On the encoding, transfer, and consolidation of spatial memories

by

Michele Nardin

June, 2022

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:

Prof. Christopher Wojtan, Chair
Prof. Jozsef Csicsvari
Prof. Gašper Tkačik
Prof. Andreas Herz

The thesis of Michele Nardin, titled *On the encoding, transfer, and consolidation of spatial memories*, is approved by:

Supervisor: Prof. Jozsef Csicsvari, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Prof. Gašper Tkačik, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Prof. Andreas Herz, LMU Munich, Planegg-Martinsried, Germany

Signature: _____

Defense Chair: Prof. Christopher Wojtan, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Michele Nardin
June, 2022

Signed page is on file

Abstract

The ability to form and retrieve memories is central to survival. In mammals, the hippocampus is a brain region essential to the acquisition and consolidation of new memories. It is also involved in keeping track of one's position in space and aids navigation. Although this space-memory has been a source of contradiction, evidence supports the view that the role of the hippocampus in navigation is memory, thanks to the formation of cognitive maps. First introduced by Tolman in 1948, cognitive maps are generally used to organize experiences in memory; however, the detailed mechanisms by which these maps are formed and stored are not yet agreed upon. Some influential theories describe this process as involving three fundamental steps: initial encoding by the hippocampus, interactions between the hippocampus and other cortical areas, and long-term extra-hippocampal consolidation. In this thesis, I will show how the investigation of cognitive maps of space helped to shed light on each of these three memory processes.

The first study included in this thesis deals with the initial encoding of spatial memories in the hippocampus. Much is known about encoding at the level of single cells, but less about their co-activity or joint contribution to the encoding of novel spatial information. I will describe the structure of an interaction network that allows for efficient encoding of noisy spatial information during the first exploration of a novel environment.

The second study describes the interactions between the hippocampus and the prefrontal cortex (PFC), two areas directly and indirectly connected. It is known that the PFC, in concert with the hippocampus, is involved in various processes, including memory storage and spatial navigation. Nonetheless, the detailed mechanisms by which PFC receives information from the hippocampus are not clear. I will show how a transient improvement in theta phase locking of PFC cells enables interactions of cell pairs across the two regions.

The third study describes the learning of behaviorally-relevant spatial locations in the hippocampus and the medial entorhinal cortex. I will show how the accumulation of firing around goal locations, a correlate of learning, can shed light on the transition from short- to long-term spatial memories and the speed of consolidation in different brain areas.

The studies included in this thesis represent the main scientific contributions of my Ph.D. They involve statistical analyses and models of neural responses of cells in different brain areas of rats executing spatial tasks. I will conclude the thesis by discussing the impact of the findings on principles of memory formation and retention, including the mechanisms, the speed, and the duration of these processes.

Acknowledgements

Becoming a scientist has been a painful yet rewarding experience. I want to thank and acknowledge many people who have determined my scientific growth, starting from my supervisor Jozsef for giving me the opportunity of joining his lab, freedom and independence in choosing the projects, and the fruitful supervision. Charlotte and Federico for patiently teaching me how to deal with and survive within neuroscience. Gašper for the guidance, lessons, and wise suggestions. Karola and Cristina, amazing collaborators, who have taught me much. Last but not least Sander, Bill, and James, who have taken the time to pursuit a project that would have never seen the light without their help and inputs.

Becoming a scientist necessitates constant support and encouragement. For this I want to thank Laura, my wonderful partner, and my entire family, especially my mother Liliana, my father Roberto, my sister Valentina, and my brother Stefano. Wonderful friends in italy and elsewhere, like Ivan, Matteo, Claudio, Luca, Giulia, Andrea, Francesco and many others that I have met during my life. I want to thank people in the Csicsvari group, friends at IST, people in the running group, chess group, and the students that took the methods of data analysis course.

Finally, I want to thank Prof. Csicsvari, Prof. Herz, and Prof. Tkačik for agreeing to act as the thesis defense committee, and Prof. Wojtan for chairing my defense.

I acknowledge the support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 665385.

About the Author

Michele Nardin completed a BSc in Mathematics at the University of Trento in 2013 and an MSc in Mathematics at the University of L'Aquila in collaboration with the Gran Sasso Science Institute of L'Aquila and the Technical University of Vienna in 2015. He then joined the group of Jozsef Csicsvari at IST Austria in 2016 where his main research interests included systems and computational neuroscience. He worked on several projects about learning, memory, sequences, representations and computations in brain networks. A strong believer in cross-disciplinarity, he likes to teach and explain complicated subjects in simple terms, for which he won the "golden sponge" award in 2020. He enjoys nature, books, runs, and food.

List of Collaborators and Publications

During my PhD I had the luck of collaborating with several incredible scientists. I will list here the collaborators who contributed[1] to the studies included in this thesis (in order of appearance): **Jozsef Csicsvari**[a], [1,2,3], **Gašper Tkačik** [a], [1], **Cristina Savin**[b], [1], **Karola Kaefer**[c], [2], **Federico Stella**[c], [2,3], **Charlotte Boccara**[d], [3], **Joseph O'Neill**[e], [3].

List of publications that appear in this thesis:

[1] Nardin M, Csicsvari J, Tkačik G, Savin C. "The structure of hippocampal CA1 interactions optimizes spatial coding across experience." (submitted, under revision)
A previous version of this work has been uploaded to Biorxiv and is available here: `https://www.biorxiv.org/content/10.1101/2021.09.28.460602v1`

[2] Nardin M, Kaefer K, Stella F, Csicsvari J. "Theta oscillations as a substrate for medial prefrontal-hippocampal assembly interactions." (submitted)
A previous version of this work has been uploaded to Biorxiv and is available here: `https://www.biorxiv.org/content/10.1101/2021.09.30.462269v1`

[3] Boccara CN*, Nardin M*, Stella F, O'Neill J, Csicsvari J. "The entorhinal cognitive map is attracted to goals." Science. 2019 Mar 29;363(6434):1443-7. (* co-first authorship)
This is the author's version of the work. It is included in this dissertation by permission of the AAAS. The definitive version was published in the journal Science on March 29th, 2019, volume 363, pages 1443–1447. DOI: `https://doi.org/10.1126/science.aav4837`

[1] Detailed author contributions are available at the beginning of each chapter.

Current affiliations:
[a] *Institute of Science and Technology Austria (ISTA), Klosterneuburg, 3400, Austria*
[b] *New York University, New York NY-10003, USA*
[c] *Donders Institute, Radboud University, Nijmegen 10 6500GL, Netherlands*
[d] *University of Oslo, Sognsvannsveien 9 Domus Medica, 0372 Oslo, Norway.*
[e] *Cardiff University, 70 Park Place, Cardiff, CF10 3AT, United Kingdom.*

Table of Contents

List of Abbreviations

CA Cornu Ammonis
FC Functional connectivity
MEC Medial entorhinal cortex
MTL Medial temporal lobe
PFC Prefrontal cortex

1

Introduction

1.1 Memory

Evolution favored the survival of animals endowed with a flexible and powerful solution: a "computational machine", integrated into living beings, in the form of a nervous system (Erwin et al., 2011; Suryanarayana et al., 2020). One of the main advantages of possessing a nervous system lies in the fact that certain aspects of previous experiences can be stored, so as to be better prepared for the next encounter with a similar situation. This ability is generally called memory.

The term "memory" is an elusive one: there have been many attempts to formally define it, but every definition has flaws and no consensus has been reached (Roediger et al., 2007). In this thesis, I will use the pragmatic approach of stating the purported function of memory, rather than a definition per se. Inspired by (Dudai et al., 2007), I will state that the function of memory is "to retain over time neurally encoded representations of the world, created or modified through experience". Although many theories of memory are based on mammalian studies, every animal species with a nervous system actively uses memory every day to survive: this is clear already for the organism with the smallest nervous system ever studied, the 302 neurons of *C. elegans* (de Bono and Maricq, 2005). Honeybees were among the first invertebrates that received formal behavioral scrutiny (Frisch et al., 1926), and the neural underpinning of their ability to quickly learn and remember the color and smell of nectar-yielding flowers was described roughly 50 years later (Menzel and Erber, 1978). Together with animal behavioral and electrophysiological studies, a long tradition of human psychological studies laid the foundations for the study of memory (Jenkins and Dallenbach, 1924). I will introduce here the different approaches and various concepts used to study memory, and then focus on studies in humans, non-human primates, and rodents, which gave rise to many theories of memory consolidation focusing on the hippocampal system.

How to study memory

Memory is usually formed through experience or learning. The main challenge is to link the behavioral indication of learning with the underlying cellular/network mechanisms. This can be studied on different levels, broadly related to the different molecular and cellular scales commonly studied.

The synaptic level. Also called "synaptic consolidation", it refers to the long-term strengthening or weakening of synaptic efficacy (Hebb, 1949; Bliss and Lømo, 1973). There exist different

forms of synaptic potentiation/depression (Magee and Grienberger, 2020): this process is generally fast-paced and finishes within minutes to hours from its initiation (Kandel, 2013). Although well accepted and with ample experimental support (Glanzman, 2010), there are still outstanding criticisms of the relevance of synaptic plasticity to learning (Gallistel and Matzel, 2013). Moreover, a number of non-synaptic mechanisms have been discovered that can be used as memory substrates (Lisman et al., 2018), which can be hard to reconcile with a pure synaptic-centered view(Abraham et al., 2019).

The engram level. An influential theory about memory discusses the concept of "engrams". In a nutshell, an engram represents the physical instantiation of memory, i.e., the group of neurons and synapses within which a memory resides (Lashley, 1950). The "search for the engram" has so far proven difficult (Bruce, 2001), although some recent successes have been reported, mostly due to novel techniques (Ryan et al., 2015; Kitamura et al., 2017; Tonegawa et al., 2018). One of the reasons why it is so difficult to locate physical engrams is the fact that memories are largely distributed across brain regions (Josselyn et al., 2015). This observation was made already at the beginning of the last century through lesions (Lashley, 1950) and has been recently observed also by groups ardently pursuing the engram theory (Roy et al., 2022).

The systems level. Systems consolidation generally refers to the reorganization of long-term memories, after initial encoding, over distributed brain circuits (Dudai et al., 2015). Many ideas of systems consolidation derive from cognitive studies on humans, which have helped shape the fields' understanding of memory consolidation (Squire, 2004). A major advantage of working with humans is that memories can be expressed verbally, and then tested using non-invasive recording techniques (Rasch and Born, 2013). On the other side, studying memory in animals is enticing, as one can record the electrical activity of single neurons during behavior. This has the obvious disadvantage that animals cannot talk, so there is a need for testing memory in different ways (Tolman, 1948). Animal studies initially contributed to the understanding of memory by studying the effects of different brain lesions in non-human primate models of amnesia (Zola-Morgan et al., 1983). Electrophysiological studies focused on responses of particular types of cells, e.g. hippocampal place cells (O'Keefe and Dostrovsky, 1971), but the advance in recording techniques allowed scientists to shed light on general memory consolidation mechanisms (Girardeau and Lopes-dos Santos, 2021). An outstanding example is a recent study that selectively inhibited the consolidation of specific experiences while leaving the others intact (Gridchyn et al., 2020).

Encoding vs. consolidation

Neural encoding refers to the ability of the brain to represent external sensory inputs, or internal physiological states, in a form amenable to further computation and later retrieval (Roediger et al., 2007). The ability to encode information is a prerequisite to forming a memory, but not every piece of information encoded will be stored and available forever. Most of it will be short-lived (i.e., short-term memory), and will be forgotten after minutes to hours, while some will be transferred to a more stable, long-term memory storage (Hebb, 1949). Some theories exist as to how memories are "tagged" for long-term storage (Benchenane et al., 2010), but there is no agreement so far (Inostroza and Born, 2013). More generally, it is not clear why certain memories should be forgotten while some others should consolidate. One could imagine a situation where every piece of information encoded is immediately stored long-term, and the problem is then shifted to retrieval (Miller and Matzel, 2000). This possibility seems to be generically disadvantageous because it would lead to a fast depletion of memory and computational capabilities. Moreover, this explanation does not agree with the literature

detailing molecular and cellular processes that, if blocked, prevent the formation of long-term memories (Martin et al., 2000). Another explanation, supported by early computational work, is that information is bound together and integrated within existing knowledge in the neocortex so as to form more useful and informative records (McClelland et al., 1995). By now, various forms of the consolidation hypothesis are widely accepted and used to explain a large body of experimental evidence on memory acquisition, loss, and recovery (Kandel, 2013; Klinzing et al., 2019; Cowan et al., 2021), although some skepticism remains (Miller, 2021).

Initial memory encoding in the hippocampus

Memory is conventionally subdivided into declarative memory, which requires awareness for retrieval, and non-declarative memory, which does not require such awareness (Squire, 2004)[1]. The distinction between the two memory systems was experimentally probed on amnesic patients in the 50s and 60s. The classical example is patient H.M., who suffered from anterograde amnesia after most of his medial temporal lobe (MTL) was surgically removed in an attempt to cure his intractable epilepsy. Scoville and Milner (1957) showed that patient H.M. could acquire new motor skills while having no memory of ever doing it. Their studies showed that the MTL is necessary to acquire episodic memories and different brain areas can act independently on different types of memory. These works inspired research for the following half-century on how MTL areas support memory. Specifically, the hippocampus has since been regarded as essential for the rapid formation of new memories and for consolidating newly acquired memories into the neocortex (Squire and Alvarez, 1995). Some mechanisms by which the hippocampus could quickly store new memories are known. The CA3 hippocampal subnetwork, thanks to its recurrent connectivity, can support attractor dynamics (Hopfield, 1982; Amit, 1989) to potentially store arbitrary input patterns (McNaughton and Morris, 1987; Treves and Rolls, 1992). In such a network, appropriately trained connections allow recalling of memories starting from noisy/corrupted versions (pattern completion). At the same time, the hippocampus needs to avoid confusion among similar memories (pattern separation). Pattern separation can be facilitated by orthogonalization, or decorrelation, of incoming cortical inputs in the dentate gyrus (Leutgeb et al., 2007). There is evidence that both pattern separation and completion are stronger in CA3 than in CA1 (Leutgeb et al., 2004; Vazdarjanova and Guzowski, 2004): this could be explained by the lower degree of recurrent connectivity in CA1, which could make it more sensitive to changes in entorhinal inputs.

Theories of systems memory consolidation

The findings on amnesic patients with hippocampal lesions gave rise to the "Standard model of systems consolidation" (McClelland et al., 1995; Squire, 2004). This theory proposes that novel memories are quickly encoded in the hippocampal and surrounding structures. The solidification of these memories is believed to rely on synaptic consolidation to bring about fast-paced (minutes to hours) changes in synaptic efficacy (Glanzman, 2010). Afterward, a process of systems consolidation kicks in, with a timescale that can vary from days to months, which shifts the responsibility of memory retention onto neocortical areas (Dudai et al., 2015). As a result, the neocortex can maintain its own accounts of the memory, and over time it becomes independent of the hippocampus. It is not known what process or condition triggers

[1]Important examples of declarative memory are "episodic" memory, i.e. the ability of recording and recollecting entire scenes from the past, and "semantic", i.e. related to general facts and acquired knowledge, not necessarily experienced (Tulving, 2002). In this thesis, I will always implicitly refer to declarative memory.

systems consolidation, but current hypotheses and recent evidence point to a central role of sleep and hippocampal reactivation (Klinzing et al., 2019).

The standard model of systems consolidation has been criticized for several reasons. The main reason is that the theory assumes that consolidation begins and ends just once for each memory. This disagrees with observations on memories that can become liable, and be modified, upon recalling them (Sara, 2000). [2] Another criticism of the standard model is the fact that the amnesia following hippocampal damage is not always temporally graded, or it is temporally graded only on certain tasks, and in most cases extends over decades (Nadel and Moscovitch, 1997). These observations support the idea that the hippocampus is necessary for retrieval, and gave rise to the "multiple trace theory" (Rosenbaum et al., 2001). This theory posits that the hippocampus is responsible for initial quick storage, acts as an "index" for distributed neocortical representations and is responsible for retrieval for as long as these memories exist (i.e., potentially up to a lifetime). This theory was later elaborated into the trace transformation theory (Winocur and Moscovitch, 2011) and is based on an additional critique of the standard model: memories initially stored in the hippocampus are not carbon-copied to the neocortex, but undergo a process of generalization. This hypothesis was supported by the finding that rats did not generalize fear conditioning across contexts the day after receiving a foot shock, but they did so after four weeks (Winocur et al., 2007). The trace transformation theory holds that the consolidation of memories in the neocortex involves a loss of time and place, and becomes more and more schematic, or fact-like semantic memory. This implies that details about past autobiographical events are not encoded in the neocortex, leading to generalization.

Another hypothesis, which is not mutually exclusive with the previous ones, is the "schema assimilation model" (Tse et al., 2007). This theory assumes that novel information is quickly consolidated into a previously established body of knowledge, i.e., a mental schema (Bartlett and Bartlett, 1932). In particular, a novel memory could be consolidated in less than 48 hours when animals were pre-trained on similar tasks (Tse et al., 2007). This theory is in agreement with human and animal lesion and psychology studies (Van Kesteren et al., 2012), and is reminiscent of transfer learning techniques in machine learning (Torrey and Shavlik, 2010). Nonetheless, this theory is in disagreement with the assumption that the neocortex is a slow learner (Takashima et al., 2009), and more research is needed to understand speed and mechanisms.

Throughout this thesis, I will discuss my PhD results in light of the existing systems consolidation hypotheses. I will describe my efforts to shed light on different aspects of systems memory consolidation by analyzing and modeling electrophysiological recordings from freely behaving rats. I will focus on the systems level of research by statistically characterizing the activity of populations of neurons and their underlying correlation network topologies. In particular, I will start by describing a network mechanism that allows the hippocampus to quickly provide informative outputs starting from novel, noisy inputs. I will then continue by describing a novel mechanism by which the hippocampus can interact with neocortical areas, which can be used for memory transfer. Finally, I will discuss results obtained from the analysis of a clear correlate of learning in the MTL. These results support the consolidation hypothesis and help to shed light on the neocortical memory dependence on the hippocampus, the different speeds of consolidation, and the existence of a schema. All these results derive from the study of spatial cognitive maps, which I introduce in the following section.

[2] This observation gave rise to the "reconsolidation" hypothesis (Dudai, 2012).

1.2 Spatial coding and cognitive maps

Moving towards an appetitive stimulus, or away from a threat, represent basic navigational behaviors that are fundamental to survival. Simpler forms of such behaviors are even hard-wired into organisms without a nervous system, such as bacteria or eukaryotes (Wadhwa and Berg, 2022; Reversat et al., 2020). Considering living beings with a nervous system, even the smallest nematodes can exhibit complex behaviors (de Bono and Maricq, 2005). In insects, simple approximations to complex computations allow for spectacular navigational behaviors. For example, foraging desert ants keep track of their own positions and manage to integrate their irregular outbound routes to then return home along straight inbound routes over distances that are thousands of times their body size (Müller and Wehner, 1988). In ants, as well as in bees, the predominant mechanism of navigation is path integration based on a skylight compass (Wehner et al., 1996). In the fruit fly *Drosophila*, researchers have discovered brain areas dedicated to the computation of allocentric heading, and the underlying network computations have been studied in depth (Seelig and Jayaraman, 2015; Kim et al., 2017). Such structures have been shown to actively guide *Drosophila* moment-to-moment behavior towards goals (Green et al., 2019). Mammals also have brain structures that keep track of their head direction (Taube et al., 1990). Although these structures are important for spatial awareness, the main area that guides navigation in mammals is the hippocampus, which is the focus of the following paragraphs.

Hippocampal role in spatial coding and navigation

Research has revealed that the hippocampus provides an extraordinarily detailed representation of an animal's location within an environment (Muller, 1996; Hartley et al., 2013). The first report of hippocampal cells which reliably fire only in specific portions of the environment appeared more than 50 years ago (O'Keefe and Dostrovsky, 1971) . Neighboring place cells fire at different locations within the environment; this ensures that, throughout the hippocampus, the entire environment is represented in the activity of the local cell population (O'Keefe and Nadel, 1978; Wilson and McNaughton, 1993). This led to the idea that within the hippocampus resides a "cognitive map" of space (O'Keefe and Nadel, 1978), a re-elaboration of a concept that was first described by Tolman (1948). Corroborating this idea, it was shown that by simultaneously recording many hippocampal cells it is possible to accurately decode the location of the animal (Wilson and McNaughton, 1993). Spatial information in the hippocampus is not only encoded but actively used for navigation: in fact, early studies on hippocampal lesioned rats showed poorer navigational performance (Morris et al., 1982; Eichenbaum et al., 1990). By now, there is enough evidence from a number of mammalian species that the hippocampus plays an important role in spatial representation, spatial memory, and navigation (Nadel, 1991; Rolls, 1999; Ekstrom et al., 2003; Sarel et al., 2017).

