field centres also became more compact for REST2 co-active neurons (Fig. 3.7C). A higher fraction of place fields were expressed by REST2 and 'gained' neurons at locations of landmarks compared to REST1 and 'lost' as well (Fig. 3.7D).

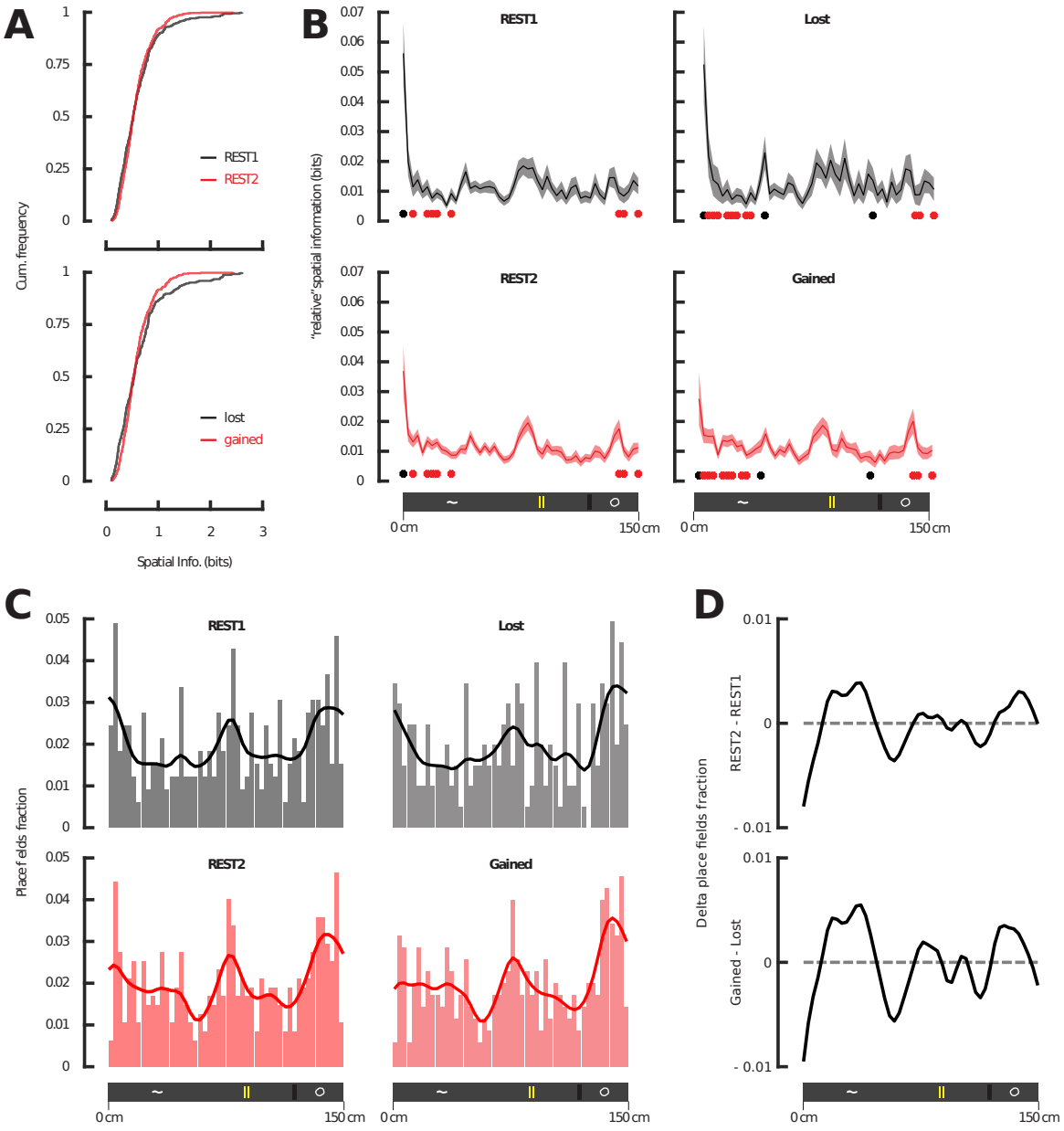
A
B
C
D
Cum. frequency
REST1
REST2
lost
gained
Spatial Info. (bits)
"relative" spatial information (bits)
REST1
Lost
REST2
Gained
0 cm
150 cm
Placefields fraction
REST2 - REST1
Gained - Lost
Delta place fields fraction

Figure 3.7: Rearrangements of ensemble neurons subsequent to behaviour are tied to features in the environment. (A) Cumulative densities of the spatial information content of co-active neurons in REST1 and REST2 (n=3 mice; n=13 sessions; n=413 REST1 neurons; n=617 REST2 neurons). Spatial information was calculated for each neuron from the RUN data, and is a measure of the amount of information conveyed by individual neurons about the animal's position during RUN. The densities were further separated by neurons that were no longer co-active following exposure ('lost') and those that became co-active ('gained') (n=169 stable/overlapping neurons). The spatial information content of 'lost' neurons was unexpectedly higher than that of 'gained' neurons(one-tailed two-sample K-S test $p<0.05$). To clarify the reasons behind this discrepancy, we skipped the summation step in Eq. 3.1. In doing so, we obtained the relative contribution of each position to the total spatial information for a given neuron. (B) The relative contribution to spatial information of co-active neurons as a function of position for all neurons in A. Lines denote mean, and shaded areas depict S.E.M. Black dots denote positions where the spatial information content was higher for REST1 neurons than REST2 and vice versa for red dots (Wilcoxon signed-rank test; $p<0.05$). For the 'lost' neurons, the majority of spatial information came from the beginning of the track, immediately after a reward was received. This over-representation accounts for the overall higher spatial information reported in A. The spatial distribution of information was more uniform for 'gained' neurons with significantly less sensitivity towards the receiving of reward. 'Gained' neurons also contained higher spatial information in the first 30 centimetres leading to the first cue, and in the last 10 centimetres leading to the reward. (C) Histograms for the spatial distribution of place field centres for REST1 and REST2 neurons that were classified as 'place cells'. Place field centres were identified for individual neurons from the RUN data (see Methods). Curves plotted over the histograms were obtained by smoothing the fractions with a 5 centimetres Gaussian kernel. The distribution for 'lost' and REST1 neurons expressed high variance over successive locations. In comparison, 'gained' and REST2 neurons showed less dispersion (coefficient of variation $c_v = \frac{\sigma}{\mu}$; $c_v = 0.506$ REST1; $c_v = 0.474$ REST2; $c_v = 0.564$ 'lost'; $c_v = 0.477$ 'gained'; n = 50 spatial bins). (D) The difference was taken between the curves in C to reflect the difference in the fractions of place fields between REST2 and REST1 (top), and 'gained' and 'lost' (bottom), as a function of position. REST2 and 'gained' neurons contained more place fields near the locations of landmarks than REST1 and 'lost' neurons. Conversely, REST1 and 'lost' neurons expressed more place fields at the beginning of the track, right after the location where rewards were delivered. These results are consistent with the results in panel B.

3.4 Discussion

In this study, we investigated the resting-state neural dynamics of the retrosplenial cortex, and identified neural patterns of activity associated with prior spatial experiences. We found that the structure of resting-state activities in retrosplenial neuronal populations is distinctly stable following 'virtual' spatial navigation. This feature impacts traditional techniques used for analysing reactivation, such as explained variance, template matching and principal components analysis, which all require strong correlations between task-related and resting-state population vectors (Tatsuno et al., 2006; Peyrache et al., 2009). In contrast, only a sparse subset of RSC neurons formed ensembles

that expressed task-relevant activities during resting epochs. With population sparseness comes the necessity to record simultaneously from large numbers of neurons in order to extract statistical effects. This is made especially difficult for microelectrode recordings that target superficial layers of the neocortex, where stable large-scale recordings are often impaired by tissue damage. In the RSC, the bulk of hippocampal back-projections terminate in layers II/III that are rich in NMDA-receptors (Monaghan and Cotman, 1985). Widely held theories postulate that the presence of rapid plasticity mechanism in the neocortex support the gradual consolidation of new and unique experiences through outflow of information encoded by the hippocampus (Teyler and DiScenna, 1986b; Teyler and Rudy, 2007; McClelland et al., 1995; McNaughton, 2010; Schwindel and McNaughton, 2011; Skelin et al., 2019). Support for the existence of such a model in the RSC has been provided by our recent works: first, the generation of spatial sequences encoded by the RSC require an intact hippocampus (Mao et al., 2018); second, a greater fraction of spatially-receptive neurons reside in superficial RSC layers rather than deep layers (Mao et al., 2017). The advantage of the two-photon technique in imaging large populations of superficial layers neurons may therefore have contributed to our observation of experience-related activity patterns during resting periods. It remains to be determined whether superficial and deep layers participate equally in reinstatement.

The presence of offline activities associated with landmark location comes as no surprise when considering the large body of studies that are consistent with such a function. Lesioning the RSC impaired the ability of the head-direction system to sustain stable representations using landmarks (Clark et al., 2010). Correspondingly, head-direction cells in the dysgranular RSC are reoriented by local landmarks within a global environment (Jacob et al., 2017). In humans, patients with lesions to the RSC were able to identify familiar landmarks, but experienced substantial impairments in remembering the spatial relationships between landmarks (Takahashi et al., 1997; Aguirre and D'Esposito, 1999). fMRI studies revealed that the RSC was most engaged during presentation of stable landmarks (Auger et al., 2012). Together, the combination of rodent and human evidence indicates that the RSC performs an active role in landmark encoding during both online and offline periods. Rather perplexing is the seemingly undervalued role of the RSC when considering the amount of information available to the structure. Indeed, during online periods, spatial coding in the RSC emerges as continuous representations over the environment (Mao et al., 2017; Alexander and Nitz, 2017). Further studies are required to resolve whether offline RSC activities imply spatial

cognitive functions beyond landmark encoding.

One major outstanding question generated from our results is the time course of memory trace events in the RSC in relation to the rest of the brain. According to prevalent theories of consolidation, novel experiences first encoded by the hippocampus are gradually merged with existing representations in the neocortex (Squire, 1992; McClelland et al., 1995; Frankland and Bontempi, 2005). As one of the main output structures of the hippocampus, we expect the RSC to reinstate experience-associated memory traces during early stages of consolidation. Answering this question is beyond the scope of the current article, as the animals used in this study had been previously exposed to the behavioural task. Meta-analytically, reactivation observed in other brain regions including the Prefrontal Cortex (Euston et al., 2007), the posterior Parietal Cortex (Wilber et al., 2017) and the visual cortex (Ji and Wilson, 2007) did not employ naïve animals either. Therefore, it can be hypothesized that consolidation occurs over a shorter time scale in the RSC; perhaps stronger reactivation would be observed during the early stages of task acquisition. Further investigation is required to uncover the time course of memory consolidation in the RSC in relationship to other regions. Altogether, our results demonstrate the involvement of the RSC in spontaneous retrieval of spatial landmark information during offline periods.

Chapter 4

Cortical Reactivation of Spatial and Non-Spatial Features Coordinates with Hippocampus to Form a Memory Dialogue[57,58]

Abstract

Episodic memories comprise diverse attributes of experience distributed across neocortical areas. The hippocampus is integral to rapidly bind these diffuse representations, as they occur, to be later reinstated during offline consolidation. However, the nature of the information exchanged during this hippocampal-cortical dialogue remains poorly understood. A recent study has shown that the secondary motor cortex carries two types of representations: place cell-like activity, which were impaired by hippocampal lesions, and responses tied to visuo-tactile cues, which became more pronounced following hippocampal lesions. Using two-photon Ca^{2+} imaging to record neuronal activities in the secondary motor cortex, we assessed the cortical retrieval of spatial and non-spatial attributes from previous explorations in a virtual environment. We show that, following navigation, spontaneous reactivations occurring during rest convey varying degrees of spatial (trajectory sequences) and non-spatial (visuo-tactile attributes) information, while reactivations of non-spatial attributes tend to precede reactivations of spatial representations surrounding hippocampal sharp-wave ripples.

4.1 Introduction

The encoding, storage and retrieval of episodic memories require a carefully orchestrated exchange of information between the hippocampus and the neocortex (Marr, 1971; Teyler and DiS-

[57] Published in Chang et al. (2023). Reproduced, here, is the authors' final draft following peer review, but prior to editorial proofreading. Reproduced with authors' permission.

[58] **Author contributions** H.C., I.M.E., M.H.M. and B.L.M. designed the experiment. H.C., I.M.E. and A.R.N. conducted the experiments and collected the data. H.C. and I.M.E. performed the surgeries. H.C. analysed the data and wrote the manuscript, which all authors helped to revised.

cenna, 1986b; Buzsáki, 1996; McClelland et al., 1995). During periods of active behaviour, unique experiences are comprised of diverse attributes that span a multitude of modalities and cortical sites. The task of associating and storing these distributed patterns of activity as they occur, so that they may be recalled at later times, has been attributed to the hippocampal network (Marr, 1971; Teyler and DiScenna, 1986b; McClelland et al., 1995). An important line of experimental evidence reinforcing this theoretical principle stems from the study of *reactivation*; during *offline* periods (i.e., sleep and quiet-wakeful states), patterns of activity that are related to previous experiences are spontaneously and repeatedly reinstated. Consistent with the notion of a distributed memory system, multiple cortical regions have been implicated in the reactivation of behavioural features, which are congruous with the known functions of their respective regions (e.g., Ji and Wilson, 2007; Jadhav et al., 2016; Ólafsdóttir et al., 2016; Wilber et al., 2017; Rothschild et al., 2017; Eckert et al., 2020). For instance, the retrosplenial cortex, a region responsible for the encoding of stable landmarks, reactivates for landmark locations (Chang et al., 2020), while the medial prefrontal cortex, implicated in cognitive flexibility and rule learning, reactivates for task rules (Peyrache et al., 2009). The reactivations of these distributed features are also temporally coordinated across cortical regions (Hoffman and McNaughton, 2002).

In rodents, a substantial portion of the variability in cellular activity of hippocampal neurons is explained by space, whereby individual neurons' discharges are correlated to specific locations in a spatial environment — these are known as *place cells* (O'Keefe and Nadel, 1978). Accordingly, reactivated patterns in the medial temporal lobe and medial prefrontal cortex manifest themselves as sequences of trajectories undertaken in a previously explored environment (Skaggs and McNaughton, 1996; Nádasdy et al., 1999; Lee and Wilson, 2002; Euston et al., 2007; Malvache et al., 2016). As is the case between cortical regions, the hippocampus reactivates in coordination with many cortical areas (Qin et al., 1997; Ji and Wilson, 2007; Jadhav et al., 2016; Ólafsdóttir et al., 2016; Rothschild et al., 2017). These coordinated reactivation events often occur in conjunction with hippocampal *sharp-wave ripples* (SWRs) — discrete high frequency events expressed by synchronous CA1 neuronal populations during offline periods — which are temporally coupled with transient cortical activities (Chrobak and Buzsáki, 1996; Siapas and Wilson, 1998; Kudrimoti et al., 1999; Sirota et al., 2003; Battaglia et al., 2004; Isomura et al., 2006; Wierzynski et al., 2009; Logothetis et al., 2012). Though a continuous gradient likely marks the timing between hippocampal

and neocortical activations (Abadchi et al., 2020), the onset of cortical reactivations typically precedes that of hippocampal SWRs by an order of ∼50-200 ms (Ji and Wilson, 2007; Peyrache et al., 2009; Rothschild et al., 2017; Wilber et al., 2017). Conversely, hippocampal population activities reliably trigger responses in the cortex (Chrobak and Buzsáki, 1996; Sirota et al., 2003; Wierzynski et al., 2009; Jadhav et al., 2016; Ólafsdóttir et al., 2016), while the content of hippocampal activities during SWRs can predict subsequent cortical patterns (Rothschild et al., 2017).

In principle, if the hippocampus were to provide a set of associative links — an 'index' — pointing to specialized patterns distributed over cortical regions (Teyler and DiScenna, 1986b; McNaughton, 2010), then it follows that the contents of the activities expressed during coordinated reactivations between the hippocampus and the neocortex should share similar features from previous experiences. Such was the case reported by multiple studies, which collectively mapped several cortical areas that reactivated for patterns complementary to those reactivated by the hippocampus (Ji and Wilson, 2007; Qin et al., 1997; Jadhav et al., 2016; Ólafsdóttir et al., 2016). However, in these studies, the shared features in question were correlated over the spatial dimension, where the contents of reactivations were linked to specific locations from previous explorations. In one exceptional study, the contents of the reactivated patterns observed in the auditory cortex were related to specific tonal stimuli (Rothschild et al., 2017). However, the exact functional relevance of the patterns reactivated in conjunction by the hippocampus could not be determined. Given that diverse cortical areas encompass a wide range of cognitive processes, the functional links that would permit associations to be formed between the hippocampus and the cortex remains to be elucidated. Specifically, the spatial and non-spatial aspects of experiences, which seemingly constitute an important functional basis for the hippocampal-cortical dialogue (Buzsáki, 1996) in a distributed memory system, require further reconciliation.

Recently, a study has reported that, during a virtual spatial navigation task, two kinds of representations were concurrently supported by primary and secondary neocortical regions: place cell-like activities which were impaired by hippocampal lesion, and responses related to visuo-tactile cues which became more pronounced following hippocampal lesion (Esteves et al., 2021). Premised on this finding, the current study aimed to explore whether these distinct representations, which are likely of hippocampal and cortical origin respectively, are reactivated concurrently by the same cortical region, and if so, whether they engage in interactions that are reflective of an exchange of

information between the hippocampus and the neocortex. Using two-photon calcium imaging, we simultaneously recorded populations of neurons in the superficial layers (LII/III) of the secondary motor cortex (M2) of Thy1-GCaMP6s transgenic mice ($n = 14$ animals; Fig. 4.1**b**; Table B.1; ~ 19 frames-per-second). Water-restricted mice were trained to navigate on a linear treadmill for a drop of sucrose water at the end of each lap (Fig. 4.1**a**). The treadmill consists of a 150 cm long belt lined with four distinct visuo-tactile cues. To characterize awake reactivation patterns, we acquired resting state activity for 10-20 min before (REST1) and after (REST2) virtual exploration (RUN) (Fig. 4.1**c**).

We found that, in the resting period following virtual navigation, awake reactivation events carry information that is, in varying degrees, related to trajectory segments in space or visuo-tactile attributes. Reactivations that are biased towards visuo-tactile features tend to occur earlier in time relative to hippocampal SWRs, whereas reactivations that are more related to spatial trajectories tend to occur later. Furthermore, concurrent reactivations of cue and trajectory information reinstate similar features from previous experiences, where the reactivated trajectory segments tend to coincide with the locations of the reactivated cues. These results are commensurate with the theorized notion that cortical reactivation of non-spatial attributes may act as partial information to seed the hippocampal retrieval of associated spatial sequences. These spatial representations are then propagated back to the cortex, hence forming a functional cortical-hippocampal-cortical loop (cf. Rothschild et al., 2017).

4.2 Methods

4.2.1 Animals and surgical procedure

All experiments were conducted in compliance with the guidelines established by the Canadian Council on Animal Care and were approved by the Animal Welfare Committee at the University of Lethbridge. Mice were single-housed under a 12 h light/dark cycle following surgery with mouse chow and water provided *ad libitum* until the beginning of experimentation. A total of 14 Thy1-GCaMP6s transgenic mice were used (aged from 2-8 months old). All mice received a 5 mm cranial window implant over the dorsal cortex (AP: -3 to +2 mm from bregma; ML: centred on midline; Fig. 4.1**b**) following the same procedures as previously described (Chang et al., 2020; Esteves et al., 2021). 11 mice were also implanted with a bipolar electrode (0.5 mm tip separation; 50.8 µm Teflon-

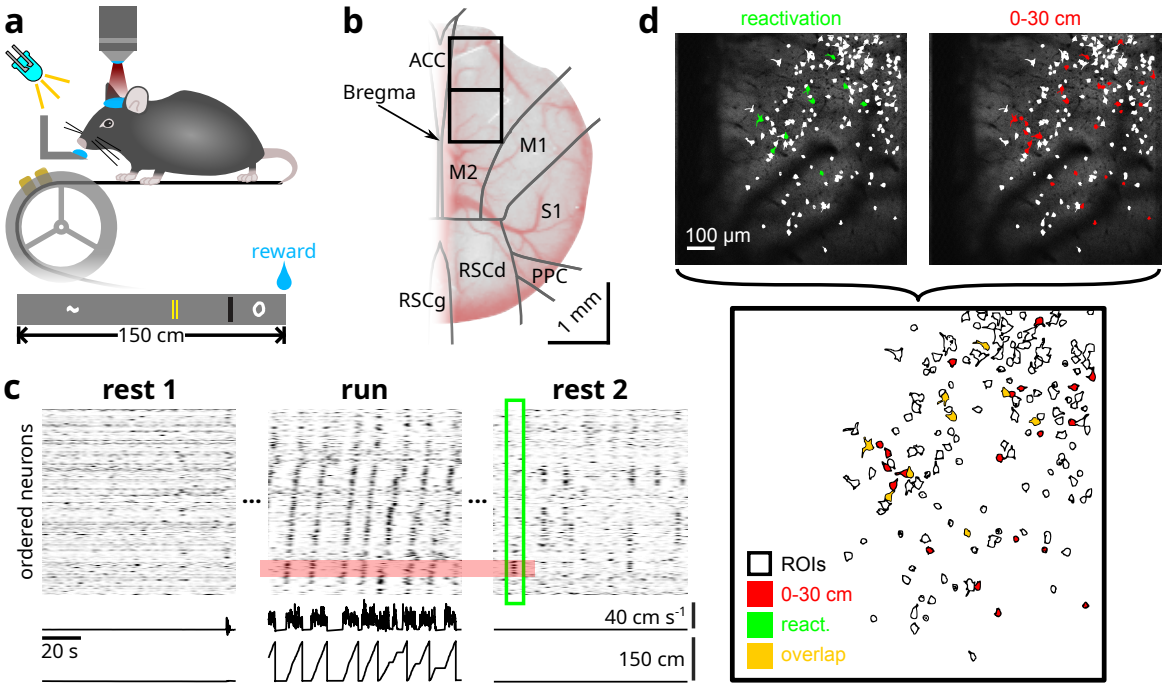

Figure 4.1: Two-photon imaging and behavioural paradigm. **a** Water-deprived mice were head-restrained over a 150 cm long treadmill belt, over which were mounted several visuo-tactile cues. An LED light illuminated the path in front of the animals so that they are able to see the incoming cues. **b** Example of a cranial window implant. Black boxes delineate the two imaging windows, of which one is chosen for each mouse on the basis of signal quality and the presence of bone-regrowth. Regional boundaries were determined from the Allen Common Coordinate Framework atlas (CCF v3). **c** Each imaging session was divided into three 10-20 min blocks. Before (REST1) and after (REST2) exploration (RUN), animals rested quietly over the belt, during which the treadmill was clamped. A 100 s segment from each imaging block is illustrated for one example session. The time courses of deconvolved $\Delta F/F_0$ for all simultaneously imaged neurons were sorted by the location of peak activity during RUN. Animal position and linear velocity are shown below. Notice that, during REST2, co-active ensembles of neurons are clustered with respect to the sequential activity patterns over spatial locations during RUN. **d** ROIs of neurons that were active during the 1 s time window surrounding the single reactivation event delineated in **c** (green), and those of neurons that were active at locations 0-30 cm, roughly corresponding to the span of the trajectory encoded by the ensemble of reactivating neurons. The overlapping ROIs between those two groups are shown. All neurons that were part of the reactivation event were also active in the first 30 cm during behaviour.

coated stainless-steel wires from A-M Systems) in dorsal CA1 stratum pyramidale ipsilateral to the imaging site. Electrodes were inserted from the posterior edge of the imaging window (AP: -3 mm; ML: 1.8 mm) with a 30° angle-of-approach along the AP axis pointed anteriorly. To determine the depth at which to lower the electrode, the signal was monitored through a speaker during lowering of the electrode. An abrupt increase in spikes signals that the lower tip has reached stratum pyramidale. Lowering the electrode by another ~ 0.5 mm should result in another burst of spikes, at which point the upper tip is within stratum pyramidale while the lower reference tip should be in stratum moleculare. Electrodes were implanted in the same hemisphere in which imaging was conducted. Two of the implanted electrodes showed poor signal quality, and the mice carrying these implants were excluded from electrophysiological analyses.

4.2.2 Behavioural paradigm

Water-restricted mice were trained to run over a 150 cm long linear treadmill for a drop of sucrose water at the end of each lap, while under head-fixation (for training protocol and treadmill design, see (Mao et al., 2017, 2018; Chang et al., 2020; Esteves et al., 2021)). The treadmill belt was lined with four distinct visuo-tactile cues (Fig. 4.1**a**) constructed out of hot glue (first and last cues), a strip of soft Velcro (second to last cue) and two strips of reflective tape (second cue). An LED light source was positioned in front of the animal to illuminate incoming cues. Before and after running, the treadmill was locked and the animals were habituated to rest quietly on the belt in addition to being trained to run. In most cases, consistent running and resting are learned after two weeks to a month of training. This typically required a week of clamping the belt for ~ 10 min before and after running. At first, animals will attempt to move the treadmill belt, but will soon learn to associate the absence of the reward port (moved away during rest periods) with immobility. With the belt clamped, the rotation encoder on the treadmill is still sensitive enough to detect motions from the animals, and these epochs were removed from subsequent analysis. Each imaging session proceeded with 10-20 min of resting (REST1), followed by ~ 8 min of running (RUN), and concluded with another 10-20 min of rest (REST2). Sessions in which animals ran less than 10 laps were discarded from further analysis.

4.2.3 Two-photon imaging and ROI tracking over days

Imaging was performed under a Thorlabs Bergamo II microscope powered by a Ti:Sapphire femtosecond pulsed laser (Coherent Chamelion Ultra II) tuned to an excitatory wavelength of 920 nm. The light beam was rasterized by Galvo-Resonant scanners bidirectionally at a frame rate of $\sim$ 19 Hz and focused onto the tissue via a 16$\times$ water immersion objective (Nikon; NA = 0.8; 80-120 mW output power measured at the sample). Emitted light signals were amplified using a GaAsP photomultiplier tube (Hamamatsu) and digitized to a resolution of 800×800 pixels at 16 bit precision. The imaging FOV consisted of a 835µm$\times$835µm square plane acquired at a depth of 100-200µm from the cortical surface to reach layers II/III. Depending on the imaging quality and the presence of bone-regrowth obstructing the window, the imaging window spanned anywhere between 0 mm and +1.67 mm over the AP axis and was centred on 0.5 mm ML (i.e., edging the superior saggital sinus; Fig. 4.1**b**).

Suite2p (Pachitariu et al., 2016) was used to identify the neuron ROIs, and the extracted fluorescent traces were deconvolved by constrained non-negative matrix factorization (Pnevmatikakis et al., 2016). The automatically detected ROIs were manually curated to remove any false-positives. In 10 animals, the same FOV was imaged across consecutive days, where specific landmarks such as blood vessels and neurons were used to guide the experimenter to manually align the FOV to that of the day-1 reference (Fig. B.6). To identify persistent neurons across two imaging sessions, the ROIs masks of the two session were binarized and registered against each other (Fig. B.9). Registration was constrained to rotation and translation only and was achieved through finding the nearest local minimum from the non-shifted images using a custom *direct search* gradient-free solver. ROIs that share 50% or more overlapping pixels were identified as the same neuron (percentage overlap calculated as the Jaccard index $\frac{O}{A+B-O} \times 100\%$, where out of A and B pixels contained by two candidate ROIs, O pixels are overlapping).

Given the nature of the present study, opting for the secondary motor cortex over the somatosensory cortex would seem rather counterintuitive. The choice of imaging the secondary motor cortex is a practical one. Because the cranial window implant is designed to be attached flush against the dorsal aspect of the skull (our microscope does not rotate), we were limited to dorsal cortical regions. Out of these regions, only the hindlimb regions of the somatosensory cortex were accessible for imaging. This leads to a marked "spatial delay" in the neuronal responses tied to visuo-tactile

cues, which could bias subsequent analyses. By contrast, the cue-related responses in the secondary motor cortex were much more centred around cue locations. Furthermore, bone-regrowth tends to occur over the lateral edges of the imaging window. This affliction scarcely impacted the midline regions, away from the anterior edge of the window, in our experience. These experimental limitations led to the pragmatic consideration of imaging the secondary motor cortex.

4.2.4 Electrophysiology and detection of SWRs

Local field potential was amplified 1000× and band-pass filtered from 0.1-10,000 Hz through a Grass 7P122G amplifier. The analogue output was digitized using a Digidata 1550B unit sampling at 192 μs intervals (Axon pCLAMP acquisition software). The signal was down-sampled to 2.6 kHz for analysis. To detect SWRs, local field potential was band-pass filtered from 150-250 Hz using a 400-order FIR filter. A RMS power envelope was extracted using an 8 ms sliding window. Amplitude regions exceeding 3 standard deviations from the mean were labelled as ripple events. 75% of this threshold was subsequently used to identify the times of onset and offset for each event. Ripple events must be at least 3 cycles long. Events that occurred less than 250 ms from a previous event were discarded (merged with the previous SWR).

4.2.5 Detection of spatially-selective cells

The methods used for classifying neurons that express spatial encoding have been described extensively in previous works (Chang et al., 2020; Esteves et al., 2021). Briefly, two criteria must be met for successful identification. For the first criterion, the *spatial information* (SI) conveyed by a neuron about the animal's location must exceed the 95$^{\text{th}}$ percentile of a shuffled distribution. Spatial information was computed as (Skaggs et al., 1993)

$$I = \sum_{i=1}^{N} p_i \frac{f_i}{f} \log_2 \frac{f_i}{f}, \tag{4.1}$$

where the average neuronal activity f_i in the i^{th} bin over the total average activity f, weighed by the spatial occupancy p_i, were evaluated over $N = 50$ spatial bins. The null distribution of SI was obtained by circularly shifting the time-course vectors of neuronal activities by a random factor 1,000 times. For the second criterion, neurons' place fields were identified by conducting a *continuous wavelet transform* over the spatial tuning curve of the neurons using a Ricker (Mexican Hat)

wavelet. The scales evaluated were $\Sigma = \{1, 2, 3, \ldots, 50\}$ corresponding to the $N = 50$ spatial bins. Local maxima exceeding 3 median absolute deviations from the wavelet coefficients at the lowest scale of the transform ($\sigma = 1$) were identified as potential place fields. If a local maximum falls within the bounds of another maximum at a higher scale (i.e., wider place field), the candidate sitting at the lower scale (with a narrower place field) is discarded. The width of a place field must be between 5 and 80% of the total length of the environment. The mean activity within a place field must be 2.5 times higher than the activity outside of place fields. Peak activity during individual trials must occur within the place field in at least a third of the trials. Cells that supported at least one place field satisfying these constrained were determined to be spatially-selective.

4.2.6 Detection of resting-state ensembles

Agglomerative clustering was used to detect groups of neurons that expressed highly synchronized activity during quiet wakeful periods (Fig. 4.2**a**) as previously described (Chang et al., 2020). First, the time-courses of simultaneously recorded neurons were Gaussian smoothed ($\sigma = 200$ ms) in order to combat temporal jittering (cf. Fig. B.20) and to increase correlation between neuron pairs. Then, the Pearson correlation matrix between the z-scored time-courses of neuron pairs was taken. The correlation coefficients r were converted into a distance metric d by $d = 1 - r$. This value, which ranges from 0 to 2, describes how similar the time-course vectors between a pair of neurons are, with 0 being completely correlated, 1 being unrelated, and 2 being completely anti-correlated. Agglomerative clustering was then performed over this distance matrix using unweighted average distance criterion. A cut-off threshold of 0.75 was applied corresponding to an average correlation coefficient of 0.25 within each cluster of neurons. Clusters needed to contain at least 5 members in order to be classified as an ensemble.

4.2.7 Classification of cue and trajectory ensembles

Irrespective of their final label, classification of ensembles began with screening for the presence of trajectories. Ensembles which contain less than 3 spatially-selective neurons were not considered. The normalized (from 0 to 1) spatial tuning curves (activity as a function of position) were extracted for each spatially-selective ensemble neuron. We identified continuous segments in space where the activity in the tuning curve of any ensemble neuron exceeded 0.5. Segments that were formed by

less than 3 spatially-selective cells were discarded. These segments were also circularly wrapped for trajectories that crossed the starting/reward location. Out of these candidate trajectories, those with a length of less than 30 cm that spanned the centre of a cue were identified. Ensemble that carried such a segment were classified as cue ensembles, while the remaining ensembles that still contained a spatial segment were labelled as trajectory ensembles.

4.2.8 Reactivation strength and extracting reactivated features

Hierarchical clustering permits the identification of neurons that are part of co-active ensembles, but otherwise does not provide information on the temporal activity of these ensembles. To further 'fine-tune' the detected ensembles and to establish the time-courses of their resting-state dynamics, a custom PCA-ICA approach was devised (Fig. B.21). First, the deconvolved firing-rates matrix $\mathbf{X}$, with columns as individual neuron's time-course vectors, is normalized to have null mean and unitary variance. Typically, PCA and ICA extract a defined number of components, which are used as the basis vectors for a reduced space into which the firing-rates vectors are projected (dos Santos et al., 2013; Peyrache et al., 2009; Le et al., 2011). Here, the ensembles detected using hierarchical clustering serve as an initial estimate of these basis vectors defined in the matrix $\widetilde{\mathbf{W}}$. For each ensemble m out of a total of M ensembles, we define a column vector $\hat{\mathbf{u}}_\mathbf{m}$ of length N in which the ensemble neurons are labelled 1 while the remaining neurons are 0. The vector is subsequently normalized to a unit vector $\mathbf{u_m} = \frac{\hat{\mathbf{u}}_\mathbf{m}}{||\hat{\mathbf{u}}_\mathbf{m}||}$ pointing to the mean direction of the ensemble neurons:

$$\widetilde{\mathbf{W}} = \begin{pmatrix} \mathbf{u_1} & \mathbf{u_2} & \mathbf{u_3} & \cdots & \mathbf{u_M} \end{pmatrix} = \begin{pmatrix} \frac{\hat{\mathbf{u}}_1}{||\hat{\mathbf{u}}_1||} & \frac{\hat{\mathbf{u}}_2}{||\hat{\mathbf{u}}_2||} & \frac{\hat{\mathbf{u}}_3}{||\hat{\mathbf{u}}_3||} & \cdots & \frac{\hat{\mathbf{u}}_\mathbf{M}}{||\hat{\mathbf{u}}_\mathbf{M}||} \end{pmatrix}. \tag{4.2}$$

Note that $\widetilde{\mathbf{W}}^\intercal\widetilde{\mathbf{W}} = \mathbf{I}$ since there are no overlapping members between ensembles detected from hierarchical clustering. In other words, $\widetilde{\mathbf{W}}$ forms an orthonormal basis, just like the eigenvectors obtained from PCA. From here, projecting $\mathbf{X}$ into $\widetilde{\mathbf{W}}$ would yield the activity vectors over time for each ensemble defined as the mean firing-rate across all member neurons over time. However, doing so would assume that each neuron member of an ensemble contributes equally to the activation of that ensemble, which is unlikely to reflect the actual dynamics of the network given that separate pairs of neurons within the same ensemble share different degrees of temporal correlation. We will rely on the ICA algorithm to 'fine-tune' this initial estimate of the basis, which will do so by

optimizing over the quality of reconstruction so that the basis accurately captures the relationships structure in the temporal dynamics of the neuronal population.

First, de-noising is performed on $\mathbf{X}$ to limit the amount of drift in the basis vectors during ICA optimization caused by spurious relationships. This is achieved by subtracting the portion of the variance captured by the initial basis estimate from $\mathbf{X}$:

$$\hat{\mathbf{X}} = \mathbf{X} - \mathbf{X}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^{\intercal}. \tag{4.3}$$

Then, PCA is conducted over $\hat{\mathbf{X}}$:

$$\text{corr}(\hat{\mathbf{X}}) = \mathbf{\Sigma}\mathbf{\Lambda}\mathbf{\Sigma}^{-1}, \tag{4.4}$$

where $\mathbf{\Sigma}$ is the matrix of eigenvectors (principal components) and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues such that $\text{diag}(\mathbf{\Lambda}) = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_N \end{pmatrix}$. The components whose associated eigenvalues are greater than the upper bound defined by the Marčenko-Pastur law $\lambda_+ = (1+\sqrt{\frac{N}{T}})^2$ are kept (dos Santos et al., 2013; Peyrache et al., 2009). Here, T is the number of time bins in $\mathbf{X}$. In doing so, the variance in the data that is contributed by noise is removed and the remaining variance unexplained by the initial estimates is kept. Concatenating the initial basis estimate obtained from hierarchical clustering $\widetilde{\mathbf{W}}$ with the L eigenvectors associated with the significant components obtained from PCA, $\mathbf{\Sigma}_{\lambda_+} = \begin{pmatrix} \boldsymbol{\sigma_1} & \boldsymbol{\sigma_2} & \boldsymbol{\sigma_3} & \cdots & \boldsymbol{\sigma_L} \end{pmatrix}$, yields the final estimate of the transform matrix:

$$\widetilde{\mathbf{W}}' = \begin{pmatrix} \widetilde{\mathbf{W}} & | & \mathbf{\Sigma}_{\lambda_+} \end{pmatrix} = \begin{pmatrix} \mathbf{u_1} & \cdots & \mathbf{u_M} & | & \boldsymbol{\sigma_1} & \cdots & \boldsymbol{\sigma_L} \end{pmatrix}. \tag{4.5}$$

Note that it can be shown that the matrix $\widetilde{\mathbf{W}}'$ is still an orthonormal basis; if $\lambda_1, \ldots, \lambda_n$ and $\mathbf{v_1}, \ldots, \mathbf{v_n}$ are the eigenvalues and eigenvectors respectively of the covariance matrix of the residuals in eq. 4.3, then $\mathbf{W}\mathbf{v_i}^{\intercal} = 0 \quad \forall \lambda_i > 0$.