Beyond the hippocampus

Over the last few decades, cells in many brain areas outside the hippocampus have been described as bearing some levels of spatial information (Sauer et al., 2022). Examples include the entorhinal cortex (Fyhn et al., 2004), retrosplenial (Mao et al., 2017), parietal (Harvey et al., 2013), prefrontal (Fujisawa et al., 2008; Ito et al., 2015), and visual (Fiser et al., 2016). I will focus here on prefrontal and entorhinal cortices, which will be discussed in chapters 2 and 3.

After the hippocampus, the medial entorhinal cortex (MEC) is the most famous brain area that contains spatially tuned neurons (Hafting et al., 2005). These cells relay the spatial location of the animal within an environment with the same precision as hippocampal place cells, but rather than having one single place field, they present multiple fields. These fields are arranged on a regular triangular regular pattern that covers the entire environment explored by the animal; such cells were therefore named "grid cells" (Hafting et al., 2005). The spacing between grid fields varies along the dorsoventral axis of the MEC, and grid cells cluster into discrete functional modules (Stensola et al., 2012). Although first reported in superficial layers, grid cells were later reported also in deeper layers (Sargolini et al., 2006). Grid cells have been described as rigid and unchanging, providing a possible metric for spatial navigation (Moser and Moser, 2008). This claim was challenged by reports of distortion of the grid after geometric manipulations of the environment (Barry et al., 2007; Stensola et al., 2015) and the response of grid cells to non-spatial signals (Lipton et al., 2007; Aronov et al., 2017). Moreover, irregularity in peak-to-peak variability (Diehl et al., 2017; Dunn et al., 2017; Ismakov et al., 2017; Stemmler and Herz, 2017) and occasional failure of grid cells to discharge in a firing field (Low et al., 2014; Nagele et al., 2020) suggest that grid cells might code for more than only distances. It is also possible that simple environments gave us a simplified view of the functioning of grid cells: it has been suggested that hippocampal spatial maps will be much more complex in real-world navigation (Derdikman et al., 2009). Partial answers to these questions come from freely flying bats: it was recently reported that grid cells are less regular and only locally organized in 3D enclosures (Ginosar et al., 2021). Important to mention is the fact that most cells in MEC, even though not grid, carry spatial information (Diehl et al., 2017; Hardcastle et al., 2017), reminiscent of non-place cells in the hippocampus (Meshulam et al., 2017; Stefanini et al., 2020). Anatomically, the MEC neighbors the hippocampus and the two areas are heavily interconnected (Boccara et al., 2015). The deep layers of the entorhinal cortex receive inputs from the CA1 hippocampal area, whereas superficial layers of MEC project to the hippocampus (Witter, 2011). Based on this connectivity, it is plausible that some of the spatial properties of hippocampal place cells are determined by the activity of neurons in the entorhinal cortex (Latuske et al., 2017). In fact, suppressing entorhinal direct inputs to CA1 partially impaired place cell coding (Brun et al., 2008). This finding is in line with other studies that found effects on hippocampal spatial coding after alterations to MEC normal functioning (Rueckemann et al., 2016; Kanter et al., 2017). Some computational studies proposed that place cells derive from grid cells and other entorhinal inputs (Rolls et al., 2006; Solstad et al., 2006). Nonetheless, these ideas go against the finding that place fields are present several days before the emergence of grid firing patterns during development (Langston et al., 2010; Wills et al., 2010) and the hippocampus does not require grid cells input to form place fields (Koenig et al., 2011; Hales et al., 2014). Moreover, the regularity of grid cells disappears if inputs from place cells are removed (Bonnevie et al., 2013), and contextual information is exchanged among the two populations (Marozzi et al., 2015). Altogether, these studies suggest that entorhinal and hippocampal spatial codes are complementary (Bush et al., 2015).

The second area discussed here is the medial prefrontal cortex, which has been described as essential for spatial navigation in a recent review (Patai and Spiers, 2021). The fact that the prefrontal cortex houses neurons that encode aspects of space was first reported more than two decades ago (Jung et al., 1998). A clear report of spatially and trajectory modulated cells came from Fujisawa et al. (2008). A recent study found that spatial representations in mPFC form rapidly also in absence of task rules, and reappears when reexposed to the same environment (Sauer et al., 2022). Although spatially modulated, this area encodes space differently from the hippocampus: on a plus maze, many mPFC cells fired similarly on opposite arms, and

the distance to the goal was easier to decode than the actual 2D information (Kaefer et al., 2020). This is in line with reports on mPFC generalization (Xu and Südhof, 2013; Bernardi et al., 2020), and could be due to the mixed selectivity of mPFC cells (Rigotti et al., 2013), whose sparseness has been shown to determine generalization (Barak et al., 2013). Until now, however, it is not clear how spatial representations in prefrontal cortical areas are established. It is possible that spatial information in the mPFC is derived from that of the hippocampus (Zielinski et al., 2019). Anatomically, this could happen through the direct projections from the ventral CA1 to mPFC (Hoover and Vertes, 2007). Inactivating this direct projection leads to deficits in the spatial coding of medial prefrontal neurons and in the performance of a spatial working memory task (Spellman et al., 2015). Moreover, the propensity of mPFC cells to code for space was found to be a function of their position along the dorsoventral axis of mPFC, which runs opposite to the density of hippocampal inputs (Sauer et al., 2022). This is in line with a report showing that encoding spatial information in several neocortical regions depends on an intact hippocampus (Esteves et al., 2021). Nonetheless, interactions among the two areas are not unidirectional, as it is known that mPFC has indirect projections to the hippocampus via the thalamic nucleus reuniens (Vertes, 2006) and the perirhinal and lateral entorhinal cortices (Delatour and Witter, 2002). This pathway influences hippocampal spatial coding in several ways: inactivating the medial prefrontal cortex or the indirect nucleus reuniens projections leads to a decrease in hippocampal place cell firing variability (Hok et al., 2013), rule-based object selectivity (Navawongse and Eichenbaum, 2013), and trajectory-dependent firing (Ito et al., 2015).

Spatial coding is part of a larger role in memory

Many findings throughout the years have claimed that the hippocampus is a brain area dedicated to spatial navigation via path integration (McNaughton et al., 2006; Moser and Moser, 2008; Hartley et al., 2013). This view contrasts with Tolman's idea of cognitive maps, which were envisioned as extending to general cognition (Tolman, 1948). Moreover, the purely spatial view is difficult to reconcile with the evidence that damage to the hippocampus results in global amnesia including both spatial and nonspatial domains (Eichenbaum et al., 1999; Squire, 2004). Additionally, place cells have been shown to encode much more than spatial coordinates; for example, stimulus events and behavioral actions (Wood et al., 2000), time (Kraus et al., 2013), sound frequency (Sakurai, 2002; Aronov et al., 2017), and other abstract relationships (Bellmund et al., 2018). Moreover, path integration seems to be a bad navigational strategy (Huth, 2013) and there is no clear evidence that the hippocampus supports path-integration computations (Navratilova and Mcnaughton, 2014). Finally, it has been reported that the hippocampus is mostly active during the initial planning of a route and not during its execution (Spiers and Maguire, 2006). The two views can be reconciled by accepting that the main contribution of the hippocampus towards navigation comes from memory (Eichenbaum et al., 1999; Eichenbaum and Cohen, 2014; Eichenbaum, 2017b). In fact, studies on navigational strategies recognized that the ability to find one's way relies on a combination of memories for the spatial layout of the environment, routes taken, and the origin and destination of a journey (Dudchenko, 2010; Wolbers and Hegarty, 2010). According to one line of research, the hippocampus supports declarative memory by binding experiences and linking them via their common elements, thereby creating a "memory space" (Eichenbaum and Cohen, 2014). This idea, first strongly advocated by Cohen and Eichenbaum (1993), has been expanded into computational models (Hasselmo, 2011) and theoretical frameworks (Buzsáki and Moser, 2013; Whittington et al., 2020). In particular, Buzsáki and Moser (2013) compared integrating multiple routes into cognitive maps to the integration of episodic memories into

memory networks. In other words, the creation of a cognitive map of space would be the same as, or at the very least a prerequisite to, creating a cognitive map of episodic memories.

Some of the arguments raised for the hippocampus could be extrapolated to the MEC, given the fact that this area can also encode non-spatial dimensions (Aronov et al., 2017; Behrens et al., 2018). Nonetheless, MEC code could more easily support path integration (Hardcastle et al., 2015). The hard fact remains that the entorhinal cortex is the main gateway between the hippocampus and most of the neocortical areas (Witter, 2011), so it is reasonable to assume that other types of memory (not only spatial) are processed in the MEC. Recent lines of research claim that the MEC supports structural abstraction, and in concert with the hippocampus allows for the storage of declarative, non-spatial information (Garvert et al., 2017; Behrens et al., 2018; Whittington et al., 2020).

The neocortex has only recently been studied from the spatial coding perspective. Different neocortical areas are thought to be responsible for other types of information and computations (Kandel, 2013). For example, prefrontal areas have been attributed to the most disparate roles, the major ones being working memory, decision making, and abstraction/generalization (Funahashi and Kubota, 1994; Miller, 2000; Xu and Südhof, 2013). Nonetheless, recent experimental evidence suggests that spatial coding is a general principle of cortical computation (Esteves et al., 2021), including prefrontal cortex (Sauer et al., 2022). The natural question arises as to whether distributed neocortical spatial coding underlines general principles of memory.

In the following chapters, I will present my PhD results and argue that the study of spatial representations and computations within and across cortical areas can shed light on general principles of memory storage.

2

Hippocampal encoding of novel memories

2.1 Aims

Theories and studies of memory consolidation agree on the fact that the hippocampus quickly encodes novel memories (Squire, 2004). As described in the introduction, this process relies on the formation of cognitive maps, which organize episodic memories in space and time (Buzsáki and Moser, 2013). Spatial cognitive maps are the most widely studied, and how quickly a cognitive map is established when an animal enters a novel environment is of primary relevance to the theory of memory. From a single-cell perspective, hippocampal cells can form stable spatial fields within minutes during a novel exploration (Wilson and McNaughton, 1993; Leutgeb et al., 2004; Frank et al., 2004). These newly acquired place fields are less informative and less reliable than the ones measured during a familiar exploration (Brun et al., 2008). This fact has been reported for hippocampal areas CA1 and CA3 (Frank et al., 2004; Leutgeb et al., 2004), and also in MEC (Barry et al., 2007). The hippocampal subarea CA1 is of particular interest for two reasons. On the one hand, it represents the main output of the hippocampal CA areas, so it is reasonable to assume that CA1 outputs ought to be as informative and easy to read as possible. On the other end, the main afferents to CA1 (i.e., CA3 and superficial MEC) have very different spatial properties (Moser et al., 2008). Given the fact that these inputs are noisier in novel than familiar explorations, there might be mechanisms in place for combating this noise and providing an output that is easily readable from downstream areas.

The aim of this chapter is to characterize the first stages of the formation of a spatial cognitive map beyond the single-cell level. In particular, the initial aim is to infer and characterize interaction patterns among CA1 cell pairs during the exploration of familiar and novel environments. Afterward, the aim is to study in a theoretical model the impact of the structure of these interaction networks on the encoding of spatial information from a population perspective, and the potential availability of information to downstream areas.

This is a study in collaboration with Jozsef Csicsvari, Gasper Tkacik, and Cristina Savin.

A previous version of this paper was posted on biorxiv on Septermber 29th, 2021, and can be retrieved here: https://doi.org/10.1101/2021.09.28.460602

The paper has been submitted to a journal for evaluation and has undergone revision; the version presented here includes modifications and additions suggested by the reviewers.

Author contributions: C.S. and G.T. designed research; M.N., J.C., G.T. and C.S. performed research; M.N. and C.S. analyzed data; M.N., G.T. and C.S. performed theoretical analyses; M.N., G.T., and C.S. wrote the paper.

The structure of hippocampal CA1 interactions optimizes spatial coding across experience

Michele Nardin[1], Jozsef Csicsvari[1], Gašper Tkačik[1], Cristina Savin[2,@]

[1] *Institute of Science and Technology Austria (ISTA), Klosterneuburg, 3400, Austria*
[2] *New York University, New York NY-10003, USA*
[@] *corresponding author: cs5360@nyu.edu*

Although much is known about how single neurons in the hippocampus represent an animal's position, how cell-cell interactions contribute to spatial coding is less well understood. Using a novel statistical estimator and theoretical modeling, both developed in the framework of maximum entropy models, we reveal highly structured CA1 cell-to-cell interactions whose statistics depend on whether the animal is in a familiar or novel environment. In both conditions the circuit interactions optimize the encoding of spatial information, but for regimes that differ in the signal-to-noise ratio of their spatial inputs. Moreover, the topology of the interactions facilitates linear decodability, making the information easy to read out by downstream circuits. These findings suggest that the efficient coding hypothesis is not only applicable to individual neuron properties in the sensory periphery, but also to neural interactions in the central brain.

2.2 Introduction

The dual role of the hippocampal formation in memory (Scoville and Milner, 1957; Eichenbaum, 2000) and spatial navigation (O'Keefe and Dostrovsky, 1971; Morris et al., 1982) is reflected in two distinct views on hippocampal coding: the place field view (Moser and Paulsen, 2001; McNaughton et al., 2006) that reduces the encoding of spatial information to tuning properties of individual neurons, and the ensemble view (Harris et al., 2003; Harris, 2005) that focuses on subsets of units that are co-activated together as the substrate for memory (Hopfield, 1982). Recent results blur the line between the single cell and the population perspective (Stefanini et al., 2020), revealing that properties of individual neurons only partially explain the circuit's contribution to spatial coding. Interactions between neurons shape collective hippocampal activity (Meshulam et al., 2017). Moreover, disrupting correlations between neurons leads to decreased decoding accuracy, in particular in CA1 (Stefanini et al., 2020). It remains unclear how experience shapes the organization of cell-to-cell interactions and what effects such changes may have on the encoding of spatial information at the level of CA1 population activity.

Experience affects the properties of single cells in many ways. While reliable position-dependent spiking is detectable after a few minutes during the very first exposure to a novel environment (Wilson and McNaughton, 1993; Leutgeb et al., 2004), the responses to a familiar environment show several systematic differences, including a reduction in overall firing, sharpening of tuning functions and sparsification of responses (Karlsson and Frank, 2008). In parallel, CA1 inhibition is weak in novel environments, transiently opening the gate for circuit reorganization via plasticity (Arriaga and Han, 2019), but it subsequently increases with experience (Nitz and McNaughton, 2004; Arriaga and Han, 2019; Geiller et al., 2020). From the perspective

of the local circuit, the main afferents to CA1 (MEC and CA3) are initially noisier (Cohen et al., 2017; Pedrosa and Clopath, 2020) and have weaker spatial tuning, which improves with familiarity (Leutgeb et al., 2004; Hafting et al., 2005; Barry et al., 2012). Since CA1 needs both inputs for detailed spatial representation (Brun et al., 2002, 2008), these results suggest that the CA1 circuit is potentially in a different dynamic regime in novel versus familiar environments, with distinct local circuit interactions and population coding properties.

Correlations among pairs of hippocampal neurons arise as a result of two effects: mechanisms that give rise to spatial tuning (i.e. *signal* correlations), and tuning-independent internal circuit dynamics (which shape *noise* correlations). Since they reflect local circuit interactions, noise correlations should depend on changes in input statistics, and be reorganized by experience. From a neural coding perspective, the structure of neural correlations can radically affect the amount of information that a population carries about stimuli (here, the animal's position) and the complexity of the readout (Cohen and Kohn, 2011; Averbeck et al., 2006). While noise correlations are generally considered to be an obstacle to optimal information coding and transfer, especially in sensory areas (Kanitscheider et al., 2015; Rumyantsev et al., 2020), there are scenarios where they can improve the quality of the overall population output (Panzeri et al., 1999; Tkačik et al., 2010; da Silveira and Berry, 2014; Brinkman et al., 2016), which might be relevant for the hippocampus.

Unlike sensory areas, where stimulus repeats make the estimation of noise correlations relatively straightforward, measuring circuit interactions and their contribution to spatial coding in the hippocampus is fraught with technical difficulties. In a two dimensional environment, the lack of stimulus repeats renders traditional approaches for estimating noise correlations inapplicable. Moreover, well documented circuit level oscillations (Colgin, 2013; Fries, 2009) act as global sources of co-modulation that obscure the fine structure of pairwise neural co-variability. The key challenge is to partition total neural covariability into an explainable component, driven by position, oscillations and other global signals, and unexplained, or 'excess' correlations, which capture local interactions.

Here we take advantage of the maximum entropy framework to develop a new statistical test for detecting excess correlations in the absence of stimulus repeats and explore their significance for the encoding of spatial information in CA1. Our method allows us to robustly detect network interactions by comparing hippocampal responses against a maximum entropy null model (Savin and Tkačik, 2017) that optimally captures the cells' place preference and population synchrony (Engel et al., 2001). When applied to CA1 tetrode recordings from rats during open field exploration in familiar and novel environments, our analysis detected structured excess correlations preferentially between principal cells with similar place selectivity and arranged into networks with high clustering coefficients (Watts and Strogatz, 1998). These highly structured excess correlations optimize the encoding of spatial information and facilitate its downstream readout in both the familiar and novel environment, with differences reflecting the different signal-to-noise ratio of spatial inputs in both environments. Taken together, our results suggest that CA1 local circuitry readjusts to changes in its inputs so as to improve population-level stimulus representation, in line with efficient coding predictions (Tkačik et al., 2010).

2.3 Results

Detecting interacting cells To investigate functional connectivity between CA1 neurons and its role in spatial information coding, we devised a procedure to infer cell-cell interactions

from simultaneous tetrode recordings of hundreds of isolated units in dorsal hippocampus of behaving rats.

Our approach starts by constructing a null model for population responses that exactly accounts for the measured spatial selectivity of each recorded neuron as well as for the moment-to-moment measured global neural synchrony, but is otherwise maximally unstructured (Fig. 2.1A, S2.1A−D). This null model is formally a maximum entropy model, which includes a rigorous propagation of uncertainty about the inferred single-cell place tuning (see Methods). Surrogate neural rasters can be sampled from the null model (Fig. 2.1A Left) and its statistics directly compared to that of real data (Savin and Tkačik, 2017)(Fig. S2.1E). For every cell pair, the model predicts the expected distribution of pairwise correlations against which the measured total correlation for that pair can be tested for significance (Fig. 2.1A Right); we report as "excess correlation" w the (normalized) amount of total correlation that is not explained by the null model. We declare cell pairs with significant excess correlations to be "interacting," likely due to underlying recurrent neural circuitry. Since our approach explicitly discounts for correlations arising from overlapping place fields and sources of global modulation (e.g, due to locking to the underlying brain oscillations or influence of behavioral covariates such as running velocity), it differs from previous attempts to use total correlations to probe the intrinsic network mechanisms (Gava et al., 2021).

We validated our detection method by constructing a synthetic dataset of spiking CA1 neurons whose responses were modulated by the position of an artificial agent and by an assumed network of interactions (Fig. 2.1B Left, see Methods). We ensured that the synthetic data matched overall firing rates, the synchrony, and the highly irregular occupancy observed in a real 20-minute exploration session(Fig. 2.1B middle). Interactions identified by our method strongly overlap with the ground truth, as measured by the area under the receiver operating characteristic (Fig. 2.1B right). The inferred excess correlations were also well aligned with the ground truth interaction strengths (Fig. S2.2A). We did not find any tendency of cells that are more (or less) similarly tuned to show higher (or lower) inferred w_{ij}s (Fig. S2.2B). Inference quality did not change when using weaker or stronger inputs (Fig. S2.2D). Introducing instability and additional noise in the tuning of single cells (Fig. S2.2F) or high-firing inhibitory-like cells (Fig. S2.2G) did not alter our ability to detect interacting pairs. In all of the cases we tested, our method outperformed a standard noise correlation detection method (Fig. S2.2E).

We next analyzed CA1 tetrode recordings of six rats exploring familiar and novel 2D environments separated by a short period of rest (Fig. 2.2A,D) (Kaefer et al., 2019; Stella et al., 2019). The two environments were of similar size (Fig. S2.3A,B) and differed in color and geometry. As expected, spatial firing fields were not related across the two environments (Fig. S2.3C). Our null model did not differ in terms of marginal log-likelihood or pseudo-R2 across environments (Fig. S2.3D,E). Putative units were filtered by using several clustering quality measures, based on the Mahalonobis distance or the inter-spike interval, (Harris et al., 2000; Schmitzer-Torbert et al., 2005; Hill et al., 2011) to ensure that they were well isolated (Fig. S2.3F, see Methods). We retained only cells active in both environments (> 0.25 spike/sec, Fig. S2.3G) (Karlsson and Frank, 2008). Considering only pairs of cells that were simultaneously recorded on different tetrodes, our final dataset includes a total of 9511 excitatory-excitatory (EE), 7848 excitatory-inhibitory (EI), and 1612 inhibitory-inhibitory (II) pairs (see Methods for animal-by-animal details). We detected both positive and negative excess correlations among cell pairs (Fig. 2.2B,C). Interestingly, cell pairs with negative excess correlations can have positive total correlations (Fig. 2.2C), illustrating the idea that total correlations are not necessarily a good predictor of local circuit interactions. Are excess

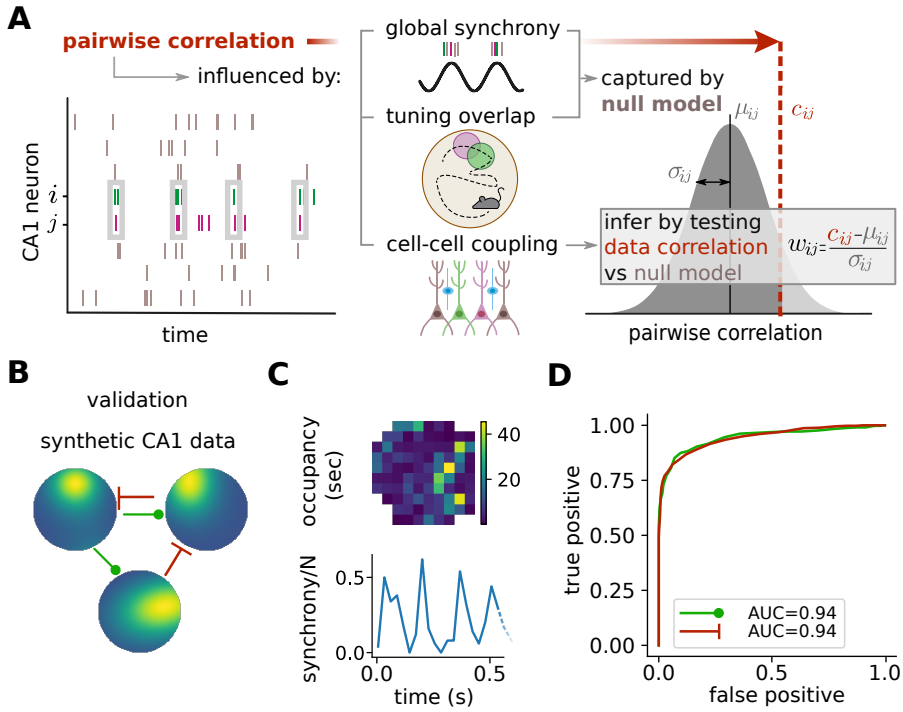

Fig. 2.1. Detecting network interactions among hippocampal CA1 cells during naturalistic behavior. **(A)** Method schematic. A null model for population responses takes into account the inferred place field tuning of each cell and the moment-to-moment global synchrony but is otherwise maximally unstructured. For each cell pair, this model predicts a null distribution for (total) pairwise correlation (gray distribution), which is compared to the correlation estimate from data (dashed red line). The normalized discrepancy between the data correlation c_{ij} and the null model expectation μ_{ij} for a pair of neurons (i, j) is referred to as "excess correlation", w_{ij}, and serves as a proxy for direct cell-cell interaction. **(B)** Method validation on synthetic data. Detection accuracy is assessed using simulated data with known positive (green) or negative (red) interactions **(C)** Synthetic data matches a real 20-minute exploration session with respect to spatial occupancy (top) and observed synchrony indices (bottom – in this plot represented as proportion of cells active in a ∼500ms example)). **(D)** Receiver-operator characteristic (ROC) shows the probability of correctly detecting positive (green) and negative (red) interactions for different detection thresholds.

correlations explained by common locking to underlying theta oscillations? We measured the similarity of cell pairs theta-locking histograms and scatter plotted against inferred w_{ij}s, and found no significant relation in the EE subnetwork (Fig. S2.4A-B, Left). The correlation was small but significant for EI cell pairs (Fig. S2.4, middle), and was strongest within the II subnetwork (Fig. S2.4, right). We checked whether pairwise excess correlations could be explained by tuning to variables not explicitly accounted for by our null model. We found a small ($r \sim 0.1$) significant correlation between w_{ij} and heading-tuning similarity (Fig. S2.4C-D, left), and no significant relation with speed-tuning similarity (Fig. S2.4C-D, right).

Interaction networks in familiar and novel environments What is the structure of the inferred excess correlations? We set the threshold to declare a cell pair as interacting at $|w| > 4.5$ (corresponding to a strict Bonferroni-correction - see Methods.).We first report a generally sparse interaction network in the excitatory-excitatory (EE) subnetwork, with $\sim 5\%$ in familiar and $\sim 8\%$ in novel environment of analyzed pairs showing significant interaction (Fig. 2.2E, Left).The fraction of interactions is larger among excitatory-inhibitory (EI) cell pairs (Fig. 2.2E, Middle), where, as expected, negative interactions dominate; the fraction is highest at $\sim 30\%$ among positive interactions in the inhibitory-inhibitory (II) subnetwork (Fig. 2.2E, Right).

We next focused on interaction changes induced by a switch from the familiar to a novel environment (Fig. 2.2D). We observed a significant increase in EE interactions, possibly arising due to decreased inhibition during novelty (Arriaga and Han, 2019; Geiller et al., 2020), which is thought to enhance learning and promote plasticity (Li et al., 2003; Whitlock et al., 2006; Cacucci et al., 2007). We indeed found putative inhibitory cells to be less synchronous and slightly less active in novel environments (Fig. S2.3H,J), in line with previous findings (Nitz and McNaughton, 2004), while excitatory neurons were more synchronous but did not differ in terms of their average firing rates (Fig. S2.3G,I). Total correlations were generally stronger in novel environments (Fig. S2.3K). Circuit modifications during spatial learning are believed to originate in altered spike transmission among connected excitatory and inhibitory neurons (Dupret et al., 2013; McKenzie et al., 2021). Consistent with this view, we found an increase in positive EI interactions, while their negative counterpart remained unchanged. This increase could not be attributed to increased reliability of monosynaptic EI connections (Fig. S2.4E), especially since cell pairs on the same tetrode were excluded from this analysis (Csicsvari et al., 1998). We did not observe significant changes in the number of II interactions.

How conserved are individual network interactions across consecutive environments? The largest overlap in detected interactions was found in the II subnetwork, where 77.5% of interactions were preserved.EI interactions showed less overlap (31.1%);the overlap was weakest (16.8%) in the EE subnetwork.All reported overlaps were statistically significant under a permutation test (1000 random shuffles of cell labels; $p < 10^{-3}$ for all subnetworks). Significance was confirmed by comparing the Jaccard similarity of the adjacency matrices of familiar and novel subnetworks against the null distributions constructed from random graphs with the same numbers of vertices and edges (1000 Erdos-Renyi graphs; $p < 10^{-3}$ for II and EI subnetworks, $p = 0.009$ for EE).

The similarity of interaction networks across the two environments extends beyond the binary presence / absence of significant interactions. Figure 2.2B compares the strength of excess correlations, w, in the familiar vs. novel environment for EE, EI, and II cell pairs. For all subnetworks, w are significantly correlated across the two environments, with the reported correlation strength related to the network overlap (Fig. 2.2A). Taken together, these findings corroborate the idea that hippocampal remapping across environments is not random at the level of cell-cell interactions.

Because spatial information is encoded predominantly by pyramidal cells (Skaggs et al., 1993; Frank et al., 2001), we analyzed the EE subnetwork in further detail (Fig. 2.2C). Our key statistical observation is shown in Fig. 2.2D: interaction probability increases nonlinearly with place field overlap for positive interactions, and is roughly constant for negative interactions. In the novel environment, the excitatory interaction probability increases ~ 3-fold over the observed range of place field overlap. In the familiar environment, the modulation with place field overlap is less pronounced, possibly indicating a shift towards a more decorrelated representation of space (Karlsson and Frank, 2008). Excluding the first few minutes of exploration, known for presenting unstable place fields, did not change our findings (Fig. S 2.5A,B). Our observation in Fig. 2.2D, which was averaged across the entire dataset, is consistent on a animal-by-animal basis (Fig. S 2.5C). Finally, we report that inference is not affected by theta phase precession (Fig. S 2.5D).

We further characterized the topology of familiar and novel excitatory networks. The number of interactions that a neuron engages in (its node degree) appears to be log-normally distributed in both environments, with clustering coefficients that are significantly higher than expected from

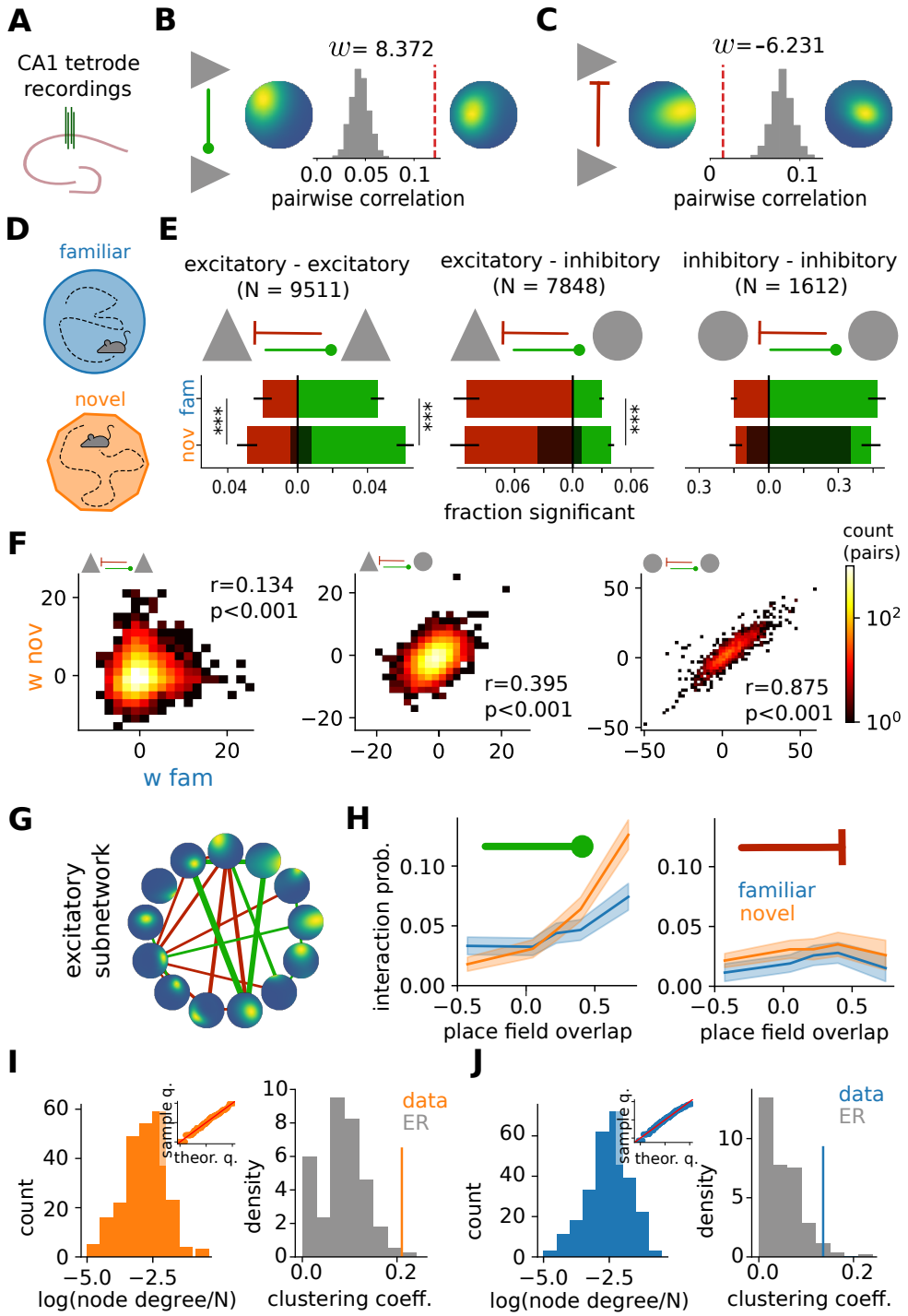

Fig. 2.2. Network interactions in familiar and novel environments. **(A)** Neural activity was recorded using tetrodes implanted in the dorsal CA1 area of the hippocampus. **(B,C)** Example pairs of pyramidal cells simultaneously recorded during free exploration of a familiar environment with significant positive (B) and negative (C) excess correlation w (gray histogram – distribution of correlation coefficients derived from the null model; red dashed line – measured raw pairwise correlation). **(D)** Experimental paradigm. 6 Animals explored a familiar environment, then rested in a sleep box (rest data not used), after which they explored a novel environment (20–40 minutes for each condition). Each animal contributed to one experimental session. **(E)** Summary of cell-cell interaction results for different cell types (triangle – pyramidal cell, circle – putative interneuron), positive (green) and negative (red) excess correlations, for both the familiar (top row, blue) and the novel (bottom row, orange) environment (stars – significant difference under binomial test at $p < 0.001$; absence of stars means $p > 0.05$). Shaded regions mark the fraction of interactions detected in the familiar environment that remain in the novel environment. Error bars indicate the standard error of the mean across N=6 animals. **(F)** Paired comparison (colormap – binned pair count) between excess correlations w_{ij} detected in familiar vs. novel environment for each cell-pair within EE (left), EI (middle) and II (right) sub-networks. *Continues on next page.*

Fig. 2.2. **(G)** Example of an estimated excitatory subnetwork. Circles show the place field selectivity of each neuron, with edges showing significant cell-cell interactions (green – positive; red – negative excess correlations); line thickness corresponds to interaction strength. **(H)** Left: interaction probability in the excitatory subnetwork increases with place field overlap ("place field overlap", measured as the Pearson correlation of 2D firing rate maps) for positive interactions (blue – familiar environment; orange – novel environment; shaded area – 99th percentile confidence interval for the mean). Right: analogous plot for negative interactions. **(I,J)** Left: distribution of $\log$ node-degree of E cells normalized by the total number of E cells in each session, for the novel (I) and familiar (J) environment. Inset: quantile-quantile plot comparing this distribution to the normal expectation. Right: excitatory subnetwork has a significantly higher clustering coefficient (orange/blue line – data) compared to the expected distribution for an Erdos-Renyi (ER) network with a matched connection density.

matched independently randomly connected (Erdos-Renyi) graphs (Fig. 2.2E,F). This effect was more pronounced during novelty (Fig. S2.6A). Accordingly, interacting excitatory triplets were over-represented, more strongly so in the novel environment (Fig. S2.6C). Finally, we found a linear relationship between the log-number of nodes and the shortest path length (i.e. the minimal distance between randomly chosen node pairs, Fig. S2.6B), which is a defining feature of small-world networks (Watts and Strogatz, 1998). Taken together, these analyses point to a highly non-random interaction structure, especially in novel environments. We described a clear cell-to-cell interaction rule (Fig. 2.2D) and topological fingerprints of the entire network. In the next sections we will explore the advantages offered by such connectivity and the underlying principles that explain both pairwise and network observations.

Effects of network interactions on spatial coding To explore how the network structure affects spatial information encoding at the population level, we constructed a statistical model of interacting excitatory cells responding to spatial inputs (Fig. 2.3A). Our model, a version of pairwise-coupled, stimulus-driven maximum entropy distribution over binary spiking units (see Methods, (Granot-Atedgi et al., 2013)) allows us to vary cell-cell excess correlations (to study the effect of network topology and interaction strength) as well as the strength of the spatial inputs (to study the effect of novel vs. familiar environment), while maintaining a fixed average firing rate for the population. For tractability, we simulated populations of 50 place cells (Fig. 2.3), and extended some of the analyses to larger networks (Fig. S2.11). We used this model to assess spatial coding at the population level as a function of the structure of network interactions.

We contrasted spatial coding in two networks that had the same spatial inputs and fixed average firing rates, but different excess correlation patterns. Interactions in the "structured" network followed the relationship between place field overlap and excess correlation w observed in real data; interactions in the "random" network were drawn from the same data-derived distribution for w, but independently of the overlap in (input) tuning (Fig. 2.3A). For each of the two choices, we further simulated the effects of familiar vs. novel environment by adjusting the strength of the feed-forward spatial input, in analogy to previous experimental results (Cohen et al., 2017; Leutgeb et al., 2004; Hafting et al., 2005; Barry et al., 2012; Brun et al., 2002, 2008). In our model, higher input strength corresponds to higher signal-to-noise ratio for the spatial drive, which is why we refer to this parameter as "input quality." We adjusted the input quality to best resemble the data in terms of various marginal statistics (spatial information, place field sparsity, peak-over-mean firing values; see Methods and Figs. S2.7,S2.8) in familiar and novel environments.

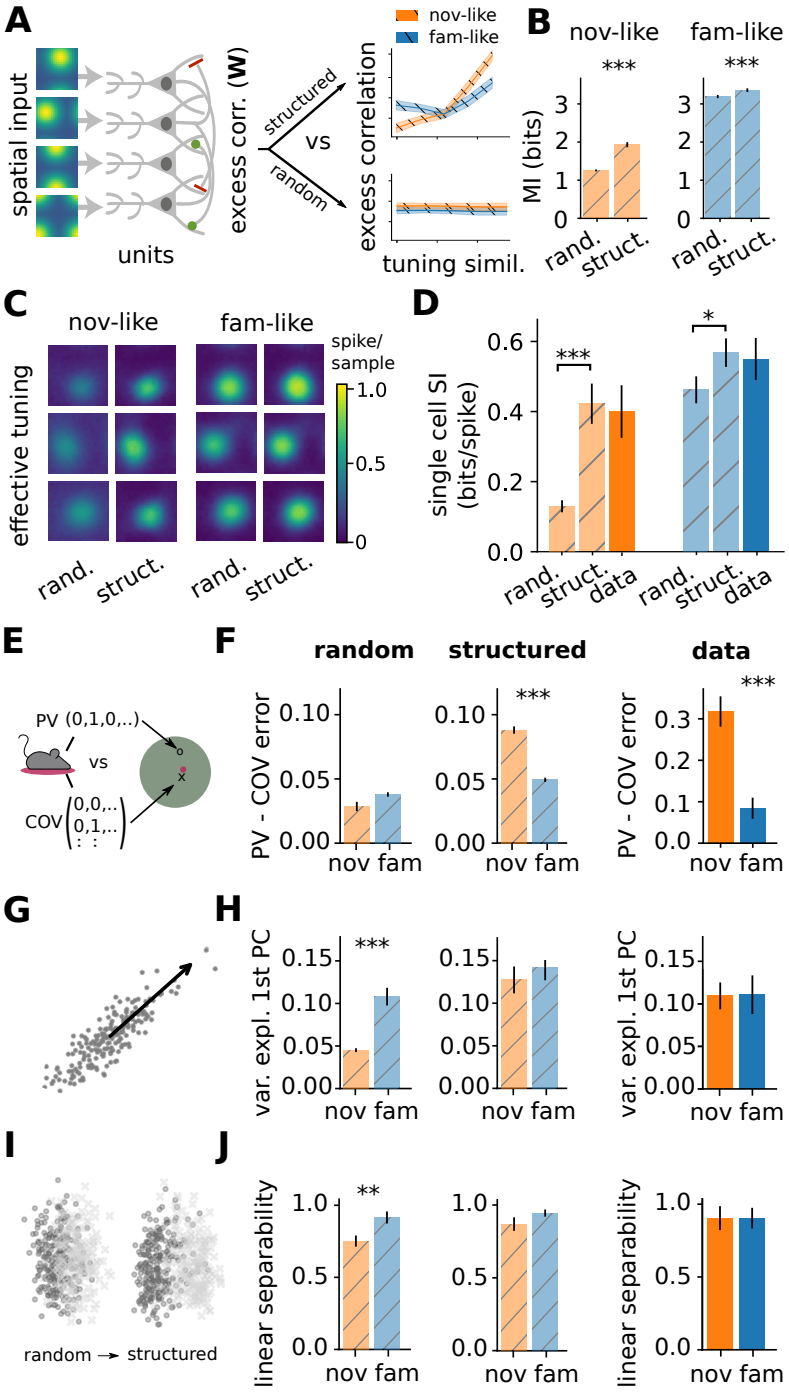

Fig. 2.3. Effects of network interactions on spatial encoding. **(A)** A schematic of the circuit model with variable excess correlations and fixed population firing rate (see Methods). Two connectivity patterns are compared: "structured" (mimicking the inferred excess correlation vs. tuning similarity relationship) vs. "random". **(B)** Estimated spatial information (MI; error bar – 99-th percentile CI for the mean) using structured and random interactions, in the novel-like and familiar-like scenario (see text). Structured interactions significantly increase the spatial information ($p < 0.001$ (***) or $p < 0.01$ (**) under a non-parametric Mann–Whitney U-test). **(C)** Effective single cell tuning for random vs. structured models in the novel-like (orange; left) and familiar-like (blue; right) scenarios. **(D)** Average single place cells spatial information (I_{spike}) for random vs. structured interactions (dashed bars) and data (solid bar) in the novel-like (orange; left) and familiar-like (blue; right) scenarios **(E)** The performance of a simple population vector ("PV") decoder is compared with the performance of a decoder which takes into account the co-variability of neurons ("COV"). **(F)** Improvement in decoding performance, measured as error of PV decoder minus error of COV decoder, evaluated on $4 \cdot 10^4$ samples for random (left), data-like structure (center) and data (right). Random 50% of data was used for train/test. Decoding error measured in spatial bins. Error bars and significance tests as in B. *Continues on next page.*

Fig. 2.3. **(G)** Fraction of variance explained by the first principal component for responses to random pairs of stimuli, used to measure the effect of structured interactions on population responses to different stimuli. **(H)** Fraction of variance explained by the first principal component of population vectors for 10^3 random pairs of locations in the maze for random (left) vs. data-like structure (center) vs. data (right). The fraction is unchanged between the novel and familiar environments on structured network and on real data, but differs significantly on the random network (error bars and significance tests as in B). **(I)** Linear separability of responses to pairs of stimuli. The schematic depicts the 2-dim PCA projection of the responses to a random pair of stimuli light grey vs dark grey when using random (left) and structured (right) excess-correlations. **(J)** Linear SVM classification accuracy of the responses to random pairs of stimuli trained on 1000 pairs of same vs. different positions for random (left) vs. data-like structure (center) vs. data (right).