The flavour of ICA used in the present study is *reconstruction ICA* (Le et al., 2011). This method was chosen over the popular FastICA algorithm for its higher computational efficiency (unconstrained optimization) and lower sensitivity towards un-whitened data. The original data matrix $\mathbf{X}$ is first projected onto the basis defined by the estimate transform matrix $\widetilde{\mathbf{W}}'$ to obtain $\mathbf{X}'$. Following this projection, only the variance accounted by the detected ensembles and the extra variance still remaining following PCA de-noising are kept. Then, reconstruction ICA is conducted by solving

the following optimization problem:

$$\underset{\mathbf{W}}{\text{minimize}} \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{x_i} - \mathbf{x_i}\mathbf{W}\mathbf{W}^\intercal||_2^2 + \sum_{i=1}^{N}\sum_{j=1}^{k} g(\mathbf{x_i}\mathbf{W_j}), \tag{4.6}$$

where $\mathbf{x_i}$ are the rows of $\mathbf{X}'$, $\mathbf{W_j}$ are the columns of $\mathbf{W}$ and $g(x) = \frac{1}{2}\log(\cosh(2x))$ is the contrast function that acts as a soft penalty term in place of the hard orthonormality constraint found in standard ICA. Here, the initial estimate of $\mathbf{W}$ passed to the solver is simply the identity matrix of size equal to the number of components in $\widetilde{\mathbf{W}}'$, given that the data has already been projected onto the basis of the estimate. Taking $\widetilde{\mathbf{W}}'\mathbf{W}^\intercal$ and extracting the first M columns of the resulting matrix, corresponding to the number ensembles detected by hierarchical cluster, yields the final principal components. From here, the extraction of the reactivation strength time-course vectors, and the characterization of the reactivated features proceeds as originally described in (Peyrache et al., 2009). That is, projecting $\mathbf{X}$ into this components space gives the reactivation strength as a function of time for each ensemble. Reactivation events were identified when the reactivation strength exceeds three standard deviations above the mean. The onset and offset of these events were delimited by 25% of this threshold. Similarly, to extract the features encoded by resting-state ensembles during RUN periods, the corresponding firing-rate matrix during RUN can be projected into this component space (Fig. B.22). Then, computing the mean ensemble activity over spatial locations gives the reactivated features.

4.3 Results

4.3.1 M2 ensemble dynamics during resting state

We began with a characterization of the resting state activity in the secondary motor cortex, while screening for evidence of reactivation of task-related patterns. At the population level, neuronal activities exhibited higher synchrony in the ultra-slow frequency range (0.05-0.5 Hz) during RUN, as compared to during awake quiescent periods (Fig. B.1**a-b,d,g**). This synchrony likely arose from the entrainement of cortical activities by locomotion (Fig. B.1**c**). In parallel, a decrease in the rate of calcium transients was also associated with running (Fig. B.1**f**), a phenomenon that has been previously described (Fisher et al., 2016). In contrast, during awake quiescence, population activities expressed stronger power in the slow oscillatory ranges (1-10 Hz) (Fig. B.1**b,e**), in accordance with

of a synchronized cortical state (see (Harris and Thiele, 2011; Fernandez et al., 2017)). Overall, the secondary motor cortex exhibited distinct population dynamics across awake behavioural states.

Consistent with previous reports (Malvache et al., 2016; Chang et al., 2020; Grosmark et al., 2021), activities during the resting periods were characterized by spontaneously co-activating groups of neurons (Fig. 4.1**c**; Fig. 4.2**a**) — henceforth referred to as *ensembles*. Between REST1 and REST2, ensembles were composed of a comparable number of neurons (Fig. B.2**a-b**) with an average of 6.92 and 7.75 members respectively. However, a greater number of ensembles were detected in REST2 (Fig. B.2**c**), which suggests that the network, following locomotion, exhibited more stereotyped patterns of activity. We tested the latter hypothesis by carrying out a *principal component analysis* on the correlation matrices of time series vectors for REST1 and REST2 (Fig. B.2**e**). The results show that fewer components were needed to explain a greater fraction of the total variance in the neuronal population of REST2, which affirms the notion that the population activity contained more recurring patterns. Moreover, REST2 ensembles activated at a lower rate than REST1 ensembles (Fig. B.2**d**). Taken together, these results suggest that following exposure to a virtual environment, the spontaneous activities of the secondary motor network became more stereotyped, with similar patterns of population activity being reinstated over time.

To understand how these synchronous activities could be related to functional patterns, we examined the characteristics of the ensemble neurons during active behaviour. In REST2, synchronous ensembles preferentially recruited neurons that expressed spatial-selectivity during active locomotion. Indeed, REST2 ensembles contained a median of 43.8% spatially-selective cells, higher than the 32.6% in REST1 ensembles (Fig. 4.2**d**). When the difference is taken between these percentages and the total fraction of spatially-selective cells in a given session, we found that the median of this difference was not significantly different from zero for REST1, while it was higher than zero for REST2 (Fig. 4.2**e**).

This suggests that the constituent neurons of REST2 ensembles were more likely to include cells that were spatially selective during RUN than REST1 ensembles. A more robust approach for testing this relationship is through a hypergeometric test, which aims to model the probability that a certain fraction is obtained by drawing at chance from a population. This test revealed that a significantly higher proportion of REST2 ensembles had a high composition of spatially-selective cells (Fig. B.3). Overall, REST2 ensemble neurons expressed higher *spatial information* content

(Fig. 4.2**f**).

This propensity for REST2 ensembles to recruit neurons with high spatial information suggest that M2 may be reactivating features related to recent experiences. The first evidence for reactivation came from explained variance analysis (Kudrimoti et al., 1999), which revealed that REST2 population activity accounted for a higher percentage of the variance in RUN than REST1 activities (Fig. 4.2**g**). Given that activity patterns during active behaviour were characterized by continuous sequences of place cell-like activity, this result suggested that most of the variance exists in the correlational structure between neurons with nearby place fields. Indeed, REST2 ensembles tend to contain spatial cells with place fields that were in proximity to each other (Fig. 4.1**c-d**; Fig. 4.2**b-c**). Next, we investigated the significance of this organization.

4.3.2 Cue and trajectory information are jointly reactivated

Our previous findings (Esteves et al., 2021) suggested that the secondary motor cortex supports two distinct encodings during virtual navigation: place cell-like activities which were impaired by hippocampal lesion, and responses associated with visuo-tactile cues which became more pronounced following hippocampal lesion (Fig. B.4). We investigated whether these separate neuronal representations may be preferentially reactivated in cortex. Intriguingly, resting-state ensembles that consisted of spatially-selective cells sharing neighbouring place fields could be classified into two separate categories based on the contents of their reactivated features. On the one hand, we noted ensembles that were composed of neurons with nearly identical spatial tuning profiles. These cells often supported multiple place fields (Fig. 4.3**a-b**) that strongly overlapped with the locations of visuo-tactile cues (Fig. 4.3**f**), while the same place fields tended to be shared between all ensemble members (Fig. 4.3**a,d**). Given their propensity to encode cue information, we dubbed these groups 'cue ensembles'. On the other hand, we noticed ensembles that were composed of members that shared place fields in proximity to one another (Fig. 4.2**b-c**; Fig. 4.3**a**; Fig. B.5), although they were not completely overlapping (Fig. 4.3**d**). Collectively, they formed short segments of continuous trajectories in space with a median length of 36 cm for REST1 and 33 cm for REST2, corresponding to 24% and 22% of the length of the environment respectively (Fig. 4.3**a**; Fig. B.6). These ensembles will be termed 'trajectory ensembles'.

Both trajectory and cue ensembles contained a significantly higher number of spatially-selective

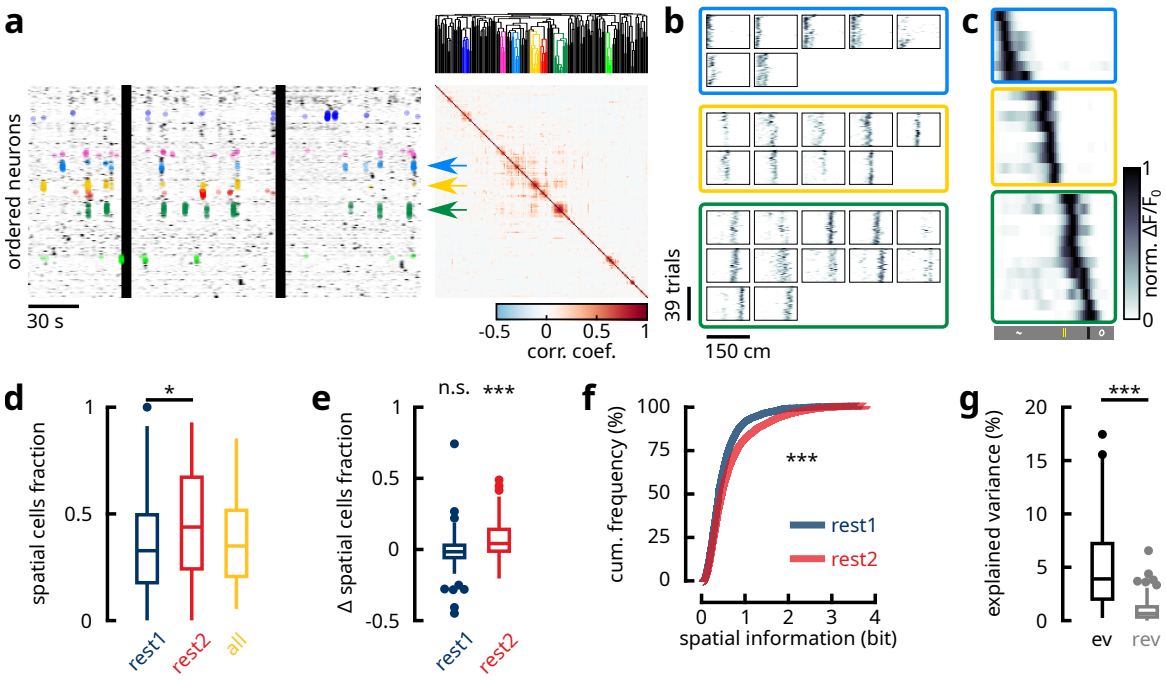

Figure 4.2: M2 resting ensembles preferentially recruit neurons expressing high spatial information. **a** Illustration of hierarchical clustering method used to detect synchronous neuronal ensembles during rest. A 1 min segment of REST2 from the same recording as Fig. 4.1**c-d** is shown. Neurons were sorted by similarity in time course vectors. Black vertical bands highlight movement epochs, which were excluded from analysis. The Pearson correlation matrix between the time-series of neuron pairs is illustrated, along with the dendrogram obtained from hierarchical clustering. Neurons part of a synchronous ensemble were grouped by the same colour. Coloured dots over the time-series mark reactivation events detected for individual ensembles. **b** The mean deconvolved $\Delta F/F_0$ as a function of position and laps for all individual neurons in the three synchronous ensembles in **a**. Neurons were sorted by place field location. Notice that neurons of the same ensemble share neighbouring place fields. **c** The mean activity of the neurons in **b** as a function of position. **d** The fraction of spatially-selective neurons that were part of REST1 and REST2 synchronous ensembles, and the overall fraction of spatial cells in all recording sessions ($n = 14$ mice; $n = 86$ sessions; $p = 0.034$ for REST1-REST2; not significant between other groups; Kruskal-Wallis H test). **e** Difference between the fraction of spatially-selective neurons within synchronous ensembles and the overall fraction of spatial cells ($n = 14$ mice; $n = 86$ sessions). REST2 ensembles contained a significantly higher fraction of spatial cells ($p < 0.001$; one-tailed Wilcoxon signed rank test for the null hypothesis of median smaller or equal to zero), while REST1 ensembles showed no significant deviation from baseline ($p = 0.92$). **f** Cumulative distribution functions of spatial information for REST1 and REST2 neurons that were part of synchronous ensembles ($n = 2672$ REST1 and $n = 3613$ REST2 neurons; $p < 0.001$ in both Kolmogorov–Smirnov test and Mann-Whitney U-test). **g** Percentage of explained variance (ev) and reverse explained variance (rev) in all imaging sessions ($n = 14$ mice; $n = 86$ sessions; $p < 0.001$; paired-sample one-tailed Wilcoxon signed rank test for the null hypothesis of median smaller or equal to zero). All box plots show the median (line), the first and last quartiles (box), the minimum and maximum values (whiskers) and the outliers (circles).

cells than the ensembles that were unclassified (Fig. 4.3**e**). This suggests that the reactivated contents supported by these groups were highly specific to spatial features encoded during locomotion. In contrast, the width of the place fields in cue ensemble neurons was markedly narrower than that of trajectory ensembles (Fig. 4.3**c**). Indeed, the median of the average width of place fields within cue ensembles was 35.4 cm, which was lower than the 45.6 cm in trajectory ensemble neurons (note that place field widths and trajectory lengths were quantified using different methods, which makes it appear as though place fields are wider than the trajectories that they form; see Methods). This indicates that cue ensemble members have inherently sharper tuning profiles, which likely reflects a primary/secondary response towards cue stimuli. Overall, trajectory ensembles were more than three times as prevalent as cue ensembles in both REST1 and REST2, although a greater fraction of REST2 ensembles could be placed into either of these two categories (Fig. 4.3**g**). These fractions are consistent with the comparatively low proportions of cue-responding neurons within the total population of spatially-selective cells, which was estimated to be 16.51% (Fig. B.7). Furthermore, a comparable number of neurons comprised the two ensemble classes (Fig. B.10**c**), signifying that the classification of these ensembles was not biased by sample size.

During locomotion, neurons belonging to cue ensembles showed stronger activation than trajectory ensemble neurons (Fig. B.8**a**). This trend was reversed, however, during resting states (Fig. B.8**c**). In parallel, the mean rates of calcium transients conveyed by trajectory ensemble neurons was slightly higher compared to cue ensemble neurons during rest, while comparable rates were reported during locomotion (Fig. B.8**b,d**). This was corroborated by a heightened rate of reactivations during REST2 for trajectory ensembles compared to cue ensembles (Fig. B.8**e-f**). Altogether, these results suggest that the reactivation of two separate types of neural representations could be observed in the secondary motor cortex.

Previous studies showed that, during replay of previous behaviour-correlated patterns, the activity sequences undergo temporal compression (Lee and Wilson, 2002; Euston et al., 2007). In relation to our current findings, an interesting hypothesis may be proposed. On the one hand, if cue ensembles contain neurons that, during behaviour, respond simultaneously to the sensing of cues, then, owing to the lack of a sequential structure, they should not undergo temporal compression during reactivation. On the other hand, trajectory sequences are expected to undergo compression are previously reported. Such was indeed the case; we reported median values of optimal compression

factors of 2× for cue ensembles, compared to 30× for trajectory ensembles (Fig. B.9). Note that, owing to the low temporal sampling rate, these estimated values are expected to fall over a broad confidence intervals range (cf. Supplementary Methods).

In spite of the tendencies for ensembles to encode cue versus trajectory information, there remains the possibility for ensembles to be composed of a mixture of "cue-responsive" and "place-responsive" cells, which themselves may be encoding for conjunctive features. To test this possibility, two models were devised and fitted to the time-series of each individual ensemble neuron. The first model assumes that the neuron's response tuning curve abides by a Gaussian function over spatial locations, acting as a first-order approximation of a place cell. The second model fits a distinct firing rate at each cue location with a constant baseline firing rate, therefore reflecting a sensory stimulus-driven neuronal response. The distribution of ratio between the goodness-of-fit of the two models (measured as the likelihood ratio) did not express bimodality, which confirms the hypothesis that ensembles or ensemble neurons encode for conjunctive features between cue and trajectory (Fig. B.10**a**). However, neurons that belonged under the class labels of cue or trajectory ensembles did show a significant bias over these ratios (Fig. B.10**a-b**), which confirms that ensembles have distinct propensities to encode for cue and trajectory. This is further corroborated by the fact that cue ensembles were more accurate at decoding individual cue identities than trajectory ensembles using a Bayesian paradigm (Fig. B.10**d**).

An additional method was employed to further validate these results. Taking advantage of the aforementioned differences in temporal compression between cue and trajectory ensembles, we conducted non-negative matrix factorisation to embed the temporal compression profiles of ensembles into a reduced features space (Fig. B.11**a-b**). The resulting projections were assigned cue and trajectory class labels by an unsupervised k-means clustering method, which showed a strong correspondence with the labels assigned by our selection criteria (Fig. B.11**b-c**). Nevertheless, the projected density varied smoothly and continuously over this features space (Fig. B.11**b**). Therefore, although ensembles can be approximately discriminated based on their tendencies for encoding cues or trajectories, they are likely to exist, in actuality, over a continuum defined by the conjunctions between these two behavioural features. However, in consideration for conciseness and interpretability, we will treat these ensembles as belonging to two distinct classes for the remainder of this article.

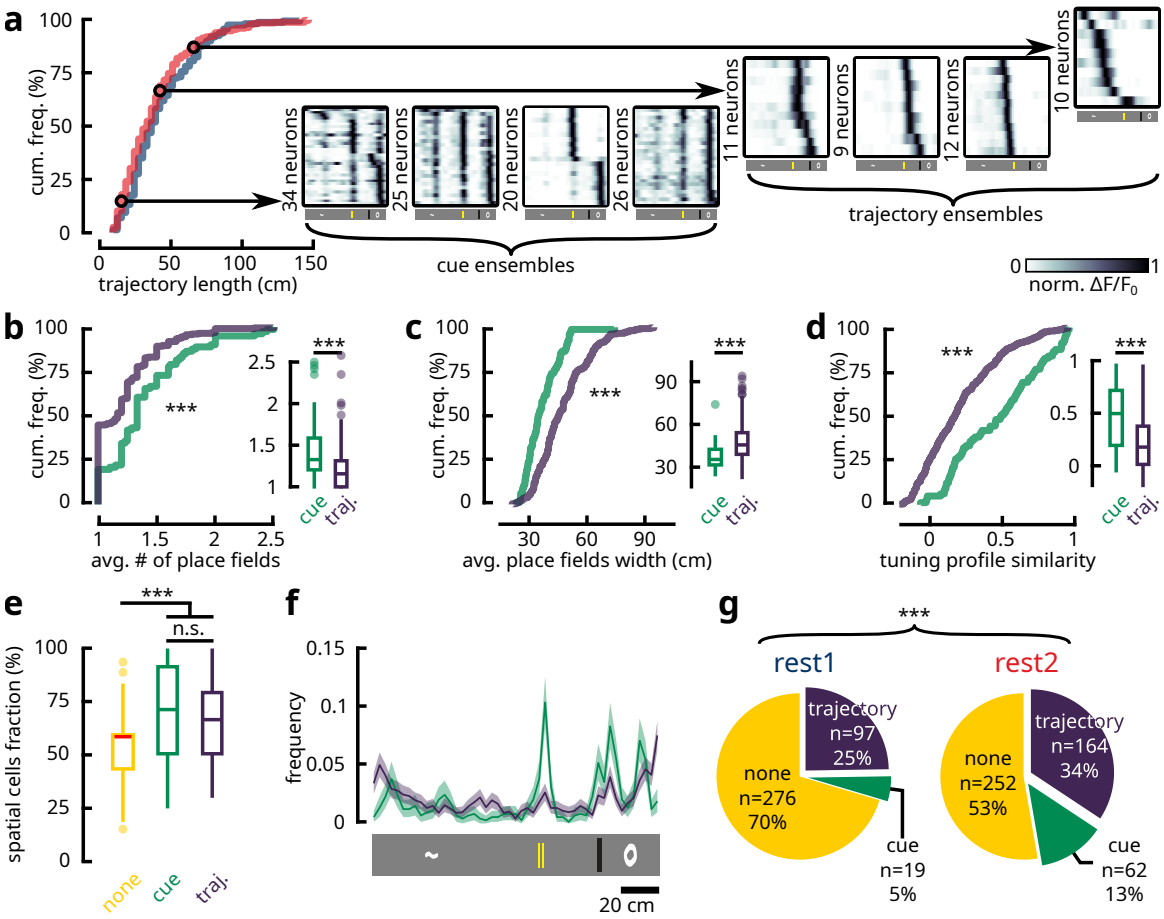
a
cum. freq. (%)
trajectory length (cm)
34 neurons
25 neurons
20 neurons
26 neurons
cue ensembles
11 neurons
9 neurons
12 neurons
10 neurons
trajectory ensembles
norm. ΔF/F0
b
avg. # of place fields
c
avg. place fields width (cm)
d
tuning profile similarity
cue
traj.
e
spatial cells fraction (%)
n.s.
none
f
frequency
20 cm
g
rest1
rest2
trajectory n=97 25%
none n=276 70%
cue n=19 5%
trajectory n=164 34%
none n=252 53%
cue n=62 13%

Figure 4.3: Resting state ensembles reactivate spatial and non-spatial features. **a** Cumulative distribution functions of the length of trajectories encoded by REST1 and REST2 ensembles ($n = 160$ REST1 trajectories; $n = 325$ REST2 trajectories; $p = 0.059$; Kolmogorov–Smirnov test). A subset of these trajectories with a length shorter than 30 cm were composed of neurons with nearly identical tuning profiles, coinciding with the locations of visuo-tactile cues. Four examples of 'cue ensembles' and 'trajectory ensembles' are shown, with the mean activity of ensemble neurons as a function of position, sorted by place field location. **b-d** Different metrics used to compare the spatial contents reactivated by cue and trajectory ensembles (two-sample two-tailed Kolmogorov-Smirnov test and two-tailed Mann-Whitney U-test; n.s. $p \geq 0.05$; *** $p < 0.001$). For each ensemble ($n = 81$ cue ensembles; $n = 261$ trajectory ensembles; $n = 528$ unclassified ensembles), neurons that were identified as spatially-selective cells were extracted. The average number of place fields (**b**) and the average width of the place fields (**c**) were computed for each ensemble. Additionally, the Pearson correlation matrix of the spatial tuning profiles between spatial cell pairs was obtained, and the average value in the upper triangle of the correlation matrix was taken (**d**). **e** The percentage of spatially-selective neurons in cue, trajectory and unclassified ensembles. Both cue and trajectory ensembles carried more spatially-selective cells than unclassified ensembles, but the fractions were comparable between the two categories of interest (Kruskal-Wallis One-way ANOVA $p < 0.001$; post-hoc tests report Bonferroni-adjusted p-values: n.s. $p \geq 0.05$, *** $p < 0.001$). For fair comparisons, unclassified ensembles with less than 3 spatially-selective cells were omitted ($n = 72$ remaining). **f** Frequency distributions of place field centres ($n = 741$ place fields for cue ensembles; $n = 1676$ place fields for trajectory ensembles) for spatially-selective cells that were recruited by cue and trajectory ensembles. Shaded area denotes 95% bootstrapped confidence interval. **g** Pie charts for the portion of trajectory, cue and unclassified (none) ensembles in REST1 and REST2. The proportions of cue and trajectory ensembles were significantly increased from REST1 to REST2 (χ^2 test; $p < 0.001$; $\chi^2 = 32.94$; df $= 2$).

4.3.3 Reactivation of cue precedes trajectory around SWRs

Having established two separate forms of reactivated features, we explored whether cue and trajectory ensembles interact differently with the hippocampus around SWRs (Fig. 4.4**a**; Fig. B.12**d**). We found that, from the onset of SWR events, cue ensemble reactivations preceded trajectory ensemble reactivations during REST2 (Fig. 4.4**b**). Similarly, from the onset of reactivation events, ripple-band power peaked at a later time during cue reactivations compared to the reactivation of trajectories (Fig. 4.4**d**). Cross-correlation analysis between the reactivation strength time vectors of cue and trajectory ensembles revealed that cue ensemble reactivations preceded trajectory ensemble reactivations by an average of 126 ms (Fig. 4.4**c**), while ripple-band power during cue ensemble reactivation also lagged 238 ms behind the spectral power during trajectory reactivations (Fig. 4.4**e**). It is important to note here that, given our image sampling rate of $\sim$ 19 Hz, the Galvo-Resonant scanners spent an average of $\sim$ 50 ms of dwell time on the tissue for each frame. Therefore, the values of time delay reported here should be interpreted with a confidence interval of $\frac{50}{3} \approx \pm 17$ ms, corresponding to the theoretical average delay between when two neurons were scanned along the slow Galvo-axis. In contrast to REST2, a less clear relationship could be determined about REST1 activities. On the one hand, no significant temporal offset was observed between cue and trajectory reactivations from the onset of ripple events (Fig. 4.4**b-c**). On the other hand, though an average delay of 263 ms separated the ripple-band powers during cue and trajectory reactivations, this effect was detected at a much lower level of statistical significance (Fig. 4.4**d-e**). Therefore, evidence for shifted temporal interactions between cue and trajectory reactivations were inconclusive with regard to REST1.

One potential confounding factor is that the reactivation of trajectories takes a longer duration, given that, in the hippocampus, the replay of place cells occurs as a temporal sequence (Lee and Wilson, 2002; Malvache et al., 2016); hypothetically, the reactivation of cues could take less time as the constituent neurons share similar spatial tuning curves and therefore need not be organized in a temporal sequence (cf. Fig. B.9). As a result, the observed delay may be influenced by the inherent timing properties of the two ensemble classes. This was not the case, however, as the median durations of the reactivation events across all three ensemble classes were comparable at $\sim$ 180 ms (Fig. 4.4**g**). To further validate this hypothesis, cross-correlation was conducted between the onset times of ensemble reactivations and those of SWRs (Fig. 4.4**f**), in which case the durations

of reactivation would no longer pose a bias. This analysis confirmed the same pattern of delay between cue and trajectory ensembles during REST2.

Overall, the majority of ensemble reactivations onsets occurred after the onset of SWRs, with a median delay of 213.2 ms (Fig. B.12**a-b**). In consideration for the continuum of cue/trajectory features, we modelled the timing differences between ensemble reactivation and SWR onsets, as a function of the likelihood ratios for ensembles' cue and trajectory encoding tendencies. It was revealed that stronger cue-bias were related to earlier reactivation onsets from SWRs, while this timing was progressively delayed with more trajectory bias (Fig. B.12**c**). Moreover, the fraction of cue and trajectory reactivation events that were associated with a SWR were comparable at 18.52% and 18.22% medians respectively in REST2 (Fig. 4.4**h**). However, a greater fraction of trajectory reactivation events were coupled with SWRs as compared to cue reactivations during REST1 (16.10% for cue and 21.93% for trajectory). In summary, these results indicate that awake reactivations of cue information following exposure to a spatial task tended to precede reactivations of trajectory information.

4.3.4 Shared features in concurrent cue/trajectory reactivation

The timing difference observed between cue and trajectory reactivations in relation to SWRs suggests that, within the same recording session, some pairs of cue-trajectory ensembles may show significant temporal relationships. To identify these interactions, we performed cross-correlations between the reactivation strength time vectors across all cue-trajectory ensemble pairs concurrently recorded within the same session (Fig. 4.5**a-b**). Out of these ensemble pairs, 15.09% (24 pairs) expressed significant temporal coupling. Consistent with the cue-trajectory delay around SWRs, the average time lag between these coupled pairs, weighed by the cross-correlation coefficients, was 93.5 ms (Fig. 4.5**c**). We reasoned that these temporally coupled ensemble pairs would share certain functional features that distinguish them from the uncoupled pairs (Fig. B.13). As a preliminary step in visualizing these potential relationships, we grouped cue ensembles into three categories based on the cue location they reactivated most strongly towards (the last two cues were grouped under the same label on account of their spatial proximity). The average reactivation strength over spatial locations in the associated trajectory ensembles appeared to follow more closely the location(s) of the cues in the temporally coupled pairs, compared to the uncoupled group (Fig. 4.5**d**; Fig. B.6).

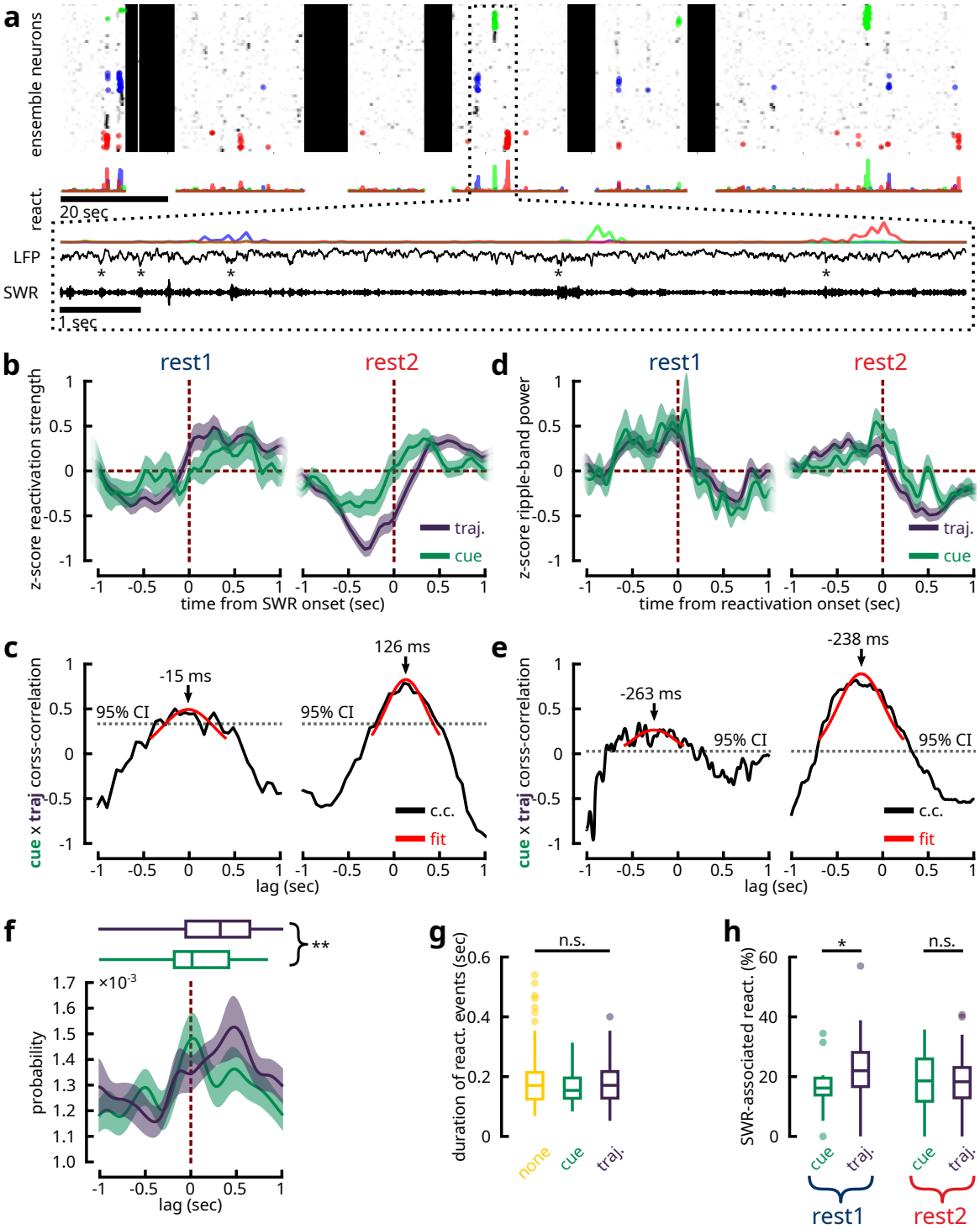
a
ensemble neurons
react.
20 sec
LFP
SWR
1 sec
b
rest1
rest2
z-score reactivation strength
traj.
cue
time from SWR onset (sec)
d
z-score ripple-band power
time from reactivation onset (sec)
c
cue x traj corss-correlation
-15 ms
126 ms
95% CI
c.c.
fit
lag (sec)
e
-263 ms
-238 ms
f
probability
×10-3
**
g
duration of react. events (sec)
n.s.
none
cue
traj.
h
SWR-associated react. (%)
*
rest1
rest2

Figure 4.4: Cue reactivations precede trajectory reactivations in relationship to hippocampal SWRs. **a** A 3 min segment of a recording with simultaneous two-photon and hippocampal LFP acquisitions. Three distinct ensembles are colour-coded with their reactivation events highlighted by dots (same as Fig. 4.2**a**). The reactivation strength over time for each ensemble is shown. Zooming in on an 8 s window, consecutive reactivations of the three ensembles can be observed, whose occurrence coincided with SWRs. The reactivation strength, the broadband LFP (downsampled to 2.6 kHz) and the bandpass filtered SWR traces (150-250 Hz) are shown. The onset times of SWRs are labelled by asterisks. **b** Average z-scored reactivation strength of trajectory and cue ensembles centred at the onset of SWRs. **c** Unbiased cross-correlation coefficients between cue and trajectory average reactivation strengths (**b**) as a function of lag period. Dotted line denotes the one-tailed 95% confidence interval. The coefficients were fitted to a Gaussian function to estimate the peak delay period. **d-e** Same as **b-c**, but for the average z-scored ripple-band power centred at the onset of ensembles reactivations. **f** Cross-correlograms between the onset time-stamps of cue and trajectory ensemble reactivations and the onset time-stamps of SWR events during REST2. Shaded area delineates 95% bootstrapped confidence intervals. Boxplots summarize the peak lag times in the cross-correlations ($n = 59$ cue, $n = 118$ trajectory ensembles; two-tailed Mann-Whitney U-test $p = 0.00268$). **g** The mean duration times for the reactivation of different ensemble classes show no significant difference across categories ($n = 62$ cue, $n = 164$ trajectory, $n = 252$ none REST2 ensembles; Kruskal-Wallis One-way ANOVA $p = 0.6421$). **h** Percentage of reactivation events associated with a SWR ($n = 16$ cue ensembles and $n = 78$ trajectory ensembles in REST1; $n = 59$ cue ensembles and $n = 118$ trajectory ensembles in REST2). During REST1, a greater number of trajectory reactivations were associated with SWRs (two-sample two-tailed Kolmogorov-Smirnov test $p = 0.010$; two-tailed Mann-Whitney U-test $p = 0.036$), while the same proportions were reported in REST2 (two-sample two-tailed Kolmogorov-Smirnov test $p = 0.189$; two-tailed Mann-Whitney U-test $p = 0.638$).

This tendency was tested by taking the Pearson correlation between the features reactivated by cue-trajectory ensemble pairs (Fig. 4.5**e**). A two-way ANOVA model indicated that temporally coupled ensemble pairs shared a significantly higher degree of similarity in their reactivated features than uncoupled pairs. For REST1 ensembles, 22.22% (8 pairs) exhibited significant temporal coupling (Fig. B.14**a**). However, no difference was found in the similarity of reactivated features between coupled and uncoupled pairs (Fig. B.14**b**). With the current sample size, this result was inconclusive as it may reflect a feature of REST1 as well as a deficiency in statistical power to discern an effect.

Drawing inspirations from the mechanisms of *pattern completion*, we further tested this relationship using a Hopfield network model (Fig. B.15**a**). Sampling still from within the same recording sessions, we trained the network to learn two reactivated trajectory features, one that was coupled with a cue ensemble and one that was not. We then presented the corresponding reactivated cue feature as a partial retrieval pattern and tested for which of the two learned representations was retrieved. This procedure was conducted over all possible combinations in sessions that satisfied the training requirements. In 73.08% of these cases, the network successfully retrieved the coupled trajectory information (Fig. B.15**b**). Analogously, the Hamming distance between coupled cue and trajectory ensemble features were shorter compared to features reactivated by uncoupled pairs (Fig. B.15**c**). Taken together, these results demonstrated that cue and trajectory ensembles that expressed interlocked timings were likely to reactivate for complementary features of previous experience.

4.3.5 Cue features exhibit higher stability across days

Lastly, we investigated whether the representations supported by cue and trajectory ensembles neurons express different degrees of stability over time. In 10 experimental animals, the same field-of-view had been imaged across consecutive recording days (Fig. B.16; Table B.1). We isolated spatially-selective cells that were part of cue or trajectory ensembles in REST2 within each imaging session, and looked for overlapping ROIs on the subsequent recording day. We observed that cue ensemble neurons maintained a stable spatial tuning profile on the following test day, while trajectory ensemble neurons saw a higher degree of remapping (Fig. 4.6**a-d**). A large density of stable trajectory ensemble neurons coincided with the locations of landmarks, suggesting that they may be cue-responsive cells organized into trajectory sequences (Fig. 4.6**a-b**).