We quantified the coding performance of our networks by estimating the mutual information between population activity and spatial position and by estimating the average decoding error. As expected, higher input quality in the familiar environment leads to overall higher information values (Fig. 2.3B) and lower decoder error (Fig. S2.9B). Less trivial are the effects of network connectivity: in both environments, structured (data-like) interactions significantly outperform random ones, with larger improvements seen in the novel environment. This suggests that network interactions among hippocampal cells adjust to maintain a high-fidelity spatial representation even when they receive lower quality, noisy inputs. Improved decodability offered by structured interactions, probed with different decoders, persists in larger networks with 100, 200, and 500 cells (Fig. S2.11). These population benefits are also reflected in a general improvement of single-cell spatial tuning (Fig. 2.3C), and single neuron spatial information (Fig. 2.3D)(Souza et al., 2018).

Do the structured interactions better predict other population-level aspects of the real hippocampal code? We assessed the importance of pairwise (co-firing) statistics for the decoding performance, highlighted by previous work (Stefanini et al., 2020) (Fig. 2.3E). For the random network, the decoding performance improvement with co-firing statistics relative to population-vector decoding is small and comparable in the novel vs. familiar environment. In contrast, for the structured network and data, the improvement is significantly larger in the novel environment (Fig. 2.3F); the improvement reaches three-fold in novel relative to the familiar environment on real data, perhaps due to the larger population size. We then contrasted the population responses to different stimuli. For each pair of locations, we measured the fraction of variance explained by the first principal component of the corresponding population activity patterns (Fig. 2.3G,H). For the random network in the novel environment, this fraction is two-fold lower than in the familiar environment. In contrast, for the structured network and data, the fraction is about 0.1, regardless of the environment. Structured interactions appear to organize neural responses in the novel environment so that the code maintains a collective correlated response even when the input quality is weak. This effect might aid downstream areas to better differentiate responses to different stimuli. Therefore, we assessed the linear separability of spatial positions based on neural population responses, as a measure of whether information is readily available to downstream brain areas (Fig. 2.3I,J). For the random network, the performance of a linear classifier trained to discriminate random positions is significantly worse in the novel environment. In contrast, the performance is restored to a high value (~ 0.9) irrespective of the environment by data-like interactions in the structured model, matching observations on real data (see Fig. S2.10 for separability of positions as a function of their mutual distance).

Taken together, our results suggest an important coding role for the interaction patterns inferred in Fig. 2.2D and the corresponding "structured" networks explored in Fig. 2.3. In comparison

to the random network, the data-like, structured network (i) encodes more information about position even when the input is of low quality; (ii) this information can be retrieved by utilizing co-firing statistics of multiple cells; (iii) selected collective statistics of place cell activity remain are preserved under change of environment. Consistent conclusions hold for the comparison between the data-like, structured network and an uncoupled population (Fig. S2.9).

CA1 interactions match predictions of an optimal coding model While Figure 2.3 suggests that interactions between cells self-organize to improve spatial information coding relative to a random or an unconnected (Fig. S2.9) network, it is not clear whether the observed organization is in any sense optimal. To address this question, we numerically optimized cell-cell interactions among a population of place cells, so as to maximize the mutual information between the population activity and spatial position (Fig. 2.4A). In essence, this amounts to finding "efficient coding" solutions for network structure given inputs to individual cells that are correlated due to place field overlaps (Tkačik et al., 2010). As before, an important control parameter is the overall magnitude (quality) of the input drive, h, which we now vary parametrically. Resource constraints were simulated by constraining the optimization to keep the average population firing rate constant and the possible couplings bounded, $|W_{ij}| \leq w_{max} = 1$ (see Methods).

As the input quality increases, the information gain due to optimal interactions decreases, indicating that optimization benefits novel environments (with noisy spatial inputs) more than familiar environments (with reliable spatial inputs) (Fig. 2.4B). We quantified performance of data-like interaction (Fig. 2.3) against optimized ones in these small networks. We found a sizable improvement (avg. $\sim 50\%$ of predicted optimal, with peaks of $> 90\%$) over null interactions both in familiar and novel environments (Fig. 2.4B). Optimal interactions improve the spatial code by reducing the entropy of the stimulus-dependent population responses ("noise entropy") while largely preserving the total output entropy (Fig. S2.12). We further find that an overlap in tuning similarity between two cells correlates with optimal pairwise interaction between them when input quality is low, but this correlation grows weaker with increasing input quality (Fig. 2.4C), consistent with theoretical expectations (Tkačik et al., 2010).

Does optimization predict a clear relationship between the tuning similarity and interaction strength for pairs of cells? Figure 2.4D shows two such relationships, for high and low input quality, predicted *ab initio* by maximizing spatial information. The optimal relationships closely resemble two analogous curves, for the familiar and novel environment, inferred from data (Fig. 2.4E). A similar resemblance is not observed if one maximizes spatial information carried by individual cells (Fig. S2.13), highlighting the importance of information coding at the population, not individual-cell, level.

As an alternative comparison to experiments, we studied the proportion of optimized couplings that reached maximal allowable strength (positive: Fig. 2.4F; negative: Fig. S2.14A). In the data, cells are deemed to be interacting when their excess correlation exceeds a threshold, and so Fig. 2.2D represents a direct counterpart to our theoretical prediction. We find a clear qualitative match that includes the decrease in proportion of strong interactions for familiar environments (Fig. S2.14B). We further observe that the proportion of optimal couplings reaching the constraint w_{max} scales nonlinearly with the tuning similarity, as in the data; the shape of the nonlinearity depends on the imposed w_{max} (Fig. S2.15). This shows how constraints shape the optimal solution, and suggests that CA1 has to prioritize interactions among similarly tuned cells to counterbalance the limit on biological resources.

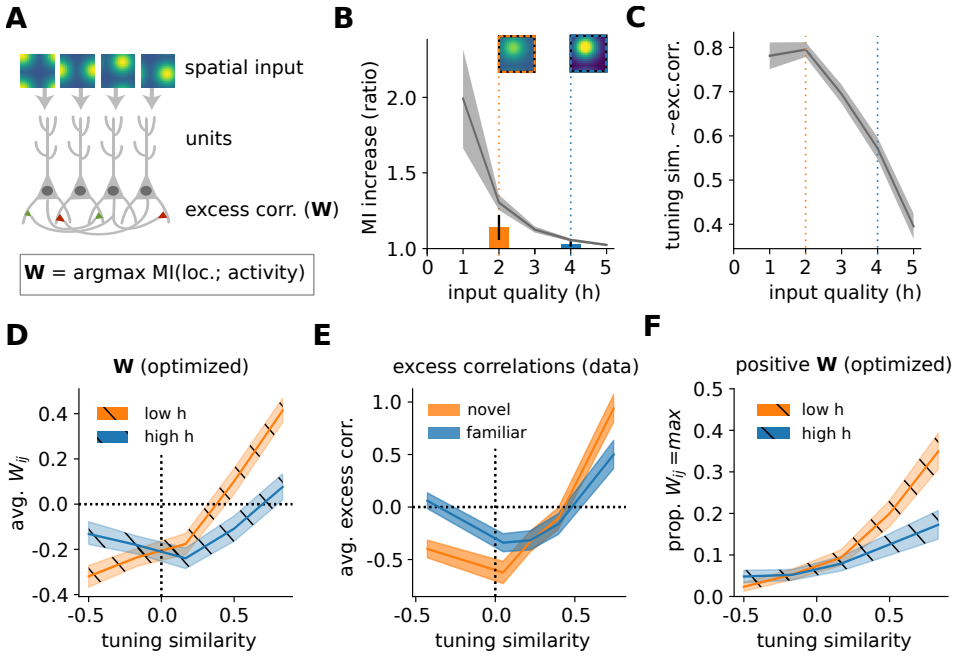

Fig. 2.4. Predicted optimal network interactions. **(A)** A schematic of the circuit model. Individual neurons, which receive spatially tuned inputs (with overall strength controlled by parameter h), are pairwise connected with interactions $\mathbf{W}$; interactions are numerically optimized to maximize the mutual information between spatial position and population responses while constraining population mean firing rates and $|W_{ij}| \leq w_{max}$ (here, $w_{max} = 1$). **(B)** Average ratio between mutual information (MI) in optimized vs non-interacting ($\mathbf{W} = 0$) networks. Dashed vertical lines denote two chosen input quality levels, together with firing rate map of an example cell ("low quality" $h = 2$, orange, resembling novel environment; "high quality" $h = 4$, blue, resembling familiar environment). In all simulation plots we show averages over 1000 replicate optimizations with random initial assignments of place fields (see Methods); shaded area – 95th percentile CI for the mean. Vertical bars represent mean $\pm$ STD of MI ratio for networks using data-like interactions. **(C)** Average alignment (Spearman's correlation) between pairwise input similarity and optimal W_{ij} as a function of input quality. **(D)** Average magnitude of optimal W_{ij} as a function of tuning similarity for the two environments. **(E)** Same as E, computed using the excitatory-excitatory excess correlations w_{ij} estimated from data. Note the vertical scale difference between (D) and (E): excess correlations w_{ij} are a statistical proxy for the true interactions W; the two are expected to be correlated but not identical (cf. Fig. S2.2A). **(F)** Proportion of optimal $W_{ij} = w_{max} = 1$ as a function of tuning similarity.

Even though our simulations use a coarse-grained and downscaled model of a real neural population (precluding exact comparisons), we observe an excellent qualitative match between theoretical predictions and the data. Taken together, these results suggest that network interactions in the hippocampus dynamically adapt to new environments so as to maximize the fidelity of the population-level spatial representation.

Central role for the nonlinear dependence of connectivity on tuning So far, our analysis of data as well as of optimized networks has identified a consistent pattern: the nonlinear dependence of interaction probability on tuning similarity (Fig. 2.2D; 2.4F). Figure 2.3 further showed that the pattern is necessary, since coding benefits were absent in randomized networks. The key remaining question is whether the observed connectivity pattern is not only necessary, but also sufficient, to convey spatial coding benefits and generate networks with data-like topology.

To address this question, we generated model networks of 50 place cells, as before, but limited their connection strengths to three possible values, $\{-J, 0, +J\}$, where $J \in [0, 1]$ could be

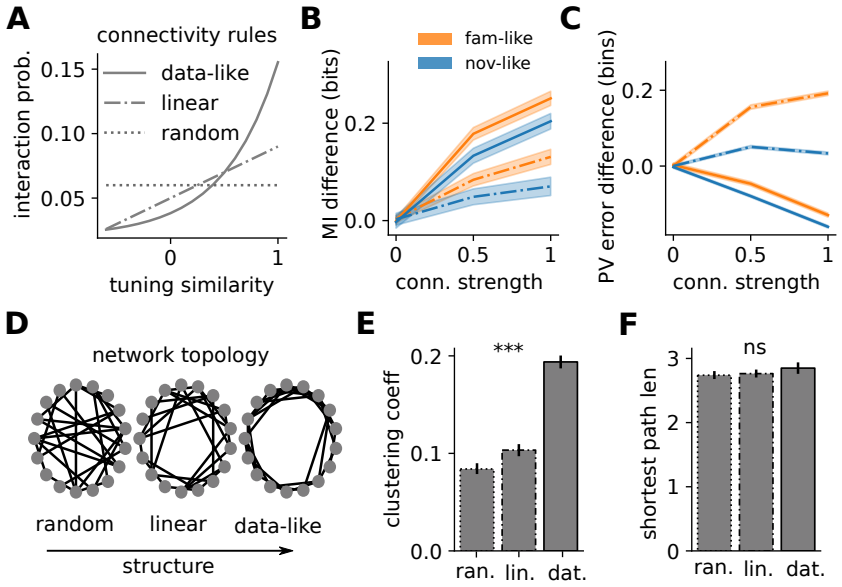

Fig. 2.5. Data-like interaction pattern is sufficient to generate small-world networks with improved spatial coding. **(A)** Connectivity rules for positive connections in a simulated place cell network with 50 units. **(B)** Mutual information (MI) increase for data-like (solid) and linear (dashed) connectivity rule relative to the random connectivity, for familiar-like (blue) and novel-like (orange) quality input. Shaded areas show the 95-th percentile confidence interval for the mean. **(C)** Average decoding error increase for data-like (solid) and linear (dashed) connectivity rule relative to random connectivity. Error measured as distance between real and decoded spatial bin. **(D)** Example network topologies obtained by using different connectivity rules from (A). Nearby nodes have high tuning similarity. **(E)** Average clustering coefficient for the three connectivity rules from A (error bars – standard error; significance – 1-way ANOVA test, $p < 0.001$ for ***, or n.s. for $p > 0.05$). **(F)** Average shortest path length for the three connectivity rules from A.

varied parametrically. We now used the interaction pattern of Fig. 2.2D as a direct *connectivity rule*: we selected 6% of pairs (as in data) to have a positive connection $+J$ and connected them according to their tuning similarity as in data (Fig. 2.5A, "data-like"). To assess the role of the nonlinearity, we compared this with networks where the connection probability was linear in tuning similarity ("linear") or where it was constant ("random"). In each of the three cases, a randomly chosen 3% of the place cell pairs (as in data) were connected with a negative strength, $-J$. As before, we fixed the average firing rate, and considered two levels of input quality, mimicking the familiar and novel environments (see Methods). This setup removed all structure (specifically, by making all connections have the same magnitude) except for that generated by the connectivity rule, allowing us to test for sufficiency.

First, we find that the data-like connectivity rule consistently improves mutual information between the population responses and position for increasing J, especially for novel-like input quality (Fig. 2.5B). This improvement is larger for the nonlinear, data-like connectivity than for the linear one. Figure S2.17 further suggests that connectivity alone accounts for a large fraction of mutual information gain, without the need for the fine-tuning of the interaction strengths. The data-like connectivity rule also improves the performance of a simple population vector decoder relative to random connectivity, in stark contrast to the linear dependence, which performs worse than the random one (Fig. 2.5C).

Finally, we asked whether different connectivity rules leave a strong signature on the network topology (Fig. 2.5D). To this end, we randomly generated 1000 networks according to the three different rules (Fig. 2.5A). The average clustering coefficient was substantially higher in networks created using the data-like rule (Fig. 2.5E) compared to both the random and linear connectivity

rules, without significantly affecting the distribution of incident edges (Fig. S2.16A) or the average shortest path length (Fig. 2.5F). Additional analysis on the clique-complexes of the connectivity graphs revealed that the 1D Betti numbers are significantly smaller for the synthetic networks generated using the data-like rule, and comparable with the data (Fig. S2.16C). These analyses are consistent with the overexpression of triangles (Fig. S2.6) and high clustering coefficients (Fig. 2.2E) observed in the data-derived network. Taken together, the nonlinear, data-like connectivity rule appears sufficient to generate small-world topologies matching data across a broad panel of network metrics.

2.4 Discussion

Statistical challenges limit our understanding of how experience shapes interactions and, consequently, information coding in a local neural circuit during animal-driven behavior. While the idea of analyzing pairwise correlations as a window into network interactions is not new (O'Neill et al., 2008; Humphries, 2017; Bassett and Sporns, 2017), the statistical problem of separating local network interactions from other factors that drive neural correlations has remained unsolved. Previous approaches based on stimulus-averaged correlations (Mathis et al., 2013), shuffles (Tocker et al., 2015), stimulus-independent Ising (Posani et al., 2017) or GLM model fits (Dunn et al., 2015) each suffer from statistical limitations (in terms of sample efficiency, strong stationarity or other model assumptions) which limit their general applicability. For this reason, most analyses of hippocampal collective behavior still rely on total correlations (Harris et al., 2003; O'Neill et al., 2008; Lopes-dos Santos et al., 2013; El-Gaby et al., 2021; Gava et al., 2021). Unfortunately, total correlations potentially conflate changes in coding with changes in nuisance variables, for instance changes in behavior: even if the representation does not change at all, a change in the animal's behavior (e.g. with experience) would be sufficient to change collective interactions defined based on total correlations. Furthermore, well documented theta oscillations, which arise from an interplay between medial septum inputs and hippocampal subcircuits (Colgin, 2013), as well as the animal's speed, which is known to substantially influence global hippocampal activity (McNaughton et al., 1983; Fuhrmann et al., 2015), can increase global synchrony and introduce spurious correlations. It is only by factoring out all these known sources of covariability, compactly captured by spike synchrony (Engel et al., 2001), that the fine structure of pairwise cell interactions can be revealed.

To reliably detect such interactions, we developed a novel statistical test rooted in the maximum entropy framework (Savin and Tkačik, 2017). When applying our detection method to tetrode recordings of hundreds of isolated units in dorsal hippocampus of freely behaving rats (Kaefer et al., 2019; Stella et al., 2019), we found stark differences between familiar and novel environments, especially in the EE subnetwork. In particular, we found increased interactions among putative pyramidal neurons in novel environments. Furthermore, we detected increased interactions between excitatory and inhibitory cells in novel environments. This effect was not explained by higher reliability of direct excitatory-inhibitory connections (Csicsvari et al., 1998). It has long been known that inhibition is generally weaker in a novel vs. a familiar environment (Nitz and McNaughton, 2004; Arriaga and Han, 2019; Geiller et al., 2020), which has been interpreted as a potential mechanism for enhancing learning by promoting synaptic plasticity in excitatory neurons (Whitlock et al., 2006; Arriaga and Han, 2019). This in turn is linked to previous reports on the sparsification of CA1 excitatory responses with increasing familiarity (Karlsson and Frank, 2008); in our data this effect was not visible as we selected excitatory cells that were active in both familiar and novel environments. Furthermore,

given that the null models capture both single cell average activity and population synchrony for each environment separately, it is unlikely that this observation can directly account for our results. Instead, our observations in the novel environment are likely to derive from an increased excitability at the dendritic level of pyramidal cells, an effect that has been observed experimentally (Sheffield et al., 2017) and has theoretically been shown as necessary for place field formation and stabilization (Pedrosa and Clopath, 2020).

Our key statistical observation could be distilled into one simple principle: a monotonic nonlinear dependence of the interaction probability on place field overlap for positive interactions among excitatory cells. This effect was observed across experience, but was more prominent during novelty. We analysed the neural coding implications of the inferred interaction structure using stimulus-dependent pairwise maximum entropy models (Granot-Atedgi et al., 2013). We found that data-like interactions offered improvements in spatial information content and decoding. These improvements are visible at different levels: at the single cell level, with a sharpening of place-fields and improved single cells' spatial information; and at the population level, by reducing the conditional entropy of stimulus-driven responses while largely preserving the total output entropy (Tkačik et al., 2010).

Coding advantages were higher during novelty: this observation argues for a mechanism employed by CA1 networks to cope with worse quality input from CA3 (Leutgeb et al., 2004) and MEC (Hafting et al., 2005; Barry et al., 2012) during novelty. The same mechanism could be used to improve spatial information in other situations where the spatial input is noisy (e.g., darkness). We also found that data-like interactions improved stimulus discriminability, corroborating previous findings (da Silveira and Berry, 2014). Moreover, our results explain why disrupting correlations between hippocampal neurons leads to decreased decoding accuracy (Stefanini et al., 2020).

Efficient coding in the place cell network yields optimal solutions in which similarly tuned neurons have a higher probability of interacting positively. This is especially prominent for lower-quality inputs in the novel environment, where the predicted relation between interaction probability and tuning similarity is clearly nonlinear, as observed in the data. Simulated networks where this observed relationship is elevated to an actual connectivity rule show that, *(i)*, the observed relationship is sufficient to improve population spatial coding, and *(ii)*, the resulting network topology shows clear small-world fingerprints (Watts and Strogatz, 1998; Bullmore and Sporns, 2009). While our results point towards small-worldness as one consequence of the particular connectivity rule that may be employed in the hippocampus (Perin et al., 2011), they do not provide any evidence that small-world networks have intrinsic coding benefits *per se* (Latora and Marchiori, 2001; Gallos et al., 2012). Further work is needed to clarify the relationship between coding and small-worldness and to experimentally probe whether small-world architecture is common in networks that need to process noisy inputs.

Even though inferred pairwise interactions do not necessarily reflect underlying synaptic connectivity directly (Das and Fiete, 2020), together with the neuron tuning function they offer an accurate statistical description of a neural population output (Schneidman et al., 2006; Tkačik et al., 2014; Meshulam et al., 2017). Moreover, pairwise interactions can be studied using well established tools from information theory, which critically rely on the differentiation between stimulus selectivity overlap and network interactions to assess the amount of information that a population carries about a stimulus (Tkačik et al., 2010). We derived and tested the efficient coding hypothesis for a network of interacting place cells, by maximizing the mutual information between the animal's location (the stimulus) and the population response, while holding individual cell tuning and overall firing rate fixed.

We found that network interactions adapt to different levels of input quality by employing different interaction vs. tuning similarity strategies. In particular, for low input quality (i.e., at low signal-to-noise ratio mimicking the novel environment) optimal network interactions are strongly aligned with the tuning similarity of the interacting cells. When input quality is higher (i.e., at higher signal-to-noise ratio mimicking the familiar environment), this relation weakens yet remains detectable. These optimality predictions closely resemble the data, suggesting that the CA1 circuit is close to an optimal operating regime across experience. Interestingly, the optimality predictions are borne out even though they are based solely on the information-theoretic efficiency of the observed neural representation, ignoring its goal-directed or decision-making value; these aspects could be relevant in other behavioral setups and could be included using existing theoretical frameworks (Chalk et al., 2016). As far as we know, this study is the first empirical test of the efficient coding hypothesis applied to network interactions, as proposed by previous theoretical work (Tkačik et al., 2010).

Theory predicts the inversion of the relative contribution of optimal interaction and tuning at very high signal-to-noise ratios (Tkačik et al., 2010). This causes the neural population to decorrelate its inputs, a regime that is characteristic for coding in the sensory periphery (Simoncelli and Olshausen, 2001). While our numerical simulations reproduce this decorrelation regime of efficient coding at very high signal-to-noise ratio inputs, our inferences and data analyses suggest that it is not relevant for the hippocampal place code. This is likely because the overall noise levels are higher in the spatial navigation circuits compared to the sensory periphery, and partially because of the intrinsic differences in the statistics of the signal to be encoded (position vs. natural images). Further work is needed to quantitatively relate the experimentally measured noise in CA1 inputs and responses to the effective "input quality" parameter that enters our predictions.

Are there previous reports where efficient coding predictions do not lead to decorrelation? A classic analysis in the retina correctly predicted that the receptive fields should lose their surrounds and switch to spatial averaging at low light (Atick and Redlich, 1990). A detailed study of retinal mosaics suggested that even during day vision receptive field centers of ganglion cells should (and do) overlap, increasingly so as the noise increases, leading to a residual redundancy in the population code (Doi and Lewicki, 2007; Borghuis et al., 2008), as reported (Puchalla et al., 2005). These findings support a more nuanced view of retinal coding (Barlow, 2001) than the initial redundancy reduction hypothesis (Barlow et al., 1961), precisely because they take into account the consequences of noise in the input and circuit processing (Linsker, 1989; Van Hateren, 1992; Karklin and Simoncelli, 2011). A recent study in fly vision focused on an interaction between two identified neurons, to find that its magnitude increased as the visual input became more and more noisy, as theoretically predicted by information maximization (Weber et al., 2012). Psychophysics of texture sensitivity that arises downstream of the primary visual cortex further suggested that the relevant neural mechanisms operate according to the efficient coding hypothesis, yet in the input-noise-dominated regime where decorrelation is not optimal (Hermundstad et al., 2014). In light of these examples and our results, efficient coding—understood more broadly as information maximization (Chalk et al., 2018) rather than solely in its noiseless decorrelating limit—should be revisited as a viable candidate theory for representations in the central brain. More generally, our approach enables a synergistic interplay between statistical analysis, information theory, graph theory and traditional neural coding, and opens new ways for investigating neural coding during complex/naturalistic behavior in other systems.

2.5 Materials and Methods

Experimental procedures

Datasets and Subjects We analyzed data from two previously published datasets (Stella et al., 2019; Kaefer et al., 2019). All procedures involving experimental animals were carried out in accordance with Austrian animal law (Austrian federal law for experiments with live animals) under a project license approved by the Austrian Federal Science Ministry. Four adult male Long-Evans rats (Janvier, St-Isle, France) were used for the experiments in (Stella et al., 2019). We further analyzed two wildtype littermate control animals from (Kaefer et al., 2019), generated by breeding two DISC1 heterozygous Sprague Dawley rats. Rats were housed individually in standard rodent cages(56X40X26 cm) in a temperature and humidity controlled animal room. All rats were maintained on a 12 hr light/dark cycle and all testing performed during the light phase. Food and water were available *ad libitum* prior to the recording procedures and bodyweight at the time of surgery was 300-375 g.

Surgery The first 4 animals (Stella et al., 2019) were implanted with microdrives housing 32 (2x16) independently movable tetrodes targeting the dorsal CA1 region of the hippocampus bilaterally. Each tetrode was fabricated out of four 10 um tungsten wires (H-Formvar insulation with Butyral bond coat California Fine Wire Company, Grover Beach, CA) that were twisted and then heated to bind them into a single bundle. The tips of the tetrodes were then gold-plated to reduce the impedance to 200-400 kU. During surgery, the animal was under deep anesthesia using isoflurane (0.5%–3% MAC), oxygen (1-2l/min), and an initial injection of buprenorphine (0.1mg/kg). Two rectangular craniotomies were drilled at relative to bregma (centered at AP =-3.2; ML = ± 1.6), the dura mater removed and the electrode bundles implanted into the superficial layers of the neocortex, after which both the exposed cortex and the electrode shanks were sealed with paraffin wax. Five to six anchoring screws were fixed on to the skull and two ground screws (M1.4) were positioned above the cerebellum. After removal of the dura, the tetrodes were initially implanted at a depth of 1-1.5 mm relative to the brain surface. Finally, the micro-drive was anchored to the skull and screws with dental cement (Refobacin Bone Cement R, Biomet, IN, USA). Two hours before the end of the surgery the animal was given the analgesic Metacam (5mg/kg). After a one-week recovery period, tetrodes were gradually moved into the dorsal CA1 cell layer (stratum pyramidale).

The last two animals (Kaefer et al., 2019) were implanted with microdrives housing 16 independently movable tetrodes targeting the right dorsal CA1 region of the hippocampus. Each tetrode was fabricated out of four 12 um tungsten wires (California Fine Wire Company, Grover Beach, CA) that were twisted and then heated to bind into a single bundle. The tips of the tetrodes were gold-plated to reduce the impedance to 300-450 kΩ. During surgery, the animal was under deep anesthesia using isoflurane (0.5-3%), oxygen (1-2 L/min), and an initial injection of buprenorphine (0.1 mg/kg). A rectangular craniotomy was drilled at -3.4 to -5 mm AP and -1.6 to -3.6 mm ML relative to bregma. Five to six anchoring screws were fixed onto the skull and two ground screws were positioned above the cerebellum. After removal of the dura, the tetrodes were initially implanted at a depth of 1-1.5 mm relative to the brain surface. Finally, the microdrive was anchored to the skull and screws with dental cement. Two hours before the end of surgery the analgesic Metacam (5 mg/kg) was given. After a one-week recovery period, tetrodes were gradually moved into the dorsal CA1 cell layer.

After completion of the experiments, the rats were deeply anesthetized and perfused through the heart with 0.9% saline solution followed by a 4% buffered formalin phosphate solution for the histological verification of the electrode tracks.

Behavioral procedures Each animal was handled and familiarized with the recording room and with the general procedures of data acquisition. For the first 4 animals (Stella et al., 2019), four to five days before the start of recording, animals were familiarized at least 30 min with a circular open-field environment (diameter = 120 cm). On the recording day, the animal underwent a behavioral protocol in the following order: exploration of the familiar circular open-field environment (40 mins), sleep/rest in rest box (diameter =26cm, 50 mins). Directly after this rest session the animals also explored a novel environment for an additional 40 min and rested after for 50 mins. The novel environment recordings were performed in the same recording room but in an enclosure of a different geometric shape but similar size (e.g., a square environment of 100cm width). The wall of both the familiar and novel environment enclosures was 30cm in height, which limited the ability of the animal to access distal room cues. In addition, in two animals a 50 mins sleep/rest session was performed before the familiar exploration.

For the last 2 animals (Kaefer et al., 2019), two to three days before the start of recording, animals were familiarized with a circular open-field environment (diameter = 80 cm). On the recording day, the animal underwent a behavioral protocol in the following order: 10 min resting in a bin located next to the open-field environment, exploration of the familiar open-field environment (20 min), sleep/rest in the familiar open-

field environment (20 min), exploration of a novel open-field environment (20 min), sleep/rest in the novel open-field environment (20 min). Whilst the familiar environment was kept constant, the novel environment differed on every recording day. The novel open-field arenas differed in their floor and wall linings, and shapes. The recordings for the familiar and novel conditions were performed in the same recording room.

During open-field exploration sessions, food pellets (MLab rodent tablet 12mg, TestDiet) were scattered on the floor to encourage foraging and therefore good coverage of the environment.

Data Acquisition A headstage with 64 or 128 channels (4 X 32 or 2 X 32 channels, Axona Ltd, St. Albans, UK) was used to preamplify the extracellular electric signals from the tetrodes.

Wide-band (0.4 Hz–5 kHz) recordings were taken and the amplified local field potential and multiple-unit activity were continuously digitized at 24 kHz using a 128-channel (resp. 64-channels) data acquisition system (Axona Ltd St. Albans, UK). A small array of three light-emitting diode clusters mounted on the preamplifier headstage was used to track the location of the animal via an overhead video camera. The animal's location was constantly monitored throughout the daily experiment. The data were analyzed offline.

Data Processing

Spike sorting The spike detection and sorting procedures were performed as previously described (O'Neill et al., 2006). Action potentials were extracted by first computing power in the 800-9000 Hz range within a sliding window (12.8 ms). Action potentials with a power >5 SD from the baseline mean were selected and spike features were then extracted by using principal components analyses. The detected action potentials were segregated into putative multiple single units by using automatic clustering software (http://klustakwik.sourceforge.net/). These clusters were manually refined by a graphical cluster cutting program. Only units with clear refractory periods in their autocorrelation and well-defined cluster boundaries were used for further analysis. We further confirmed the quality of cluster separation by calculating the Mahalanobis distance between each pair of clusters (Harris et al., 2000). Afterwards, we also applied several other clustering quality measures and selected only cells which passed stringent measures. In particular we implemented: isolation distance and l-ratio (Schmitzer-Torbert et al., 2005), ISI violations (Hill et al., 2011) and contamination rate. We employed the code available on Github: https://github.com/cortex-lab/sortingQuality. The criteria for the cells to be considered for analysis were the following:

- Isolation distance $> 10-$th percentile
- ISI violations < 0.5
- contamination rate $< 90-$th percentile

Periods of waking spatial exploration, immobility, and sleep were clustered together and the stability of the isolated clusters was examined by visual inspection of the extracted features of the clusters over time. Putative pyramidal cells and putative interneurons in the CA1 region were discriminated by their autocorrelations, firing rate, and waveforms, as previously described (Csicsvari et al., 1999a).

Data inclusion criteria We set a minimum firing rate of > 0.25 Hz for each cell on average, across both familiar and novel environments. Tetrodes in the above mentioned experiments were 0.4mm apart, guaranteeing that no two tetrodes would pick up the same cell (Henze et al., 2000). Nonetheless, we monitored for that possibility by measuring the cross-correlogram of cells on different tetrodes, and found no suspicious pairs of cells that could be duplicates. The final dataset consisted of 294 putative excitatory and 128 putative inhibitory cells across 6 animals. Considering only pairs of units recorded on different tetrodes, the dataset includes a total of 9511 excitatory-excitatory (EE) pairs, 7848 excitatory-inhibitory (EI) and 1612 inhibitory-inhibitory (II) pairs.

Spiking data was binned in 25.6 ms time windows, reflecting the sampling rate for positional information. We excluded bins where:

- the animal was static (speed < 3cm/s)
- sharp-wave ripple oscillatory activity was high, i.e. periods with power in the band $150 \sim 250$ Hz in the top 5th percentile (Csicsvari et al., 2000; O'Neill et al., 2006)
- theta oscillatory activity was particularly low, with power in the band $5 \sim 15$ Hz in the lowest 5th percentile; it is known that hippocampal theta oscillations support encoding of an animal's position during spatial navigation and reduces overall synchrony of population (Buzsáki and Moser, 2013; Mizuseki and Buzsaki, 2014).

Detailed dataset information For each animal, only one experimental session was recorded and analyzed, giving us a total of 6 familiar and 6 novel explorations. In detail, each animal/session provided us with the following numbers of cells and cell pairs:

animal	#excitatory	#inhibitory
mjc161	29	19
mjc163	111	42
mjc169	48	20
mjc186	58	28
mjd12	10	10
mjd34	38	9

animal	#EE	#EI	#II
mjc161	385	526	169
mjc163	5833	4472	825
mjc169	1085	917	185
mjc186	1556	1531	357
mjd12	29	96	42
mjd34	623	306	34

Null model of population responses and detection of excess correlations

Maximum entropy null model We construct a null model for population responses $(y_1, \ldots, y_N)$ that takes into account the position of the animal, $\mathbf{s}$ and the population synchrony, $k = \sum_i^N y_i$, but is otherwise maximally variable. We use this model to generate a large ensemble of surrogate datasets, that match the data with respect to tuning but without additional noise correlations. Using these surrogates allow us to estimate an empirical distribution of (total) pairwise correlations under the null model, which we then compare to data.

Under the assumption that spike counts have mean $\lambda(\mathbf{s}, k)$ with Poisson noise, the distribution of the joint neural responses under the null model factorizes as:

$$p_{ind}(\mathbf{y}\,|s, K) = \prod \text{Poisson}(y_i|\lambda_i(\mathbf{s}, k)). \quad (2.1)$$

One important caveat is that the population synchrony depends on the neural responses themselves, which introduces the additional constraint that $k = \sum_i y_i$ for each of these surrogate draws, something that we enforce by rejection sampling (Press et al., 1992). The only remaining step is to estimate the tuning function of each cell, $\lambda_i(\mathbf{s}, k)$, which we achieve using a nonparametric approach based on Gaussian Process (Rasmussen, 2003) priors.

Tuning function estimation Here we briefly describe the key steps of the approach, and refer the reader to (Savin and Tkacik, 2016) for further details. The data is given as T input pairs, $\mathcal{D} = \{\mathbf{x}_t, \mathbf{y}_t\}_{t=1,2,\ldots,T}$, where $\mathbf{x}_t$ denotes the input variables, defined on a $3-$dimensional lattice for the $2d-$position of the animal in the environment and population synchrony, defined as $k = \sum_{n=1}^N y_t^{(n)}$; $\mathbf{y}_t$ denotes spike counts of N neurons in the $t-$th time bin ($dt = 25.6$ms).

Neural activity of each single neuron is modeled as an inhomogeneous Poisson process with firing rate dependent on input variables, $\lambda(\mathbf{x}_t)$. We use a Gaussian Process (GP) prior to specify the assumption that the neuron's tuning is a smooth function of the inputs, with an exponential link function, $f = \log\lambda$, $f \sim \mathcal{GP}(\mu, k)$, with mean function $\mu(\cdot)$ and covariance function $k(\cdot,\cdot)$. In particular, we use a product of squared exponential (SE) kernels for the covariance function:

$$k(\mathbf{x}, \mathbf{x}') = \prod_{d=1}^{3} k_d(x_d, x_d') = \prod_{d=1}^{3} \rho_d \exp(x_d - x_d')/2\sigma_d^2, \quad (2.2)$$

This allows the prior covariance matrix to be decomposed as a Kronecker product $K = K_1 \otimes K_2 \otimes K_3$, dramatically increasing the efficiency of the fitting procedure (Flaxman et al., 2015).

The parameters $\theta = \{\mu, \rho, \sigma\}$ are fitted for each cell by maximizing the marginal likelihood of the data given parameters. Given estimated parameters, $\hat{\theta}$, we infer the predictive distribution $p(f_*|\mathcal{D}, \mathbf{x}_*, \hat{\theta})$ for a set of input values $\mathbf{x}_*$ (defined below). This distribution can be computed by marginalizing over $\mathbf{f}$:

$$p(f_*|\mathcal{D}, \mathbf{x}_*, \hat{\theta}) = \int p(f_*|\mathcal{D}, \mathbf{x}_*, \hat{\theta}, \mathbf{f}) p(\mathbf{f}|\mathcal{D}, \hat{\theta}) d\mathbf{f} \quad (2.3)$$

This distribution is intractable, but can be approximated by using a Laplace approximation for $p(\mathbf{f}|\mathcal{D}, \hat{\theta})$ so that ultimately $p(f_*|\mathcal{D}, \mathbf{x}_*, \hat{\theta}) \approx \mathcal{N}(\mu_{f_*}, \sigma_{f_*})$. Finally, thanks to the exponential link function, the inferred firing rate of an individual input point $\lambda(\mathbf{x}_*) = \exp(f_*)$ is log-normally distributed, whose mean and variance can be computed as:

$$\mathbb{E}(\lambda(\mathbf{x}_*)) = \exp(\mu_{f_*} + \sigma_{f_*}^2/2) \quad (2.4)$$

and

$$\text{Var}(\lambda(\mathbf{x}_*)) = \exp(\sigma_{f_*}^2 - 1)\exp(2\mu_{f_*} + \sigma_{f_*}^2) \quad (2.5)$$

We chose input points $\mathbf{x}_* = (\mathbf{s}, k)$ that corresponded to the binned 2D location $\mathbf{s}$ of the animal (5cm bins) and binned population synchrony k (10 equally weighted bins, each containing 10% of the data, i.e. the bin edges correspond to the $(0\text{th}, 10\text{th} \ldots, 100\text{th})$ percentiles).

Generating surrogate data At each moment in time, given the position $\mathbf{s}$ and population synchrony k, the GP tuning estimate provides a distribution over possible firing rates for cell i, $\lambda_i(\mathbf{s}, k)$, as a log normal distribution, $\log\lambda_i \sim \mathcal{N}(\mu_{f_*}, \sigma_{f_*})$. This captures uncertainty about the tuning of the cell, given the data. Our method allows us to propagate this uncertainty throughout all the subsequent analyses by sampling firing rates from this distribution (formally, by treating neural responses as arising from a doubly stochastic point model). We generate surrogate spike counts in two steps. First, we sample the mean firing from this $p(\lambda_i|s, K)$ distribution. Second, for each λ_i sample, we draw the corresponding spike count from $\text{Poisson}(\lambda_i)$. Applying this procedure for all cells and all time points generates a surrogate dataset from the unconstrained null model. We enforce the constraint $\sum_i y_i = k$ by discarding and redrawing samples that do not satisfy it. In rare cases (less than 2% of data),

it was not possible to replicate the desired k statistic, i.e. achieving the desires k required more than 500 re-samplings. Such time bins were excluded from subsequent analysis (both for for real data and all surrogates). We generate a total of 1000 surrogate datasets.