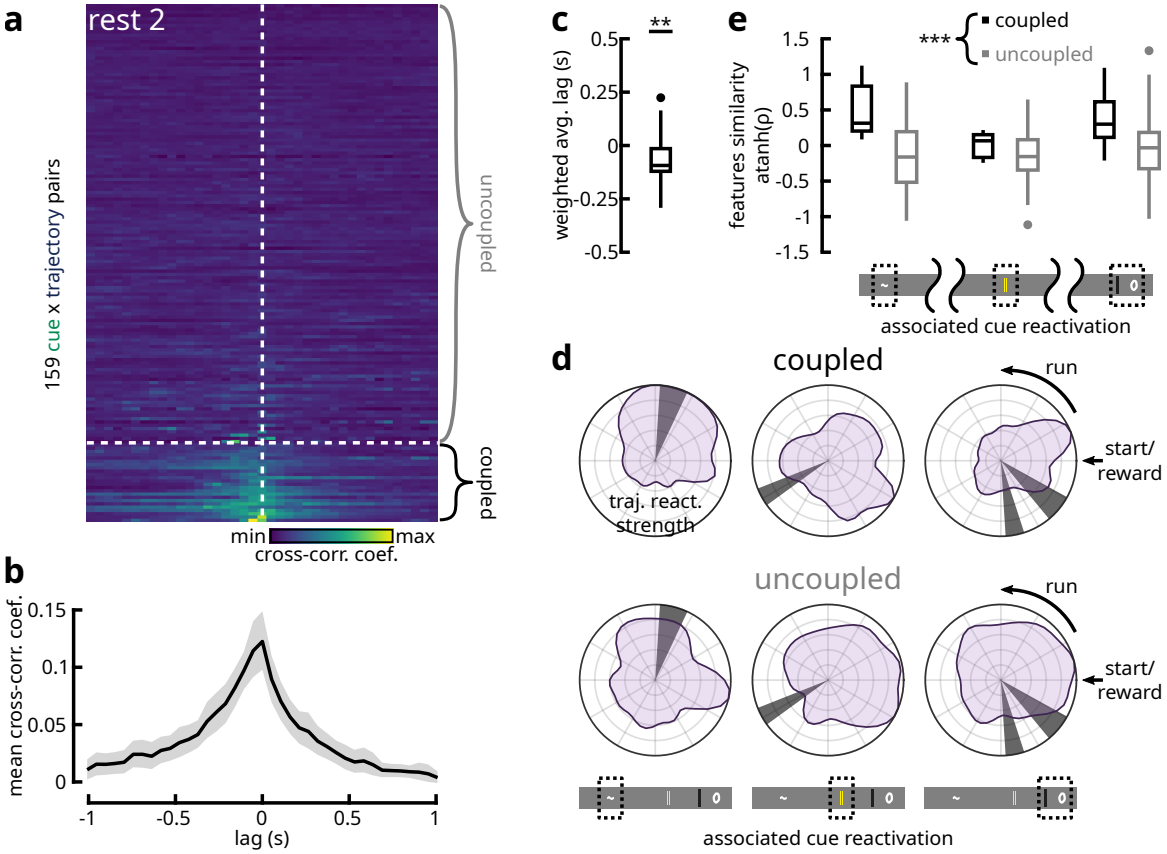

Figure 4.5: Temporally coordinated cue-trajectory ensemble pairs reactivate for similar features of previous experience. **a** Cross-correlation coefficients over a ± 1 s window between the reactivation-strength time-course vectors of cue-trajectory ensemble pairs found within the same REST2 session. Out of 159 such pairs, 24 showed significant temporal coordination. These pairs were labelled as 'coupled', whereas the remaining non-interacting pairs were referred to as 'uncoupled'. **b** Average correlation-coefficient over time lags for 'coupled' cue-trajectory ensemble pairs. Shaded area denotes bootstrapped 95% confidence intervals. **c** Average time lag, weighted by the cross-correlation coefficient, between 'coupled' pairs suggested a significant delay from the reactivation of cue ensembles to that of the trajectory ensembles (Wilcoxon sign rank test for the null hypothesis of zero median; $p = 0.0079$). **d** We separate the ensemble pairs into three groups, based on the cue location that the cue ensemble was most strongly reactivating for (the last two cues were labelled under the same group due to their spatial proximity). The average reactivation strength of the associated trajectory ensembles as a function spatial location are illustrated in polar coordinate space. Qualitatively, the reactivated trajectory features tend to follow the associated cues more faithfully in the 'coupled' pairs compared to the 'uncoupled' pairs. **e** Pearson correlation coefficients between the reactivated cue and trajectory features in 'coupled' and 'uncoupled' pairs (atanh-tranformed for normality). 'Coupled' pairs expressed a greater degree of similarity in reactivated features than 'uncoupled' pairs (two-way type II ANOVA with cue and temporal coupling as between factors; no significant interactions between factors; $p = 0.2299$; no effect of cue group on feature similarities; $p = 0.1291$; significant effect of temporal coupling on features similarity; $p < 0.001$; Shapiro's normality test $p = 0.2754$; Levene's equality of variances test; $p = 0.1736$).

Given the tendency for cue features to be more stable, we reasoned that the same cue ensembles may be recruited during resting state across days. To test this possibility, we quantified the proportion of overlapping neurons between ensembles across days for all possible combinations of REST1 and REST2 (i.e., between REST1 ensembles across days, between REST2 ensembles across days and between REST1 ensembles and REST2 ensembles across days) (Fig. B.17**a-d**). During REST2, there was a slight tendency for similar cue ensembles to be found on the subsequent recording day, more often so than trajectory ensembles (Fig. B.17**d-e**). However, no difference in persistence was found across other REST combinations (Fig. B.17**a-c,e**). Overall, cue and trajectory ensembles expressed a similar degree of persistence across days, with an average of 22.73% and 20.59% persistent ensembles respectively (Fig. B.17**f**). Despite this lack of preference, REST2 contained a substantially higher fraction of persistent cue and trajectory ensembles across days (Fig. B.17**g-h**).

Within the same session, we found no preference between the recruitment of cue and trajectory ensembles across resting periods either (Fig. 4.6**e**), with 22.85% and 26.22% of cue and trajectory ensembles being reported as persistent respectively (Fig. 4.6**f**). Overall, these results indicate that cue ensembles consist of neurons that encode for more stable features across time, while trajectory ensemble neurons are more susceptible to remapping. However, this persistence in encoded features did not translate into a persistent recruitment of cue ensembles across recording days or across rest periods. Therefore, the process that determines the membership of offline ensembles is likely unbiased by functional features. Following locomotion, similar resting state ensembles are more likely to be found across days. Taken together, these results indicate that a recent exposure to a familiar experience invokes the reactivation of similar groups of neurons, despite the remapping of certain features.

4.3.6 Patchy topographic organization in resting state ensembles

Lastly, we were interested in whether cue and trajectory ensemble neurons express differences in topographic arrangements. Naïvely, because sensory information tend to be organised in a topographic fashion in primary cortices (including in the motor cortex) (Van Essen and Glasser, 2018), while the outflow of spatial information from the hippocampus would likely be distributed without discernible topography (cf. França and Monserrat, 2019), we hypothesized that cue ensembles may exhibit a higher level of topographic clustering, whereas trajectory ensemble neurons may be

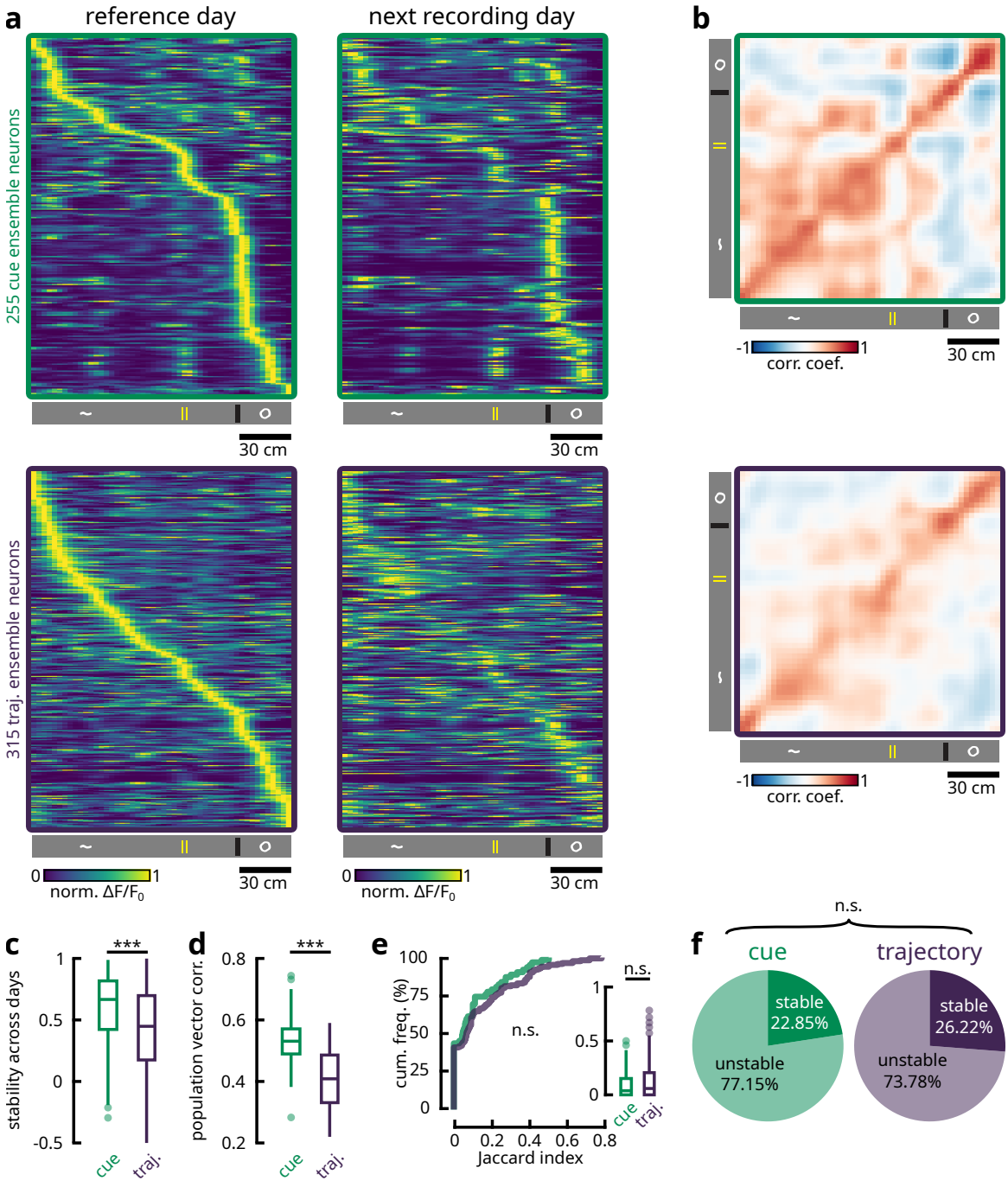
a
reference day
next recording day
255 cue ensemble neurons
315 traj. ensemble neurons
30 cm
0
1
norm. ΔF/F0
b
-1
1
corr. coef.
c
stability across days
***
cue
traj.
d
population vector corr.
e
cum. freq. (%)
Jaccard index
n.s.
f
cue
trajectory
stable
22.85%
unstable
77.15%
stable
26.22%
unstable
73.78%

Figure 4.6: Cue ensemble neurons support more stable representations across days. **a** The same FOV was recorded over consecutive experimental days in 10 animals. Spatially-selective neurons that were part of cue and trajectory ensembles during REST2 were identified in the following recording day. Neurons' average activities as a function of position were sorted by the location of peak activity on the reference day. The same sorting was kept for tunings on the subsequent recording day. **b** Pearson correlation matrices of population vectors over positions (columns in **a**) between the reference day (y-axis) and the following recording day (x-axis). **c** Pearson correlation coefficients of neurons' spatial tuning profiles (rows in **a**) across recording days. Cue ensemble neurons maintained more stable tunings over days (one-tailed Mann-Whitney U-test; $p < 0.001$). **d** Pearson correlation coefficients of population vectors (columns in **a**) across recording days. Populations vectors express a higher degree of similarity across days for cue ensemble neurons (paired-sample one-tailed Wilcoxon signed rank test; $p < 0.001$; $n = 50$ spatial bins). **e** The Jaccard index was computed to measure the persistence of ensembles between REST1 and REST2. For each REST2 ensembles, we calculated the number of member neurons overlapping with each REST1 ensembles. The Jaccard index was given as $J = \frac{O}{A+B-O}$, where for O overlapping members, there are A and B total members in the REST1 and REST2 ensembles respectively. For each REST2 ensemble, we keep the maximum Jaccard index value following paired comparisons with all REST1 ensembles. Cue ($n = 62$) and trajectory ($n = 166$) ensembles were equally persistent across resting states (two-sample two-tailed Kolmogorov-Smirnov test $p = 0.725$; two-tailed Mann-Whitney U-test $p = 0.340$). **f** Following on **e**, we tested for the significance of the proportion of overlap by one-tailed Fisher's exact test against the null hypothesis that the overlapping fraction is not higher than chance. P-values lower than $\alpha = 0.001$ were considered as persistent ensembles. In REST2, a similar fraction of cue and trajectory ensembles remained persistent from REST1 ensembles (χ^2 test; $p = 0.028$; $\chi^2 = 4.827$; $\mathrm{df} = 1$).

more diffuse. We began by examining the distribution of cells' tendencies for encoding either cue or spatial locations, for all spatially-selective neurons. In this map, we observed a few patches of heightened densities of cells encoding for either feature (Fig. B.18**a-b**). However, no clear organization could be distinguished, suggesting that the topographic arrangement may be localized to patches. Four 100×100 μm windows were drawn over these patches (Fig. B.18**b**). Neurons found within the two windows over the regions with a higher prevalence of position-correlated responses form a uniform representation of spatial locations, with population vectors decorrelating smoothly over distance. In contrast, neurons contained within the two regions exhibiting preference for cue-responses form a discontinuous representation biased by the locations of cues. Overall, these results suggest that a patchy topographic arrangement may bias the tuning tendencies of secondary motor neurons towards cues and positions.

Next, we examined the topographic distributions of ensemble neurons for different categories of ensembles (Fig. B.18**c**). Overall, cue ensemble neurons were more likely to be found over the lateral aspect (similar to Fig. B.18**a**), compared to trajectory ensemble neurons and neurons of unclassified ensembles, which were uniformly distributed (Fig. B.18**d**). To quantify the difference in these topographic distributions, the Kullback-Leibler divergence was taken between the probability density maps of each ensemble class and the density map of all neuronal ROIs (Fig. B.18**e**). Under this measure, the distribution of cue ensemble neurons was orders more different from the overall distribution of ROIs, compared to unclassified and trajectory ensemble neurons. An immediate hypothesis that can be generated from these results is that neurons that are part of a cue ensemble should be more clustered topographically due to their lateral confinement, whereas trajectory ensemble neurons should be more dispersed, assuming a random recruitment strategy for ensemble neurons. This was, however, not the case, as both cue and trajectory ensembles exhibit a similar degree of clustering in topographic space (Fig. B.18**f-g**). Nevertheless, over $\sim 70\%$ of ensembles of either category showed significant clustering, as opposed to dispersed. Therefore, resting state ensembles in M2 appear to be intrinsically organised into topographic patches. Taken together, these results suggest that both the distribution of cue-tending and position-tending cells, and the recruitment of resting state ensemble neurons follow a patchy arrangement in topographic space, though neither cue nor trajectory representations have a well-defined topographic organization.

4.4 Discussion

We found that, following a virtual spatial navigation task, two distinct types of experience-related representations were reactivated in the motor cortex. On the one hand, synchronous ensembles of neurons during resting state consisted of cells that responded to visuo-tactile landmarks. On the other hand, short segments of continuous trajectories in space were formed from other sets of synchronous neuronal ensembles. Around the onset of SWRs detected in the ipsilateral hippocampus, the reactivation of cue information, on average, preceded that of trajectory information. The same pattern of delay was found between cue-trajectory ensemble pairs within the same recording sessions, whose patterns of reactivation showed significant temporal coupling. Such ensemble pairs reactivated for related features of previous experiences, whereby reactivated trajectories tended to occur around the locations of reactivated cues. Our findings illustrate a functional and timing-dependent relationship surrounding the offline retrieval of memories in cortical structures.

The existence of two parallel functional representations in the secondary motor cortex had been alluded to previously by multiple studies. In terms of visuo-tactile responses, the M2 is reciprocally connected with the somatosensory cortex as well as the visual cortical areas (Zingg et al., 2014; Barthas and Kwan, 2017). Accordingly, M2 has been found to respond to tactile stimulations in rodents (Manita et al., 2015), while lesioning the structure evokes somatosensory neglect (Vargo et al., 1988). Interestingly, optogenetic inactivation of M2 fibres projecting to S1 during NREM sleep caused impairments in a novel object recognition task requiring discrimination of tactile textures (Miyamoto et al., 2016). This finding demonstrated M2's involvement in the consolidation of somatosensory features. With regard to spatial coding, a number of recent articles have identified neuronal responses similar to those of place cells in multiple regions of the neocortex (Esteves et al., 2021; Haggerty and Ji, 2015; Harvey et al., 2012; Ji and Wilson, 2007; Mao et al., 2017; Minderer et al., 2019; Qin et al., 1997). Crucially, the formation of these representations was severely disrupted (including in M2) following bilateral lesions to the dorsal hippocampus (Mao et al., 2018; Esteves et al., 2021). These findings suggest that the hippocampal outflow of spatial information may pervade widely distributed regions of the neocortex.

Importantly, we found that both forms of representations were reactivated in M2 during quiet-wakefulness. A majority of previous investigations into cortical patterns of reactivation have uncovered features describing singular dimensions of behaviour, usually those related to the known

functions of the structure. To name a few examples, it had been reported that the prefrontal cortex reactivated for task rules (Peyrache et al., 2009), the posterior parietal cortex for egocentric parameters (Wilber et al., 2017), the retrosplenial cortex for environmental landmarks (Chang et al., 2020) and the primary motor cortex for movement sequences (Eckert et al., 2020). From these findings, one might conclude that cortical reactivations recapitulate aspects of previous experiences specific to the modality of the region. However, there are a few studies that demonstrated reactivation of spatial features in extra-hippocampal regions, namely in the visual (Ji and Wilson, 2007), the posterior parietal (Qin et al., 1997), the medial prefrontal (Euston et al., 2007; Jadhav et al., 2016) cortices, the deep layers of the medial entorhinal cortex (Ólafsdóttir et al., 2016) and the ventral striatum (Pennartz et al., 2004), where the reactivation of such features were interlocked with hippocampal reactivations of the same behavioural experiences. Our current results help to reconcile these separate reports by suggesting that spatial and non-spatial features may be jointly expressed in the same cortical structure during offline periods. Nevertheless, given the limitations of the head-fixed preparation, it is hard to determine the exact behavioural states the animals were in during the resting periods. Further research is needed to describe whether these patterns of reactivation express different dynamics during sleep or during particular states in awake quiescence.

Lending support to the presence of such a dichotomy is the temporal delay observed between the reactivation of cues and trajectory information. Classifying the reactivating ensembles into two categories based on their encoded features revealed that cue information was reactivated earlier on average than trajectory information. This finding complements a previous report in which patterns of activity in the auditory cortex during NREM sleep preceding SWRs could accurately predict the ensuing CA1 activities, while these CA1 patterns could in turn explain the subsequent activities in the auditory cortex (Rothschild et al., 2017). In that study, the offline activities in the auditory cortex were explained by distinct patterns of sounds, though the exact nature and contributions of hippocampal activities with regard to the cortex could not be determined. Here, our results suggest that the early and late reactivations surrounding SWR events may in fact constitute distinct functional representations. Given this temporal dissociation, it may be hypothesized that the earlier reactivation of cue-centric information reflect locally-encoded attributes in the secondary motor cortex, whereas the later reactivation of trajectory sequences may be driven by hippocampal outflow (cf. Abadchi et al., 2020). This notion is corroborated by the heightened stability of cue representations across

recording days compared to trajectory information, which is consistent with the higher rates of synaptic turnover in hippocampus compared to neocortex (Attardo et al., 2015).

The idea that local cortical patterns could serve as triggers for the retrieval of mnemonic information from the hippocampus places under scrutiny the potential for other cortical regions in contributing to this dialogue. In fact, rodent studies have demonstrated that specific auditory stimuli presented during NREM sleep could bias the content of hippocampal replay (Bendor and Wilson, 2012; Rothschild et al., 2017), while the presentation of olfactory cues had the same effect in a human fMRI study (Rasch et al., 2007). These sensory modalities, with the addition of visuo-tactile representations elaborated presently, imply that widely distributed neocortical sites could each serve as the initiator for hippocampal reactivation, where the subsequently evoked hippocampal patterns would permeate global cortical modules linking together the different attributes associated with episodic experiences. This model could explain the prevalence of trajectory reactivations over cue reactivations in our data. In particular, reactivated trajectory patterns observed in the secondary motor cortex could have been initiated by a separate region of the brain.

From the output side, it may be inferred from the present results that the information received by cortical sites downstream from the hippocampus during offline reactivation comprises sequential place cells activations analogous to the patterns of replay observed directly in the hippocampus (Skaggs and McNaughton, 1996; Nádasdy et al., 1999; Lee and Wilson, 2002; Malvache et al., 2016). Similarly, previous reports on coordinated reactivations between the hippocampus and the cortex for the same behavioural experiences have largely involved spatial features (Qin et al., 1997; Ji and Wilson, 2007; Ólafsdóttir et al., 2016; Jadhav et al., 2016). This observation calls into question the relevance of the 'place-code' for cortical processing and consolidation. One possibility asserts that the hippocampal 'place-code' encompasses a diverse range of non-spatial information as well (Eichenbaum, 1996; Eichenbaum et al., 1999; O'Keefe and Nadel, 1978), which is reflected in the modulation of firing rates within otherwise stable place fields by varying sensory/task conditions, a phenomenon known as *rate remapping* (Leutgeb et al., 2005; Colgin et al., 2010). As such, the 'place-code' provides a vehicle through which arbitrary associations could be formed between the diverse attributes of episodic experiences spanning disparate modalities. In the present study, cue and trajectory ensembles that expressed temporally coordinated reactivations encoded for similar attributes of a recent experience. It is therefore possible, hypothetically, that the overlap between the

spatial and the non-spatial dimensions is the 'semantic link' that allows cortical cue features to be used to retrieve hippocampal patterns (Tse et al., 2007; Wiltgen and Silva, 2007; McNaughton, 2010) (Fig. 4.7). Specifically, this retrieval process can be facilitated by *pattern completion* mechanisms supported by the CA3 recurrent network, which forms the basis of a *content-addressable* memory system (Marr, 1971; McNaughton and Morris, 1987; McNaughton and Nadel, 1990; McClelland et al., 1995; Teyler and DiScenna, 1986b). The hippocampal output in turn elicits previously associated attributes of experience found across wide regions of the neocortex. This globally-coherent retrieval would hence allow cortico-cortical associations to form in such a way as to extract the statistical regularities found in those experiences.

This proposed hypothetical model has implications for both modulating online behaviour and mediating offline mnemonic processes. On the one hand, the targeted outflow of hippocampal spatial information onto cortical sites may provide a contextual and/or mnemonic frame of reference for guiding active behaviours (Ranganath and Ritchey, 2012). Particularly, in the M2, it has been shown that population activities convey spatial/contextual information with a putative involvement in guiding action and decision-based planning (Olson et al., 2020). Similarly, the interactions between the medial prefrontal cortex and the hippocampus appear important for navigation tasks that rely on memories of past trajectories (Viena et al., 2018). On the other hand, the offline reactivation of recent experiences may drive changes in sensory encoding and functional representations in cortical areas. A recent study has shown that reactivation events in the lateral visual cortex more faithfully replicated the sensory responses of neurons in the future, rather than the preceding stimulus-evoked responses (Nguyen et al., 2022). In parallel, neurons in the primary visual cortex gradually acquire sharper discrimination between distinct grating patterns with training (Failor et al., 2021). Such descriptions of integrative processes that occur during both online and offline periods have also been hinted by our data. In particular, additional modelling has suggested that the features encoded by resting state ensembles exist over a continuum where varying degrees of cue and trajectory information is expressed. This conjunctive aspect may not only reflect a gradual consolidation of spatial information into existing cortical representations (cf. Esteves et al., 2023), but also a mechanism for imparting contextual information to sensory representations to guide online behaviour. Moreover, the heightened stability of cue representations over days, as compared to spatial information, may indicate a difference in the rate of consolidation or experience-driven drifts in functional representa-

tions in the cortex. Overall, a hypothesis can be proposed, stating that spatial information integrates with non-spatial information in the cortex to both modulate online behaviour through conjunctive coding and drive representational changes during offline periods. It remains to be determined how the functional encodings in cortical neurons may be gradually shaped by the outflow of hippocampal place information and the degree to which this process may rely on the hippocampus.

Taken together, the current findings demonstrate a temporally organized sequence of cortico-hippocampal exchanges during offline reactivation, which is graded by the specific contents of the behavioural features being reactivated. These reactivated features likely arise separately from the hippocampus and the neocortex, and accordingly involves spatial and non-spatial aspects of previous experiences. As such, neocortical recollections of the specific 'attributes' of prior experiences may seed the reactivation of hippocampal sequences (i.e., 'episodes'), which in turn is propagated back to neocortical regions. Pattern completion is proposed as a likely mechanism through which these separate dimensions of behaviour may be linked in order to facilitate the targeted retrieval of recent experiences from the hippocampus by the neocortex.

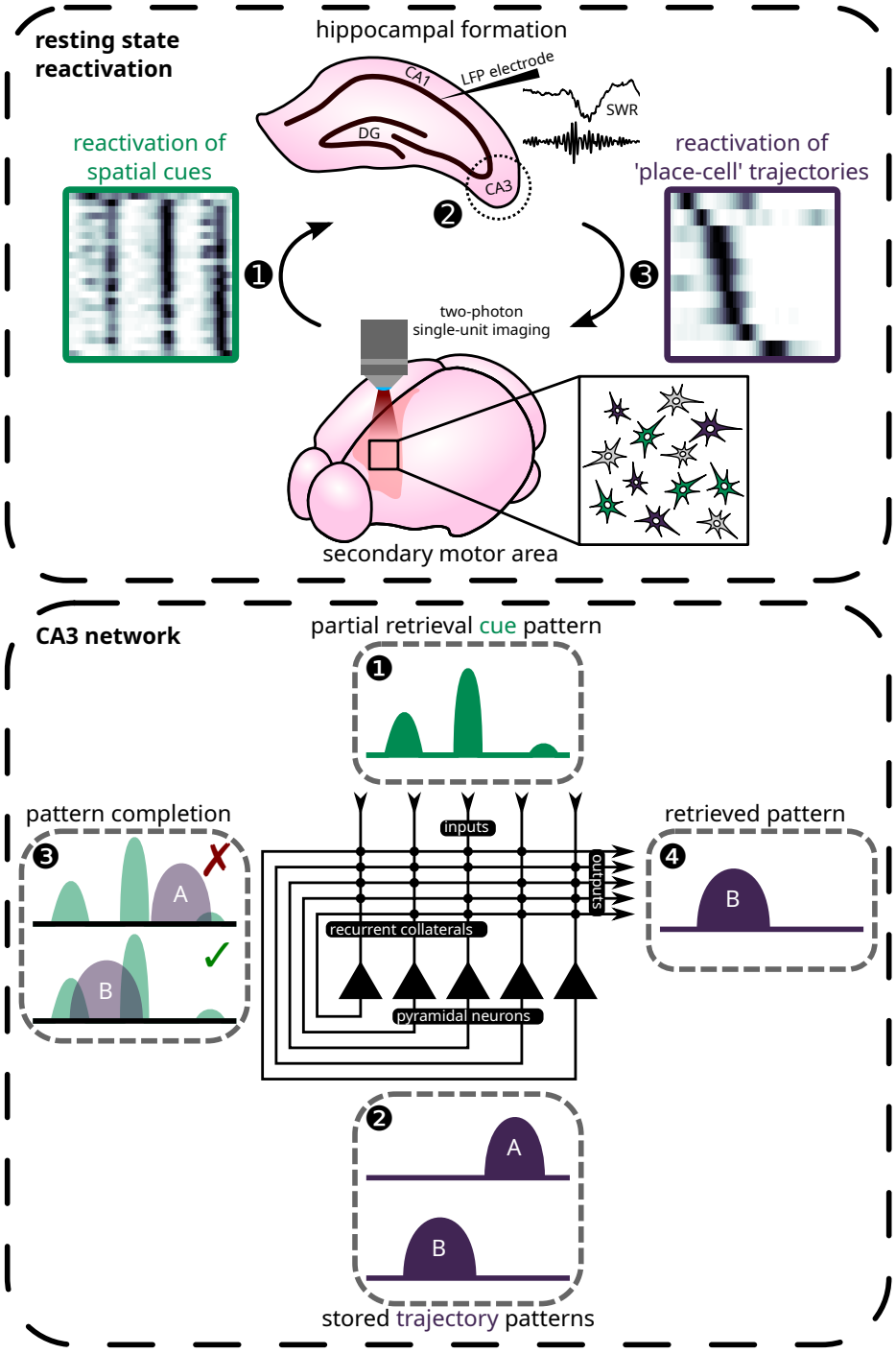

Figure 4.7: Proposed hypothetical model for temporal dynamics of reactivation based on the theory of *pattern completion*. See Discussion for explanation.

Chapter 5

General Discussion[59]

In the present thesis, I described three studies that, collectively, mapped the expression and reinstatement of spatial representations in the cortex during online and offline periods. Detailed interpretations for the results of each study have been offered in their respective discussion sections. In general, the present work builds upon the original discoveries made by Qin et al. (1997); Ji and Wilson (2007), which showed that (1) cortical areas support spatial coding and that (2) these cortical spatial representations reactivate in a coherent manner with the hippocampal place code. The works conducted for this thesis further expand these findings by suggesting that this cortical spatial code relies on an intact hippocampus to form (Chapter 2), that not all regions reactivate for spatial trajectories (Chapter 3) and that both spatial and non-spatial representations are reactivated in some cortical areas (Chapter 4).

These results, though novel and interesting, are very difficult to interpret in the context of existing theories. In fact, given our current knowledge on the neurobiology of space and memory (summarised in the General Introduction), it is safe to say that no mainline theories had predicted the existence of place cells in the cortex and their subsequent reactivations during offline periods. The most conservative approach would be to state that the hippocampal outflow of allocentric information provides a unified scaffold for binding distributed cortical representations, allowing for "coordinate transforms" to occur seamlessly from primary to associational regions. The reactivation of these spatial representations during subsequent offline periods, therefore, reflect the gradual consolidation and integration of this code into cortical representations. There is, in fact, good evidence in literature to support such an idea, namely the conjunctive egocentric and allocentric codes found across multiple cortical areas (see Wang et al., 2020, for review). Nonetheless, this narrative can

[59]Upon reviewing this General Discussion, Bruce had the following to say: "...it looks like you might have followed Alice. Which side of the mushroom did you eat?" The readers have been warned.

only superficially integrate into the rich literature and theories surrounding *learning and memory* and is, frankly, a bit boring. This is indeed the stance that I would adopt if the present manuscript was to be published in a serious scientific journal. However, given the personal nature of a dissertation, I will allow myself a scientific indiscretion, here, to offer some heavy speculations over the issues of *Why are there place cells all over the cortex?* and *How are they implicated in memory?*

In particular, I believe that the key to understanding the phenomena reported in this thesis lies with unravelling the utility behind having a heavily *redundant* and *distributed* code for space.

5.1 Redundancy and Holographic Memories

Almost a century later, the same problems of redundancy and non-locality that have puzzled Lashley can also be seen in our data. The *place code*, which for a long time researchers had no reasons to suspect to be found outside of the hippocampus and associated structures, have been uncovered in multiple regions of the neocortex. Our study, in combination with many other articles that appeared in recent years (Fiser et al., 2016; Haggerty and Ji, 2015; Harvey et al., 2012; Ji and Wilson, 2007; Qin et al., 1997; Mao et al., 2017; Minderer et al., 2019; Pakan et al., 2018; Saleem et al., 2018), collectively suggested that spatial coding, the formation of which relies on an intact hippocampus, is a ubiquitous feature in the cortex. Moreover, in a recent article that I co-authored, we demonstrated that this spatial code can, over time, be consolidated into cortical networks such that they are gradually recovered following hippocampal damage[60] (Esteves et al., 2023).

The advantage of such a storage scheme is apparent from Lashley's works; storing many copies of an important information at multiple locations makes that information resilient to damage[61]. However, such a strategy would appear highly inefficient both from a computational and a metabolic

[60] It is worth noting that, in this study, bilateral hippocampal lesions were confined only to the dorsal portions. Signs of consolidation may not persist with complete lesions (cf. Martin et al., 2005).

[61] During my undergraduate studies, I came across a most fascinating demonstration of *redundancy* in humans. As part of a lecture on neuropsychology, a guest speaker played a video clip of a patient undergoing a *Wada test*. This test is often administered to patients suffering from intractable epilepsy in the medial temporal lobe, wherein surgical intervention is necessary. By introducing a barbiturate (usually sodium amobarbital) through the left or the right internal carotid artery, the left and right hemispheres are selectively inhibited. Psychological assessments on language, memory and other faculties are then used to determine hemispheric dominance and to predict the post-procedural deficits on cognitive functions. In one patient, who was self-reported as ambidextrous, inhibition of either hemispheres did not lead to a significant speech impairment. However, the patient, visibly in an altered state of mind, answered many of the questions by substituting synonyms for the words in her previous responses (e.g., "pop" instead of "soda"). It seemed as though, following loss of substantial cortical functions, alternative strategies were spontaneously devised to solve the same problems. Related phenomena are observed in patients with Broca's aphasia, in the form of *verbal paraphasia*. When asked a question, patients would often provide single-word answers that, though lacking in eloquence and clarity, does convey the gist of the ideas that they are attempting to express.

standpoint. In fact, even in Marr's auto-associational network, storage capacity is a huge limitation and one that is yet to be resolved for the model to account for the seemingly large amount of information stored by the hippocampus (cf. Battaglia and Treves, 1998). If we scale this problem up to the proportion of the entire cortical mantle, we encounter a hard theoretical limit. To elucidate the reason for this redundancy, I will resuscitate a somewhat antiquated computational theory of memory: the *Holographic Theory of Memory*. In the 70-80s, this theory enjoyed a short period of popularity owing to its ability to explain Lashley's extraordinary results. I believe that this same theory, many years later, can still provide a useful, if not highly speculative, framework for understanding the ubiquity of the place code.

As a preliminary, let us dissect the structure of *associative* and *content-addressable* memories, which we briefly explored in the General Introduction. This is best accomplished by paraphrasing a toy model described by McNaughton and Morris (1987). Suppose we have two vectors, **a** and **b**, each storing some numerical patterns A and B. These patterns are metaphorical of two patterns of activity that jointly occurred in the neocortex. Because they tend to occur together very frequently, they likely belong to the same semantic category or concept. It would, therefore, be useful to *associate* them with one another, such that evoking one pattern leads to the recall of the other pattern, as a mechanism for recognising and linking semantic concepts. In our toy model, we perform this association by taking the outer product between vectors **a** and **b** to obtain a weight matrix:

$$\mathbf{a} \otimes \mathbf{b} = \mathbf{a}\mathbf{b}^{\mathsf{T}} = \mathbf{W}. \tag{5.1}$$

It can be shown with relative ease that pattern A can be reconstructed from this matrix by presentation of pattern B, and vice versa:

$$\mathbf{a}\mathbf{b}^{\mathsf{T}} = \mathbf{W} \tag{5.2}$$

$$\mathbf{a}\mathbf{b}^{\mathsf{T}}\mathbf{b} = \mathbf{W}\mathbf{b} \tag{5.3}$$

$$\mathbf{a} = \frac{\mathbf{W}\mathbf{b}}{|\mathbf{b}|^2}. \tag{5.4}$$

I.e., the product between the weight matrix **W** and the input vector **b**, normalised by the sum of squares of the input vector, reconstructs the associated vector for pattern A. In fact, this property can

be generalised in order for the weight matrix to store multiple pairs of associates:

$$\mathbf{AB} = \mathbf{W} \tag{5.5}$$

$$\mathbf{ABB}^{+} \approx \mathbf{WB}^{+} \tag{5.6}$$

$$\mathbf{A} \approx \mathbf{WB}^{+}, \tag{5.7}$$

where $\mathbf{A}$ is an $m \times p$ matrix and $\mathbf{B}$ is a $p \times n$ matrix with p pairs of input patterns to be associated and $\mathbf{B}^{+}$ is the Moore-Penrose pseudoinverse of matrix $\mathbf{B}$. In general, when the rank of the weight matrix is lower than p (i.e., when the number of patterns to be stored exceeds the theoretical limits), reconstruction is no longer accurate, with the magnitude of error increasing with the number of extra patterns to be stored. Note that a single pattern/row in $\mathbf{B}$ would retrieve its paired associate as a column in $\mathbf{A}$ and not another pattern:

$$\mathbf{AB} = \mathbf{W} \tag{5.8}$$

$$\mathbf{a}_1\mathbf{b}_1^{\intercal} + \mathbf{a}_2\mathbf{b}_2^{\intercal} + \cdots + \mathbf{a}_p\mathbf{b}_p^{\intercal} = \mathbf{W} \tag{5.9}$$

$$(\mathbf{a}_1\mathbf{b}_1^{\intercal} + \mathbf{a}_2\mathbf{b}_2^{\intercal} + \cdots + \mathbf{a}_p\mathbf{b}_p^{\intercal})\mathbf{b}_1 = \mathbf{W}\mathbf{b}_1 \tag{5.10}$$

$$\mathbf{a}_1\mathbf{b}_1^{\intercal}\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2^{\intercal}\mathbf{b}_1 + \cdots + \mathbf{a}_p\mathbf{b}_p^{\intercal}\mathbf{b}_1 = \mathbf{W}\mathbf{b}_1 \tag{5.11}$$

$$\mathbf{a}_1|\mathbf{b}_1|^2 + (\mathbf{a}_2\mathbf{b}_2^{\intercal}\mathbf{b}_1 + \cdots + \mathbf{a}_p\mathbf{b}_p^{\intercal}\mathbf{b}_1) = \mathbf{W}\mathbf{b}_1. \tag{5.12}$$

Assume that $\mathbf{a}_n$ and $\mathbf{b}_n$ are i.i.d. with null mean and that all these patterns in $\mathbf{B}$ are linearly independent from each other (i.e., they do not interfere with each other), then the scalar product $\mathbf{b}_m^{\intercal}\mathbf{b}_n \approx 0$ for all $m \neq n$, and

$$\mathbf{a}_1 \approx \frac{\mathbf{W}\mathbf{b}_1}{|\mathbf{b}_1|^2}. \tag{5.13}$$

In addition, pattern A can still be retrieved when the input pattern B is incomplete or degraded. We can simulate this by contaminating $\mathbf{B}$ with centred Gaussian noise and use it to retrieve the pattern

$\hat{\mathbf{A}}$:

$$\hat{\mathbf{A}} = \mathbf{W}(\mathbf{B}^{+} + \varepsilon), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{5.14}$$

$$= \mathbf{W}\mathbf{B}^{+} + \mathbf{W}\varepsilon \tag{5.15}$$

$$= \mathbf{A} + \mathbf{W}\varepsilon. \tag{5.16}$$

The quality of the retrieval can be quantified as the sum of squares of the residuals $\mathbf{R}$ between the original matrix $\mathbf{A}$ and the retrieved matrix $\hat{\mathbf{A}}$:

$$\mathbf{R} = \mathbf{A} - \hat{\mathbf{A}} \tag{5.17}$$

$$= \mathbf{A} - (\mathbf{A} + \mathbf{W}\varepsilon) \tag{5.18}$$

$$= -\mathbf{W}\varepsilon. \tag{5.19}$$

By Cochran's theorem,

$$\sum_{i}^{m}\sum_{j}^{n}\mathbf{R}_{i,j}^2 = \sum_{i}^{m}\sum_{j}^{n}\mathbf{W}_{i,j}^2\varepsilon^2 \sim \frac{\sigma^2}{\sum_i^m\sum_j^n \mathbf{W}_{i,j}^2}\chi_r^2, \tag{5.20}$$

where $r = \text{rank}(\mathbf{W})$ is the rank of matrix $\mathbf{W}$. That is, the error in retrieval is χ-squared distributed and scales with the variance of the noise contaminants. In the absence of noise, $\sigma^2 = 0$ and $\hat{\mathbf{A}} = \mathbf{A}$, i.e. the original pattern is retrieved perfectly.