Inference of excess correlations We use the pairwise correlations between neural responses as the test statistic and compare it to the distribution of pairwise correlations expected under the null model that assumes that the firing rate of cells is only driven by the stimulus and the synchrony of the population, without further pairwise interactions.

Given the Pearson correlation coefficient between the activities of cells i and j computed on real data, c_{ij}, and c_{ij}^{γ} the same quantity computed on a surrogate dataset $\{\mathbf{y}_{1:t}^{\gamma}\}$ for $\gamma = 1, 2, \dots 1000$. We define the quantity we refer to as "excess correlations" as:

$$w_{ij} = \frac{c_{ij} - < c_{ij}^{\gamma} >}{\sigma(c_{ij}^{\gamma})} \tag{2.6}$$

where $< \cdot >$ denotes the sample average and σ the sample standard deviation of c_{ij}^{γ}. Assuming that the c_{ij}^{γ} distribution is normal, this quantity is closely related to confidence bounds, and p-values (via the error function). An excess correlation is deemed significant if $|w_{ij}| > 4.5$, which corresponds to a p-value threshold of $p = 0.05$ with a Bonferroni correction for the 7500 multiple comparisons.

Validation To validate our method, we construct an artificial dataset with know interactions, by sampling from a coupled stimulus dependent MaxEnt model. We consider $N = 50$ neurons and binary activations $\mathbf{y} = (y_1, \dots y_N)^{\mathsf{T}}$ for any given time window. The distribution of responses $\mathbf{y}$ given a location-stimulus s and synchrony level k is

$$p(\mathbf{y}|s,k) \propto \exp\Big(\sum_i f_i(s) y_i + \sum_{i>j} W_{ij} y_i y_j - \sum_i (y_i - k/N) \Big) \tag{2.7}$$

where $s \in \{s_1, \dots, s_K\}$ is a spatial position chosen from a set of discrete locations uniformly spaced in the environment, and the feedforward input to each cell, $f_i = f_i(s)$, is as described in methods subsection (2.5). We try to match the general statistics of the data as closely as possible. In particular, we match the true time-dependent occupancy, s_t, observed in a 20 minutes exploration session, and the corresponding time-dependent synchrony observed in the same session, k_t, by sampling one population activity vector (after adequate burn-in time) at each time point $\mathbf{y}(t) \sim P(\mathbf{y}|s_t, k_t)$ using Gibbs sampling (Geman and Geman, 1984).

Given this artificial dataset, we analyze it with the same processing pipeline that we use for the neural recordings and compare the estimated interactions w_{ij} with the ground truth couplings W_{ij}, which are randomly and independently drawn from $\mathcal{N}(0,1)$. Furthermore, we generate data with the same constraints but without any interactions. We asses the ability of our statistical test to detect true interactions using the receiver operating characteristic (ROC), and estimate false positive rates for our statistical test.

Hippocampal population responses with adjustable network structure

Stimulus dependent MaxEnt model In order to explore the effects of the noise correlation structure on the coding properties of the hippocampal system, we employed a statistical model of the collective behavior of a population of place cells that allowed us to vary the couplings among cells while keeping fixed the output firing rate. A similar, stimulus dependent maxent model was introduced in (Granot-Atedgi et al., 2013), and more recently was used in (Meshulam et al., 2017) to prove that correlation patterns in CA1 hippocampus are not due to place encoding only, but also to internal structure and pairwise interactions. Our model includes spatially-selective inputs with adjustable strength, $h \in \mathbb{R}$, and noise correlations modelled as a matrix $\mathbf{W}$ describing the strength of interaction between cell pairs. Additionally, we constrained average population firing rates to be the same for each possible choice of h and $\mathbf{W}$, as a way of implementing metabolic resource constraints.

More specifically, consider N neurons with binary activations $\mathbf{y} = (y_1, \dots y_N)^{\mathsf{T}}$. The distribution of responses $\mathbf{y}$ given a location-stimulus s we considered is

$$p(\mathbf{y}|s) \propto \exp\Big(h \sum_i f_i(s) y_i + \sum_{i>j} W_{ij} y_i y_j - h_0 \sum_i y_i \Big) \tag{2.8}$$

where $s \in \{s_1, \dots, s_K\}$ is a spatial position chosen from a set of discrete locations uniformly spaced in the environment (the unit square, $[0,1] \times [0,1]$). The feedforward input to each cell, $f_i = f_i(s)$, is modelled as a $2-$D Gaussian bump with continuous boundary conditions, mean randomly drawn from a uniform on $[0,1] \times [0,1]$ and fixed covariance $0.1\mathbb{I}$. The parameter h_0 allows us to fix the average population firing rate to 20% of the population size, and is found by grid optimization. Once the input tuning f_i is fixed for each cell, we select the connections W_{ij} for each cell pair by sampling from the data-inferred excess correlations of cell pairs with similar input tuning, and then scaling according to the results found during method validation (Fig S 2.2). We did so separately for familiar and for novel environments.

Inference of input quality from data We fixed the appropriate parameter h, separately for familiar-like and novel-like connections, by matching average marginal statistics of firing rate maps obtained from the model to the data. We utilized three measures: single cell spatial information, sparsity and gain, which are described in detail in Methods subsection (2.5). Separately for familiar and novel environments, we randomly chose 50 tuning functions and selected pairwise connections (rescaled from data, see previous subsection) depending on tuning similarity. We then simulated population activity according to real occupancy (as in validation) for $h \in \{0.1, 0.2, \ldots, 5\}$, computed firing rate maps and their statistics. We selected the h value that best matched the data (Fig. S2.9).

Optimization of connections for fixed input and fixed firing rate Given h,$\{f_i(\cdot)\}$, we optimize the connections $\mathbf{W}$ so as to maximize the mutual information between population activity and spatial position, $MI(\mathbf{y}; s) = \sum_{\mathbf{y},s} p(\mathbf{y}|s)p(s) \log \frac{p(\mathbf{y}|s)}{p(\mathbf{y})}$, via Sequential Least SQuares Programming (SLSQP) (Virtanen et al., 2020). We further constrain the population average firing to 20% of the neural population, and each W_{ij} is restricted to lay in $[-1, 1]$. Both reflect biological resource constraints on the optimal solution.

Most simulations use $N = 10$ neurons, which allows the mutual information to be computed in closed form (by enumerating all possible patterns). Reported estimates are obtained by averaging across 1000 randomly initialized networks (different $f_i(\cdot)$ centers, and initial conditions for the optimization). To ensure that our results generalize to large networks, we also performed limited numerical simulations for $N = 20$ (only for $h = 2$ and $h = 4$, averaging over 10 networks.

Comparison of data-like interactions with optimization results We compared the performance (in terms of MI) of optimized networks vs networks with data-like interactions (Fig. 2.4B). To do so, we randomly initialized a small network with 10 neurons, as explained in the previous subsection, and for each pair assigned an interaction strength according to the relation found in data (as in Fig. 2.3). Finally, we optimized the interactions so as to maximize the mutual information between population spiking and stimuli.

Optimal coding for large networks The exact computation of the mutual information $MI(\mathbf{y}; s)$ is very resource intensive and only applicable to small networks ($N \leq 20$). To investigate the effects of noise correlations at larger scales we need to rely on efficient approximations. The mutual information between population binary responses $\mathbf{y}$ and location-stimulus s can be written as

$$\begin{aligned} MI(\mathbf{y}; s) &= \sum_{\mathbf{y},s} p(s|\mathbf{y})p(\mathbf{y}) \log p(s|\mathbf{y}) \\ &\quad - \sum_{\mathbf{y},s} p(s|\mathbf{y})p(\mathbf{y}) \log p(s) \\ &= H(s) - H(s|\mathbf{y}), \end{aligned} \tag{2.9}$$

where H denotes (conditional) entropy. Assuming that $p(s)$ is a uniform distribution over stimuli, we have $H(s) = 2 \log B$, where B is the number of bins used to discretize each dimension of the $2-$dim environment. We generally use $B = 16$. The challenge is to compute $H(s|\mathbf{y})$. For a given $\mathbf{y}$, denote with $\hat{h}(\mathbf{y}) := -\sum_s p(s|\mathbf{y}) \log p(s|\mathbf{y})$. Then we have:

$$\begin{aligned} H(s|\mathbf{y}) &= -\sum_{\mathbf{y},s} p(s|\mathbf{y})p(\mathbf{y}) \log p(s|\mathbf{y}) \\ &= \sum_{\mathbf{y}} p(\mathbf{y})\hat{h}(\mathbf{y}) \\ &= \sum_{s} p(s) \sum_{\mathbf{y}} p(\mathbf{y}|s)\hat{h}(\mathbf{y}) \end{aligned} \tag{2.10}$$

We used the last expression and estimated $H(s|\mathbf{y})$ by drawing 10^6 samples from $p(\mathbf{y}|s)$ for each stimulus s using Gibbs sampling (Geman and Geman, 1984). We reported the estimated average across stimuli and confidence intervals in the figures. The quantity $\hat{h}(\mathbf{y}) = -\sum_s p(s|\mathbf{y}) \log p(s|\mathbf{y})$ is the entropy of the posterior distribution on stimuli given a certain binary vector. The main obstacle to computing $\hat{h}$ is that, for each stimulus s, we need to know the proportionality constant $Z_s = \sum_{\mathbf{y}} p(\mathbf{y}|s)$ (i.e. the partition function), that makes the probability (2.8) sum up to 1. We computed Z_s exhaustively for $N \leq 20$ by enumerating all the possible binary vectors. For $N \geq 20$ we estimated it using a simple Monte Carlo method by randomly drawing 10^9 independent $N-$dim binary samples for each stimulus, and then regularizing by applying a mild 2D gaussian smoothing ($\sigma = 0.5$ bins) on the log-transformed Z_s among neighboring stimuli.

"Topology" model simulations We aimed at characterizing the influence of higher order structure on the coding of the network. We used the same model as in eq. [2.8] with 50 place cells, but allowed connections to be either $-J$, 0 or $+J$, where $J \in [0, 1]$ is the connection strength. We employed three different strategies to select the units to connect, as described in the main text, based on their tuning similarity. We kept fixed the number of positive ($+J$) and negative ($-J$) couplings to 6% and 3% respectively. For each choice of tuning, connectivity rule and strength J we used the parameter h_0 to enforce the population average firing to be 20% of the population size.

Analysis of experimental data

Single cell tuning characterization To describe the tuning properties of single cells we employed several

measures:

- gain: peak firing rate over mean, estimated from the tuning function of a cell,
- sparsity: $< \lambda_x >_x^2 \; / \; < \lambda_x^2 >_x$, where λ_x denotes the average firing at location x, is a measure of how compact the firing field is relative to the recording apparatus (Jung et al., 1994),
- spatial information: $< \frac{\lambda_x}{\lambda} \log \frac{\lambda_x}{\lambda} >_x$, where $\lambda =< \lambda_x >_x$, is the leading term of the MI between average spiking and discretized occupancy for small time windows (Skaggs et al., 1993; Souza et al., 2018).

Decoding of spatial position from data We subdivided the environment in equally spaced $2-$dimensional bins with bin side length of 20 cm. This choice was due to the fact that, to properly estimate the average co-activation of cells one needs many samples and a finer subdivision of the environment made this task extremely difficult. We randomly subdivided the data in two parts, 75% for training and 25% for decoding. With the training data we estimated, for each bin separately, the average activation and the covariance of the neurons activity. With the remaining 25% of the data, we computed for each non-overlapping 10 consecutive 25.6 ms time bins the activation (denoted by population vector or PV) and the covariance (COV). We then simply compared them to all the expected PV and COV measured over the training set in different bins and picked the most similar one in terms of Pearson correlation. For Fig. S 2.11 we also used a bayesian maximum a posteriori (MAP) decoder. We inferred the conditional probability of spiking of cell i given location s, $P(y_i|s)$, on $4 \cdot 10^5$ samples. We then tested the decoder on additional 10^5 samples. Cells were assumed to be conditionally independent of each other, hence $P(\mathbf{y}|s) = \prod_i P(y_i|s)$. We used a flat prior, so for a population vector $\mathbf{y}$ the decoded position is:

$$\hat{s}_{MAP} = \arg\max_s P(s|\mathbf{y}) = \arg\max_s P(\mathbf{y}|s) \quad (2.11)$$

PCA, linear separability of pairs of stimuli We wanted to investigate the linear separability of population responses to different locations. We randomly selected 500 times two distinct locations in the environment and selected all the 250ms population responses in a 10 cm surrounding of the two positions. We then found the best hyperplane that separated the two sets of responses by using a soft-margin linear SVM with hinge loss, and reported the training error. We also computed the principal components of the population responses to both locations together, and reported the variance explained by the first PC.

Network analysis

Graph theoretical measures All the measures were carried out using the library NetworkX (release 2.4) in Python 3.7. We considered unweighted and non directed graphs where each cell was a vertex and an edge connected each cell pair that had a significant interaction ($|w_{ij}| > 4.5$). A graph $G = (V, E)$ formally consists of a set of vertices V and a set of edges E between them. An edge e_{ij} connects vertex v_i with vertex v_j. The neighbourhood for a vertex v_i is defined as its immediately connected neighbours: $N_i = \{v_j : e_{ij} \in E \vee e_{ji} \in E\}$ and its size will be denoted by $k_i = |N_i|$.

We measured:

1. **Clustering coefficient:** this measure represents the average clustering coefficient of each node, which is defined as the fraction of existing over possible triangles that include that node as a vertex. Formally, the local clustering coefficient c_i for a vertex v_i is given by the proportion of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them, hence measuring how close its neighbourhood is to forming a clique. If a vertex v_i has k_i neighbours, $\frac{k_i(k_i-1)}{2}$ edges could exist among the vertices within the neighbourhood. Thus, the local clustering coefficient for vertex v_i can be defined as

$$c_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

and the average clustering coefficient as

$$c_G = \frac{1}{n} \sum_{v_i \in V} c_i$$

2. **Average shortest path length:** this measure can be computed only if the graph is connected. If not, we computed this measure on the largest connected subgraph.

$$a_G = \sum_{u,v \in V} \frac{d(u,v)}{n(n-1)}$$

where u, v are distinct vertices, $d(u, v)$ is the shortest path length between u, v and n is the size of the graph G.

Triangles analyses We tested for the over-expression of particular interaction patterns by counting the number of triangles (i.e 3 all-to-all interacting cells) composed by 3 inhibitory cells, 2 inhibitory and 1 excitatory, 1 inhibitory and 2 excitatory or 3 excitatory cells. We tested these counts against the counts from the same networks with shuffled edges. We employed an edge-shuffling procedure that preserved both the

total number of edges and the number of incident edges per node, separately for the EE, EI and II subnetworks (i.e. an edge connecting two excitatory cells could be exchanged only with another edge connecting two excitatory edges etc). To do this, we randomly selected two edges of each subnetwork, say AB and CD. If $A \neq C \neq D$ and $B \neq C \neq D$ we removed the two edges and inserted the "swapped" ones, AC and BD. We repeated this procedure 100 times for each subnetwork to yield one shuffled network. We repeated this procedure 1000 times, which gave us a null distribution to test the original counts against. In Supp. Fig. 2.6 we reported the counts of each pattern, separately for familiar and novel environments, normalized against our null distribution.

Betti numbers We computed the Betti numbers of the clique-complex induced by the graphs. These are distinct from the graphs Betti numbers (Giusti et al., 2015). A clique in a graph is an all-to-all connected set of vertices. The clique complex $X(G)$ of an undirected graph G is an abstract simplicial complex (that is, a family of finite sets closed under the operation of taking subsets), formed by the sets of vertices in the cliques of G. Intuitively, the clique-topology can be characterized by counting arrangements of cliques which bound holes. Formally, the dimensions of the homology groups $Hm(X(G), \mathbb{Z}_2)$ yield the Betti numbers b_m (Giusti et al., 2015). Given our low connectivity (9%), b_m was almost always zero for $m \geq 2$. On the other side, b_0 simply counts the number of connected components, so in our analysis we focused on b_1. This is the number of cycles, or holes, that are bounded by 1-dim cliques. Graphically, these are 4 edges that form a square, or 5 edges that form a pentagon etc. Notice that 3 edges that form a triangle don't count towards b_1, because they represent a 2-dim clique (i.e. 3 vertices that are all-to-all connected). This is why a higher clustering coefficient (i.e. more triangles) implies a lower b_1.

Supplementary figures

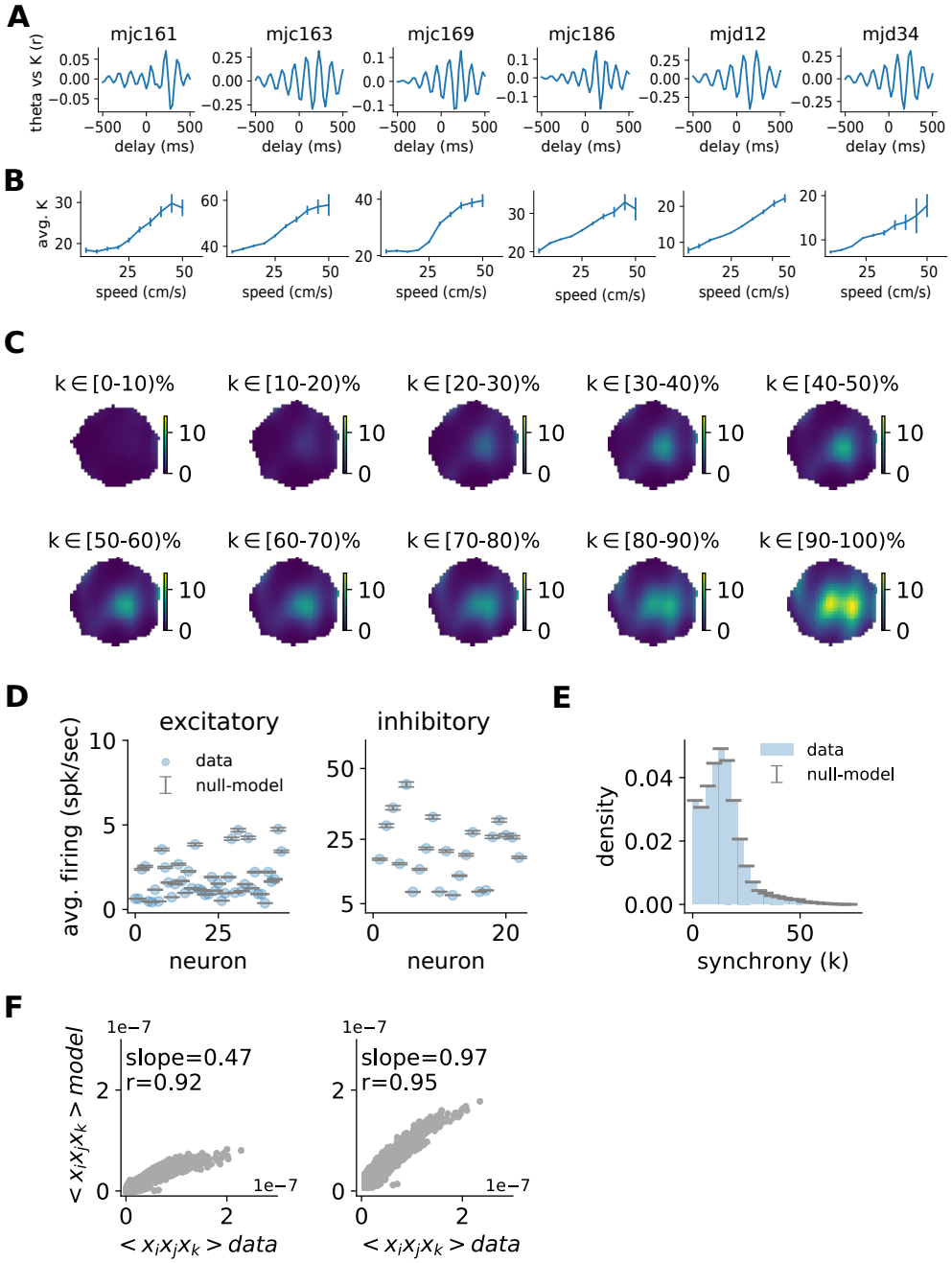

Fig. S. 2.1. Null model. Our null model accounts for the measured 2D spatial selectivity of each recorded neuron as well as the extra dimension given by the moment-to-moment global neural synchrony of both excitatory and inhibitory cells. We have chosen this constraint because it captures both underlying theta oscillations (A) and animal speed (B). **(A)** Pearson cross-correlation of theta oscillation amplitude (LFP filtered using Butterworth 2nd order in $6-12$ Hz and then downsampled) vs summed activity of all neurons (i.e. synchrony) K in 25.6 ms time windows. One panel for each of the 6 animals during free exploration of a familiar environment. **(B)** Average synchrony K as a function of speed. Speed was binned in 5cm/s bins. Error bars represent $99th$ CI. **(C)** Example 3D model inferred for one place cell during exploration of a familiar environment. Notice how the firing rate map depends on both 2D spatial location and synchrony K. **(D,E)** Null model marginal statistics of an example session. (D) The null model captures average firing rate of both putative excitatory (left) and inhibitory (right) cells. Error bars represent $99th$ CI. (E) Distribution of synchrony. **(F)** We measured the coincident firing of all the triplets of cells $< x_i x_j x_k >$ from data (same example session as (D), (E)) in 25.6 ms time windows, and compared with the expected coincident firing of triplets from a model that accounts for 2D spatial selectivity and without dependence on K (left) or with K (right). Notice how the 3D model captures the majority of true triplet cofiring, which are generally underestimated by a simple 2D null model. Regression analyses, both $ps < 0.0001$.