Taken together, this toy model informs us that, when pairs of patterns are associated together inside a covariance-like matrix, the result is an *associational memory* system that exhibits the following properties: *addressable* (one pattern can retrieve its associated pattern), *pattern completion* (incomplete or degraded input patterns can retrieve their associated patterns), *pattern separation* (the input pattern retrieves the correct associated pattern amongst all stored patterns) and some limited storage capacity.

We are particularly interested in the storage capacity[62] of this model, which informs us of the efficiency and the amount of redundancy that exist in the weight matrix $\mathbf{W}$. For simplicity, we create a *content-addressable* memory (i.e., pattern is associated with itself) with patterns A to be stored.

[62]Note that, in eq. 5.20, the mean and the variance of the χ-squared distribution, defined as r and $2r$ respectively, increase with the rank of matrix $\mathbf{W}$. It follows that the memory store is less susceptible to corruption when a smaller number of patterns are encoded.

We begin by storing a single pattern vector:

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^{\mathsf{T}} \tag{5.21}$$

$$\mathbf{a}\mathbf{a}^{\mathsf{T}} = \mathbf{W} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix}. \tag{5.22}$$

Notice that rows 2 through 4 are just multiples of row 1, which contains the original pattern. We can divide rows 2 through 4 by their respective row number:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}. \tag{5.23}$$

The original pattern is effectively stored four times, so $\mathbf{W}$ is said to be four times redundant. Suppose we now store 2 patterns:

$$\mathbf{a} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}^{\mathsf{T}} \tag{5.24}$$

$$\mathbf{a}\mathbf{a}^{\mathsf{T}} = \mathbf{W} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}. \tag{5.25}$$

Now, the same pattern is repeated in rows 1 and 3, as well as rows 2 and 4. The stored information is redundant twice. It is not until we fill the weight matrix up with four linearly independent patterns (i.e., saturate the storage capacity[63]) that we no longer have redundancy. In general, for this type of linear storage system, redundancy is only eliminated when the number of patterns is equal to

[63] A complete analysis on the storage capacity of Hopfield networks can be found in Hertz et al. (1991, Section 2.5).

the length of individual patterns (cf. rank-nullity theorem). Intuitively, however, we might expect the stored information to become more resilient as redundancy increases. We can investigate this relationship in our toy model:

$$\mathbf{A}\mathbf{A}^\intercal = \mathbf{W} \tag{5.26}$$

$$\mathbf{A} = \mathbf{W}(\mathbf{A}^\intercal)^+. \tag{5.27}$$

We simulate a "lesion" on our memory $\mathbf{W}$ by contaminating it with some noise ε:

$$(\mathbf{W}+\varepsilon)(\mathbf{A}^\intercal)^+ = \mathbf{W}(\mathbf{A}^\intercal)^+ + \varepsilon(\mathbf{A}^\intercal)^+ \tag{5.28}$$

$$\hat{\mathbf{A}} = \mathbf{A} + \varepsilon(\mathbf{A}^\intercal)^+. \tag{5.29}$$

We measure the degradation in memory recall (i.e., error of reconstruction) by taking the Frobenius Norm between the original patterns $\mathbf{A}$ and the recalled patterns $\hat{\mathbf{A}}$:

$$||\mathbf{A}-\hat{\mathbf{A}}||_\mathrm{F} = ||\mathbf{A}-(\mathbf{A}+\varepsilon(\mathbf{A}^\intercal)^+)||_\mathrm{F} \tag{5.30}$$

$$= ||-\varepsilon(\mathbf{A}^\intercal)^+||_\mathrm{F} \tag{5.31}$$

$$= ||\varepsilon(\mathbf{A}^+)^\intercal||_\mathrm{F} \tag{5.32}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{p}\sqrt{|\sum_{k=1}^{n}\varepsilon_{ik}\mathbf{A}^+_{kj}|^2}. \tag{5.33}$$

We see that the amount of error in the recall of information increases approximately linearly with the number of pattern p stored in memory. Therefore, there exists a trade-off between storage efficiency and resilience. The more redundant information is kept in storage, the more resilient this storage system is to damage. Similarly, the more saturated the storage capacity becomes, the less corruption it is able to sustain.

There is, however, a more interesting way to create redundancy, therefore resilience, in data structure. This type of memory system, known as *holographic memories*, originated as a discovery in physics and, over time, became adopted by *learning and memory* theorists who sought a conceptual tool to explicate the parallel and distributed quality of memories (reviewed in Willshaw, 1981; Kohonen, 1988, Chapter 10). Suppose we have a coherent light source (Fig. 5.1**a**). The beam is

split in half, with one half projected onto some object and the other half reflected by a mirror. The two diffracted beams then converge back onto a single photographic plate, which holds a record of the interference pattern produced by the two wavefronts. Once the photographic plate is developed, projecting a single beam from the position of the mirror onto the plate would create a virtual image of the original object. This apparatus constitutes, therefore, an optical device for creating and recalling *associative memories*. In this case, the photographic plate acts as the memory store. Lights diffracted off of the object and the mirror comprise patterns A and B respectively. Once committed to the memory store, presentation of pattern B elicits recall of pattern A and vice versa.

A key insight comes from examining the behaviour of this memory after incurring "lesions" upon it. We simulate this condition by taking the photographic plate and shattering it into pieces. If any shard is subsequently picked up and presented with the reconstruction beam, the original pattern would be reconstructed in its entirety, albeit at a lower resolution. To understand how this occurs, we must first realise that, by shining light on the surface of an object, the geometry of the object is no longer described by a cloud of points in space. Instead, the photons diffracted off of the object participate in constructive and destructive interference with itself, as well as with the beam reflected from the mirror. The patterns that are imprinted onto to photographic plate, therefore, are a record of the amplitude and phase information of interfering electromagnetic waves. In other words, the object underwent transformation from the spatial domain to the frequency domain. This process is similar to taking the *Fourier transform* of an image (Fig. 5.1**b**), which takes information encoded in the spatial domain and stores that information in the frequency domain. If one is to incur "lesions" focally on this spectrogram, the result is the loss of certain frequency and phase information at particular angles. However, an overall complete pattern of the original image can be reconstructed in space by taking the inverse Fourier transform. Therefore, localised lesions do not translate into localised deficits; memory is parallel and distributed.

We are now in a position to offer some speculations over the nature of the cortical spatial code. Based on our studies, there appears to exist a place code that is largely diffuse and redundant in multiple regions of the cortex. This code appears to rely on an intact hippocampus for its initial formation. During offline periods, it is spontaneously reinstated in the cortex. Based on the framework delineated here, it is possible that patterns of activity in the cortex undergo a domain transform as they are encoded by the hippocampus, whereby high-dimensional sensory information

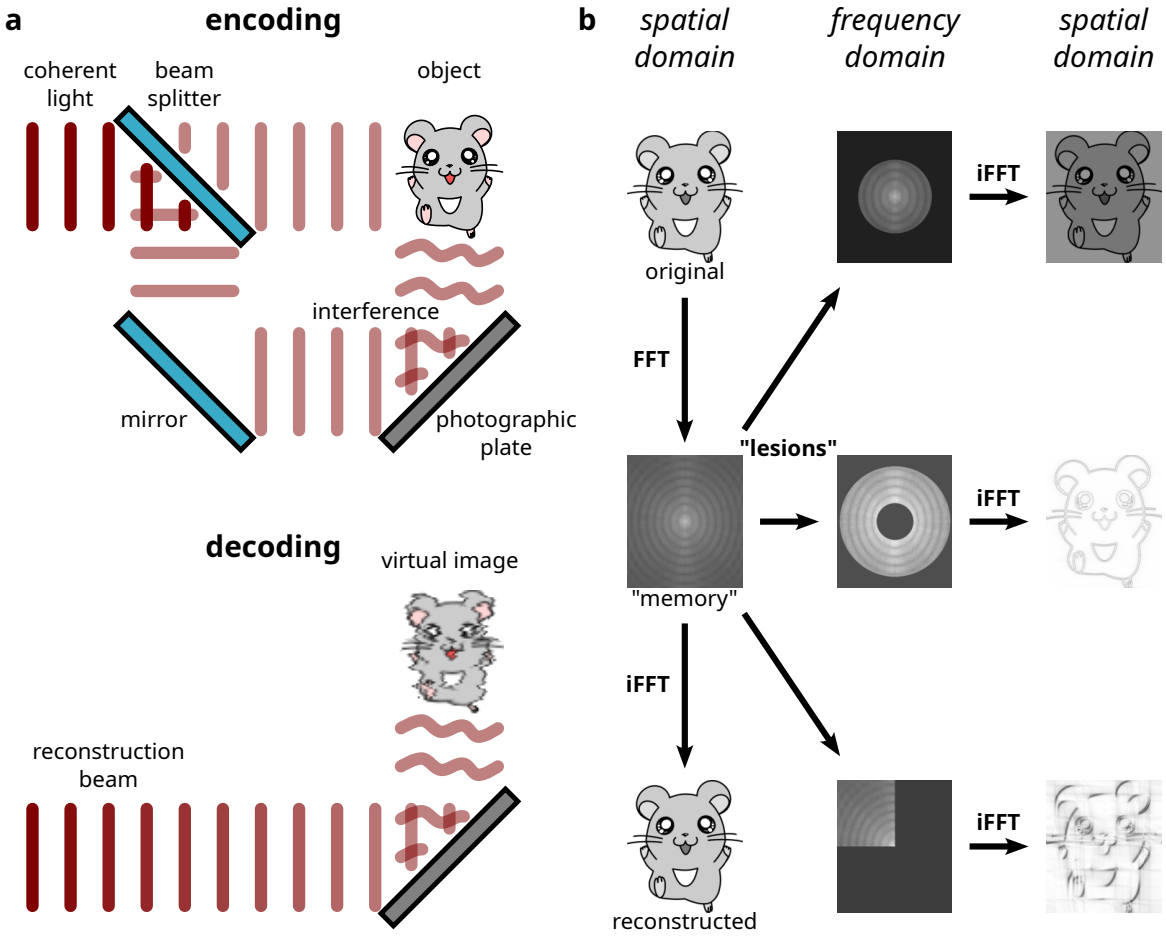

Figure 5.1: Holographic associative memories. **a** Storage and retrieval of holographic memories. Replace the mirror with a second object, subsequent light diffraction off of the second object retrieves a virtual image of the first object, hence forming an associative memory. **b** Storage and retrieval of "Fourier memories". A degraded form of the complete pattern can be retrieved following focal "lesions" to the memory store.

is transformed into the spatial domain supported by the hippocampal cognitive map. Information, henceforth redundant and domain-transformed, becomes resilient to focal lesions if they are to be over time consolidated into distributed cortical stores. The existence of a parallel and distributed spatial code may, therefore, constitute the neural correlate of Lashley's engrams. This interpretation is, however, incomplete. In our toy models, memory retrieval requires either computing the inverse of a matrix or taking an inverse Fourier transform. Nevertheless, our experimental observations suggest that this spatial code is both expressed and reinstated verbatim during online and offline periods. If these cortical spatial representations are reflective of some form of encoded memories, then a decoding step is clearly missing.

5.2 A case for the Cognitive Map

Memory encoding and storage typically proceeds as follow: We begin with some input data. Then, a weight matrix is calculated. The input data is *encoded* by projecting it into a new features space defined by the weight matrix. To decode the data, we compute the inverse of the weight matrix and project the encoded data backwards. In the example above, this is analogous to first taking a Fourier transform of the data for encoding, then taking the inverse Fourier transform for decoding. The whole process proceeds as follow:

$$\mathbf{SW} = \mathbf{M} \qquad \textit{encoded} \tag{5.34}$$

$$\mathbf{SWW}^{-1} = \mathbf{MW}^{-1} \tag{5.35}$$

$$\mathbf{S} = \mathbf{MW}^{-1} \qquad \textit{decoded,} \tag{5.36}$$

where $\mathbf{S}$ is the input matrix, $\mathbf{W}$ is the weight matrix and $\mathbf{M}$ is the encoded "memory".

A conceptual issue arises, however, when considering how biological agents could implement such an algorithm. Whereas we can train artificial neural networks using the *backpropagation algorithm*, neurons do not seem capable of computing derivatives and sending these error values back via their dendrites (but cf. Lillicrap et al., 2016). Similarly, computing the inverse of a matrix is, as far as I know, not a calculation that can be readily performed by the brain. A quick glance over the prominent computational theories on memory reveals that a solution to this problem is not well defined. In fact, Marr (1971) did not provide a definitive answer to how inputs are mapped back

from the output side. Models such as the *restricted Boltzmann machine* or *Hopfield network* indeed provide a solution: the input layer serves also as the output layer. Hence, there is a one-to-one correspondence between input patterns and reconstructed patterns. However, in the brain, there is no guarantee for post-synaptic neurons to form connections back to the same pre-synaptic neurons. How information is encoded from cortex to the hippocampus, then retrieved from hippocampus to cortex, therefore becomes unresolved.

A solution possibly lies in modulating cortical cells with the hippocampal place code. Suppose a cortical pattern of activity $\mathbf{c}^p$ maps onto the hippocampal pattern $\mathbf{h}^p$ via the encoding matrix $\mathbf{W}_\text{e}$. Some computation occurs in the hippocampus, following which the hippocampal pattern at the next time point $\mathbf{h}^{p+1}$ then maps back onto the cortex $\mathbf{c}^{p+1}$ via the decoding matrix $\mathbf{W}_\text{d}$:

$$\mathbf{W}_\text{e}\mathbf{c}^p = \mathbf{h}^p \tag{5.37}$$

$$\mathbf{W}_\text{d}\mathbf{h}^{p+1} = \mathbf{c}^{p+1}. \tag{5.38}$$

In this way, the hippocampus and the cortex form a closed feedback loop. Suppose that the weights of matrix $\mathbf{W}_\text{d}$ are fixed. We apply a simple Hebbian learning rule to matrix $\mathbf{W}_\text{e}$ such that cortical cells at one time stamp $\mathbf{c}^p$ that activate in conjunction with hippocampal neurons at the next time stamp $\mathbf{h}^{p+1}$ will have their synapses strengthened:

$$\mathbf{W}_\text{e} = \frac{1}{P-1}\sum_{p=1}^{P-1}\mathbf{h}^{p+1}(\mathbf{c}^p)^\intercal \tag{5.39}$$

$$\mathbf{W}_\text{e}^\intercal = \frac{1}{P-1}\sum_{p=1}^{P-1}\mathbf{c}^p(\mathbf{h}^{p+1})^\intercal \tag{5.40}$$

$$= \frac{1}{P-1}\sum_{p=1}^{P-1}\mathbf{W}_\text{d}\mathbf{h}^p(\mathbf{h}^{p+1})^\intercal \tag{5.41}$$

$$= \mathbf{W}_\text{d}\frac{1}{P-1}\sum_{p=1}^{P-1}\mathbf{h}^p(\mathbf{h}^{p+1})^\intercal. \tag{5.42}$$

A simple solution can be found if we make a few assumptions based on the functional properties of the hippocampus. First, the activity patterns of the hippocampus are sparse; a relatively small fraction of place cells are active at any given location/time. Suppose there are N neurons, out of which n are active at any one time. Second, on a moment-to-moment basis, a significant portion of

hippocampal neurons that are active at one time remain active at the next time point. This is because place fields have a certain width, and one set of active place cells would tend to remain active over discrete displacements made by a navigating agent. Call the number of persistent neurons α such that $n \geq \alpha > 0$. Finally, for simplicity, each neuron will only assume the values 0 or 1. Then, the diagonal elements are given as:

$$\frac{1}{P-1}\sum_{p=1}^{P-1}\mathbf{h}_i^p\mathbf{h}_i^{p+1} = \frac{1}{P-1}\sum_{p=1}^{P-1}\frac{1}{N}\alpha = \frac{\alpha}{N}. \tag{5.43}$$

Off-diagonal elements are (for all $i \neq j$):

$$\frac{1}{P-1}\sum_{p=1}^{P-1}\mathbf{h}_i^p\mathbf{h}_j^{p+1} = \frac{1}{P-1}\sum_{p=1}^{P-1}(\frac{n}{N})^2 \tag{5.44}$$

$$= (\frac{n}{N})^2 \tag{5.45}$$

$$\approx 0 \quad \text{if} N \gg n \,\text{and}\, \alpha \approx n. \tag{5.46}$$

$$\therefore \mathbf{W}_{\mathrm{e}}^{\intercal} = \mathbf{W}_{\mathrm{d}}\frac{\alpha}{N}\mathbf{I} \tag{5.47}$$

$$\mathbf{W}_{\mathrm{e}}^{\intercal} = \frac{\alpha}{N}\mathbf{W}_{\mathrm{d}} \tag{5.48}$$

$$\mathbf{W}_{\mathrm{e}}^{\intercal}\mathbf{W}_{\mathrm{d}}^{-1} = \frac{\alpha}{N}\mathbf{I}. \tag{5.49}$$

This analysis informs us that, by modulating cortical activities with a sparse place code from the hippocampus, the inverse matrix problem can be naturally solved without the need for some special biological mechanism or network architecture. This is because, when the cortex and the hippocampus form a closed feedback loop, sets of cortical neurons that are driven by a sparse set of hippocampal neurons are more likely, on the returning path, to have their synapses strengthened with the hippocampal neurons, which they receive modulations from. In this sense, the hippocampal place code serves as a "functional tag" to establish one-to-one correspondences between cortical and hippocampal cells. Indices are, therefore, established on-the-fly during behaviour, given some simple associative learning rule[64], such that a level of symmetry is available later on for contents to

[64] A potential candidate for this type of learning are the mossy fibre inputs onto CA3 pyramids. Granule cells synapse sparsely but strongly onto pyramidal cells, while their synapses exhibit short-term potentiation (cf. Vyleta et al., 2016; McNaughton and Morris, 1987). These "detonators", in combination with perforant path inputs, could help establish

be addressed and retrieved by the cortex. According to this conjecture, the index and the contents of memories are one and the same.

Some experimental data in support of this model can be found in Qin et al. (1997); Ji and Wilson (2007). During behaviour, hippocampal-cortical neuron pairs that expressed a significant degree of temporal coherence remained associated during the post-task sleep period. This coupling, however, was not present during the pre-task sleep, suggesting that these associations were formed during behavioural exposure. Importantly, the cortical neurons reported in these studies also exhibited spatial responses, with the degree of temporal association between hippocampal-cortical neuron pairs, therefore, dictated by the amount of overlap between their place fields. If we are to assume that these cortical spatial responses have a hippocampal origin (cf. Chapter 2), then cortical cells that are modulated by certain place cells could indeed be tagged for association with those same place cells. Further experimentation is necessary to confirm whether these functional connections are indeed symmetrical. In Ji and Wilson (2007), the cross correlation between hippocampal-cortical cell pairs has a broad peak centred around 0 lag. These results are commensurate with the idea that sometimes the cortex leads the hippocampus, whereas other times, the hippocampus leads the cortex, with an overall symmetrical and bidirectional association between cell pairs.

5.3 Reconciling Space and Memory: a "Semantic Linking" Hypothesis

There is no definitive explanation to the experimental observations reported in this thesis. In fact, it is not completely unfair to say that no existing theories would have predicted these results. Take the *indexing theory* for instance. It predicts that the hippocampus stores an index pointing to the original patterns of activity that occurred in the cortex. Suppose this index is the place code, which does account for a large portion of the cellular activities in the hippocampus. Our results suggest that it is not so much the original patterns that are retrieved than the index itself. Nevertheless, the evidence that we have at present are as follow:

1. During online periods, hippocampal spatial information is broadcast seemingly *verbatim* to distributed neocortical regions (cf. Chapter 2).

sparse one-to-one mappings between hippocampal cells and cortical cells. It has been reported that mossy fibre inputs are important during the acquisition of a spatial task (Lassalle et al., 2000). However, their inactivation did not lead to subsequent impairments in recall and consolidation. Therefore, they may indeed participate in the online formation of (symmetric) hippocampal-neocortical indices. Once these indices are formed, memory retrieval can occur in the absence of their involvement.

2. During offline periods, spatial information is reinstated, again *verbatim*, in some (cf. Chapter 4) but not all (cf. Chapter 3) neocortical regions.

From the perspective of systems level consolidation, these findings imply, therefore, that the encoding and decoding of mnemonic contents occur both within the spatial domain.

In light of these findings, I propose the following hypothesis: During online periods, the hippocampus broadcast a sparse spatial code to vastly distributed regions of the cortex to partition their activity into discrete locations. By modulating cortical neurons with the place code, the hippocampus functionally "tags" their downstream targets and indexes them on-the-fly in such a way as to form symmetric channels for the encoding and decoding of information/memories. In this sense, the establishment of a "symmetric index" occurs during online periods, in conjunction with the encoding of the contents of memory episodes. More precisely, the index and the contents of memories are one and the same. During offline periods, these same spatial representations are spontaneously retrieved by presentation of partial cues in the form of non-spatial information encoded locally in the cortex. These spontaneous retrievals can be initiated by any cortical region (e.g., RSC or M2). However, subsequent retrieval of "raw" spatial information may only target specific cortical areas (e.g., only M2 but not RSC). This regional bias could potentially be demarcated by the *default mode network* (e.g., RSC) and the *task positive network* (e.g., M2) (cf. Kaefer et al., 2022). Over time, this heavily distributed and redundant place code is consolidated into neocortical representations. By transforming high-dimensional cortical features into a unified "spatial domain", memory becomes resilient to focal damages. Therefore, the hippocampal-neocortical dialogue is a mnemonic dialogue that occurs over space, with contents being dynamically indexed and encoded and subsequently consolidated into parallel and distributed engrams.

This hypothesis is, of course, overwhelmingly speculative at best. Given the novelty of these results and the many potential pitfalls (e.g., head-fixed as opposed to freely moving, reactivation of conjunctive spatial and non-spatial features as opposed to well separated, etc.), to even propose such a hypothesis seems ill-justified. Nevertheless, it is not everyday that one has the chance to write a thesis and if there is one function that the brain performs, which I can be sure of, that would be to desperately find explanations where no explanations are to be found.

Bibliography

Abadchi, J. K., Nazari-Ahangarkolaee, M., Gattas, S., Bermudez-Contreras, E., Luczak, A., McNaughton, B. L., and Mohajerani, M. H. (2020). Spatiotemporal patterns of neocortical activity around hippocampal sharp-wave ripples. *eLife*, 9:e51972.

Aguirre, G. K. and D'Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain*, 122(9):1613–1628.

Alexander, A. S. and Nitz, D. A. (2015). Retrosplenial cortex maps the conjunction of internal and external spaces. *Nature Neuroscience*, 18(8):1143–1151.

Alexander, A. S. and Nitz, D. A. (2017). Spatially Periodic Activation Patterns of Retrosplenial Cortex Encode Route Sub-spaces and Distance Traveled. *Current Biology*, 27(11):1551–1560.e4.

Allen, T. A., Salz, D. M., McKenzie, S., and Fortin, N. J. (2016). Nonspatial sequence coding in ca1 neurons. *Journal of Neuroscience*, 36(5):1547–1563.

Alyan, S. and McNaughton, B. (1999). Hippocampectomized rats are capable of homing by path integration. *Behavioral neuroscience*, 113(1):19.

Amaral, D. G. and Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience*, 31(3):571–591.

Ammassari-Teule, M. and Passino, E. (1997). The dorsal hippocampus is selectively involved in the processing of spatial information even in mice with a genetic hippocampal dysfunction. *Psychobiology*, 25(2):118–125.

Arns, M., Sauvage, M., and Steckler, T. (1999). Excitotoxic hippocampal lesions disrupt allocentric spatial learning in mice: effects of strain and task demands. *Behavioural brain research*, 106(1-2):151–164.

Attardo, A., Fitzgerald, J. E., and Schnitzer, M. J. (2015). Impermanence of dendritic spines in live adult ca1 hippocampus. *Nature*, 523:592–596.

Auger, S. D., Mullally, S. L., and Maguire, E. A. (2012). Retrosplenial cortex codes for permanent landmarks. *PloS one*, 7(8):e43620.

Balduzzi, D., Frean, M., Leary, L., Lewis, J., Ma, K. W.-D., and McWilliams, B. (2017). The shattered gradients problem: If resnets are the answer, then what is the question? In *International Conference on Machine Learning*, pages 342–350. PMLR.

Barthas, F. and Kwan, A. C. (2017). Secondary motor cortex: Where 'sensory' meets 'motor' in the rodent frontal cortex. *Trends in Neurosciences*, 40:181–193.

Bassett, J. P., Tullman, M. L., and Taube, J. S. (2007). Lesions of the Tegmentomammillary Circuit in the Head Direction System Disrupt the Head Direction Signal in the Anterior Thalamus. *The Journal of Neuroscience*, 27(28):7564.

Battaglia, F. P., Sutherland, G. R., and McNaughton, B. L. (2004). Hippocampal sharp wave bursts coincide with neocortical "up-state" transitions. *Learning and Memory*, 11(6):697–704.

Battaglia, F. P. and Treves, A. (1998). Stable and rapid recurrent processing in realistic autoassociative memories. *Neural Computation*, 10(2):431–450.

Bayley, P. J. and Squire, L. R. (2002). Medial temporal lobe amnesia: Gradual acquisition of factual information by nondeclarative memory. *Journal of Neuroscience*, 22(13):5741–5748.

Bendor, D. and Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience*, 15:1439–44.

Bengio, Y. et al. (2009). Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127.

Blair, H. T., Cho, J., and Sharp, P. E. (1998). Role of the Lateral Mammillary Nucleus in the Rat Head Direction Circuit: A Combined Single Unit Recording and Lesion Study. *Neuron*, 21(6):1387–1397.

Bonin, V., Histed, M. H., Yurgenson, S., and Reid, R. C. (2011). Local Diversity and Fine-Scale Organization of Receptive Fields in Mouse Visual Cortex. *Journal of Neuroscience*, 31(50):18506–18521.

Bonnevie, T., Dunn, B., Fyhn, M., Hafting, T., Derdikman, D., Kubie, J. L., Roudi, Y., Moser, E. I., and Moser, M. B. (2013). Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience 2013 16:3*, 16(3):309–317.

Botez, M. I. (1987). *Neuropsychologie clinique et neurologie du comportement*. Les presses de l'Université de Montréal.

Bower, M. R. (2005). Sequential-Context-Dependent Hippocampal Activity Is Not Necessary to Learn Sequences with Repeated Elements. *Journal of Neuroscience*, 25(6):1313–1323.

Bower, M. R., Stead, M., Bower, R. S., Kucewicz, M. T., Sulc, V., Cimbalnik, J., Brinkmann, B. H., Vasoli, V. M., Louis, E. K. S., Meyer, F. B., et al. (2015). Evidence for consolidation of neuronal assemblies after seizures in humans. *Journal of Neuroscience*, 35(3):999–1010.

Brandon, M. P., Bogaard, A. R., Libby, C. P., Connerney, M. A., Gupta, K., and Hasselmo, M. E. (2011). Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science*, 332(6029):595–599.

Brandon, M. P., Koenig, J., Leutgeb, J. K., and Leutgeb, S. (2014). New and distinct hippocampal place codes are generated in a new environment during septal inactivation. *Neuron*, 82(4):789.

Brooks, D. N. and Baddeley, A. (1976). What can amnesic patients learn? *Neuropsychologia*, 14(1):111–122.

Brun, V. H., Leutgeb, S., Wu, H. Q., Schwarcz, R., Witter, M. P., Moser, E. I., and Moser, M. B. (2008). Impaired Spatial Representation in CA1 after Lesion of Direct Input from Entorhinal Cortex. *Neuron*, 57(2):290–302.

Brzdak, P., Wójcicka, O., Zareba-Koziol, M., Minge, D., Henneberger, C., Wlodarczyk, J., Mozrzymas, J. W., and Wójtowicz, T. (2019). Synaptic potentiation at basal and apical dendrites of hippocampal pyramidal neurons involves activation of a distinct set of extracellular and intracellular molecular cues. *Cerebral Cortex*, 29(1):283–304.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Burgess, N., Barry, C., and O'Keefe, J. (2007). An oscillatory interference model of grid cell firing. *Hippocampus*, 17(9):801–812.

Burgess, N., Jackson, A., Hartley, T., and O'Keefe, J. (2000). Predictions derived from modelling the hippocampal role in navigation. *Biological cybernetics*, 83(3):301–312.

Burgess, N. E., Jeffery, K. J., and O'Keefe, J. E. (1999). *The hippocampal and parietal foundations of spatial cognition.* Oxford University Press.

Buzsáki, G. and Llinás, R. (2017). Space and time in the brain. *Science*, 358(6362):482–485.

Buzsáki, G. and Tingley, D. (2018). Space and time: the hippocampus as a sequence generator. *Trends in cognitive sciences*, 22(10):853–869.

Buzsáki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral Cortex*, 6:81–92.

Calton, J. L., Stackman, R. W., Goodridge, J. P., Archey, W. B., Dudchenko, P. A., and Taube, J. S. (2003). Hippocampal Place Cell Instability after Lesions of the Head Direction Cell Network. *Journal of Neuroscience*, 23(30):9719–9731.

Catani, M., Dell'Acqua, F., Bizzi, A., Forkel, S. J., Williams, S. C., Simmons, A., Murphy, D. G., and de Schotten, M. T. (2012). Beyond cortical localization in clinico-anatomical correlation. *cortex*, 48(10):1262–1287.

Chang, H., Esteves, I. M., Neumann, A. R., Mohajerani, M. H., and McNaughton, B. L. (2023). Cortical reactivation of spatial and non-spatial features coordinates with hippocampus to form a memory dialogue. *Nature Communications*, 14:7748.

Chang, H., Esteves, I. M., Neumann, A. R., Sun, J., Mohajerani, M. H., and McNaughton, B. L. (2020). Coordinated activities of retrosplenial ensembles during resting-state encode spatial landmarks. *Philosophical Transactions of the Royal Society B*, 375(1799):20190228.

Chen, G., King, J. A., Burgess, N., and O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1):378–383.

Chen, G., Manson, D., Cacucci, F., and Wills, T. J. (2016). Absence of visual input results in the disruption of grid cell firing in the mouse. *Current Biology*, 26(17):2335–2342.

Chen, L. L., Lin, L.-H., Green, E. J., Barnes, C. A., and McNaughton, B. L. (1994). Head-direction cells in the rat posterior cortex. *Experimental Brain Research*, 101(1):8–23.

Chirimuuta, M. (2021). Reflex theory, cautionary tale: misleading simplicity in early neuroscience. *Synthese*, 199(5-6):12731–12751.

Chrobak, J. J. and Buzsáki, G. (1996). High-frequency oscillations in the output networks of the hippocampal–entorhinal axis of the freely behaving rat. *The Journal of Neuroscience*, 16:3056 LP – 3066.

Claparède, E. (1911). Récognition et moïté. In Flournoy, T. and Claparède, E., editors, *Archives de psychologie*, volume 11, pages 79–90. Librairie Kündig, Genève.

Clark, B. J., Bassett, J. P., Wang, S. S., and Taube, J. S. (2010). Impaired head direction cell representation in the anterodorsal thalamus after lesions of the retrosplenial cortex. *The Journal of Neuroscience*, 30(15):5289–302.

Clark, R. E., Broadbent, N. J., and Squire, L. R. (2005). Hippocampus and remote spatial memory in rats. *Hippocampus*, 15(2):260–272.

Cohen, N. J. and Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. MIT press.

Cohen, N. J. and Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210(4466):207–210.

Colgin, L. L., Leutgeb, S., Jezek, K., Leutgeb, J. K., Moser, E. I., McNaughton, B. L., and Moser, M.-B. (2010). Attractor-map versus autoassociation based attractor dynamics in the hippocampal network. *Journal of Neurophysiology*, 104:35–50.

Cooper, B. G. and Mizumori, S. J. Y. (1999). Retrosplenial cortex inactivation selectively impairs navigation in darkness. *NeuroReport*, 10(3):625–30.

Cowansage, K., Chang, A., Mayford, M., Shuman, T., Dillingham, B., and Golshani, P. (2014). Direct Reactivation of a Coherent Neocortical Memory of Context. *Neuron*, 84(2):432–441.

Deacon, R. M., Croucher, A., and Rawlins, J. N. P. (2002). Hippocampal cytotoxic lesion effects on species-typical behaviours in mice. *Behavioural brain research*, 132(2):203–213.

Deacon, R. M. and Rawlins, J. N. P. (2005). Hippocampal lesions, species-typical behaviours and anxiety in mice. *Behavioural brain research*, 156(2):241–249.

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

dos Santos, V. L., Ribeiro, S., and Tort, A. B. (2013). Detecting cell assemblies in large neuronal populations. *Journal of Neuroscience Methods*, 220:149–166.

Dragoi, G. and Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, 469(7330):397–401.

Dragoi, G. and Tonegawa, S. (2013). Distinct preplay of multiple novel spatial experiences in the rat. *Proceedings of the National Academy of Sciences*, 110(22):9100–9105.

Drieu, C., Todorova, R., and Zugaro, M. (2018). Nested sequences of hippocampal assemblies during behavior support subsequent sleep replay. *Science*, 362(6415):675–679.

Drieu, C. and Zugaro, M. (2019). Hippocampal sequences during exploration: mechanisms and functions. *Frontiers in Cellular Neuroscience*, 13:232.

Du, P., Kibbe, W. A., and Lin, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17):2059–2065.

Duff, M. C., Covington, N. V., Hilverman, C., and Cohen, N. J. (2020). Semantic memory and the hippocampus: Revisiting, reaffirming, and extending the reach of their critical relationship. *Frontiers in human neuroscience*, 13:471.

Duff, M. C., Hengst, J., Tranel, D., and Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature neuroscience*, 9(1):140–146.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature neuroscience*, 3(11):1184–1191.

Eckert, M. J., McNaughton, B. L., and Tatsuno, M. (2020). Neural ensemble reactivation in rapid eye movement and slow-wave sleep coordinate with muscle activity to promote rapid motor skill learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375.

Eichenbaum, H. (1996). Is the rodent hippocampus just for 'place'? *Current Opinion in Neurobiology*, 6(2):187–195.

Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron*, 44(1):109–120.

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., and Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, 23(2):209–226.

Esteves, I. M., Chang, H., Neumann, A. R., and McNaughton, B. L. (2023). Consolidation of cellular memory representations in superficial neocortex. *Iscience*, 26(2).

Esteves, I. M., Chang, H., Neumann, A. R., Sun, J., Mohajerani, M. H., and McNaughton, B. L. (2021). Spatial information encoding across multiple neocortical regions depends on an intact hippocampus. *Journal of Neuroscience*, 41(2):307–319.

Euston, D. R., Tatsuno, M., and McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *science*, 318(5853):1147–1150.

Failor, S. W., Carandini, M., and Harris, K. D. (2021). Visuomotor association orthogonalizes visual cortical population codes. Preprint at `https://www.biorxiv.org/content/10.1101/2021.05.23.445338v3`.

Fairen, A., DeFelipe, J., and Regidor, J. (1984). Nonpyramidal neurons. In *Cerebral Cortex: Cellular components of the cerebral cortex*, pages 206–241. New York: Plenum.

Faraji, J., Lehmann, H., Metz, G. A., and Sutherland, R. J. (2008). Rats with hippocampal lesion show impaired learning and memory in the ziggurat task: a new task to evaluate spatial behavior. *Behavioural brain research*, 189(1):17–31.

Feldman, D. E. (2012). The spike-timing dependence of plasticity. *Neuron*, 75(4):556–571.

Feldman, J. A. (1981). A connectionist model of visual memory. In Hinton, G. E. and Anderson, J. A., editors, *Parallel Models of Associative Memory*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Ferbinteanu, J. and Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron*, 40(6):1227–1239.

Fernandez, L. M., Comte, J.-C., Le Merre, P., Lin, J.-S., Salin, P.-A., and Crochet, S. (2017). Highly dynamic spatiotemporal organization of low-frequency activities during behavioral states in the mouse cerebral cortex. *Cerebral cortex*, 27(12):5444–5462.

Ferrier, D. (1875). Experiments on the brain of monkeys.—no. i. *Proceedings of the Royal Society of London*, 23(156-163):409–430.

Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., and Keller, G. B. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nature neuroscience*, 19(12):1658–1664.

Fisher, S. P., Cui, N., McKillop, L. E., Gemignani, J., Bannerman, D. M., Oliver, P. L., Peirson, S. N., and Vyazovskiy, V. V. (2016). Stereotypic wheel running decreases cortical activity in mice. *Nature communications*, 7(1):13138.

Flaherty, C. F., Hamilton, L. W., Gandelman, R. J., and Spear, N. E. (1977). *Learning and Memory*. Rand McNally College Publishing Company.

Fortin, N. J., Agster, K. L., and Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature neuroscience*, 5(5):458–462.