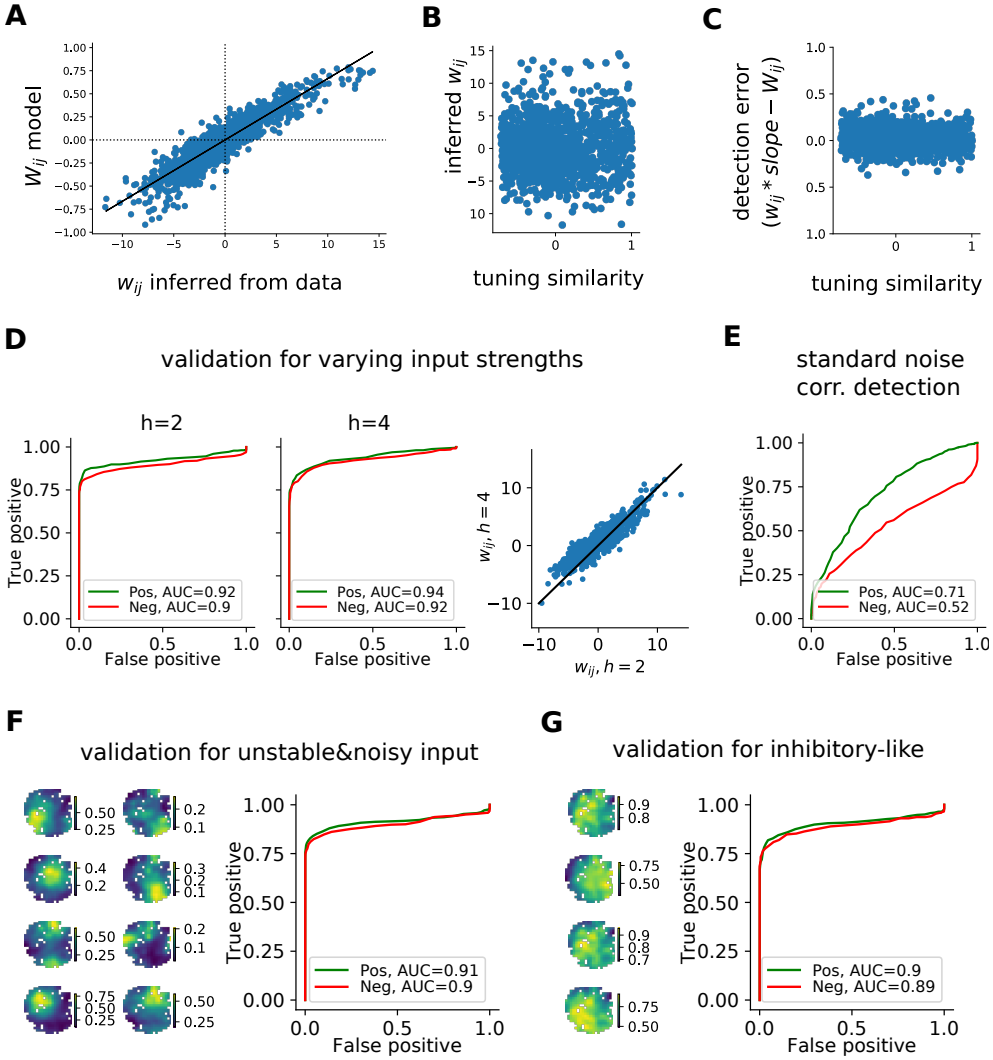

Fig. S. 2.2. Further data on validation and null model. **(A)** Scatter plot of ground truth W_{ij} values used in the model for validation vs w_{ij} inferred from artificial data. Black line: linear regression (slope=0.066, intercept=-0.009, r=0.918, p=0.0). **(B)** Scatter of inferred w_{ij} vs tuning similarity. Tuning similarity is computed as Pearson correlation of the tuning functions $corr(f_i(s), f_j(s))$ for each pair i, j of cells in the model (see Methods). Notice the absence of bias towards detection for cells with higher or lower tuning similarity. Linear regression: r = 0.023, p = 0.24. **(C)** W_{ij} detection error inferred as the difference between w_{ij} (scaled by the appropriate slope) and the true W_{ij}. Notice the absence of bias towards highly similarly tuned pairs. Linear regression: r = 0.009, p = 0.61. **(D)** ROC for lower (h=2, left) and higher (h=4, middle) input strengths. Right: scatter plot of inferred w_{ij} for the two input strengths. Linear regression r=0.913, p<0.0001. **(E)** Standard noise correlation detection. For each spatial bin where the synthetic agent spent more than 10 seconds the Pearson correlation was computed among cells and then averaged. **(F)** Novel-like scenario: unstable and noisy multi-component tuning. Input strength was drawn random from $\mathcal{N}(2, 1)$. Number of components chosen as Poisson with average=2. At each time point, component weights were re-drawn from a Dirichlet distribution with parameter alpha=2. The average population firing was always kept to 10% of the population. Left: example inferred tuning functions. Right: ROC. **(G)** Inhibitory-like scenario: high-firing and weak spatial tuning, noisy multi-component tuning. Number of components chosen as Poisson with average=5. At each time point, component weights were re-drawn from Dirichlet distribution with parameter alpha=2. The average population firing was kept to 50% of the population. Left: example inferred tuning functions. Right: ROC.

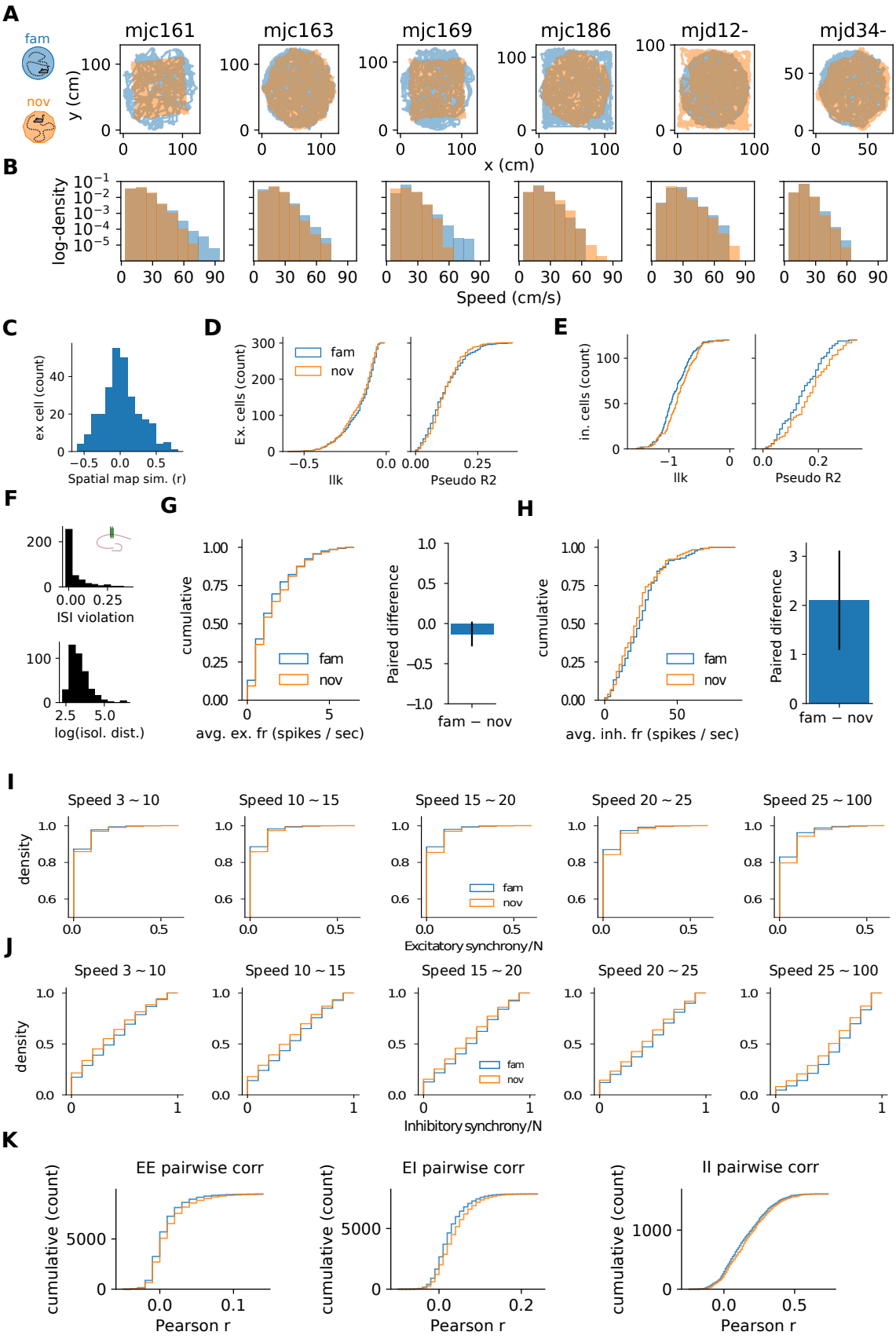

Fig. S. 2.3. Animal behavior and CA1 neurons marginal statistics.

Fig. S. 2.3. Animal behavior and CA1 neurons marginal statistics. **(A)** Comparison of familiar (blue) vs novel (orange) occupancy and enclosure sizes (x and y axes, measured in cm), animal by animal. Each animal contributed to one familiar and one novel recording only. **(B)** Distribution of running speed during familiar (blue) vs novel (orange) explorations. KS test, all ps<0.01. Mann-Withney U-test, all ps $<$ 0.01. **(C)** Distribution of spatial firing field similarity of familiar vs novel environments, measured as Pearson r of 2D firing rate maps. **(D)** Distribution of marginal log-likelihood (left) and pseudo R2 (right) of excitatory single cells null models in familiar (blue) vs novel (orange). KS tests, both $ps > 0.1$. **(E)** Same as (D), for putative inhibitory neurons. KS tests, both $ps > 0.01$. **(F)** Top: distribution of ISI violation scores after spike sorting for the data included in the analyses. Bottom: same for the Isolation Distance measure. **(G)** Left: distribution of average firing rates of putative CA1 excitatory neurons in familiar (blue) and novel (orange) environment (KS test: p=0.505, N = 294). Right: paired difference across environments (familiar − novel), paired t-test: p=0.091. Error bars represent 95th CI for the mean. **(H)** Same as (G) for putative inhibitory neurons (N=128). KS test: p=0.209, paired t-test: p=$6e-05$. **(I)** Distribution of synchrony measured as cells active over population size (k/N) in 25.6 ms time windows of excitatory neurons for different behavioral speed: $[3, 10), [10, 15), [15, 20), [20, 25), [25, 100)$ cm/sec for familiar (blue) and novel (orange).Data was pooled across the 6 animals. All KS test $p < 0.0001$. **(J)** Same as (I) for putative inhibitory neurons.All KS test $p < 0.0001$. **(K)** Distribution of (raw) pairwise correlations among EE (left), EI (middle) and II (right) pairs. All KS tests had p<0.001.

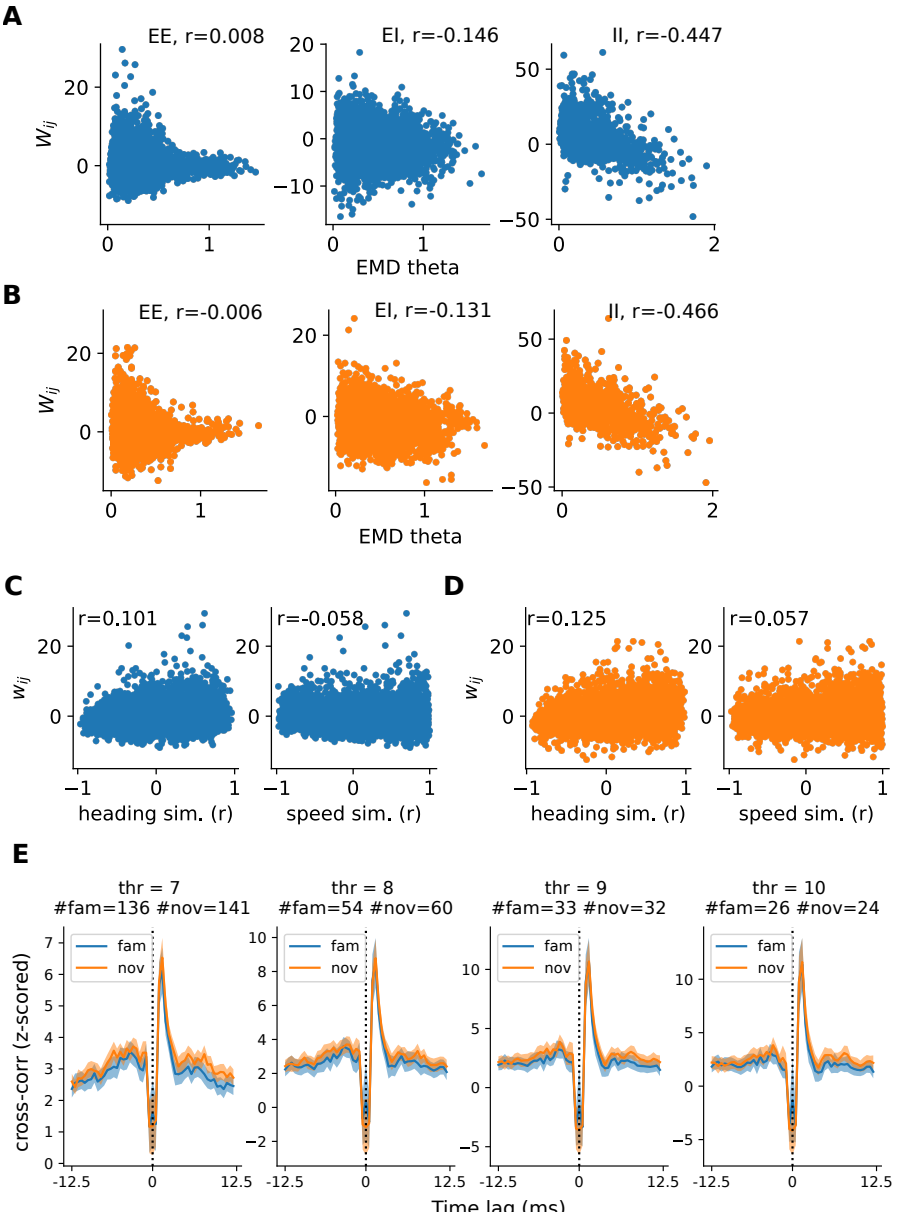

Fig. S. 2.4. Excess correlations vs theta locking/speed/heading similarity. Efficacy of excitatory-inhibitory monosynaptic connections. **(A)** For each cell we computed a theta-phase histogram, which measures the tendency of a cell to fire at a particular theta phase. We then compared the dissimilarity of theta-phase histograms among each pair of cells by means of an earth mover distance (EMD). We then scatter plotted the inferred w_{ij} vs dissimilarity of theta selectivity (t-test for Spearman rank correlations: EE $p > 0.05$, EI $p < 0.001$, EE $p < 0.001$). **(B)** Same as (A), for EE, EI and II pairs in novel environemnts. (t-test for Spearman rank correlations: EE $p > 0.05$, EI $p < 0.001$, EE $p < 0.001$). **(C)** Left: For each cell we measured the average firing rate for each possible heading angle in 30deg bins. We then measured the similarity in heading tuning for each pair of EE cells, and scatter plotted against inferred w_{ij} (t-test for Spearman rank correlations: $p < 0.01$). Right: we measured the average firing as a function of speed in 5cm/s speed bins. We then measured the similarity of speed tuning for each pair of EE cells, and scatter plotted against inferred w_{ij} (t-test for Spearman rank correlations: $p = 0.028$). Reported in the figure r-values for Spearman rank correlations. **(D)** Same as (C), for EE pairs in novel environments. Left: t-test for Spearman rank correlations: $p < 0.01$. Right: t-test for Spearman rank correlations: $p = 0.032$. **(E)** Firing of cells was binned into 0.5ms time windows, and cross-correlogram was normalized by subtracting the mean and dividing by the STD of cross-correlograms computed on randomly shifted data 100 times. The pairs that had peak (normalized) cross-correlogram $> thr$ STDs, for $thr \in [7, 8, 9, 10]$, were labelled as monosynaptically connected. Thresholds were selected according to literature (Csicsvari et al., 1998). From left to right, each panel shows results for increasing thresholds and also reports the number of monosynaptic connected pairs in each environment. Shaded areas represent 95th CI for the mean. Mann-Withney U test for the peaks, each panel p>0.05.

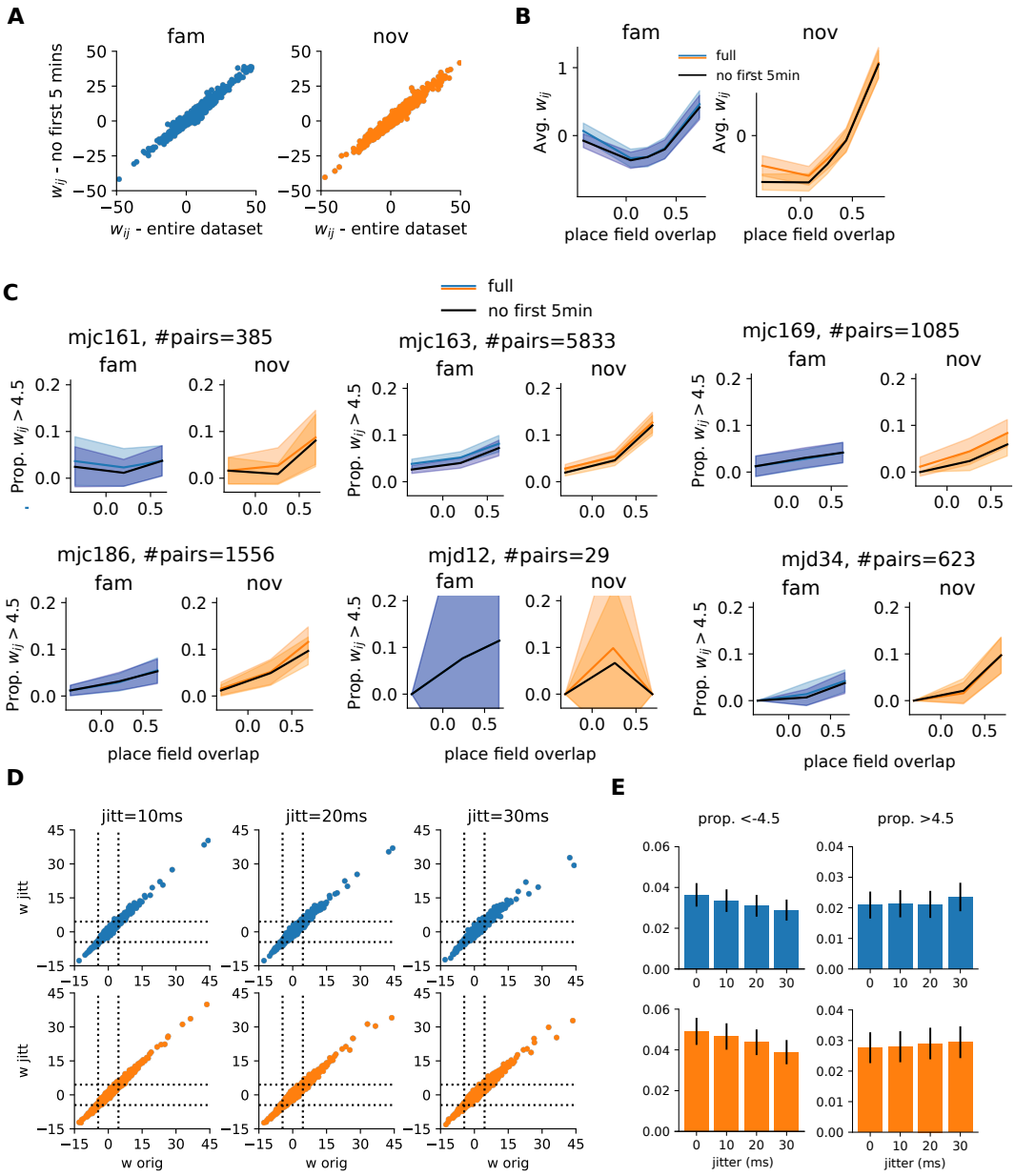

Fig. S. 2.5. Stability of main results. The spatial tuning of single cells emerges during the first few minutes of exploration of a novel environment (Wilson and McNaughton, 1993). To ensure that this instability does not affect our results, we detected excess correlations removing the first 5 minutes of exploration (panels A - C). Moreover, effects such as theta phase precession are not equal in familiar and novel environments and could induce fine grained synchrony among cells that our model cannot detect (O'Keefe and Recce, 1993). To test whether phase precession influences our ability to infer excess correlations, we introduced a small amount of random jitter on each spike (panels D - E) **(A)** Scatter plot of inferred w_{ij} on the entire dataset vs reduced dataset (excluding initial 5 minutes of exploration). **(B)** Average w_{ij} on full dataset (colored solid line) vs no first 5 minutes (black solid line). Shaded area corresponds to $99th$ CI. **(C)** Analysis as in Fig. 2.2H for each animal individually on full dataset and excluding first 5 minutes. **(D)** Individually for each spike of each excitatory cell we added a random jitter$\sim \mathcal{U}(-X, X)$ms, for $X \in [10, 20, 30]$. The figure shows scatter of inferred w_{ij} on original dataset vs jittered for familiar (top) and novel (bottom). Dotted lines represent significance (i.e., $z < -4.5$ and $z > 4.5$). **(F)** Proportion of excitatory pairs that pass significancy thresholds for different levels of jitter.