Foster, D. J. (2017). Replay Comes of Age. *Annual Review of Neuroscience*, 40(1):581–602.

Foster, D. J. and Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683.

França, T. F. and Monserrat, J. M. (2019). Hippocampal place cells are topographically organized, but physical space has nothing to do with it. *Brain Structure and Function*, 224(9):3019–3029.

Frankland, P. W. and Bontempi, B. (2005). The organization of recent and remote memories. *Nature reviews neuroscience*, 6(2):119–130.

Friston, K., Moran, R. J., Nagai, Y., Taniguchi, T., Gomi, H., and Tenenbaum, J. (2021). World model learning and inference. *Neural Networks*.

Fuhrmann, F., Justus, D., Sosulina, L., Kaneko, H., Beutel, T., Friedrichs, D., Schoch, S., Schwarz, M. K., Fuhrmann, M., and Remy, S. (2015). Locomotion, Theta Oscillations, and the Speed-Correlated Firing of Hippocampal Neurons Are Controlled by a Medial Septal Glutamatergic Circuit. *Neuron*, 86(5):1253–1264.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202.

Gabrieli, J. D., Cohen, N. J., and Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and cognition*, 7(2):157–177.

Gardner, R. J., Lu, L., Wernle, T., Moser, M.-B., and Moser, E. I. (2019). Correlation structure of grid cells is preserved during sleep. *Nature Neuroscience*, 22(4):598–608.

Geisler, C., Robbe, D., Zugaro, M., Sirota, A., and Buzsáki, G. (2007). Hippocampal place cell assemblies are speed-controlled oscillators. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19):8149–8154.

Gener, T., Perez-Mendez, L., and Sanchez-Vives, M. V. (2013). Tactile modulation of hippocampal place fields. *Hippocampus*, 23(12):1453–62.

Geva-Sagiv, M., Las, L., Yovel, Y., and Ulanovsky, N. (2015). Spatial cognition in bats and rats: From sensory acquisition to multiscale maps and navigation. *Nature Reviews Neuroscience*, 16(2):94–108.

Glisky, E. L., Schacter, D. L., and Tulving, E. (1986a). Computer learning by memory-impaired patients: Acquisition and retention of complex knowledge. *Neuropsychologia*, 24(3):313–328.

Glisky, E. L., Schacter, D. L., and Tulving, E. (1986b). Learning and retention of computer-related vocabulary in memory-impaired patients: Method of vanishing cues. *Journal of Clinical and Experimental Neuropsychology*, 8(3):292–312.

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

Grosmark, A. D., Sparks, F. T., Davis, M. J., and Losonczy, A. (2021). Reactivation predicts the consolidation of unbiased long-term cognitive maps. *Nature Neuroscience*, 24:1574–1585.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806.

Haggerty, D. C. and Ji, D. (2015). Activities of visual cortical and hippocampal neurons co-fluctuate in freely moving rats during spatial behavior. *elife*, 4:e08902.

Hales, J. B., Schlesiger, M. I., Leutgeb, J. K., Squire, L. R., Leutgeb, S., and Clark, R. E. (2014). Medial entorhinal cortex lesions only partially disrupt hippocampal place cells and hippocampus-dependent place memory. *Cell reports*, 9(3):893–901.

Harland, B., Grieves, R. M., Bett, D., Stentiford, R., Wood, E. R., and Dudchenko, P. A. (2017). Lesions of the head direction cell system increase hippocampal place field repetition. *Current Biology*, 27(17):2706–2712.

Harris, K. D. and Thiele, A. (2011). Cortical state and attention. *Nature reviews neuroscience*, 12(9):509–523.

Hartley, T., Burgess, N., Lever, C., Cacucci, F., and O'keefe, J. (2000). Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus*, 10(4):369–379.

Harvey, C. D., Coen, P., and Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62–68.

Hassabis, D., Kumaran, D., Vann, S. D., and Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, 104(5):1726–1731.

He, L. (2023). the closing ceremony for 31st wmc finals. `https://www.worldmemorychampionships.com/the-closing-ceremony-for-31st-wmc-finals/`.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Hertz, J., Krogh, A., Palmer, R. G., and Horner, H. (1991). *Introduction to the theory of neural computation*. American Institute of Physics.

Hill, A. (1978). First occurrence of hippocampal spatial firing in a new environment. *Experimental neurology*, 62(2):282–297.

Hinman, J. R., Brandon, M. P., Climer, J. R., Chapman, G. W., and Hasselmo, M. E. (2016). Multiple Running Speed Signals in Medial Entorhinal Cortex. *Neuron*, 91(3):666–679.

Hinman, J. R., Penley, S. C., Long, L. L., Escabí, M. A., and Chrobak, J. J. (2011). Septotemporal variation in dynamics of theta: speed and habituation. *Journal of neurophysiology*, 105(6):2675–2686.

Hinton, G. E. (2012). A practical guide to training restricted boltzmann machines. *Neural Networks: Tricks of the Trade: Second Edition*, pages 599–619.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554.

Hoffman, K. L. and McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science*, 297(5589):2070–2073.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books, Inc., New York.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106.

Hughlings Jackson, J. (1873). On the anatomical investigation of epilepsy and epileptiform convulsions. *British medical journal*, 1(645):531.

Insausti, R., Herrero, M. T., and Witter, M. P. (1997). Entorhinal cortex of the rat: Cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. *Hippocampus*, 7(2):146–183.

Isomura, Y., Sirota, A., Özen, S., Montgomery, S., Mizuseki, K., Henze, D. A., and Buzsáki, G. (2006). Integration and segregation of activity in entorhinal-hippocampal subregions by neocortical slow oscillations. *Neuron*, 52:871–882.

Itskov, P. M., Vinnik, E., and Diamond, M. E. (2011). Hippocampal representation of touch-guided behavior in rats: persistent and independent traces of stimulus and reward location. *PloS one*, 6(1):e16462.

Itskov, P. M., Vinnik, E., Honey, C., Schnupp, J., and Diamond, M. E. (2012). Sound sensitivity of neurons in rat hippocampus during performance of a sound-guided task. *Journal of neurophysiology*, 107(7):1822–1834.

Jacob, P. Y., Casali, G., Spieser, L., Page, H., Overington, D., and Jeffery, K. (2017). An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. *Nature Neuroscience*, 20(2):173–175.

Jadhav, S. P., Rothschild, G., Roumis, D. K., and Frank, L. M. (2016). Coordinated excitation and inhibition of prefrontal ensembles during awake hippocampal sharp-wave ripple events. *Neuron*, 90(1):113–127.

Jayasumana, S., Hartley, R., Salzmann, M., Li, H., and Harandi, M. (2013). Kernel methods on the riemannian manifold of symmetric positive definite matrices. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 73–80.

Ji, D. and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience*, 10(1):100–107.

Jung, M. W., Wiener, S. I., and McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, 14(12):7347–7356.

Kaefer, K., Stella, F., McNaughton, B. L., and Battaglia, F. P. (2022). Replay, the default mode network and the cascaded memory systems model. *Nature Reviews Neuroscience*, 23(10):628–640.

Kentros, C., Hargreaves, E., Hawkins, R. D., Kandel, E. R., Shapiro, M., and Muller, R. V. (1998). Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science*, 280(5372):2121–2126.

Kirkcaldie, M., Watson, C., Paxinos, G., and Franklin, K. (2012). Straightening out the mouse neocortex. In *Australian Neuroscience Society Annual Conference*.

Knierim, J. J. and Neunuebel, J. P. (2016). Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiology of learning and memory*, 129:38–49.

Koenig, J., Linder, A. N., Leutgeb, J. K., and Leutgeb, S. (2011). The spatial periodicity of grid cells is not sustained during reduced theta oscillations. *Science*, 332(6029):592–595.

Kohonen, T. (1988). *Self-Organization and Associative Memory*. Springer-Verla.

Kolb, B., Sutherland, R. J., and Whishaw, I. Q. (1983). A comparison of the contributions of the frontal and parietal association cortex to spatial localization in rats. *Behavioral neuroscience*, 97(1):13.

Krupic, J., Burgess, N., and O'Keefe, J. (2012). Neural representations of location composed of spatially periodic bands. *Science*, 337(6096):853–857.

Kudrimoti, H. S., Barnes, C. A., and McNaughton, B. L. (1999). Reactivation of Hippocampal Cell Assemblies: Effects of Behavioral State, Experience, and EEG Dynamics. *The Journal of Neuroscience*, 19(10):4090–4101.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L., and Pennartz, C. M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biology*, 7(8):e1000173.

Lashley, K. S. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain*. University of Chicago Press.

Lashley, K. S. (1950). In search of the engram. In *Physiological Mechanisms in Animal Behaviour*, Symposia of the society for experimental biology. Cambridge Univ. Press.

Lassalle, J.-M., Bataille, T., and Halley, H. (2000). Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the morris navigation task. *Neurobiology of learning and memory*, 73(3):243–257.

Le, Q. V., Karpenko, A., Ngiam, J., and Ng, A. Y. (2011). Ica with reconstruction cost for efficient overcomplete feature learning. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, pages 1–9.

Leahey, T. H. (1997). *A history of psychology: Main currents in psychological thought*. Prentice-Hall, Inc.

Lee, A. K. and Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36(6):1183–1194.

Lengyel, M., Szatmáry, Z., and Érdi, P. (2003). Dynamically detuned oscillations account for the coupled rate and temporal code of place cell firing. *Hippocampus*, 13(6):700–714.

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., and Moser, M.-B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309(5734):619–623.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., and Burgess, N. (2009). Boundary Vector Cells in the Subiculum of the Hippocampal Formation. *The Journal of Neuroscience*, 29(31):9771.

Li, X.-G., Somogyi, P., Ylinen, A., and Buzsáki, G. (1994). The hippocampal ca3 network: an in vivo intracellular labeling study. *Journal of comparative neurology*, 339(2):181–208.

Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, 7(1):13276.

Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience*, 33(10):2017–2031.

Lisman, J., Buzsáki, G., Eichenbaum, H., Nadel, L., Ranganath, C., and Redish, A. D. (2017). Viewpoints: how the hippocampus contributes to memory, navigation and cognition. *Nature neuroscience*, 20(11):1434–1447.

Liu, C., Todorova, R., Tang, W., Oliva, A., and Fernandez-Ruiz, A. (2023). Associative and predictive hippocampal codes support memory-guided behaviors. *Science*, 382(6668):eadi8237.

Logothetis, N. K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H. C., Besserve, M., and Oeltermann, A. (2012). Hippocampal–cortical interaction during periods of subcortical silence. *Nature*, 491:547–553.

Lowenstein, D. H. (2018). Seizures and epilepsy. In Jameson, J. L., Fauci, A. S., Kasper, D. L., Hauser, S. L., Longo, D. L., and Loscalzo, J., editors, *Harrison's Principles of Internal Medicine, 20e*. McGraw-Hill Education, New York, NY.

MacDonald, C. J., Carrow, S., Place, R., and Eichenbaum, H. (2013). Distinct hippocampal time cell sequences represent odor memories in immobilized rats. *Journal of Neuroscience*, 33(36):14607–14616.

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology*, 42(3):225–238.

Maguire, E. A., Valentine, E. R., Wilding, J. M., and Kapur, N. (2003). Routes to remembering: the brains behind superior memory. *Nature neuroscience*, 6(1):90–95.

Malvache, A., Reichinnek, S., Villette, V., Haimerl, C., and Cossart, R. (2016). Awake hippocampal reactivations project onto orthogonal neuronal assemblies. *Science*, 353(6305):1280–1283.

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., Ota, K., Matsubara, C., Inutsuka, A., Sato, M., Ohkura, M., Yamanaka, A., Yanagawa, Y., Nakai, J., Hayashi, Y., Larkum, M. E., and Murayama, M. (2015). A top-down cortical circuit for accurate sensory perception. *Neuron*, 86:1304–1316.

Manns, J. R. and Eichenbaum, H. (2009). A cognitive map for object memory in the hippocampus. *Learning & memory*, 16(10):616–624.

Manns, J. R., Howard, M. W., and Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, 56(3):530–540.

Mao, D., Kandler, S., McNaughton, B. L., and Bonin, V. (2017). Sparse orthogonal population representation of spatial context in the retrosplenial cortex. *Nature communications*, 8(1):243.

Mao, D., Neumann, A. R., Sun, J., Bonin, V., Mohajerani, M. H., and McNaughton, B. L. (2018). Hippocampus-dependent emergence of spatial sequence coding in retrosplenial cortex. *Proceedings of the National Academy of Sciences*, 115(31):8015–8018.

Markus, E. J., Barnes, C. A., Bruce, ., McNaughton, L., Gladden, V. L., and Skaggs, W. E. (1994). Spatial Information Content and Reliability of Hippocampal CA1 Neurons: Effects of Visual Input. *HZPPOCAMPUS*, 4(4):410–421.

Markus, E. J., Qin, Y.-L., Leonard, B., Skaggs, W. E., McNaughton, B. L., and Barnes, C. A. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *Journal of Neuroscience*, 15(11):7079–7094.

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal society of London. Series B. Biological sciences*, 176(1043):161–234.

Marr, D. (1971). Simple memory: a theory for archicortex. *Proceedings of the Royal society of London. Series B. Biological sciences*, 262:23–81.

Martin, S. J., De Hoz, L., and Morris, R. G. (2005). Retrograde amnesia: neither partial nor complete hippocampal lesions in rats result in preferential sparing of remote spatial memory, even after reminding. *Neuropsychologia*, 43(4):609–624.

Mauguiere, F. and Corkin, S. (2015). Hm never again! an analysis of hm's epilepsy and treatment. *Revue neurologique*, 171(3):273–281.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419.

McFarland, W. L., Teitelbaum, H., and Hedges, E. K. (1975). Relationship between hippocampal theta activity and running speed in the rat. *Journal of comparative and physiological psychology*, 88(1):324.

McNaughton, B., Chen, L., and Markus, E. (1991). "dead reckoning," landmark learning, and the sense of direction: a neurophysiological and computational hypothesis. *Journal of Cognitive Neuroscience*, 3(2):190–202.

McNaughton, B. L. (2010). Cortical hierarchies, sleep, and the extraction of knowledge from memory. *Artificial Intelligence*, 174(2):205–214.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., Kudrimoti, H., Qin, Y., Skaggs, W., Suster, M., et al. (1996). Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *Journal of Experimental Biology*, 199(1):173–185.

McNaughton, B. L., Barnes, C. A., and O'keefe, J. (1983). The Contributions of Position, Direction, and Velocity to Single Unit Activity in the Hippocampus of Freely-moving Rats. *Exp Brain Res*, 52:41–49.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M.-B. (2006). Path integration and the neural basis of the'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678.

McNaughton, B. L., Leonard, B., and Chen, L. (1989). Cortical-hippocampal interactions and cognitive mapping: A hypothesis based on reintegration of the parietal and inferotemporal pathways for visual processing. *Psychobiology*, 17(3):230–235.

McNaughton, B. L., Mizumori, S., Barnes, C., Leonard, B., Marquis, M., and Green, E. (1994). Cortical representation of motion during unrestrained spatial navigation in the rat. *Cerebral Cortex*, 4(1):27–39.

McNaughton, B. L. and Morris, R. G. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in neurosciences*, 10(10):408–415.

McNaughton, B. L. and Nadel, L. (1990). *Hebb-Marr networks and the neurobiological representation of action in space.*, pages 1–63. Lawrence Erlbaum Associates, Inc.

Mehta, M. R., Barnes, C. A., and McNaughton, B. L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences*, 94(16):8918–8921.

Miller, V. M. and Best, P. J. (1980). Spatial correlates of hippocampal unit activity are altered by lesions of the fornix and entorhinal cortex. *Brain research*, 194(2):311–323.

Milner, B. (1962). Les troubles de la mémoire accompagnant des lésions hippocampiques bilatérales. In Passouant, P., editor, *Physiologie de l'hippocampe*. Centre national de la recherche scientifique, Quai Anatole-France, Paris.

Milner, P. M. (1999). Cell assemblies: Whose idea? *Psycoloquay*, 10:53.

Minderer, M., Brown, K. D., and Harvey, C. D. (2019). The spatial structure of neural encoding in mouse posterior cortex during navigation. *Neuron*, 102(1):232–248.

Minsky, M. and Papert, S. (1969). *Perceptrons*. MIT Press.

Mishkin, M., Suzuki, W. A., Gadian, D. G., and Vargha-Khadem, F. (1997). Hierarchical organization of cognitive memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1360):1461–1467.

Miyamoto, D., Hirai, D., Fung, C. C., Inutsuka, A., Odagawa, M., Suzuki, T., Boehringer, R., Adaikkan, C., Matsubara, C., Matsuki, N., Fukai, T., McHugh, T. J., Yamanaka, A., and Murayama, M. (2016). Top-down cortical input during nrem sleep consolidates perceptual memory. *Science*, 352:1315–1318.

Miyashita, T. and Rockland, K. S. (2007). GABAergic projections from the hippocampus to the retrosplenial cortex in the rat. *European Journal of Neuroscience*, 26(5):1193–1204.

Mizumori, S., McNaughton, B., Barnes, C., and Fox, K. (1989). Preserved spatial coding in hippocampal ca1 pyramidal cells during reversible suppression of ca3c output: evidence for pattern completion in hippocampus. *Journal of Neuroscience*, 9(11):3915–3928.

Mizuseki, K. and Buzsáki, G. (2013). Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell reports*, 4(5):1010–1021.

Moita, M. A., Rosis, S., Zhou, Y., LeDoux, J. E., and Blair, H. T. (2003). Hippocampal place cells acquire location-specific responses to the conditioned stimulus during auditory fear conditioning. *Neuron*, 37(3):485–97.

Monaghan, D. and Cotman, C. (1985). Distribution of N-methyl-D-aspartate-sensitive L-[3H]glutamate-binding sites in rat brain. *The Journal of Neuroscience*, 5(11):2909–2919.

Morris, R. G., Garrud, P., Rawlins, J. a., and O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, 297(5868):681–683.

Moser, M. B., Rowland, D. C., and Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor Perspectives in Biology*, 7(2).

Muir, G. M. and Bilkey, D. K. (2001). Instability in the place field location of hippocampal place cells after lesions centered on the perirhinal cortex. *Journal of Neuroscience*, 21(11):4016–25.

Muller, R. U. and Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, 7(7):1951–68.

Mumby, D. G., Gaskin, S., Glenn, M. J., Schramek, T. E., and Lehmann, H. (2002). Hippocampal damage and exploratory preferences in rats: memory for objects, places, and contexts. *Learning & memory*, 9(2):49–57.

Murtagh, F. and Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., and Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, 19(21):9497–9507.

Nadel, L. (1968). Dorsal and ventral hippocampal lesions and behavior. *Physiology & Behavior*, 3(6):891–900.

Nadel, L. and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2):217–227.

Nadel, L., Winocur, G., Ryan, L., and Moscovitch, M. (2007). Systems consolidation and hippocampus: two views. *Debates in Neuroscience*, 1:55–66.

Nerad, L., Liu, P., and Bilkey, D. K. (2009). Bilateral NMDA lesions centered on the postrhinal cortex have minimal effects on hippocampal place cell firing. *Hippocampus*, 19(3):221–7.

Nguyen, N. D., Lutas, A., Fernando, J., Vergara, J., McMahon, J., Dimidschstein, J., and Andermann, M. L. (2022). Cortical reactivations predict future sensory responses. Preprint at `https://www.biorxiv.org/content/10.1101/2022.11.14.516421v1`.

Nitz, D. (2009). Parietal cortex, navigation, and the construction of arbitrary reference frames for spatial information. *Neurobiology of learning and memory*, 91(2):179–185.

Nitz, D. A. (2006). Tracking route progression in the posterior parietal cortex. *Neuron*, 49(5):747–756.

Nitz, D. A. (2012). Spaces within spaces: rat parietal cortex neurons register position across three reference frames. *Nature neuroscience*, 15(10):1365–1367.

O' Keefe, J. and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature 1996 381:6581*, 381(6581):425–428.

Ocampo, A. C., Squire, L. R., and Clark, R. E. (2018). The beneficial effect of prior experience on the acquisition of spatial memory in rats with ca1, but not large hippocampal lesions: a possible role for schema formation. *Learning & Memory*, 25(3):115–121.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1):78–109.

O'Keefe, J. and Conway, D. H. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. *Experimental brain research*, 31:573–590.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175.

O'Keefe, J. and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford University Press, Oxford.

O'Keefe, J. and Recce, M. L. (1993). Phase Relationship Between Hippocampal Place Units and the EEG Theta Rhythm. *Hyppocampus*, 3(3):317–330.

O'Keefe, J. and Speakman, A. (1987). Single unit activity in the rat hippocampus during a spatial memory task. *Experimental Brain Research*, 68(1):1–27.

Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D., and Spiers, H. J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. *Elife*, 4:e06063.

Olson, J. M., Li, J. K., Montgomery, S. E., and Nitz, D. A. (2020). Secondary motor cortex transforms spatial information into planned action during navigation. *Current Biology*, 30(10):1845–1854.

O'Keefe, J. and Krupic, J. (2021). Do hippocampal pyramidal cells respond to nonspatial stimuli? *Physiological Reviews*, 101(3):1427–1456.

Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L. F., Dipoppa, M., Rossi, L. F., Carandini, M., and Harris, K. D. (2016). Suite2p : beyond 10 , 000 neurons with standard two-photon microscopy. *bioRxiv*, pages 1–30.

Packard, M. G. and Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual review of neuroscience*, 25(1):563–593.

Pakan, J. M., Currie, S. P., Fischer, L., and Rochefort, N. L. (2018). The impact of visual cues, reward, and motor feedback on the representation of behaviorally relevant spatial locations in primary visual cortex. *Cell reports*, 24(10):2521–2528.

Parron, C. and Save, E. (2004). Evidence for entorhinal and parietal cortices involvement in path integration in the rat. *Experimental brain research*, 159(3):349–359.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321(5894):1322–1327.

Pavlides, C. and Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *Journal of neuroscience*, 9(8):2907–2918.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford University Press, London.

Penfield, W. (1954). Temporal lobe epilepsy. *British Journal of Surgery*, 41(168):337–343.

Penfield, W. (1955). The permanent record of the stream of consciousness. *Acta Psychologica*, 11:47–69.

Penfield, W. and Milner, B. (1958). Memory deficit produced by bilateral lesions in the hippocampal zone. *Archives of Neurology and Psychiatry*, 79(5):475–497.

Pennartz, C. M., Lee, E., Verheul, J., Lipa, P., Barnes, C. A., and McNaughton, B. L. (2004). The ventral striatum in off-line processing: Ensemble reactivation during sleep and modulation by hippocampal ripples. *Journal of Neuroscience*, 24:6446–6456.

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., and Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature neuroscience*, 12(7):919–926.

Pfeiffer, B. E. and Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–79.

Phillips, C., Zeki, S., and Barlow, H. (1984). Localization of function in the cerebral cortex: past, present and future. *Brain*, 107(1):328–361.

Pnevmatikakis, E. A., Soudry, D., Gao, Y., Machado, T. A., Merel, J., Pfau, D., Reardon, T., Mu, Y., Lacefield, C., Yang, W., Ahrens, M., Bruno, R., Jessell, T. M., Peterka, D. S., Yuste, R., and Paninski, L. (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron*, 89(2):285–99.

Poldrack, R. A. and Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews*, 32(2):197–205.

Poulter, S., Hartley, T., and Lever, C. (2018). The neurobiology of mammalian navigation. *Current Biology*, 28(17):R1023–R1042.

Qin, Y.-L., McNaughton, B. L., Skaggs, W. E., and Barnes, C. A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1360):1525–1533.

Quirk, G. J., Muller, R. U., and Kubie2, J. L. (1990). The Firing of Hippocampal Place Cells in the Dark Depends on the Rat's Recent Experience. *The Journal of Neuroscience*, 7(6):2008–2017.

Ranck, J. B. (1973). Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats: Part I. Behavioral correlates and firing repertoires. *Experimental Neurology*, 41(2):462–531.

Ranganath, C. (2022). *Oxford Handbook of Memory: Episodic Memory*. PsyArXiv.

Ranganath, C. and Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature reviews neuroscience*, 13(10):713–726.

Rasch, B., Büchel, C., Gais, S., and Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, 315:1426–1429.

Rich, P. D., Liaw, H. P., and Lee, A. K. (2014). Large environments reveal the statistical structure governing hippocampal representations. *Science*, 345(6198):814–817.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Rothschild, G., Eban, E., and Frank, L. M. (2017). A cortical–hippocampal–cortical loop of information processing during memory consolidation. *Nature neuroscience*, 20(2):251–259.

Rumelhart, D. E., McClelland, J. L., and Group, T. P. R. (1986). *Parallel distributed processing*, volume 1. MIT press.

Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D., and Carandini, M. (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature*, 562(7725):124–127.

Samsonovich, A. and McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15):5900–5920.

Save, E., Nerad, L., and Poucet, B. (2000). Contribution of multiple sensory information to place field stability in hippocampal place cells. *Hippocampus*, 10(164-76).

Schlesiger, M. I., Cannova, C. C., Boublil, B. L., Hales, J. B., Mankin, E. A., Brandon, M. P., Leutgeb, J. K., Leibold, C., and Leutgeb, S. (2015). The medial entorhinal cortex is necessary for temporal organization of hippocampal neuronal activity. *Nature Neuroscience*, 18(8):1123–32.

Schwindel, C. D. and McNaughton, B. L. (2011). Hippocampal-cortical interactions and the dynamics of memory trace reactivation. In *Progress in Brain Research*, volume 193 of *Slow Brain Oscillations of Sleep, Resting State and Vigilance*, pages 163–177. Elsevier.

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11–21.

Serre, T. (2014). Hierarchical models of the visual system. *Encyclopedia of computational neuroscience*, 6:1–12.

Sheintuch, L., Geva, N., Baumer, H., Rechavi, Y., Rubin, A., and Ziv, Y. (2020). Multiple maps of the same spatial context can stably coexist in the mouse hippocampus. *Current biology*, 30(8):1467–1476.

Sheintuch, L., Rubin, A., Brande-Eilat, N., Geva, N., Sadeh, N., Pinchasof, O., and Ziv, Y. (2017). Tracking the same neurons across multiple days in ca2+ imaging data. *Cell reports*, 21(4):1102–1115.

Sherrington, C. S. (1906). *The integrative action of the nervous system*, volume 35. Yale University Press.

Siapas, A. G. and Wilson, M. A. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep of sws and during behavioral immobility. these high-frequency oscillations constitute a major mode of hippo-campal activity (buzsaki et al they provide. *Neuron*, 21:1123–1128.

Silva, D., Feng, T., and Foster, D. J. (2015). Trajectory events across hippocampal place cells require previous experience. *Nature neuroscience*, 18(12):1772–1779.

Sirota, A., Csicsvari, J., Buhl, D., and Buzsáki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences*, 100:2065 LP – 2069.

Skaggs, W. E. and McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271(5257):1870–1873.

Skaggs, W. E. and McNaughton, B. L. (1998). Spatial firing properties of hippocampal ca1 populations in an environment containing two visually identical regions. *Journal of Neuroscience*, 18(20):8455–8466.

Skaggs, W. E., McNaughton, B. L., and Gothard, K. M. (1993). An Information-Theoretic Approach to Deciphering the Hippocampal Code. In Hanson, S. J., Cowan, J. D., and Giles, C. L., editors, *Advances in Neural Information Processing Systems 5*, pages 1030–1037. Morgan-Kaufmann.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., and Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2):149–172.

Skelin, I., Kilianski, S., and McNaughton, B. L. (2019). Hippocampal coupling with cortical and subcortical structures in the context of memory consolidation. *Neurobiology of Learning and Memory*, 160:21–31.

Smith, D. M., Barredo, J., and Mizumori, S. J. Y. (2012). Complimentary roles of the hippocampus and retrosplenial cortex in behavioral context discrimination. *Hippocampus*, 22(5):1121–1133.

Solstad, T., Moser, E. I., and Einevoll, G. T. (2006). From grid cells to place cells: A mathematical model. *Hippocampus*, 16(12):1026–1031.

Solstad, T., Yousif, H. N., and Sejnowski, T. J. (2014). Place cell rate remapping by ca3 recurrent collaterals. *PLoS computational biology*, 10(6):e1003648.

Spiers, H. J., Hayman, R. M., Jovalekic, A., Marozzi, E., and Jeffery, K. J. (2015). Place field repetition and purely local remapping in a multicompartment environment. *Cerebral Cortex*, 25(1):10–25.

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, 9(3):206–221.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99(2):195–231.

Squire, L. R. (2009). The legacy of patient h.m. for neuroscience. *Neuron*, 61(1):6–9.

Squire, L. R. and Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus*, 8(3):205–211.

Squire, L. R. and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253(5026):1380–1386.

Stark, C., Stark, S., and Gordon, B. (2005). New semantic learning and generalization in a patient with amnesia. *Neuropsychology*, 19(2):139.

Steinvorth, S., Levine, B., and Corkin, S. (2005). Medial temporal lobe structures are needed to re-experience remote autobiographical memories: evidence from hm and wr. *Neuropsychologia*, 43(4):479–496.

Stella, F., Baracskay, P., O'Neill, J., and Csicsvari, J. (2019). Hippocampal reactivation of random trajectories resembling brownian diffusion. *Neuron*, 102(2):450–461.

Stensola, H., Stensola, T., Solstad, T., FrØland, K., Moser, M. B., and Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature 2012 492:7427*, 492(7427):72–78.

Strange, B. A., Witter, M. P., Lein, E. S., and Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience*, 15(10):655–669.

Sugden, A. U., Zaremba, J. D., Sugden, L. A., McGuire, K. L., Lutas, A., Ramesh, R. N., Alturkistani, O., Lensjø, K. K., Burgess, C. R., and Andermann, M. L. (2020). Cortical reactivations of recent sensory experiences predict bidirectional network changes during learning. *Nature neuroscience*, 23(8):981–991.

Sutherland, R. J., Kolb, B., and Whishaw, I. Q. (1982). Spatial mapping: definitive disruption by hippocampal or medial frontal cortical damage in the rat. *Neuroscience letters*, 31(3):271–276.

Sutherland, R. J., Sparks, F. T., and Lehmann, H. (2010). Hippocampus and retrograde amnesia in the rat model: a modest proposal for the situation of systems consolidation. *Neuropsychologia*, 48(8):2357–2369.

Sutherland, R. J., Whishaw, I. Q., and Kolb, B. E. (1988). Contributions of cingulate cortex to two forms of spatial learning and memory. *The Journal of Neuroscience*, 8(6):1863–1872.

Takahashi, N., Kawamura, M., Shiota, J., Kasahata, N., and Hirayama, K. (1997). Pure topographic disorientation due to right retrosplenial lesion. *Neurology*, 49(2):464–9.

Tanni, S., de Cothi, W., and Barry, C. (2022). State transitions in the statistically stable place cell population correspond to rate of perceptual change. *Current Biology*, 32(16):3505–3514.e7.

Tary, J. B., Herrera, R. H., and Van Der Baan, M. (2018). Analysis of time-varying signals using continuous wavelet and synchrosqueezed transforms. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2126):20170254.

Tatsuno, M., Lipa, P., and McNaughton, B. L. (2006). Methodological Considerations on the Use of Template Matching to Study Long-Lasting Memory Trace Replay. *Journal of Neuroscience*, 26(42):10727–10742.

Terrazas, A., Krause, M., Lipa, P., Gothard, K. M., Barnes, C. A., and McNaughton, B. L. (2005). Self-motion and the hippocampal spatial metric. *Journal of Neuroscience*, 25(35):8085–8096.

Teyler, T. J. and DiScenna, P. (1986a). The hippocampal memory indexing theory. *Behavioral neuroscience*, 100(2):147.

Teyler, T. J. and DiScenna, P. (1986b). The Hippocampal Memory Indexing Theory. *Behavioral Neuroscience*, 100(2):147–154.

Teyler, T. J. and Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory: Updating the index. *Hippocampus*, 17(12):1158–1169.

Tizard, B. (1959). Theories of brain localization from flourens to lashley. *Medical history*, 3(2):132–145.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4):189–208.

Trettel, S. G., Trimper, J. B., Hwaun, E., Fiete, I. R., and Colgin, L. L. (2019). Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nature Neuroscience*, 22(4):609–617.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., and Morris, R. G. (2007). Schemas and memory consolidation. *Science*, 316:76–82.

Tulving, E., Hayman, C., and Macdonald, C. A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: a case experiment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(4):595.

Tulving, E., Hayman, C. A. G., and MacDonald, C. A. (1972). Episodic and semantic memory. In Tulving, E. and Donaldson, W., editors, *Organization of memory*. Academic Press.

Ulanovsky, N. and Moss, C. F. (2007). Hippocampal cellular and network activity in freely moving echolocating bats. *Nature Neuroscience 2007 10:2*, 10(2):224–233.

Van Essen, D. C. and Glasser, M. F. (2018). Parcellating cerebral cortex: how invasive animal studies inform noninvasive mapmaking in humans. *Neuron*, 99(4):640–663.

Van Groen, T. and Wyss, J. M. (1992). Connections of the retrosplenial dysgranular cortex in the rat. *Journal of Comparative Neurology*, 315(2):200–216.

Vann, S. D. and Aggleton, J. P. (2004). Testing the importance of the retrosplenial guidance system: Effects of different sized retrosplenial cortex lesions on heading direction and spatial working memory. *Behavioural Brain Research*, 155(1):97–108.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., and Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, 277(5324):376–380.

Vargo, J. M., Corwin, J. V., King, V., and Reep, R. L. (1988). Hemispheric asymmetry in neglect produced by unilateral lesions of dorsomedial prefrontal cortex in rats. *Experimental Neurology*, 102:199–209.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Vedder, L. C., Miller, A. M., Harrison, M. B., and Smith, D. M. (2017). Retrosplenial Cortical Neurons Encode Navigational Cues, Trajectories and Reward Locations during Goal Directed Navigation. *Cerebral Cortex*, 27(7):3713–3723.

Viena, T. D., Linley, S. B., and Vertes, R. P. (2018). Inactivation of nucleus reuniens impairs spatial working memory and behavioral flexibility in the rat. *Hippocampus*, 28(4):297–311.

Vogt, B. A. and Miller, M. W. (1983). Cortical connections between rat cingulate cortex and visual, motor, and postsubicular cortices. *Journal of Comparative Neurology*, 216(2):192–210.

Vyleta, N. P., Borges-Merjane, C., and Jonas, P. (2016). Plasticity-dependent, full detonation at hippocampal mossy fiber–ca3 pyramidal neuron synapses. *Elife*, 5:e17977.

Waaga, T., Agmon, H., Normand, V. A., Nagelhus, A., Gardner, R. J., Moser, M. B., Moser, E. I., and Burak, Y. (2022). Grid-cell modules remain coordinated when neural activity is dissociated from external sensory cues. *Neuron*, 110(11):1843–1856.e6.

Wang, C., Chen, X., and Knierim, J. J. (2020). Egocentric and allocentric representations of space in the rodent brain. *Current opinion in neurobiology*, 60:12–20.

Whishaw, I. Q. (1998). Place learning in hippocampal rats and the path integration hypothesis. *Neuroscience & Biobehavioral Reviews*, 22(2):209–220.

Whishaw, I. Q., Hines, D. J., and Wallace, D. G. (2001). Dead reckoning (path integration) requires the hippocampal formation: evidence from spontaneous exploration and spatial learning tasks in light (allothetic) and dark (idiothetic) tests. *Behavioural brain research*, 127(1-2):49–69.

Whitlock, J. R., Pfuhl, G., Dagslott, N., Moser, M.-B., and Moser, E. I. (2012). Functional split between parietal and entorhinal cortices in the rat. *Neuron*, 73(4):789–802.

Wierzynski, C. M., Lubenov, E. V., Gu, M., and Siapas, A. G. (2009). State-dependent spike-timing relationships between hippocampal and prefrontal circuits during sleep. *Neuron*, 61:587–596.

Wikenheiser, A. M. and Redish, A. D. (2015). Hippocampal theta sequences reflect current goals. *Nature neuroscience*, 18(2):289–294.

Wilber, A. A., Clark, B. J., Forster, T. C., Tatsuno, M., and McNaughton, B. L. (2014). Interaction of egocentric and world-centered reference frames in the rat posterior parietal cortex. *Journal of Neuroscience*, 34(16):5431–5446.

Wilber, A. A., Skelin, I., Wu, W., and McNaughton, B. L. (2017). Laminar organization of encoding and memory reactivation in the parietal cortex. *Neuron*, 95(6):1406–1419.

Wills, T. J., Cacucci, F., Burgess, N., and O'Keefe, J. (2010). Development of the hippocampal cognitive map in preweanling rats. *science*, 328(5985):1573–1576.

Willshaw, D. (1981). Holography, associative memory, and inductive generalization. In Hinton, G. E. and Anderson, J. A., editors, *Parallel Models of Associative Memory*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Willshaw, D., Dayan, P., and Morris, R. G. (2015). Memory, modelling and marr: a commentary on marr (1971)'simple memory: a theory of archicortex'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1666):20140383.

Willshaw, D., Hallam, J., Gingell, S., and Lau, S. L. (1997). Marr's theory of the neocortex as a self-organizing neural network. *Neural Computation*, 9(4):911–936.

Willshaw, D. J. and Buckingham, J. (1990). An assessment of marr's theory of the hippocampus as a temporary memory store. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 329(1253):205–215.

Wilson, M. A. and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124):1055–1058.

Wilson, M. A. and McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679.

Wiltgen, B. J. and Silva, A. J. (2007). Memory for context becomes less specific with time. *Learning and Memory*, 14:313–317.

Winter, A. (2012). Wilder penfield and the recording of personal experience. In *Memory: Fragments of a Modern History*. University of Chicago Press.

Witter, M. P. (2018). Connectivity of the hippocampus. *Hippocampal microcircuits: a computational modeler's resource book*, pages 5–28.

Wolters, M. A. and Braun, W. J. (2018). A practical implementation of weighted kernel density estimation for handling shape constraints. *Stat*, 7(1):e202.

Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720):613–616.

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., and Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27(3):623–633.

Wyss, J. M. and Van Groen, T. (1992). Connections between the retrosplenial cortex and the hippocampal formation in the rat: A review. *Hippocampus*, 2(1):1–11.

Xu, H., Baracskay, P., O'Neill, J., and Csicsvari, J. (2019). Assembly responses of hippocampal ca1 place cells predict learned behavior in goal-directed spatial tasks on the radial eight-arm maze. *Neuron*, 101(1):119–132.

Yates, F. A. (1966). *The Art of Memory*. University of Chicago Press.

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270.

Zeithamova, D. and Bowman, C. R. (2020). Generalization and the hippocampus: More than one story? *Neurobiology of Learning and Memory*, 175:107317.

Zhang, K., Ginzburg, I., McNaughton, B. L., and Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *Journal of Neurophysiology*, 79(2):1017–1044.

Zilli, E. A. (2012). Models of grid cell spatial firing published 2005–2011. *Frontiers in neural circuits*, 6:16.

Zingg, B., Hintiryan, H., Gou, L., Song, M. Y., Bay, M., Bienkowski, M. S., Foster, N. N., Yamashita, S., Bowman, I., Toga, A. W., and Dong, H. W. (2014). Neural networks of the mouse neocortex. *Cell*, 156:1096–1111.

Ólafsdóttir, H. F., Carpenter, F., and Barry, C. (2016). Coordinated grid and place cell replay during rest. *Nature Neuroscience*, 19:792–794.

Appendix A

Supplementary Materials for Chapter 3

Table A.1: Imaging Sessions

Animal	Session	#Simult. recorded neurons	#Laps covered	#REST1 ensembles	#REST2 ensembles
1	1	352	20	3	4
	2	164	49	2	2
	3	302	27	4	4
	4	287	38	1	1
	5	213	63	4	3
2	1	290	45	2	4
	2	144	54	3	4
	3	264	43	2	2
	4	226	79	3	1
	5	268	70	1	3
	6	126	36	0	2
3	1	164	8*	-	-
	2	199	8*	-	-
	3	290	11	2	2
	4	276	7*	-	-
	5	222	15	0	2

* Session rejected due to low number of running laps.

A.1 Summary statistics for Bayesian analysis of reactivation

Bivariate linear regression model fitted to the fraction of significantly decoded frames (response variable) as a function of distance from landmarks and distance from reward (predictor variables). This analysis was conducted over the results obtained from the two different shuffling methods: cell identity shuffling and time bins shuffling. In the former case, the response variable for REST2 was significantly anti-correlated with the distance from landmarks and the distance from reward, while in REST1 the two predictors were not correlated with the decoded positions. The same conclusions were derived from time bin shuffling. However, a positive correlation between the distance from cues and the decoded positions was observed. This is likely due to REST1 ensembles containing higher fractions of neurons that expressed their place fields at the very beginning of the track and at the location between the first and second landmark (at around 50 centimetres from the start) which were located away from landmarks (see Fig. 3.7 in main text).

A) Cell identity shuffle

REST1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	147.99	17.887	8.2732	9.9908e-11
distance from reward	-0.69257	0.66554	-1.0406	0.30338
distance from cues	6.7057	4.0263	1.6655	0.10247

Number of observations: 50 spatial bins, Error degrees of freedom: 47
Root Mean Squared Error: 61.7
R-squared: 0.12, Adjusted R-Squared: 0.0826
F-statistic vs. constant model: 3.21, p-value = 0.0496

REST2

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	289.83	42.893	6.757	1.911e-08
distance from reward	-6.4317	1.5959	-4.03	0.00020268*
distance from cues	-29.064	9.6551	-3.0102	0.0041893*

Number of observations: 50 spatial bins, Error degrees of freedom: 47
Root Mean Squared Error: 148
R-squared: 0.281, Adjusted R-Squared: 0.251
F-statistic vs. constant model: 9.19, p-value = 0.000426

B) Time bins shuffle

REST1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	179.13	23.482	7.6284	9.1828e-10
distance from reward	-0.037	0.87369	-0.042349	0.9664
distance from cues	14.002	5.2856	2.6492	0.010955*

Number of observations: 50 spatial bins, Error degrees of freedom: 47
Root Mean Squared Error: 81.1
R-squared: 0.155, Adjusted R-Squared: 0.119
F-statistic vs. constant model: 4.3, p-value = 0.0192

REST2

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	228.77	39.499	5.7917	5.5545e-07
distance from reward	-8.1345	1.4697	-5.535	1.3508e-06*
distance from cues	-18.415	8.8911	-2.0712	0.043856*

Number of observations: 50 spatial bins, Error degrees of freedom: 47
Root Mean Squared Error: 136
R-squared: 0.395, Adjusted R-Squared: 0.369
F-statistic vs. constant model: 15.4, p-value = 7.4e-06

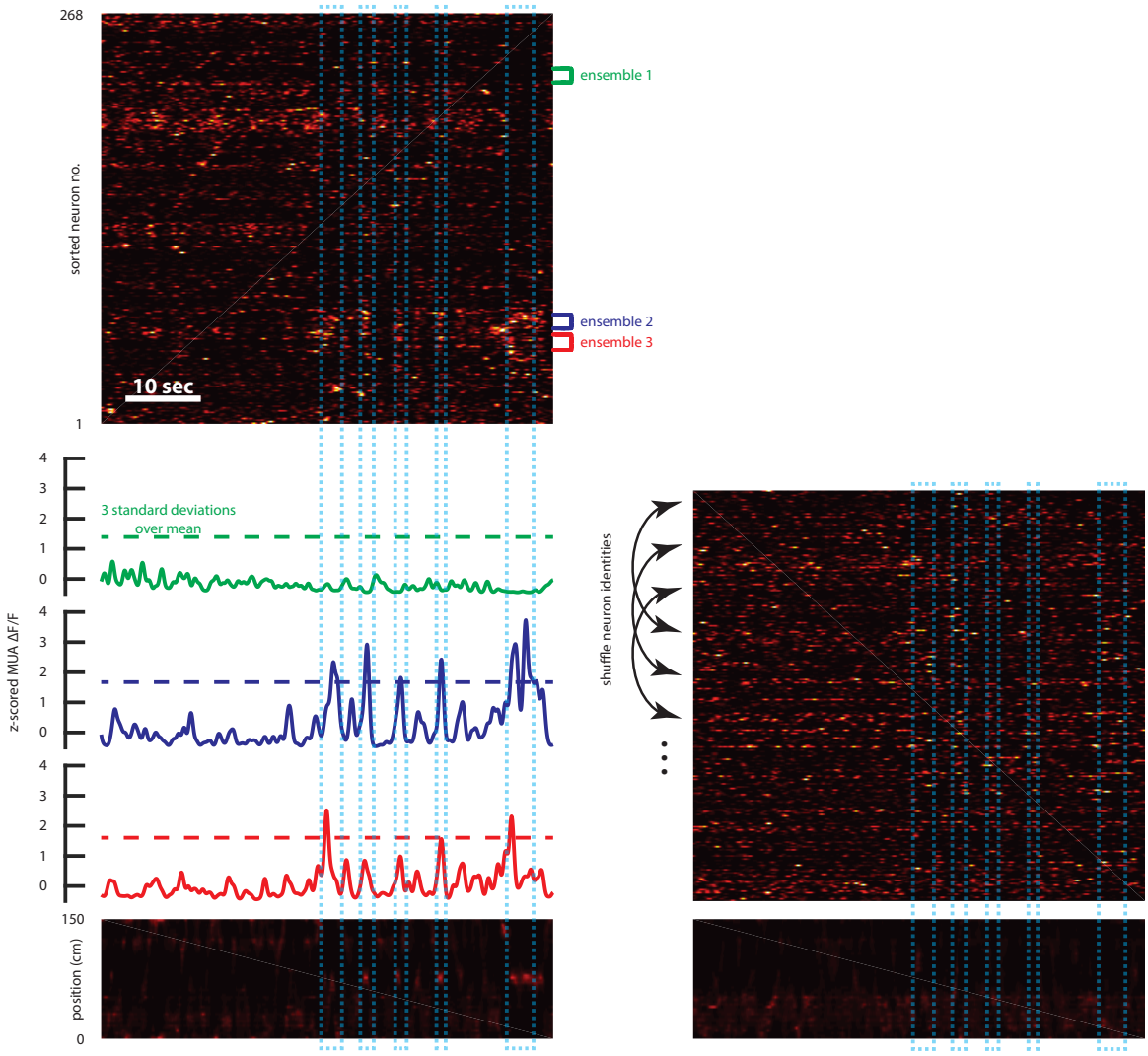

Figure A.1: Bayesian decoding method for inferring position during rest. Data presented in Fig. 3.5 of the main text shall be used here to illustrate in detail the Bayesian decoding method. Three ensembles were simultaneously detected in this example REST2 recording. The multi-unit activity (MUA) was calculated from the z-scored deconvolved ΔF/F of each ensemble. SCEs generated by individual ensembles were detected by setting a threshold of 3 standard deviations above the mean of the ensemble MUA (regions bounded by blue dashed-line boxes). For each SCE, the entire population vector associated with the event is fed to the decoder. Under the null hypothesis, the Bayesian model will output a location whenever a given set of neurons participate in concerted activity. If the spatial distribution of neural activity is not uniform (i.e. more neurons in the population are active at certain spatial locations), then by chance the decoder will output the location with most bias. To control for this, we shuffled the identities of the neurons (y-axis), while preserving the temporal relationships between neurons (x-axis). In other words, the same ensembles and SCEs will be detected from the shuffled data, though the spatial locations encoded by the neurons will be changed. In this recording, a high fraction of neurons expressed activity in the first half of the linear track (see Fig. A.2). Unsurprisingly, the locations decoded from the shuffled data showed a similar spatial bias.

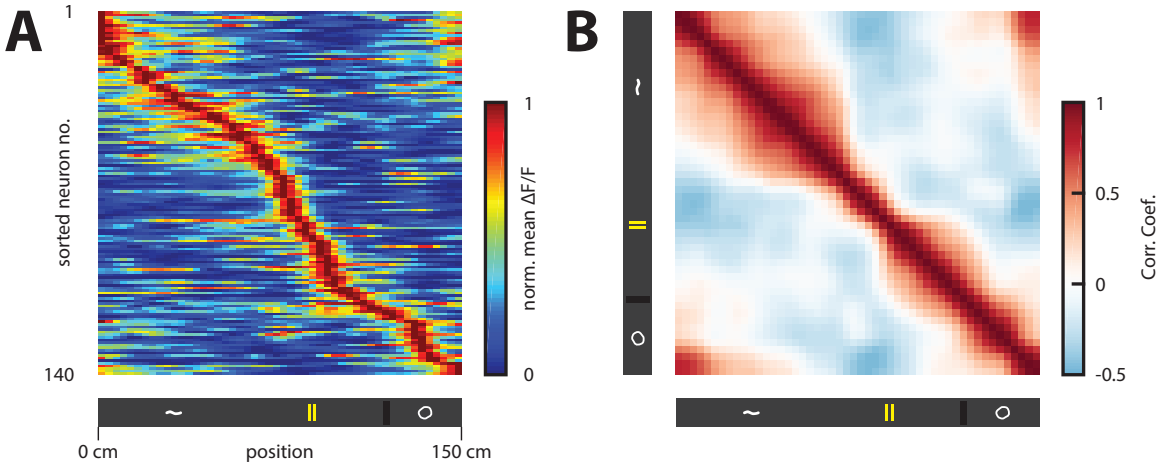

Figure A.2: Spatial coding characteristics. (A) Average normalized activity of 140 neurons classified as 'place cells' (out of the 268 neurons in Fig. A.1) as a function of position. For the first half of the linear track, neurons express broader tuning functions. This result ties well with the demonstration in Fig. A.1. (B) Correlation matrix of positional population vectors (shown in A) as a function of position.

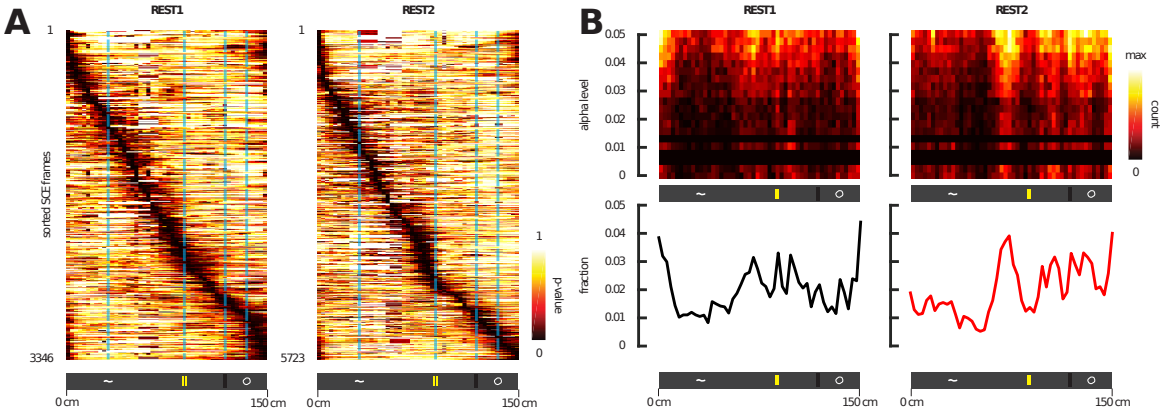

Figure A.3: Time bins shuffle method. Same as in Fig. 3.6 of the main text, but reporting results obtained by the 'time bins shuffling' method. (A) For each time frame associated with a SCE, p-values were obtained as the fraction of probabilities obtained from shuffling that were greater than the probabilities obtained from the true data, at every given position. In other words, we evaluated the degree to which a decoded position during a SCE is likely due to chance (the lower the p-value, the less likely the decoding was obtained by chance). These time frames were sorted according to the location of most significant decoding. Results from all SCE-associated time frames (obtained from all three animals, all 13 sessions) are shown for REST1 and REST2. (B) Joint density distributions of the fraction of significantly decoded positions at different α levels (p-values). α levels were binned over a logarithmic scale. The fractions of significant decoding ($p<0.05$; analogous to summing over the columns) as a function of position is shown below. Notice in REST1, the fractions are more uniformly distributed, while in REST2, the fractions are more concentrated near the locations of landmarks.

A.2 Results for individual animal

Fig. A.4-A.6 perform the same analysis shown in Fig. 3.6 of the main text on each individual animal. For each time frame associated with a SCE, p-values were obtained as the fraction of Bayesian decoded probabilities obtained from shuffling that were greater than the probabilities obtained from the true data, at every given position. In other words, we evaluated the degree to which a decoded position during a SCE is likely due to chance. These time frames were sorted according to the location of most significant decoding. Results from all SCE-associated time frames are shown for REST1 and REST2. The fractions of SCEs that expressed decoded probabilities at α levels below 0.05 are plotted below for every given position over the linear track.

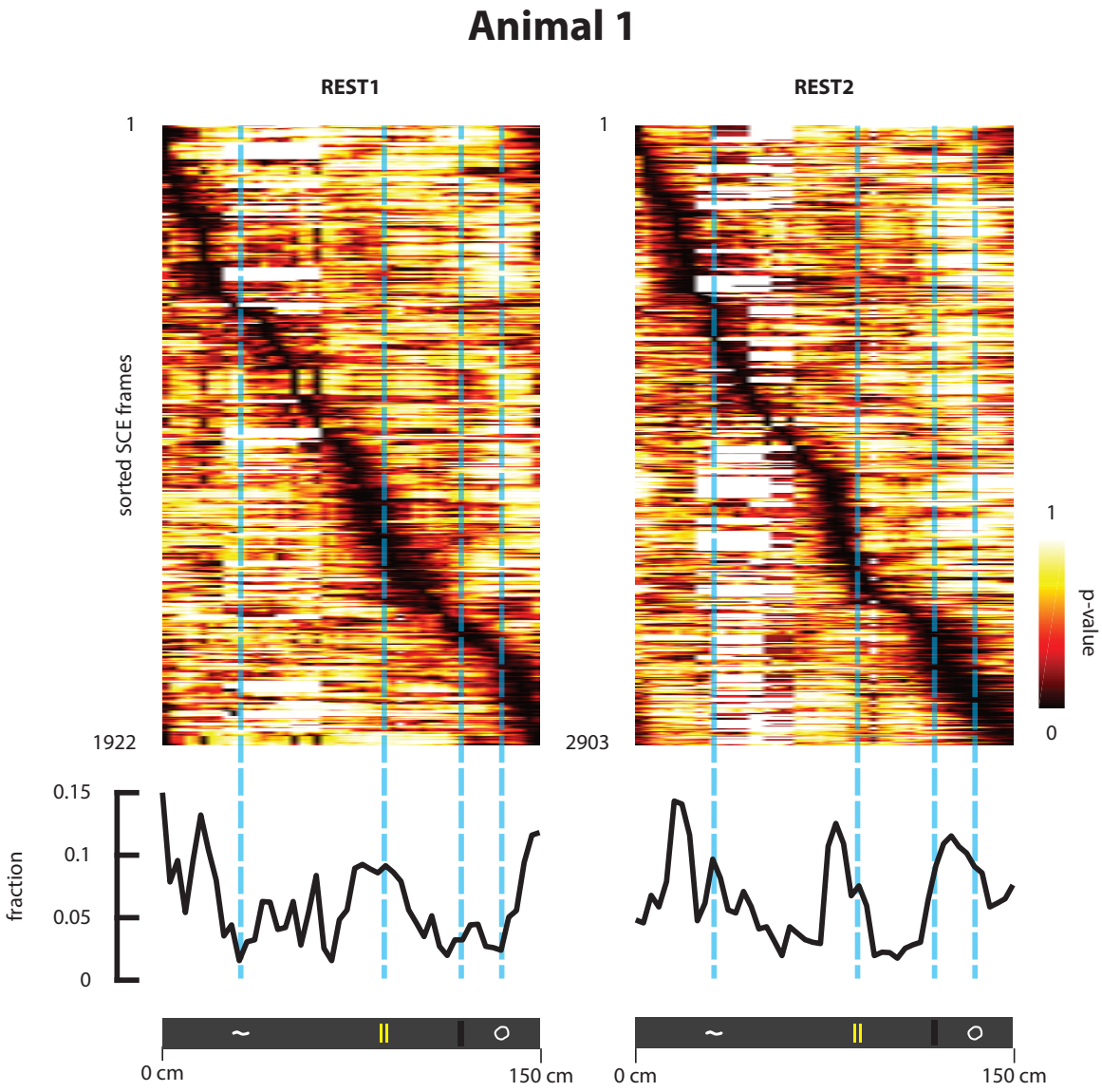

Figure A.4: Animal 1.

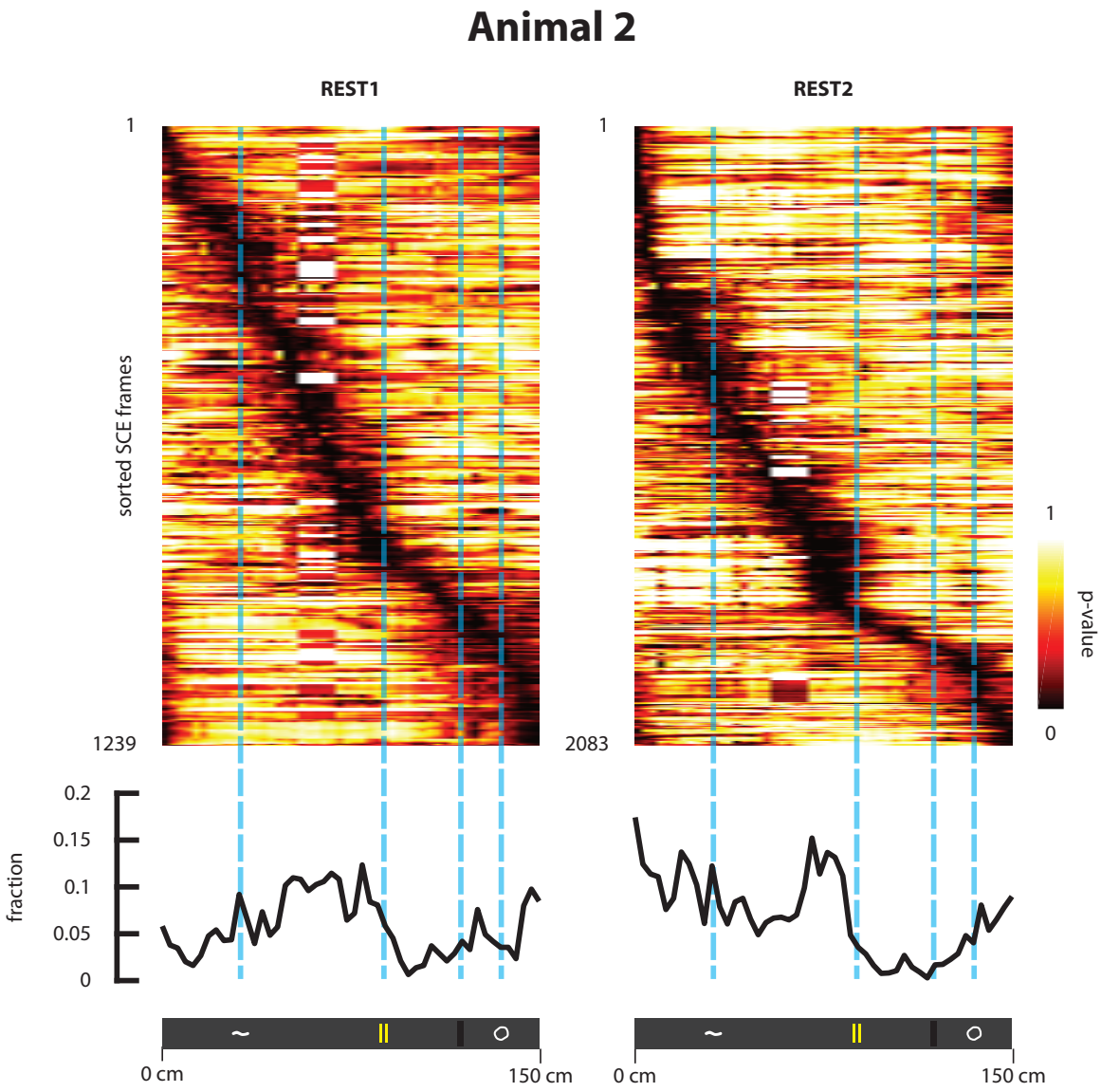

Figure A.5: Animal 2.

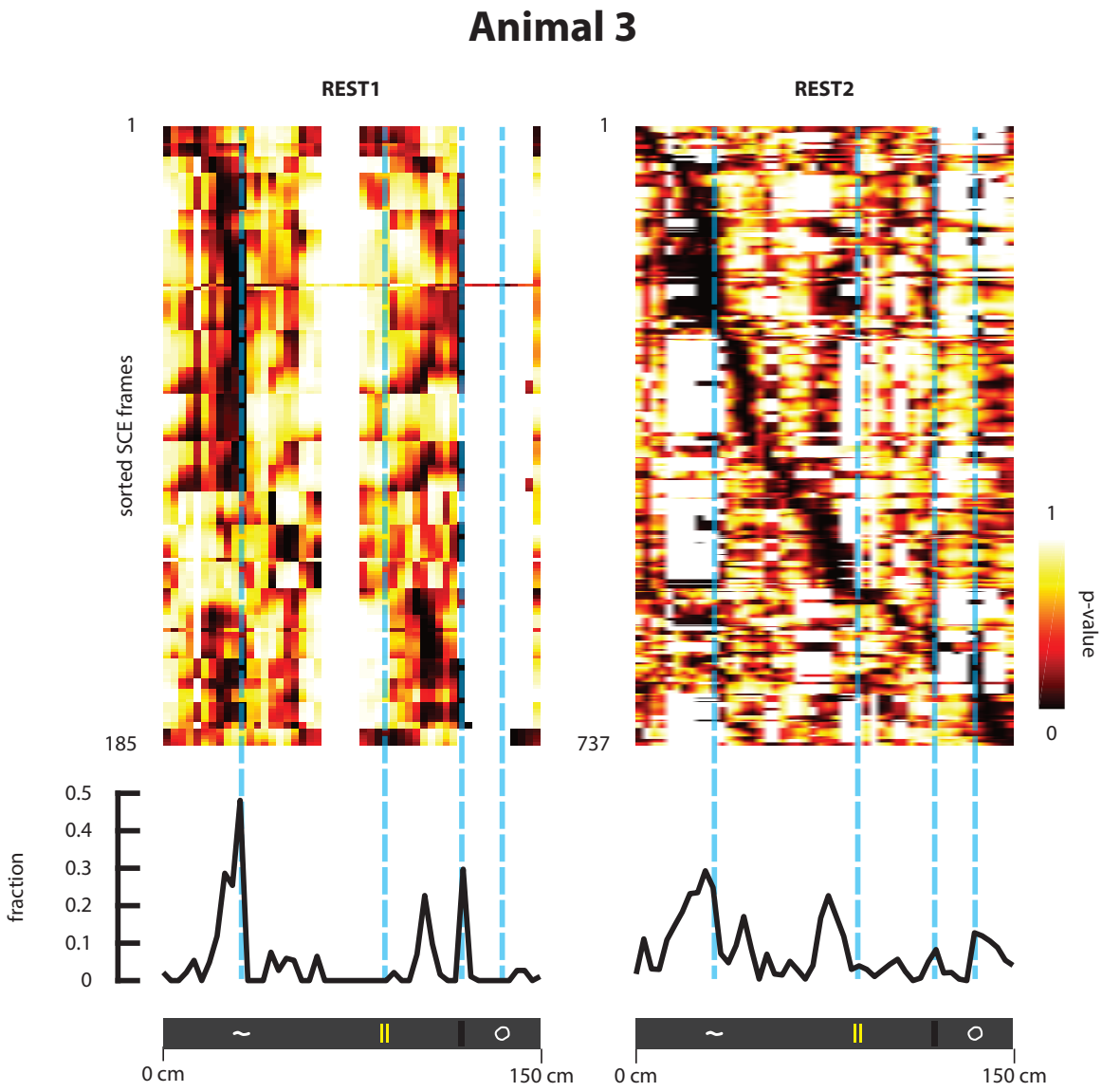

Figure A.6: Animal 3.

Appendix B

Supplementary Materials for Chapter 4

B.1 Supplementary Figures

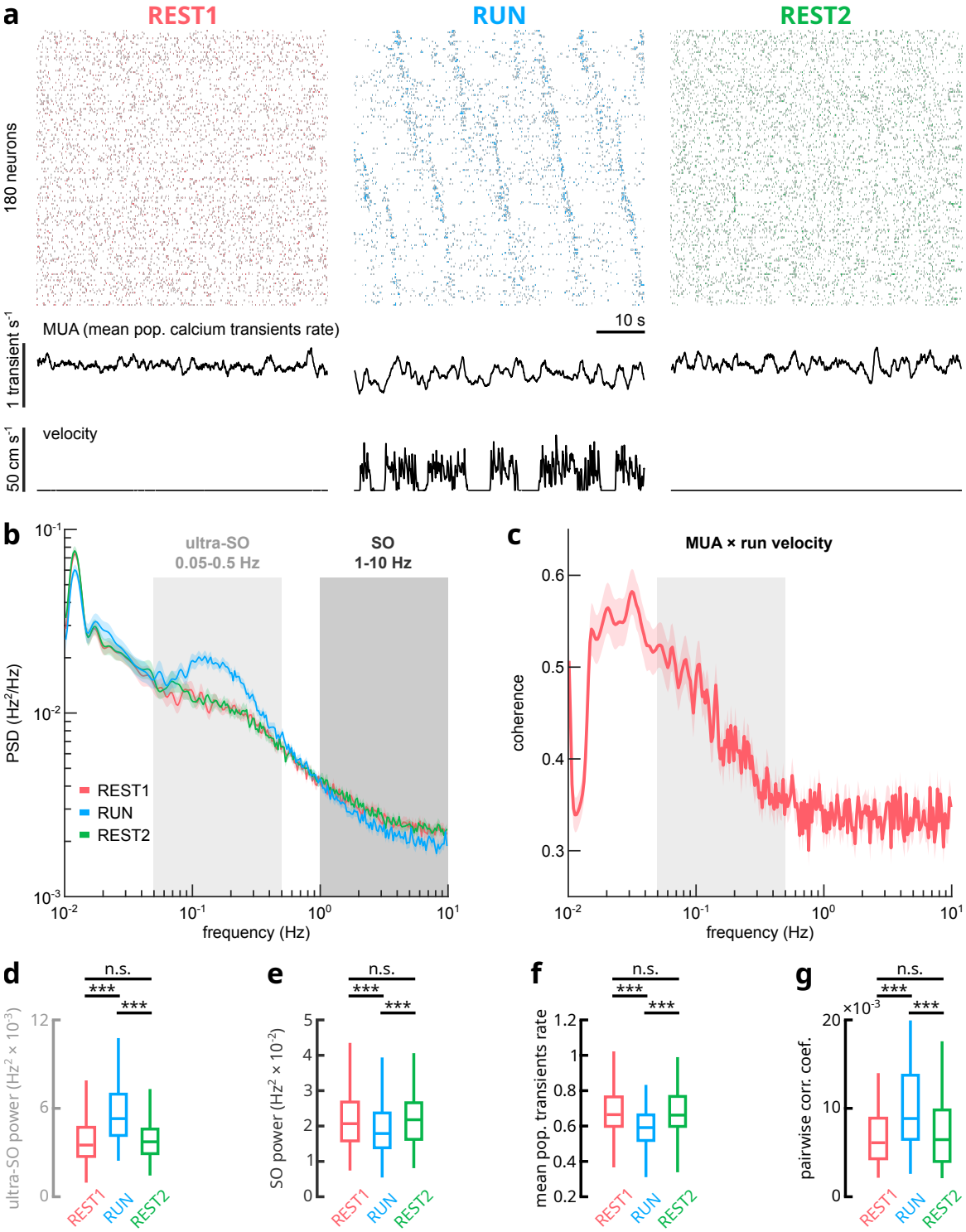

Figure B.1: Locomotion entrains slow population dynamics. **a** Binary matrices of calcium events detected by deconvolution of the fluorescent traces. Neurons are sorted by the location of peak activity. 1 min segments from each behavioural epoch are shown (same example session as Fig. 4.1**c**). The multi-unit activity (MUA), estimated as the rate of transients per neuron averaged over the population at each imaging frame, and the run velocity are illustrated below. **b** Power spectral density (PSD) of the MUA in each behavioural epoch, averaged over all recording sessions ($n = 86$; shaded area denotes $\pm$S.E.M.). PSD was estimated by Welch's method, using 200 s Hann windows with 75% overlap. **c** Coherence between RUN MUA and velocity over the treadmill. Spectral densities were estimated using the same methods as **b**. **d-f** The MUA power in the ultra-slow oscillations (0.05-0.5 Hz) and slow oscillations (1-10 Hz) bands, as well as the mean MUA rates (transients s^{-1}), are compared across behavioural epochs (pairwise Wilcoxon signed rank tests; n.s. $p \geq 0.05$; *** $p < 0.001$; p-values were Bonferroni-adjusted). **g** Pearson correlation coefficients were obtained between the time-series of each pair of neurons within a session, and the average across all pairs was taken.

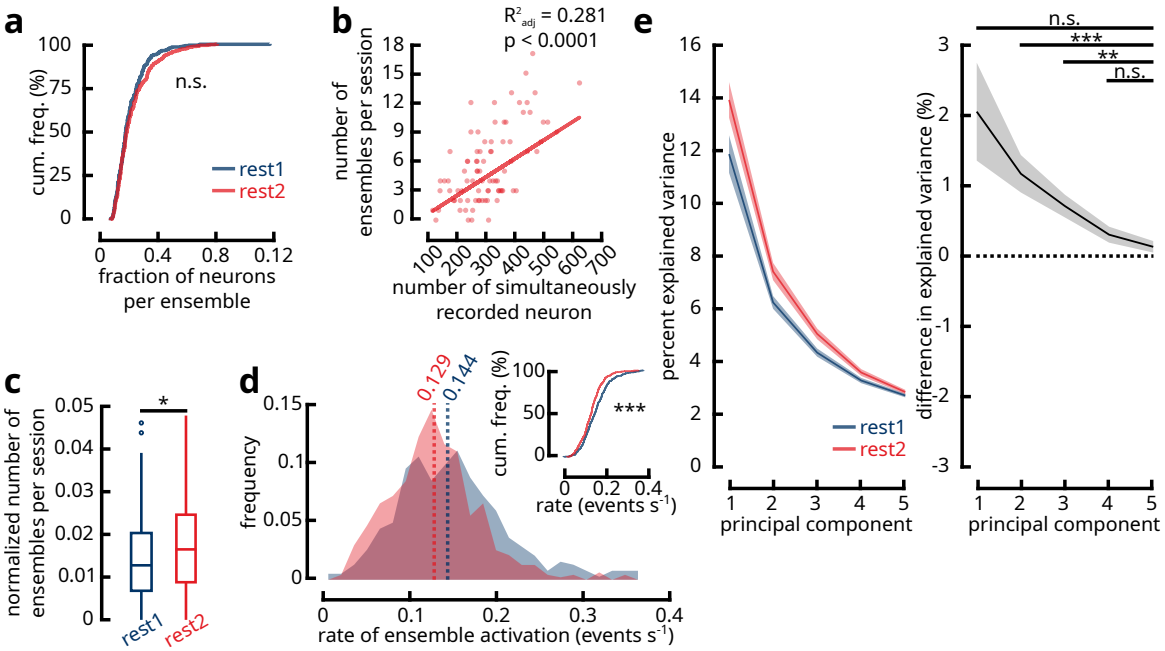

Figure B.2: REST2 exhibits more synchronous and sparse population dynamics. **a** Cumulative frequency distributions of the fraction of neurons that belong to an ensemble with respect to the total number of simultaneously recorded neurons within a session for REST1 ($n = 86$ sessions; $n = 392$ ensembles) and REST2 ($n = 478$ ensembles) (two-sample two-tailed Kolmogorov–Smirnov test; $p = 0.074$). **b** Number of ensembles per recording session vs. the total number of simultaneously acquired neurons for REST2 ($n = 86$ sessions). **c** Given that the number of ensembles within a recording is dependent upon the total number of recorded neurons, the former quantity needs to be normalized for appropriate comparisons. Depicted is the number of ensembles in REST1 and REST2 expressed as a fraction of the total number of neurons. A higher number of synchronous ensembles are found in individual REST2 sessions ($n = 86$ sessions; Paired-samples Wilcoxon signed-rank test $p = 0.001$). **d** Distributions of the average rate of activation of REST1 ($n = 392$) and REST2 ($n = 478$) ensembles. The median activation rate of REST1 ensembles was higher than that of REST2 ensembles (two-tailed Mann-Whitney U-test $p < 0.001$; two-sample two-tailed Kolmogorov-Smirnov test $p < 0.001$). **e** PCA was conducted over the correlation matrices of time series vectors of simultaneously recorded neurons in REST1 and REST2 separately. The first five principal components were extracted and the percentage of the total variance explained by each component was examined (lines show mean±s.e.m.). For these paired samples ($n = 86$ sessions), we took the difference in percentage explained variance between REST2 and REST1 for each component to test whether REST2 components reliably explain more of the variance (one-way repeated measures ANOVA with Greenhouse-Geisser correction; residuals approximately normal; significant difference between components with $p = 0.019$). Post-hoc tests using the fifth component as a baseline reference suggests that the second and third components in REST2 accounted for more of the variance in the population dynamics than their REST1 counterparts (n.s. $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; p-values were Bonferroni-adjusted).

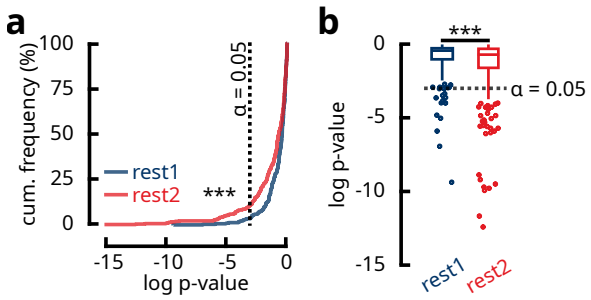

Figure B.3: Hypergeometric modelling of the fraction of spatially-selective cells in rest ensembles. For each ensemble ($n = 392$ REST1 ensembles; $n = 478$ REST2 ensembles), the number of spatially-selective cells is counted out of the total number of cells that are part of the ensemble. These values are compared to the total number of spatial cells, as well as the total cell count, within the corresponding recording session, using a one-tailed Fisher's exact test. Intuitively, the p-value returned from this test reflects the likelihood of obtaining a certain number of spatial cells in an ensemble of a given size by drawing at chance from the sample population. Small p-values indicate that the fraction of spatial cells within an ensemble is higher than what is expected at chance level. P-values were log-transformed for better detection of low-probability events. **a** Empirical cumulative distribution functions of $\log(p_{\text{values}})$ for REST1 and REST2 ensembles. A greater number of REST2 ensembles contained a large fraction of spatially-selective neurons (two-tailed two-sample Kolmogorov-Smirnov test; $p < 0.001$). Dotted line represents significant α level. **b** Same as **a**, but represented as boxplots (two-tailed Mann-Whitney U-test; $p < 0.001$).

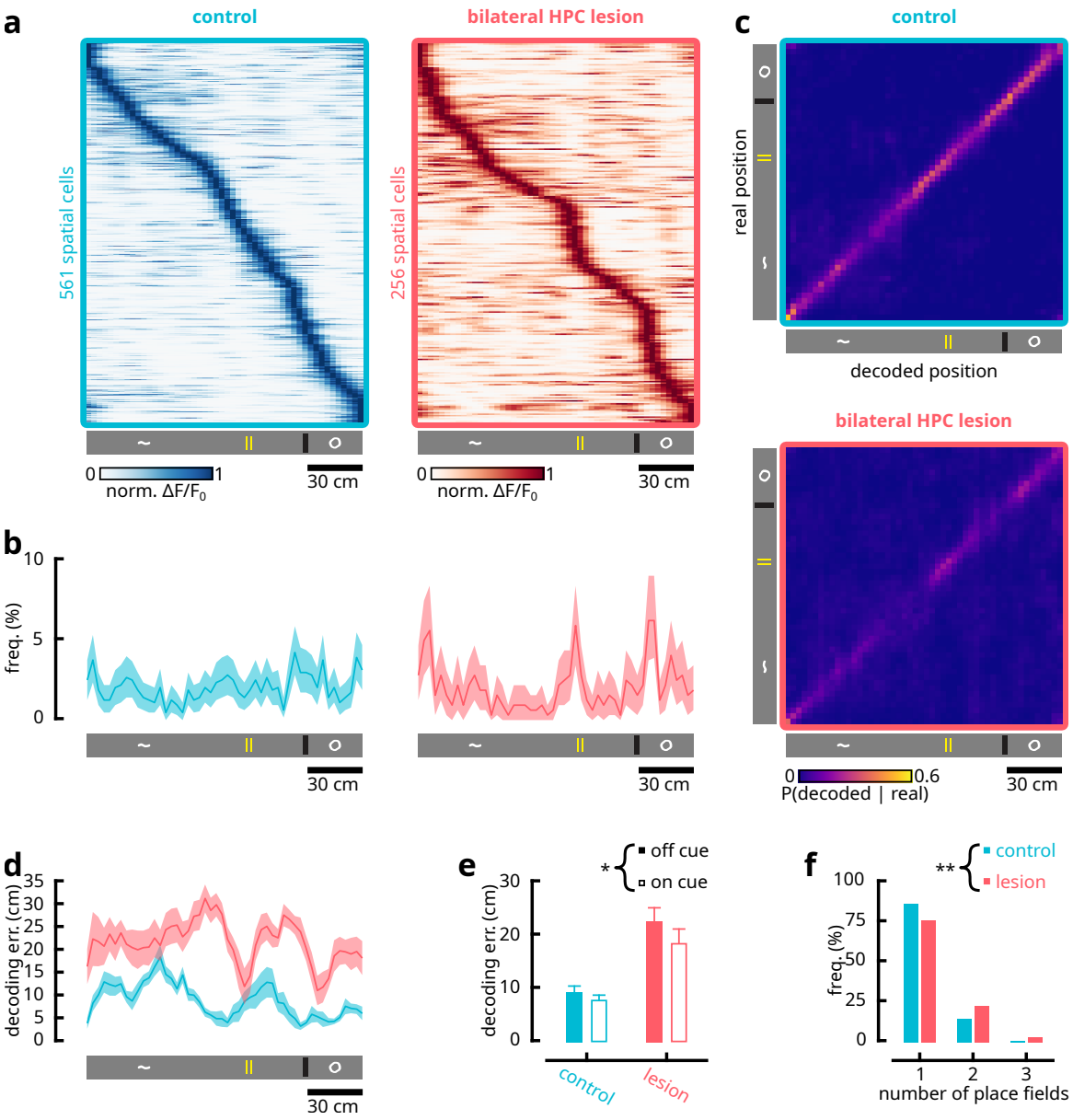
a
control
bilateral HPC lesion
561 spatial cells
256 spatial cells
0
1
norm. ΔF/F0
30 cm
b
freq. (%)
10
5
0
30 cm
c
control
real position
decoded position
bilateral HPC lesion
0
0.6
P(decoded | real)
30 cm
d
decoding err. (cm)
35
30
25
20
15
10
5
0
30 cm
e
decoding err. (cm)
30
20
10
0
*
off cue
on cue
control
lesion
f
**
control
lesion
freq. (%)
100
75
50
25
0
1
2
3
number of place fields

Figure B.4: Representation of visuo-tactile cues in the secondary motor cortex is more pronounced following bilateral lesion of the dorsal hippocampus. Data from Esteves et al. (2021). **a** The average response (normalized between 0 to 1) of spatially-selective neurons as a function of spatial location for control (blue) and hippocampal-lesioned (red) animals. Neurons sorted by their peak average firing location. Note that the population representation of space in control animals is approximately uniform, while this representation is biased towards the locations of cues in the lesioned group. **b** Histogram distribution of place field centres in control (blue; $n = 650$ place fields) and lesioned (red; $n = 327$ place fields) groups. Shaded areas represent 95 % bootstrapped confidence intervals. **c** Normalized confusion matrices for real and Bayesian decoded positions (see Esteves et al. (2021) for methods). Decoding error assessed by *leave-one-out* cross-validation over trials. Notice that accuracy is higher over cue locations in the lesion group. **d** Average decoding as a function of spatial location ($n = 10$ sessions in $n = 4$ control animals; $n = 8$ sessions in $n = 4$ lesion animals). Shaded area denote S.E.M. **e** Average decoding error inside and outside cue locations. Robust two-way mixed ANOVA with 20 % trimmed-means (R package 'WRS'). Main effects of hippocampal lesion ($p = 0.010$) and cue location ($p = 0.028$). No significant interaction ($p = 0.080$). **f** Frequency in the number of place fields per individual spatial neurons. Spatial cells in the lesioned group tend to support a higher number of place fields (χ^2 test; $p = 0.001$).

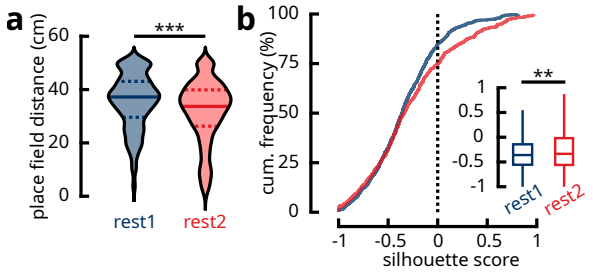

Figure B.5: REST2 ensembles tend to be composed of spatially-selective neurons sharing neighbouring place fields. **a** The average distance between place field centres of neurons within the same REST1 or REST2 ensemble ($n = 392$ REST1 ensembles; $n = 478$ REST2 ensembles; $p < 0.001$; Mann-Whitney U-test). The kernel densities for the violin plots were estimated using a 2 cm Gaussian window. The median (solid line), first and last quartiles (dashed lines) are shown. **b** Cumulative distribution functions of the silhouette coefficients of the place field centre locations of REST1 and REST2 neurons that belonged to synchronous ensembles. The silhouette coefficient, in the current context, measures how similar the neurons within the same ensemble are as opposed to neurons outside the ensemble. Similarity is based on the Euclidean distance between place field centres. Silhouette values range from -1 to 1, where high values suggest strong cohesion between a neuron and the other neurons within the same ensemble, and clear separation of the neuron from to the rest of the neuronal population. A significantly larger portion of REST2 neurons contain positive silhouette values compared to REST1 neurons ($n = 2672$ REST1 ensemble neurons; $n = 3613$ REST2 ensemble neurons; Kolmogorov–Smirnov test $p < 0.001$; Mann-Whitney U-test $p = 0.007$; outliers omitted).

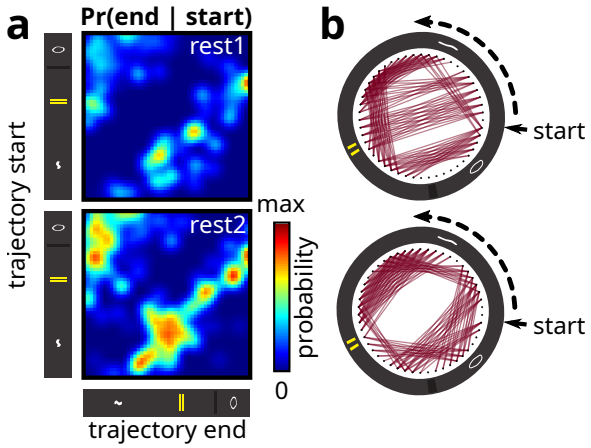

Figure B.6: Trajectory ensembles encode short spatial segments that tend to span the locations of cues. **a** Conditional probability matrices of the end location of trajectories given their starting location ($n = 160$ REST1 trajectories; $n = 325$ REST2 trajectories). Notice that the densities aggregate over a diagonal slightly offset from the central diagonal of the matrix, meaning that the majority the trajectories consist of short segments over space. Within these densities, a large fraction tend to span the locations of cues or the space delimited by two cues. Both of these trends are expressed prominently in REST2 trajectory ensembles, while REST1 ensembles show no clear organization. **b** The same probabilities in **a** represented in graph form. Short arrow point to the beginning of the track. Dashed arrow indicates the running direction.

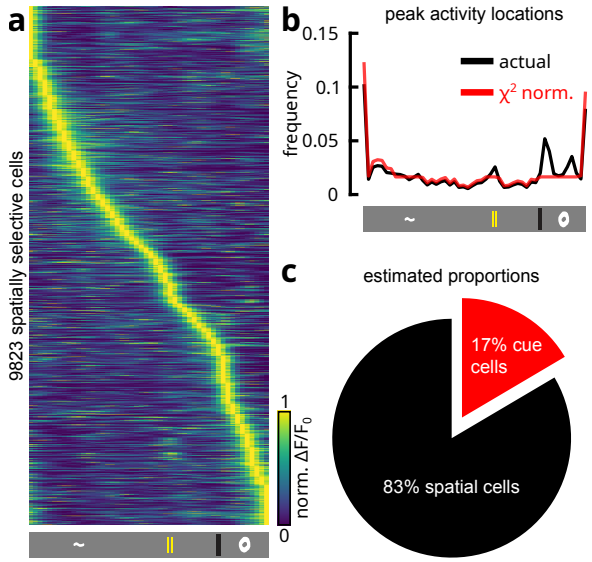

Figure B.7: Estimating the proportions of cue-responsive cells. **a** The average activity as a function of position for all neurons that passed the criteria for spatial selectivity (all animals; all sessions). The criteria used for identifying these neurons do not discriminate between location-encoding and cue-encoding neurons (see Methods). Assuming that location-encoding cells are uniformly distributed over space, the extra densities observed at cue locations would reflect cells that respond to cue sensation. **b** The distribution of locations of peak activity in **a**. Using a minimum χ^2 estimation procedure (see Supplementary Methods), neurons were iteratively removed at the locations of cues until the distribution was maximally uniform. **c** The fraction of cue-responding cells was estimated by counting the number of cells removed at cue locations in **b**.

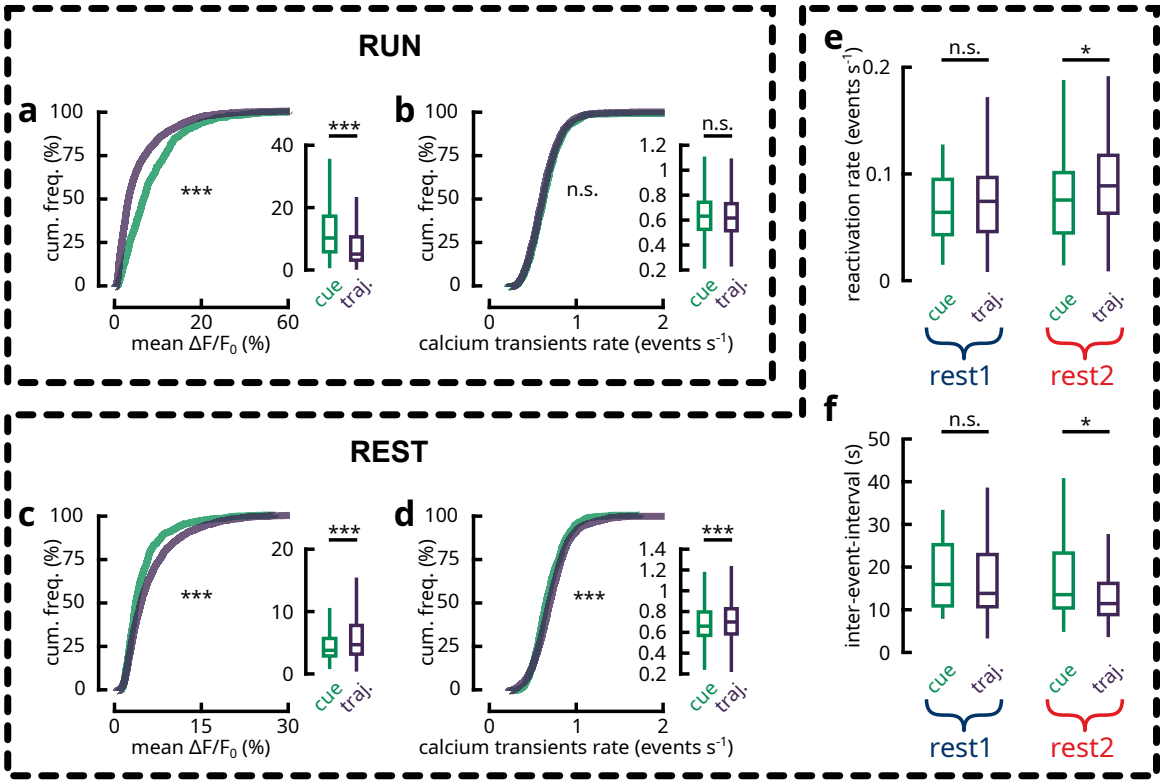

Figure B.8: Online and offline dynamics of cue and trajectory ensembles. **a-b** Mean $\Delta F/F_0$ and calcium transient rates for cue ($n = 751$) and trajectory ($n = 2056$) ensemble cells during RUN epochs. Cue ensemble neurons exhibit stronger activities than trajectory ensemble neurons in terms of calcium transients, albeit the number of transients per unit time were comparable between the two groups (two-sample two-tailed Kolmogorov-Smirnov test and two-tailed Mann-Whitney U-test; n.s. $p \geq 0.05$; *** $p < 0.001$). **c-d** Same as **a-b**, but during REST epochs. Here, the trends are reversed, whereby trajectory ensemble neurons show stronger activation and slightly more elevated rates of calcium transients. **e-f** Average rate of reactivation events and the inter-event-interval between reactivations for cue ($n = 19$ in REST1; $n = 62$ in REST2) and trajectory ($n = 97$ in REST1; $n = 164$ in REST2) ensembles. Trajectory ensembles reactivate more frequently during REST2 (Mann-Whitney U-test; $p = 0.034$), in accordance with **d**, but at equal rates as cue ensembles during REST1 ($p = 0.732$). Similarly, the average interval between trajectory reactivations was shorter in REST2 (Mann-Whitney U-test; $p = 0.021$), but not during REST1 ($p = 0.704$). All outliers omitted.

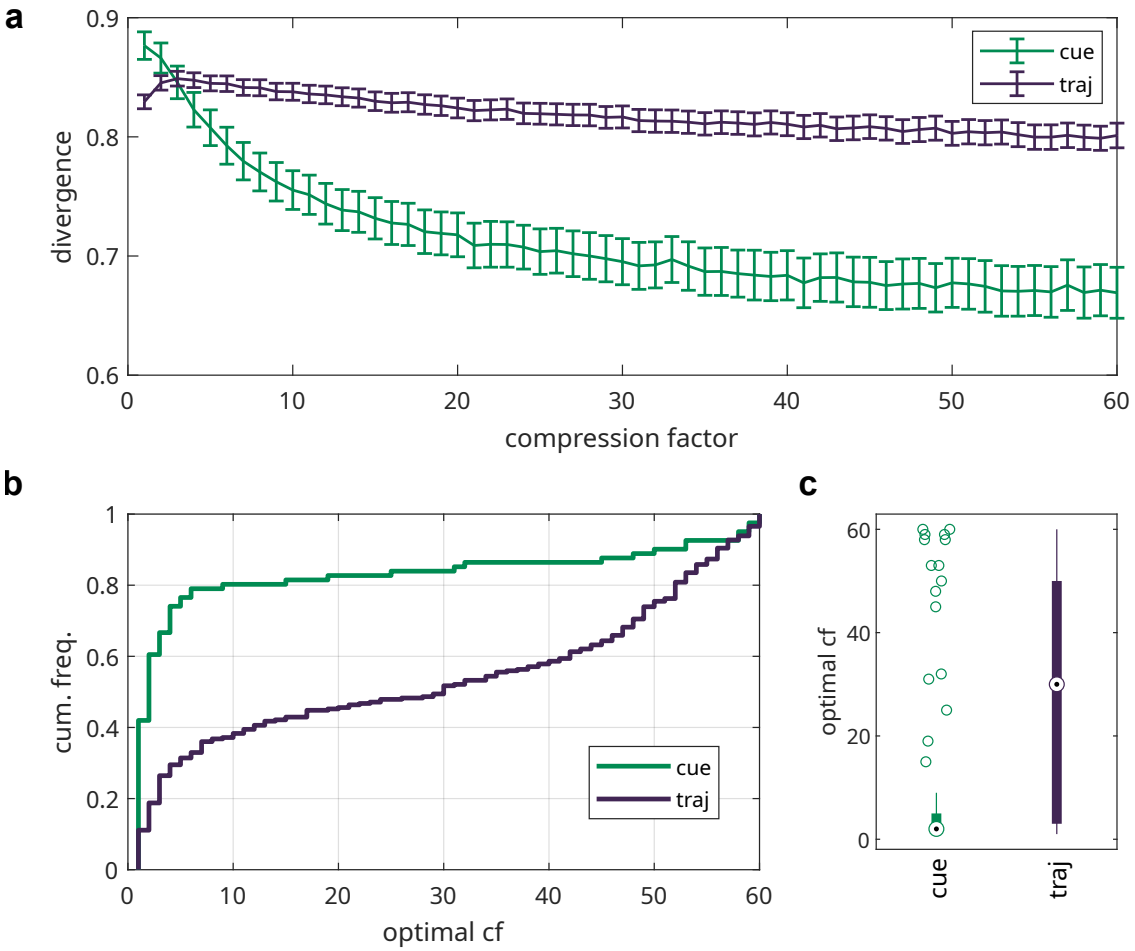

Figure B.9: Time compressed reactivation of trajectory, but not cue features. **a** Divergence as a function of compression factor for cue ($n = 81$) and trajectory ($n = 261$) ensembles (REST1 and REST2 pooled; mean $\pm$ SEM). Notice that cue ensembles express peak divergence with no compression, while divergence sharply decreases with higher compression. In comparison, trajectory ensembles benefit from some degree of compression. These results are commensurate with the notion that trajectory ensembles contain neurons that firing sequentially during RUN, whereas cue ensemble neurons fire in synchrony during behaviour. **b** Empirical cumulative distribution functions of the compression factor where peak divergence is found for the sample data in **a**. Reactivation of trajectory ensembles is more compressed temporally (median of 30$\times$), whereas cue reactivations are predominantly not compressed (median of 2$\times$; two-sample two-tailed Kolmogorov-Smirnov test; $p < 0.001$). **c** Same as **b** in boxplot (two-tailed Mann-Whitney U-test; $p < 0.001$).

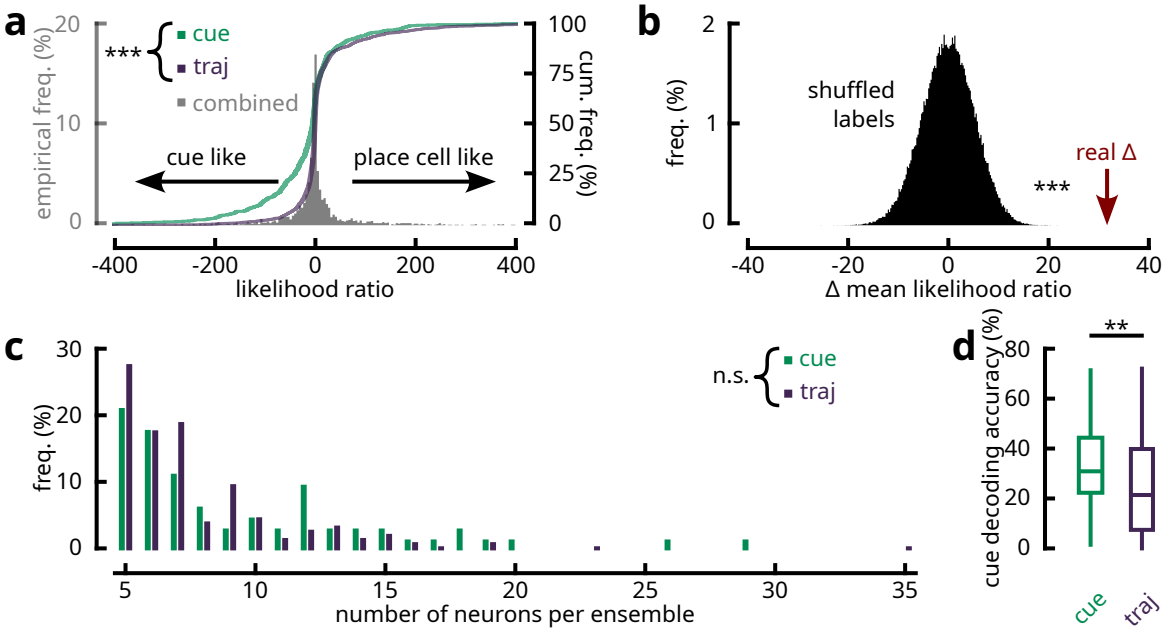

Figure B.10: Cue and trajectory information are conjunctively encoded by resting-state ensembles; however, ensembles express varying degrees of bias for each separate behavioural feature. **a** Two models were fitted to the activities of ensemble neurons (see Methods): one is driven by sensory cues and the other is modelled after a *place-cell*'s tuning curve. Taking the likelihood ratio between the two models gives a neuron's proclivity for either type of response tuning. The distribution of these ratios express no bimodality, suggesting that ensemble neurons encode for a conjunction of cue and place responses. Labelling the neurons based on their cue/trajectory ensemble membership showed however that ensembles are biased in the type of behavioural features their encode for ($n = 601$ cue ensemble neurons; $n = 1314$ trajectory ensemble neurons; two-sample Kolmogorov–Smirnov test $p < 0.001$). **b** Distribution of the difference between mean likelihood ratios between cue and trajectory ensemble neurons, obtained by permutating the membership labels. The real difference is identified by the red arrow, confirming the bias for cue and trajectory ensembles in encoding their respective features ($p < 0.001$). **c** Distribution for the number of neurons per cue or trajectory ensemble. Cue and trajectory ensembles supported a similar number of neurons (χ^2 test; $p = 0.17$), meaning that any detected differences in the encoding of behavioural parameters were not due to bias in sample size. **d** Bayesian decoding accuracy for the identity of individual cues using cue and trajectory ensemble neurons. Cue ensemble neurons were more accurate at identifying the identity of the four cues (two-tailed Mann-Whitney U-test; $p = 0.0012$), likely due to their tendency of supporting multiple place fields with different firing rates at different cue locations. However, this accuracy was still low at a median of $\sim$30 %. This could be due to the responses of cue ensemble neurons being targetted at specific cues rather than all four cues indiscriminately.

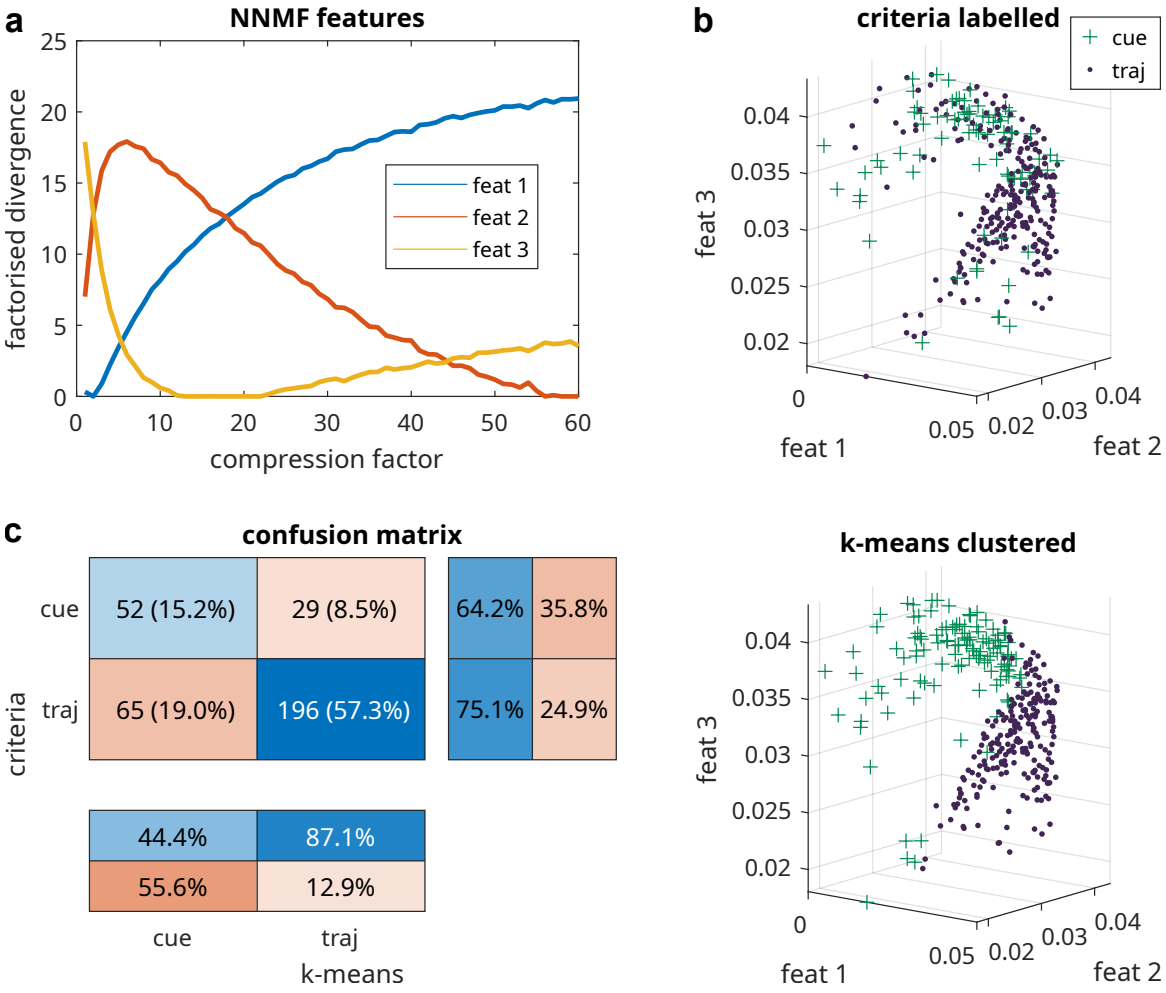

Figure B.11: Cue and trajectory ensembles emerge as separate categories under unsupervised clustering of temporal compression. **a** The three features extracted by NNMF are shown. Notice that feature 3 peaks in the beginning and sharply decreases with higher compression factors, closely resembling the trend observed for cue ensembles in Fig. B.9a. Similarly, feature 2 has a later peak and decays at a slower rate, following a similar pattern as trajectory ensembles. **b** Ensembles projected into the coordinates given by the coefficients matrix. Class labels are assigned either by the selection criteria defined in Methods or by k-means clustering. In accordance with our hypothesis, cue ensembles have higher weights over the axis of feature 3, which corroborate with their affinity to lower temporal compression. In contrast, trajectory ensembles have higher weights over feature 2, which reinforces the notion that trajectory reactivations are temporally compressed. Visually, the clusters isolated by the selection criteria are comparable to those identified by k-means. Note, however, that this point cloud forms a continuous density as opposed to two visually separate clusters, which corroborates with the notion that a continuum likely describes the affinity of ensembles for either cue or trajectory. **c** Confusion matrix comparing the labels assigned by criteria and the labels determined by k-means. The two labelling methods agreed strongly on the identity of trajectory ensembles. Cue ensembles labelled under k-means were, however, less than half as likely to be also considered as cue ensembles under selection criteria. These are likely reactivation of short trajectory segments that fell outside of the cue zone, which NNMF and k-means are unable to distinguish due to not having knowledge of the locations of cues. Overall, cue and trajectory labels assigned by criteria and by unsupervised clustering show a significant degree of overlap (χ^2 test; $p < 0.001$; $\chi^2 = 42.405$; df = 1).

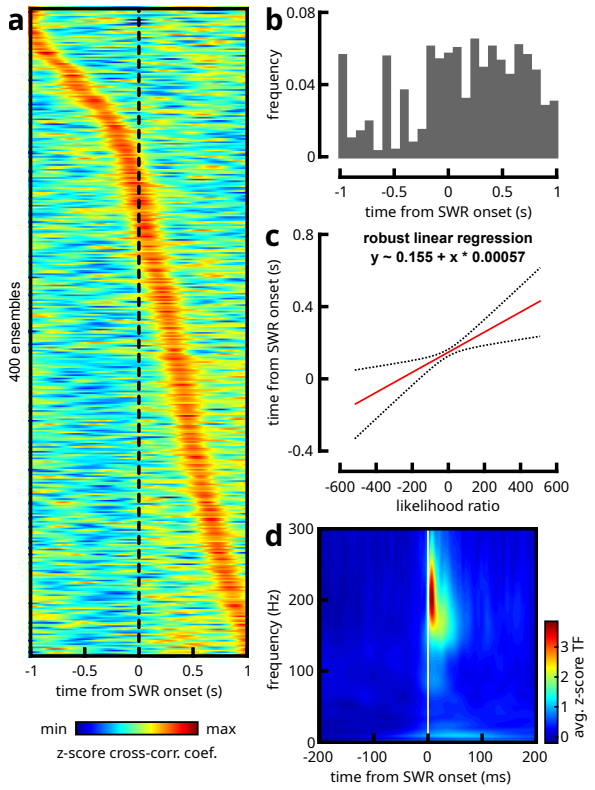

Figure B.12: Timing relationship between reactivation and SWR events is biased by cue versus position encoding. **a** Z-scored cross-correlogrammes (as in Fig. 4.4**f**) between SWR onset times and ensemble reactivation onset times. The majority of ensemble reactivations follow from SWRs. **b** Histogram of peak cross-correlation time lags in **a**. **c** A regression analysis was sought to model the relationship between cells' tendency for encoding either cue or positional features and the timing of their activity with hippocampal SWRs. From **b**, it is clear that the distribution of timings is not normal. To compensate for this limitation, a robust linear regression was conducted. The time lag of peak cross-correlation was modelled as a function of the likelihood ratio, which was used as a measure of cue versus place coding (see eq. 12). Only cells that were part of resting-state ensembles and that passed the criteria for spatial selectivity were considered ($n = 1309$ neurons), while the corresponding peak time lag was derived from the associated ensemble. The fitted model has an intercept at 0.155 s ($p < 0.001$) and a positive slope of 0.57 ms per unit of likelihood ratio ($p = 0.0043$; $R^2_{\text{adj}} = 0.0034$). The fitted model is commensurate with the notion that more cue-tuned cells belong to ensembles that reactivate earlier in time in relation to SWRs, while this timing is progressively delayed as cells are more position-tuned. **d** Time-frequency plot obtained by continuous wavelet transform centred at the onset of SWRs in one example recording session.

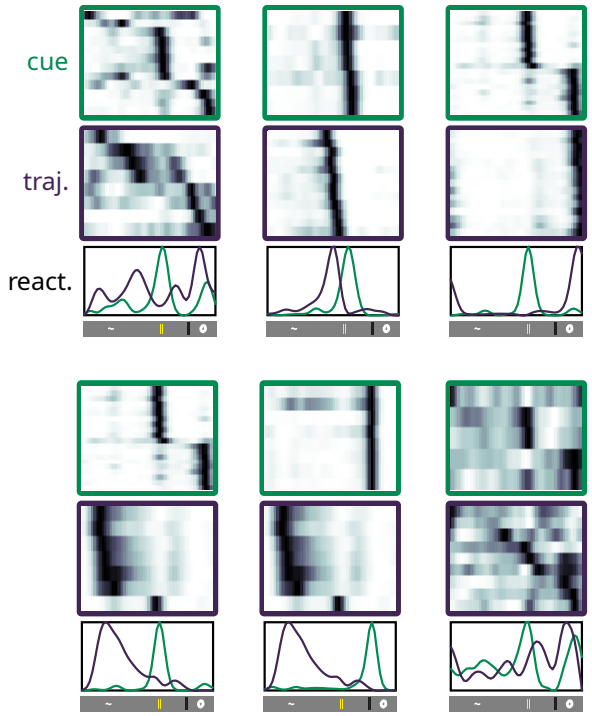

Figure B.13: Six examples of temporally coupled cue-trajectory ensemble pairs. The average neuronal activities as a function of spatial location for all neurons belonging to a cue or trajectory ensemble are shown. Neurons were sorted by location of peak activity. The corresponding reactivation strengths as a function of spatial location for the corresponding cue and trajectory ensembles are illustrated.

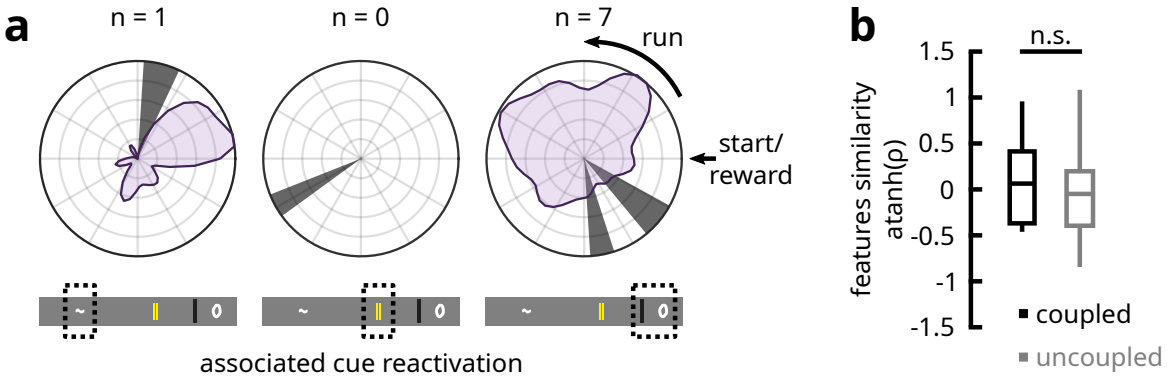

Figure B.14: Cue and trajectory ensemble pairs from REST1 that are temporally-coupled. **a** Same as Fig. 4.5**d** for coupled REST1 ensembles. Out of the eight pairs of cue and trajectory ensembles, seven were associated with the later two cues, while one is associated with the first cue. **b** Given the small sample size, the cue identify was omitted (cf. Fig. 4.5**e**), and only the main effect of temporal-coupling was tested. No significant difference in the similarity of reactivated features between coupled and uncoupled ensemble pairs was detected (two-tailed two-sample t-test on atanh-transformed Pearson correlation coefficients; $p = 0.494$). However, given the low statistical power owing to the restricted sample size, it cannot be concluded that REST1 ensembles do not reactivate for complementary features.

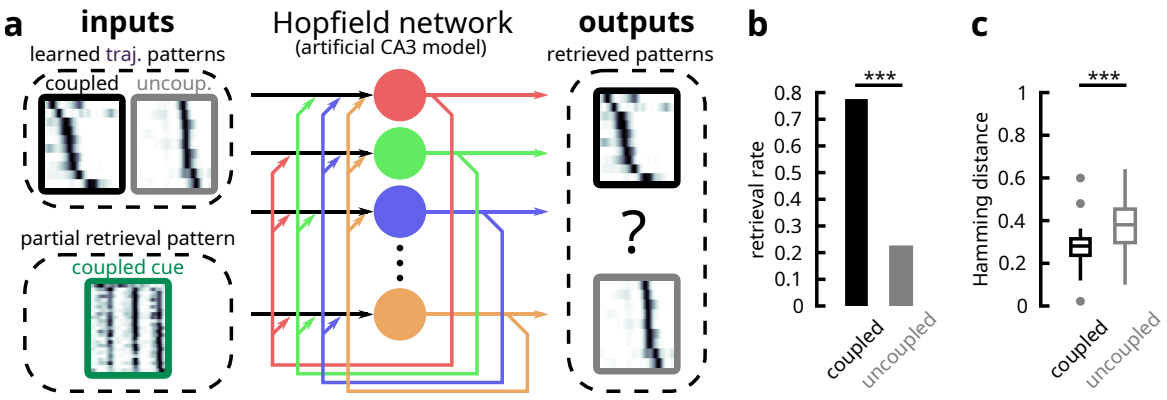

Figure B.15: Hopfield network modelling for the retrieval of trajectories by seeding with cue information. **a** Illustration of the Hopfield network model. Within each session where a 'coupled' pair was present, the reactivated trajectory features from the 'coupled' ensemble and from one 'uncoupled' ensemble were used as training patterns fed to the Hopfield network. The 'coupled' cue ensemble pattern was subsequently used as a partial/degraded pattern to retrieve one of the two stored patterns. **b** Temporally 'coupled' trajectory features were more likely to be retrieved from the Hopfield network by the associated cue features than the 'uncoupled' trajectory features (binomial test; $p < 0.001$). **c** The Hamming distances between the reactivated features of 'coupled' pairs were shorter than those of the 'uncoupled' pairs (one-tailed Mann-Whitney U test; $p < 0.001$), which again suggests a greater degree of similarity in reactivated features between temporally-coordinated cue/trajectory ensemble pairs.

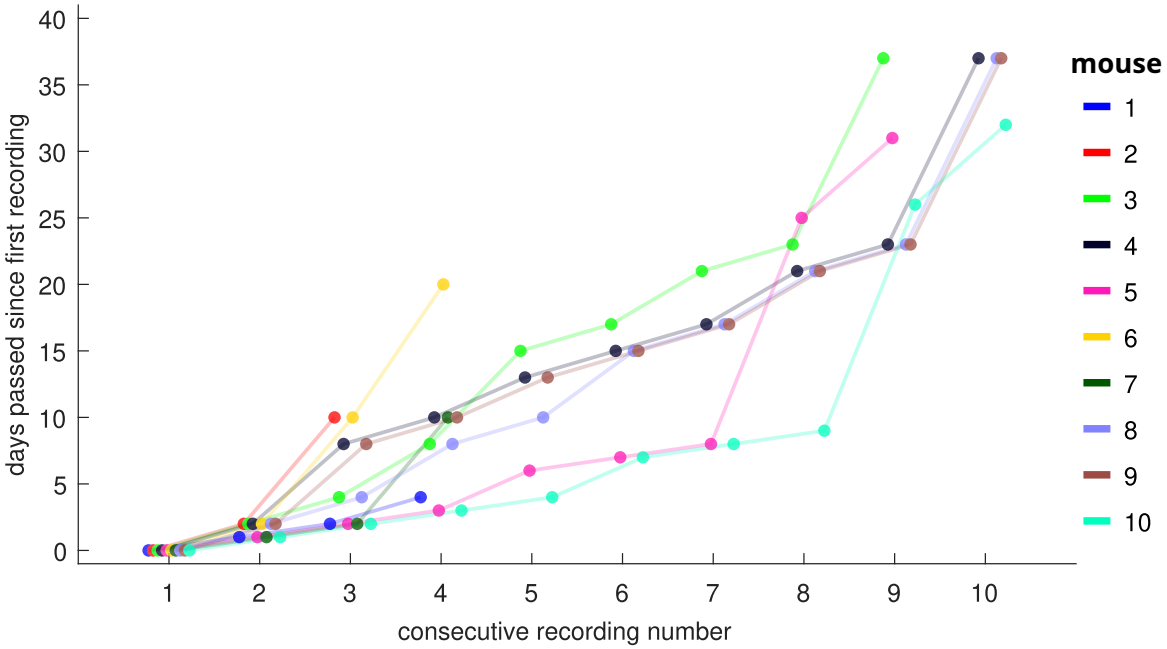

Figure B.16: The number of days elapsed since the first recording for each consecutive recording sessions in the ten animals in which the same neuron ROIs were tracked across time.

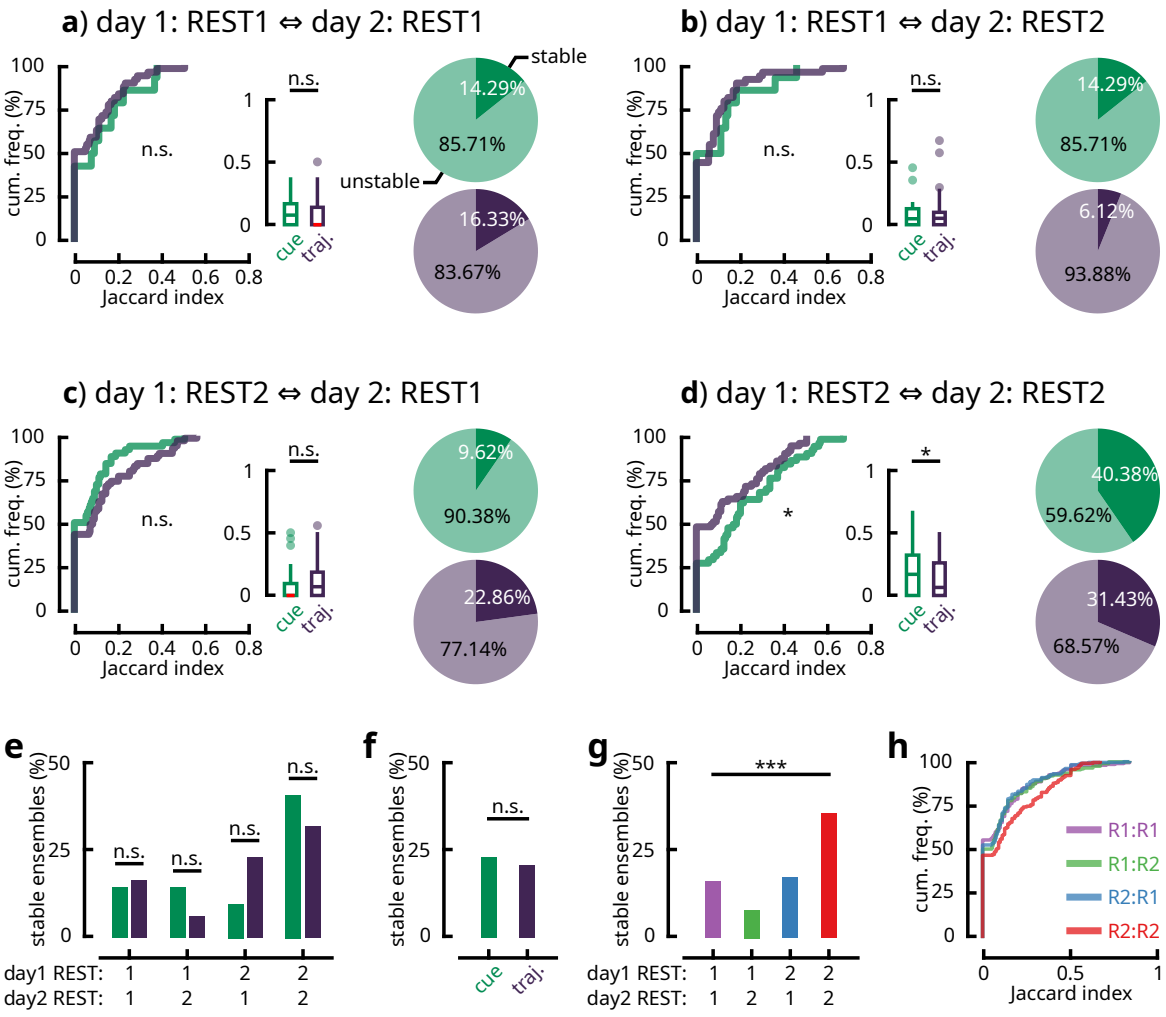
a) day 1: REST1 ⇔ day 2: REST1
b) day 1: REST1 ⇔ day 2: REST2
c) day 1: REST2 ⇔ day 2: REST1
d) day 1: REST2 ⇔ day 2: REST2
cum. freq. (%)
Jaccard index
n.s.
*
cue
traj.
stable
unstable
14.29%
85.71%
16.33%
83.67%
6.12%
93.88%
9.62%
90.38%
22.86%
77.14%
40.38%
59.62%
31.43%
68.57%
e
f
g
h
stable ensembles (%)
day1 REST:
day2 REST:
***
R1:R1
R1:R2
R2:R1
R2:R2

Figure B.17: Resting state ensembles observed following active locomotion persist across recording days. **a-d** The proportions of overlapping ensemble members (quantified as Jaccard distances) and the percentage of persistent ensembles across consecutive recording days were determined using the same procedures described in Fig. 4.6**e-f**. All combinations of REST1/REST2 ensembles across days were evaluated (e.g., in **a**, REST1 ensembles on the reference day were matched with REST1 ensembles on the subsequent recording day). Out of these combinations, cue ensembles ($n = 14$ in REST1; $n = 52$ in REST2) from REST2 on the reference day were slightly more stable than trajectory ensembles ($n = 49$ in REST1; $n = 70$ in REST2) in REST2 on the subsequent day (two-sample two-tailed Kolmogorov-Smirnov test $p = 0.024$; two-tailed Mann-Whitney U-test $p = 0.029$). Otherwise, no differences had been observed in the other combinations. Overall, a greater fraction of both cue and trajectory ensembles persisted in REST2 across recording days (**d**). **e** To further corroborate the results in **a-d**, pairwise χ^2 tests were performed between the proportions of persistent cue and trajectory ensembles over all REST combinations (df $= 1$; significance $\alpha = 0.05$). No differences between the fractions of stable cue and trajectory ensembles were found across all conditions, including in REST2-REST2, even prior to adjusting for multiple comparisons. Therefore, the increased persistence of cue ensembles across days following locomotion, observed in **d**, was marginal. **f** At the group level (i.e., taking the sum across all REST combinations), there was no difference in persistence between cue and trajectory ensembles either ($\chi^2 = 0.231$; df $= 1$; $p = 0.631$). **g** There was however a main effect of across-days REST combinations on the stability of ensembles, without accounting for cue/trajectory labels ($\chi^2 = 23.148$; df $= 3$; $p < 0.001$). **h** Plotting the empirical cumulative distribution functions of Jaccard distance for all REST combinations revealed that the proportions of stable cue or trajectory ensembles in REST2-REST2 was markedly higher than in other combinations. Taken together, these results suggest that following locomotion, similar resting state ensembles tend to be recruited across days, although no preference is assigned to either cue or trajectory ensembles.

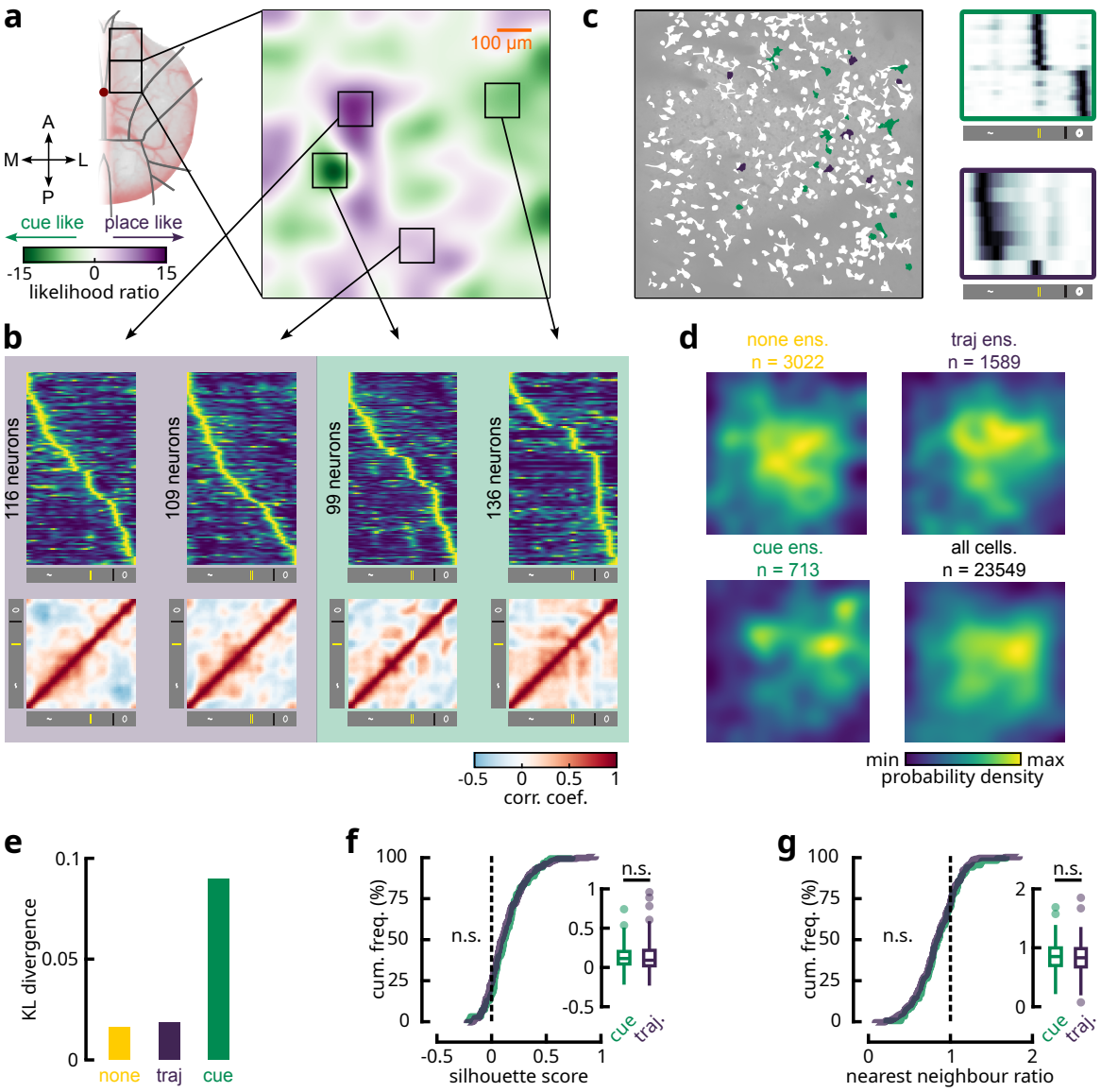
a
100 µm
A
M
L
P
cue like
place like
-15
0
15
likelihood ratio
b
116 neurons
109 neurons
99 neurons
136 neurons
-0.5
0
0.5
1
corr. coef.
c
d
none ens.
n = 3022
traj ens.
n = 1589
cue ens.
n = 713
all cells.
n = 23549
min
max
probability density
e
KL divergence
0.1
0.05
0
none
traj
cue
f
cum. freq. (%)
100
75
50
25
0
n.s.
n.s.
1
0.5
0
-0.5
cue
traj.
-0.5
0
0.5
1
silhouette score
g
cum. freq. (%)
100
75
50
25
0
n.s.
n.s.
2
1
0
cue
traj.
0
1
2
nearest neighbour ratio

Figure B.18: Topographic organisation of cue and spatially-selective cells. **a** Topographic map of the distribution of cells' tendency for encoding either cues or spatial positions. **b** Four 100 µm × 100 µm windows were drawn over patches with higher densities of either cue or locations responses. The average responses over positions of the neurons, whose centroids fall within the boundaries of these windows, were obtained, along with the correlation matrices of the resulting population vectors. **c** Mask of neuronal ROIs in one example recording session. Individual examples of a cue and of a trajectory ensemble, and their corresponding ROIs are shown. **d** Density maps of the locations of neurons belonging to each resting ensemble class (REST1 and REST2 pooled), as well as all neuronal ROIs, were obtained by regular kernel density estimation using the same parameters as in **a**. **e** Kullback-Leibler divergence between the probability density maps of each ensemble class and the density map of all neuronal ROIs in **d**. **f-g** Averaged silhouette scores and nearest neighbour ratios across neurons of a given ensemble ($n = 75$ cue ensembles; $n = 196$ trajectory ensembles). Overall, there was no difference in the degree of clustering between cue and trajectory ensemble neurons (two-sample two-tailed Kolmogorov-Smirnov test $p = 0.513$ for silhouette score, $p = 0.936$ for NNI; two-tailed Mann-Whitney U-test $p = 0.392$ for silhouette score, $p = 0.511$ for NNI). However, on both measures of clustering, over $\sim$70 % of ensemble neurons of either category tend to be more clustered in topographic space, rather than dispersed (Wilcoxon signed rank test $p < 0.001$ for median silhouette score equals 0, $p < 0.001$ for median NNI equals 1).

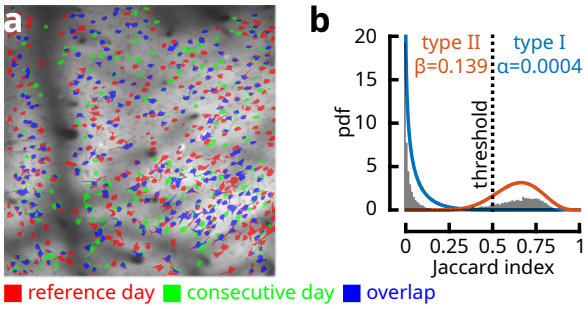

Figure B.19: Identification of the same neuronal ROIs across consecutive days of recording. **a** Registered ROI masks for neurons detected on the reference day (red) and on the subsequent day of recording (green), in one example session. Overlapping pixels are coloured in blue. **b** Probability density function of the percentage of overlap (measured by Jaccard distance) between neuron ROIs across days, for all neuron pairs with at least one overlapping pixel. The histogram reveals a bimodal distribution consisting of persistent and differing cells. These two clusters were separated by K-means and were each fitted to a beta distribution by maximum likelihood estimation. The beta distribution, commonly used to model percentages and probabilities, is defined over the interval $[0, 1]$ and aptly describes Jaccard distances. With the overlapping threshold set at 50 %, the left-side beta distribution (blue), which represents unstable neurons, estimates a false positive rate of 0.04 %. Meanwhile, the statistical power given by the right-side distribution (orange) is estimated at 86.1 %. Therefore, a 50 % overlapping threshold for distinguishing persistent neurons across days is a highly robust criterion for the present dataset, at a reasonable cost to statistical power. This model was inspired by Sheintuch et al. (2017).

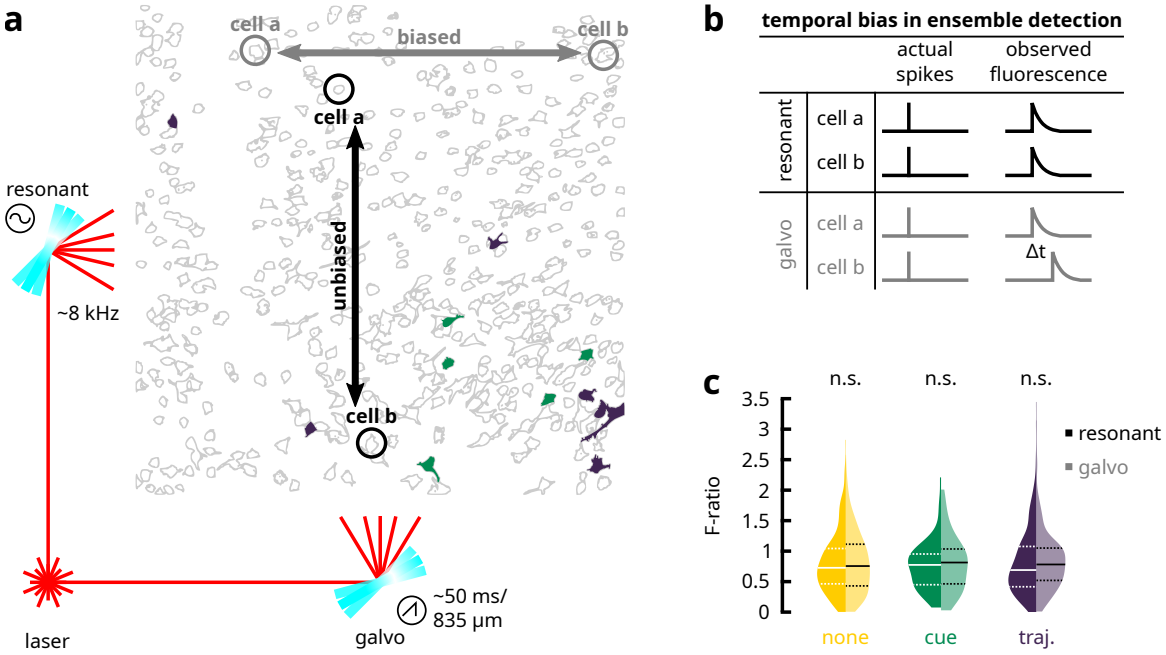

Figure B.20: Detection of offline ensembles is not biased by the scan mirrors. **a** The laser beam is directed to the two axes of the field-of-view by a galvanometer scan mirror and a resonant scan mirror respectively. Examples of two offline ensembles (a cue ensemble in green and a trajectory ensemble in purple) are superimposed over all detected ROIs. **b** A direct consequence of this preparation is that neurons that are farther apart over the galvo axis (up to the midway point between the scanning path and the mirror flyback) are susceptible to shifted/delayed temporal dynamics. This can introduce a bias in the detection of offline ensembles, where synchronous ensembles found along the resonant axis are more likely to be detected compared to synchronous neurons along the galvo axis. Temporal smoothing had been performed to compensate for this issue. **c** To verify that this bias did not impact the detection, the following statistical hypothesis was proposed: the variance in the positions of ensemble neurons along the galvo axis, normalised by the total variance of all ROIs (i.e., the F-ratio), should be equal to that along the resonant axis. This null hypothesis held for all detected offline ensembles, irrespective of category (paired-sample two-tailed Wilcoxon signed rank tests; $p = 0.5$ for none; $p = 0.36$ for cue; $p = 0.273$ for trajectory). Therefore, synchronous neuron pairs are as likely to be detected over the span of the resonant axis as they are over the galvo axis.

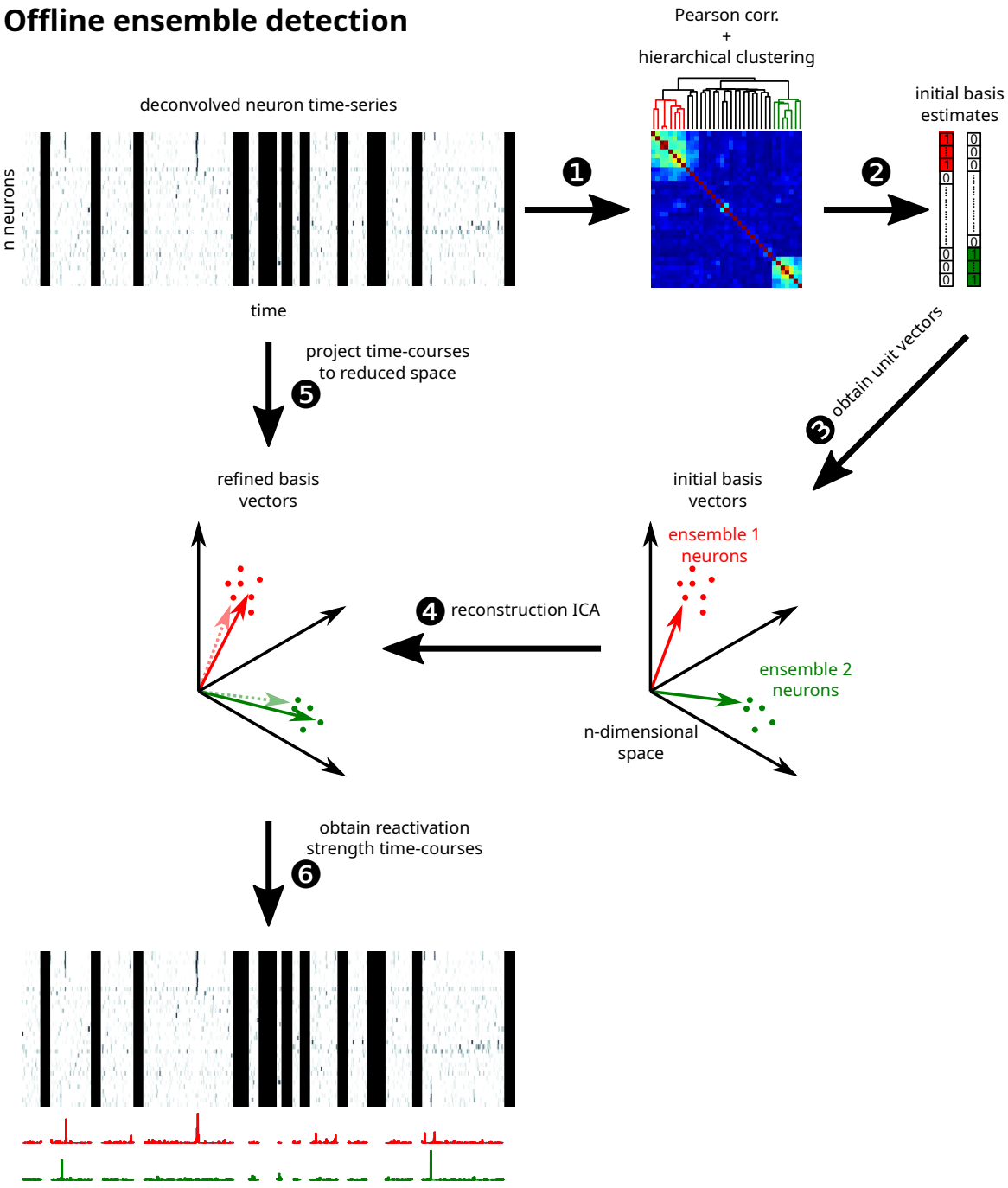

Figure B.21: Methods used for detecting synchronous resting-state ensembles and for establishing the time-courses of their reactivation. See Methods for detailed description. 1) Hierarchical clustering is conducted over the Pearson correlation matrix of the neuronal time-series for the resting period. 2) For each ensemble, a binary column vector of the same length as the total number of neurons is created. Members belonging to the ensemble are labelled as '1', while the remaining are '0'. 3) Normalizing these vectors by their norm yields a set of unit vectors that together form an orthonormal basis. 4) Using reconstruction ICA, these basis vectors are fine-tuned in such a way as to retain most of the variance in the original data, and in doing so capture the relative contributions/weights of each member neuron to an ensemble's temporal dynamics. 5) The original neuronal time-series matrix is projected into the new space defined by the basis. 6) This projection yields the reactivation strength of each ensemble as a function of time.

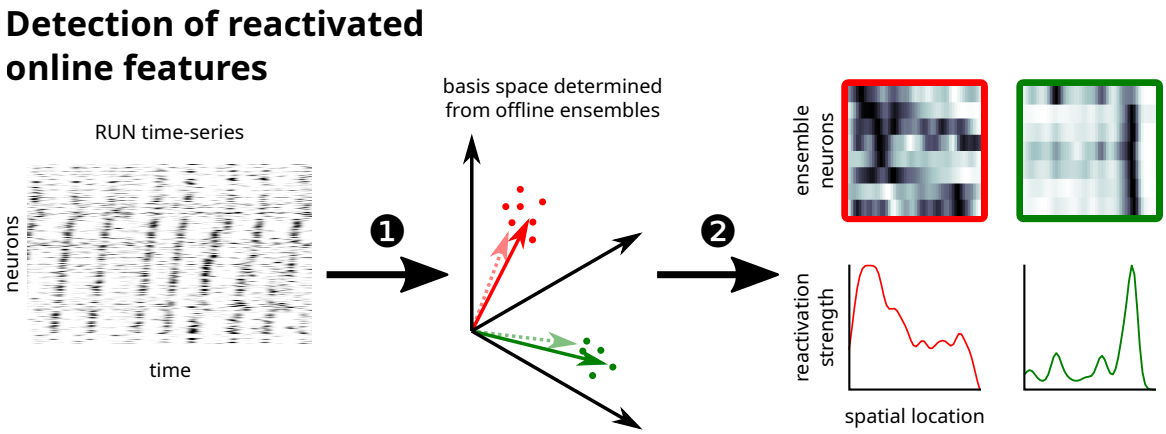

Figure B.22: By projecting the neuronal time-series matrix during RUN into the space defined by the basis estimated from resting network dynamics (1), the ensemble activation strength as a function of animal location can be obtained (2).

B.2 Supplementary Tables

Table B.1: Experimental conditions for individual animal subjects

				number of ensembles							
				REST1				REST2			
animal	num. sessions	LFP	ROI tracking	none	cue	traj.	all	none	cue	traj.	all
1	6	n/a	n/a	12	0	7	19	13	0	10	23
2	5	n/a	n/a	24	2	10	36	16	3	34	53
3	1	n/a	n/a	6	1	2	9	0	0	2	2
4	4	avail.	avail.	10	0	3	13	9	1	2	12
5	3	poor	avail.	19	1	13	33	16	8	13	37
6	1	avail.	n/a	4	0	1	5	6	1	3	10
7	9	avail.	avail.	19	1	4	24	38	1	15	54
8	10	avail.	avail.	30	5	9	44	12	17	12	41
9	9	avail.	avail.	26	0	4	30	13	4	3	20
10	4	avail.	avail.	14	3	8	25	26	6	16	48
11	4	poor	avail.	9	1	15	25	7	10	17	34
12	10	avail.	avail.	15	1	0	16	12	1	8	21
13	10	avail.	avail.	50	1	7	58	52	0	10	62
14	10	avail.	avail.	38	3	14	55	32	10	19	61
total	86	9	10	276	19	97	392	252	62	164	478

B.3 Supplementary Methods

B.3.1 Estimating proportions of cue/spatial cells

Ideally, cells that respond exclusively to visuo-tactile cues would have been identified by introducing laps over a blank belt and comparing the responses of neurons following the addition of cues. This was, however, not the case in the current set of experiments. Therefore, only a rough estimation of the proportions of cue-responsive cells can be made. For this estimation, we assume that cortical spatial cells are uniformly scattered over locations, consistent with the behaviour of hippocampal place cells (cf. Rich et al., 2014). At the locations of cues, extra cells will be found responding to the sensory cues. Therefore, if we obtain the locations at which neurons show peak firing, the

resulting distribution is expected to be approximately uniform, with elevated density at the locations of cues (Fig. B.7**a-b**).

It follows that, if we are to iteratively remove neurons whose peak responses occurred at cue locations, then the distribution of activity peaks will gradually approach a uniform distribution. We used the χ^2 statistic to measure the uniformity of this distribution:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - \frac{N}{n})^2}{\frac{N}{n}}, \tag{B.1}$$

where O_i is the number of peaks in bin i, for which there are $n = 50$ bins and N peaks in total. $\frac{N}{n}$ is the expected number of peaks at each location under the assumption of uniformity. The objective is to minimize χ^2 by removing peaks falling within the locations of cues (i.e., to fit parameters O_i for all i bins coinciding cue locations and adjusting N accordingly). We solved this minimization problem by the interior-point method. Finally, taking the fraction between the number of cells removed following "uniformization" of the distribution of peaks and the total number of cell yields the estimated proportion of cue-responsive cells (Fig. B.7**c**).

In the first and last bins, an unusually elevated number of peaks are found, likely reflecting cells responding to reward (Fig. B.7**a-b**). These bins were omitted from the optimization.

B.3.2 Temporal Compression

Given the low sampling rate of our imaging data, estimating the temporal compression during replay events may not proceed in the same manner as detailed in (Euston et al., 2007) (i.e., template matching) without the risk of incurring a substantial bias. This is because the median duration of reactivation events is on the orders of ∼200 ms (Fig. 4.4**g**), which corresponds to only 4 imaging frames, too few to accurately represent any putative sequences in the reactivation events. Instead, a surrogate analysis was derived from our PCA-ICA method. Following on Equation 6, we extracted the principal components corresponding to each resting-state ensemble. We reasoned that, by downsampling (i.e., compressing) the RUN time-series matrix, these resting-state principal components would explain a higher fraction of the variance in the RUN data, if the RUN data corresponding to the rest ensembles is organised in a sequential form. In other words, if a component captures a trajectory ensemble, then temporal compression of the RUN time-series would lead to the sequence collapsing into synchronous activities similar to those observed during rest, from which the component was derived. In contrast, cue ensembles should not benefit from such compression, as the neurons would already fire in synchrony during the RUN periods in response to cue sensations. To quantify this relationship, we proposed a measure of *divergence*, given as

$$d = 1 - |\frac{r^2_{\text{RUN}}}{r^2_{\text{REST}}} - 1|, \tag{B.2}$$

where r^2_{RUN} and r^2_{REST} correspond to the proportion of variance explained by a given PCA-ICA component during RUN and REST, respectively. This measure has an upper bound of 1, where a value close to 1 indicates that the fraction of variance explained by a given component is similar between RUN and REST behavioural states. This divergence was calculated at each compression of the RUN time-series. Downsampling was conducted using an averaging filter of the same order as the compression factor.

To further validate our criteria for the classification of cue and trajectory ensembles, we sought an unsupervised clustering method that is not biased by assumptions. The temporal compression analysis (Fig. B.9) provides an elegant way to discover these putative ensemble classes; if cue and

trajectory ensembles are distinguished by the presence of a sequential structure, then in principle they should organise into two separate clusters in a feature space that describes their affinity to temporal compression. To test this hypothesis, *non-negative matrix factorisation* (NNMF) was performed over the divergence matrix of ensembles (a $c \times e$ matrix where the rows correspond to different compression factors, the columns to individual ensembles and the values are the divergence scores reported in Fig. B.9). Note that, though our divergence measure does not have a lower bound, no negative values were obtained in practice. NNMF factorises the divergence matrix into two matrices. The first matrix (the 'feature' matrix) extracts an arbitrary number of components that accounts for regular patterns in the divergence measures across all ensembles. The second matrix (the 'coefficients' matrix) attributes weights to individual ensemble for each of these distinctive features. NNMF was conducted using three components, in consideration of the three putative ensemble classes proposed (none, cue and trajectory).

B.3.3 Modelling cue- and position-correlated responses

In order to quantify the individual contributions of visuo-tactile cues and spatial locations to neuronal responses, two candidate models were fitted to the activity time-courses of neurons. In the first model, which aims to reflect the individual neuronal responses that form a trajectory sequence, it was assumed that the firing rate profile of trajectory ensemble neurons follow that of 'place cells'. Accordingly, a Gaussian function was used to describe a neuron's firing rate as a function of location:

$$\lambda(x) = A\exp(-\frac{(x-B)^2}{2C^2}). \tag{B.3}$$

Here, the neuron's activity as a function of position $\lambda(x)$ follows a Gaussian function with amplitude A, place field centre B and standard deviation C. Parameters A through C were fitted by maximizing the likelihood function formulated below (eq. B.5) using an *interior-point* approach. The optimization was subjected to the following constraints: $A \geq 0$ and $0 \leq B \leq 150$ (corresponding to the length of the belt).

In the second model for cue responses, the mean activity rates $(\lambda_1, \ldots, \lambda_4)$ in each individual space defined by visuo-tactile cues (a, b, c, d), along with the baseline activity outside of cue locations (λ_0), were obtained such that the position-dependent response function is given by:

$$\lambda(x) = \begin{cases} \lambda_1 & \text{if } x \in a \\ \lambda_2 & \text{if } x \in b \\ \lambda_3 & \text{if } x \in c \\ \lambda_4 & \text{if } x \in d \\ \lambda_0 & \text{if } x \notin a \cup b \cup c \cup d \end{cases}. \tag{B.4}$$

As such, neurons can support multiple place fields in the form of a distinct firing rate at each unique cue location. The quality of the fitted models at explaining the observed neuronal responses was assessed from a likelihood function. Assuming that neurons' activity patterns follow an inhomogeneous Poisson process, it can be shown that the log-likelihood of a candidate model given the activity and position vectors from the observed data is expressed as:

$$l(\lambda; n_1, \ldots, n_T, x_1, \ldots, x_T) = \sum_{i=1}^{T} n_i \log(\lambda(x_i)) - \lambda(x_i) - \log(\Gamma(n_i + 1)). \tag{B.5}$$

Here, the time-course vectors of neuronal activity and animal position are represented by n_i and

x_i respectively, and the activity rate function $\lambda(\cdot)$ is given by one of the candidate models. It is important to note that it is inaccurate to describe neuronal responses, as observed from calcium fluorescent transients, through a Poisson distribution, given that the sampling rate of the technique is too slow to discern individual spikes. However, it has been shown previously that this assumption, as a first-order approximation, does model accurately the relationships between calcium activity and behavioural parameters (Mao et al., 2018; Esteves et al., 2021). In the present scenario, where the time-course vector n_i resides over a positive continuous domain, the integer factorial was substituted for the gamma function following the relationship $\Gamma(n+1) = n!$.

The likelihood ratio was taken between the likelihoods under the two candidate models to evaluate which of the two models is better supported by the observed data:

$$L_{\mathrm{R}} = -2(l_{\mathrm{cue}} - l_{\mathrm{place}}). \tag{B.6}$$

Under this ratio, higher values imply that the 'place cell' model is a stronger candidate than the 'cue' model and *vice versa* for lower values.

B.3.4 Hopfield network modelling

The reactivated features (i.e., the reactivation strength as a function of spatial position) were discretized into 50 spatial bins and normalized between 0 to 1. These vectors were then binarized at a threshold of 0.5, where bins with activity higher than 0.5 are set to +1 and the remaining are set to -1. At the learning stage, a weight matrix $w_{i,j}$ is constructed from a pair of trajectory patterns that are to be stored by the Hopfield network (cf. Hopfield, 1982):

$$w_{i,j} = \frac{1}{2}\sum_{p=1}^{2} v_i^p v_j^p, \tag{B.7}$$

where v_i^p is the value of the pattern p in bin i, and the diagonal of the weight matrix is set to 0. At the retrieval stage, the retrieval cue pattern s_j with $N = 50$ spatial bins is presented:

$$s_j(t+1) = \mathrm{sgn}(\sum_{j}^{N} w_{i,j} s_j(t)), \tag{B.8}$$

where t is the iteration number and $\mathrm{sgn}(\cdot)$ is the sign function (turning s back into a binary vector with values of +1 and -1). Equation B.8 is repeated until one of the two stored trajectory patterns is retrieved or when the number of iterations exceeds 50 in which case the trial is considered to have failed to converge.

B.3.5 Topographic Organisation

Out of 86 imaging sessions ($n = 14$ mice), 74 sessions ($n = 11$ mice) were conducted over the posterior window. Given these unequal sample sizes, the current topographic analysis will only focus on the posterior aspect of the secondary motor cortex, in consideration for adequate statistical power. A weighted kernel density estimation procedure (Wolters and Braun, 2018) was used to construct a topographic map of the distribution of cells' tendency for encoding cues versus spatial positions. First, the centroid coordinates of neuronal ROIs were obtained and pooled across all animals and sessions. The likelihood ratios for each neuron, that passed the criteria for spatial selectivity, were then calculated for each centroid following Equation 12 ($n = 7871$ neurons). A weighted average of the likelihood ratio was then computed at each topographic bin (400×400 bins), where the weights are assigned by a Gaussian kernel with a bandwidth of 50 μm.

Two measures of clustering/dispersion in topographic space were computed. On the one hand, the silhouette score is given as

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \tag{B.9}$$

where, for a given ensemble neuron i, a_i is the mean Euclidean distance between it and all other neurons belonging to the same ensemble, and b_i is its mean distance from all other neurons that are not part of the ensemble. The silhouette score ranges between -1 and 1, with higher positive values corresponding to higher clustering and lower negative values signifying more dispersion.

On the other hand, the nearest neighbour ratio (NNI) was computed as

$$\text{NNI} = \frac{D_o}{D_e}, \tag{B.10}$$

where D_o is the average Euclidean distance between the nearest neighbours amongst neurons of a given ensemble, while D_e is the expected mean distance under random sampling. D_e was obtained by randomly drawing a sample of the same size as the number of neurons within the ensemble from all the available ROIs of a recording session. This random sampling procedure was performed 1000 times and the average was taken as the value of D_e. NNI values smaller than 1 tend towards clustering, while values greater than 1 suggest dispersion.