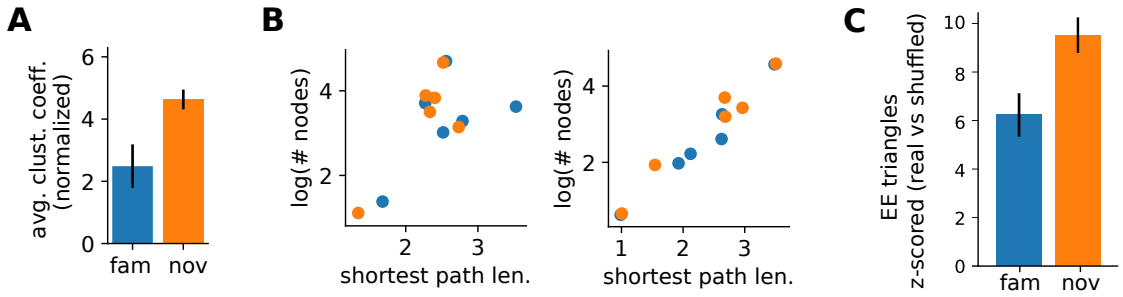

Fig. S. 2.6. Small worldness of EE subnetwork. **(A)** Average clustering coefficient of excitatory subnetworks normalized against the same values computed on ER random graphs with matching edges density (Fig 2). Z-test against $\mu = 0$, fam: p<0.05, nov: p<0.001. Paired t-test: p<0.01 (N=6 animals). Error bars represent $99th$ CI.. **(B)** Left: log-nodes number vs shortest path length in the largest connected component of excitatory subnetworks with standard significancy threshold at $|w| > 4.5$ (two dots per animal: familiar (blue) and novel (orange)). Linear regression: r = 0.66, p=0.014. Right: same as left for excitatory subnetworks with higher significancy threshold at $|w| > 6$. Linear regression: r = 0.97, p<0.0001. **(C)** Overexpression of triangles in real networks against random shuffling of the edges that preserved the number of incident edges onto each single node (see Methods). Z-test, fam: p<0.001, nov: p<0.001. Error bars represent $99th$ CI.

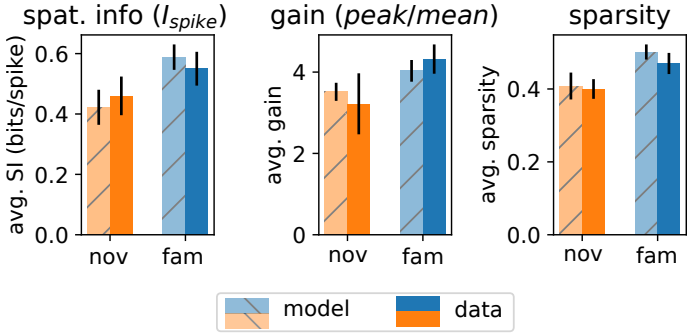

Fig. S. 2.7. Average marginal statistics of place cells in hippocampus match circuit model. The interactions in the model were drawn from the inferred couplings observed in data and rescaled according to Supp. Fig. 2.2A. Afterwards, we fixed the input strength by picking the parameters that allowed the model to best match the marginal statistics observed in data $h =\sim 2$ for novel, $h =\sim 4$ for familiar. All the measures were computed on traditional 2D firing rate maps (see Methods). (left) single cell spatial information, (center) firing rate map gain, measured as peak over mean (right) firing rate sparsity. Error bars represent 99th confidence interval of the mean. Data: N=294, model: N=50.

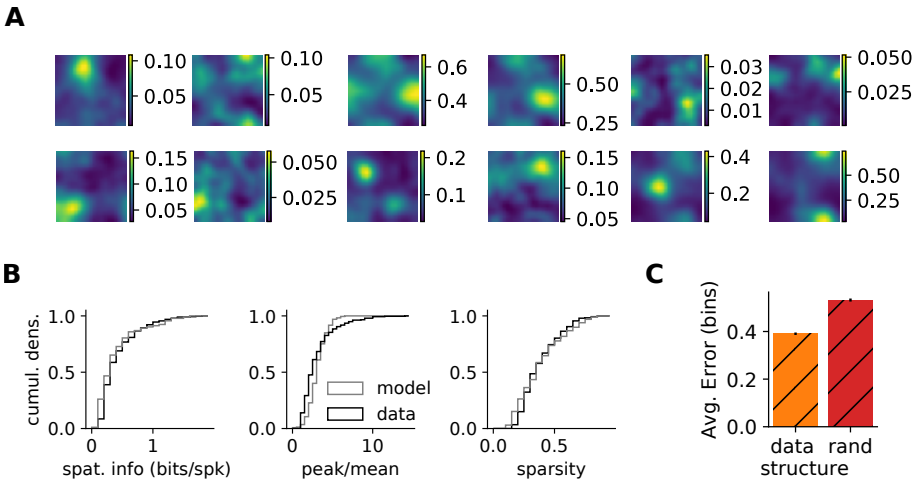

Fig. S. 2.8. Marginal statistics of larger place cell populations can be matched in distribution, and decoding benefits persist. We considered a population of 200 synthetic place cells and introduced heterogeneity in the distribution of input strengths ($\sim \mathcal{N}(2,1)$), number of components ($\sim Pois(3)$) and components weights ($\sim Dirichlet(2)$). The interactions in the model were drawn from the inferred couplings observed in data and rescaled according to Supp. Fig. 2.2A. **(A)** Example firing rate maps computed on $4 \cdot 10^4$ samples. Color bars indicate average number of spikes per sample. **(B)** Distribution of marginal statistics for data (black) and model (grey). All the measures were computed on traditional 2D firing rate maps. Left: single cell spatial information (KS test, p=0.410), middle: firing rate map gain, measured as peak over mean (KS test, p=0.0126), right: firing rate sparsity (KS test, p=0.348). Data: N=294, model: N=200. **(C)** PV decoding error computed on $4 \cdot 10^4$ samples (cross-validated 50/50) for data-like interactions vs randomized ones.

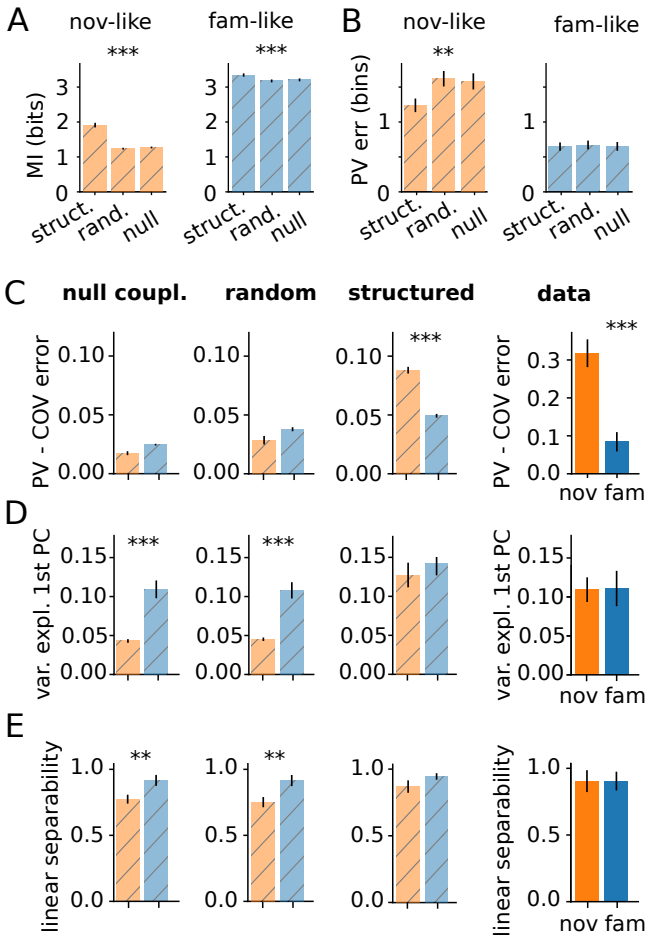

Fig. S. 2.9. Comparison with null couplings. **(A)** Estimated spatial information (MI; error bar – 99th percentile CI for the mean) using structured, random and null interactions, in the novel-like and familiar-like scenario (see text). Structured interactions significantly increase the spatial information ($p < 0.001$ (***) or $p < 0.01$ (**) under a non-parametric Mann–Whitney U-test). **(B)** Decoding error using a simple population vector approach (PV; error bar – 99th percentile CI for the mean) using structured, random and null interactions, in the novel-like and familiar-like scenario. Structured interactions significantly decrease the average decoding error in novel environments ($p < 0.01$ (**) under a non-parametric Mann–Whitney U-test). **(C)** Improvement in decoding performance by taking into account co-variability of cells ("COV" decoder) relative to a simple population vector ("PV") decoder, evaluated on $4 \cdot 10^4$ samples). (error bars and significance tests as in B). **(D)** Fraction of variance explained by the first principal component of population vectors for 10^3 random pairs of locations in the maze. The fraction is unchanged between the novel and familiar environments on structured network and on real data, but differs significantly on the random and null networks (error bars and significance tests as in B). **(E)** Linear separability measured as SVM classification accuracy of random pairs of stimuli (trained on 1000 pairs of same vs. different positions). The separability is unchanged between the novel and familiar environments on structured network and on real data, but differs significantly on the random and null networks.

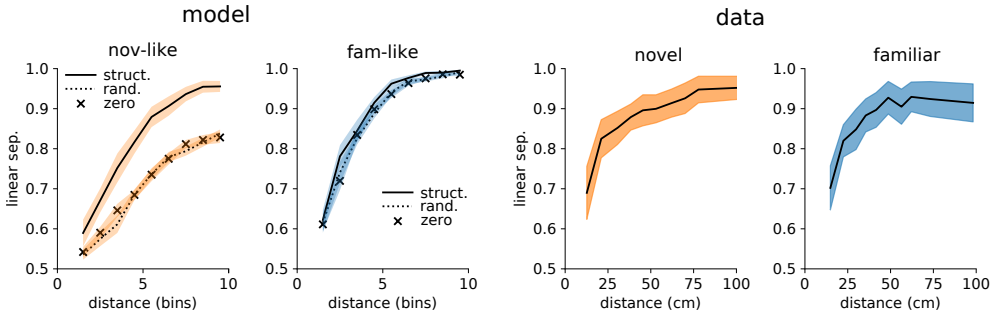

Fig. S. 2.10. Linear separability as a function of distance. Left: linear separability of responses to stimuli at a given distance for data-like interaction structure (solid line), random (dotted) or zero (x) for novel-like (orange) and familiar-like (blue) input quality. Right: linear separability of responses to stimuli at a given distance for data novel environments (orange) and familiar (blue). Shaded area represents $99th$ CI for the mean.

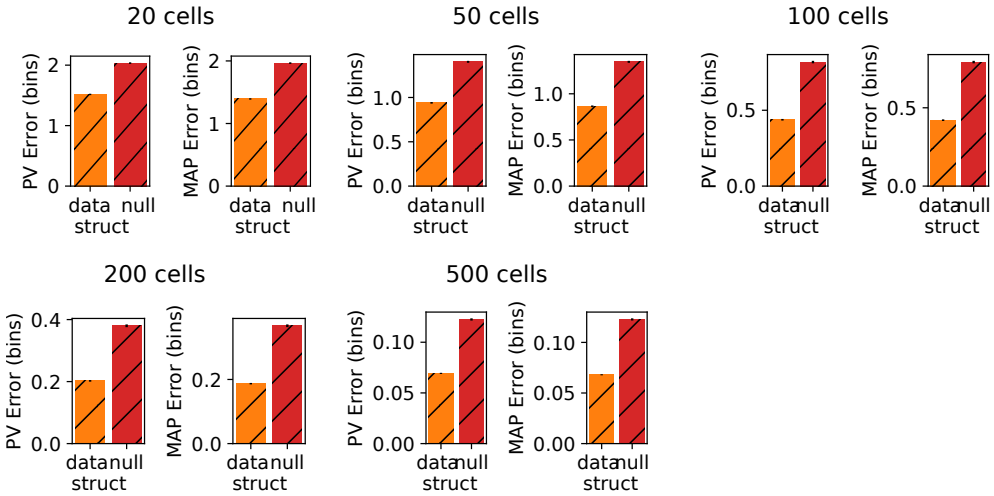

Fig. S. 2.11. Decoding advantages for larger networks. We re-scaled the interaction strengths according to their network size $(1/N)$. We trained a PV decoder and a MAP bayes decoder on $4 * 10^5$ random samples, and then tested the decoding accuracy on 10^5 samples. Decoding accuracy measured in bins (synthetic environments were unit squares and binned in 20 X 20). We run all these experiments with a novel-like input quality. Reported average $\pm$ 99th confidence interval for the mean.

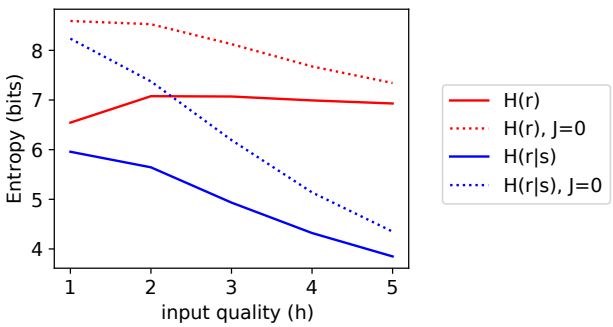

Fig. S. 2.12. Total vs noise entropy in optimized networks. Average total output entropy $H(r)$ (red line) vs noise entropy $H(r|s)$ (blue line) for noetworks of 10 place cells with optimal interactions (solid line) vs same networks with 0 interactions (dotted lines).

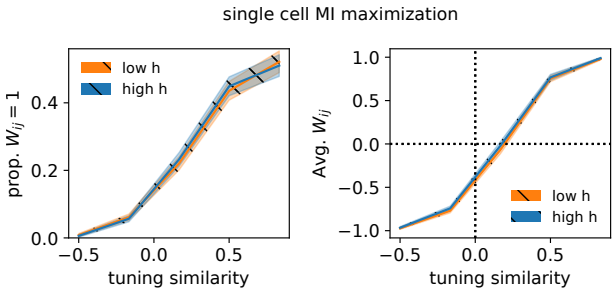

Fig. S. 2.13. Single cell MI optimization. Optimizing the mutual information between single cells stimulus-dependent (marginalized) activity and location-stimulus led to the same result for each level of input quality – almost linear relation between place field overlap and optimal predicted W_{ij}. Shaded area represents $99th$ confidence intervals.

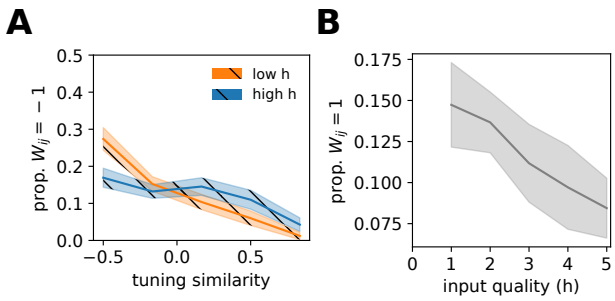

Fig. S. 2.14. Negatively coupled optimized connections and proportion of strongest. A Proportion of cell pairs to reach minimum allowed W_{ij} as a function of tuning similarity. **B** Proportion of cell pairs that reached maximum $W_{ij} = 1$ (after optimization) decreased for increasing input quality h. Shaded area represents $99th$ CI for the mean.

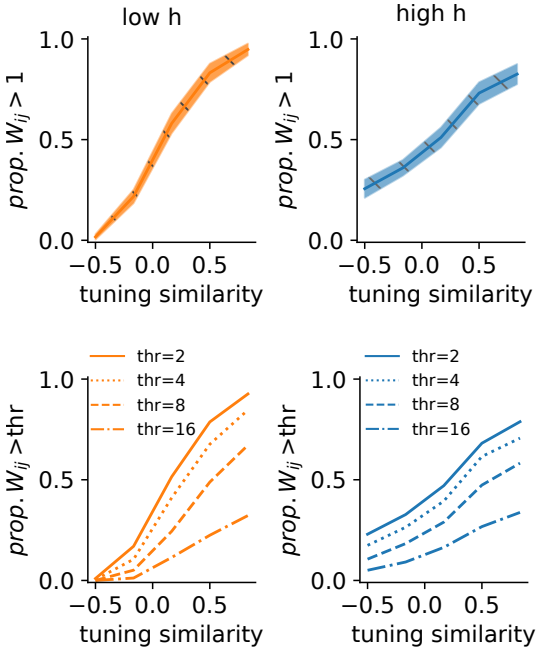

Fig. S. 2.15. Non constrained maximization does not show nonlinear coupling preferences. Top row: Proportion of couplings that exceed 1 after optimization. Couplings were optimized so to maximize the mutual information between population responses and stimuli. The average population firing rate was constrained but W_{ij}s were not. Linear regression: low h: r=0.91, p<0.001; high h: r=0.94, p<0.001. Shaded area represents $99th$ CI for the mean. Bottom row: mean proportion of couplings that exceed different thresholds also do not show the nonlinear relation we observed in the constrained case.

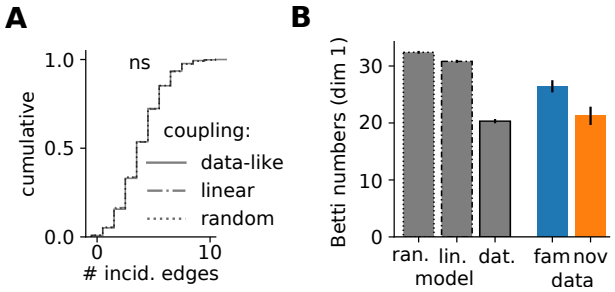

Fig. S. 2.16. Additional topology measures on networks generated with different rules. (A) Distribution of incident edges with the three different connectivity-rules. KS test, for each comparison $p > 0.05$. **(B)** Betti numbers of the clique complex induced by the graph (b_1) for 1-dim holes. Using the data-like nonlinear coupling strategy increased the chance of creating triangles, hence diminishing the number of 1-dim cavities. Error bars: $99th$ CI for the mean.

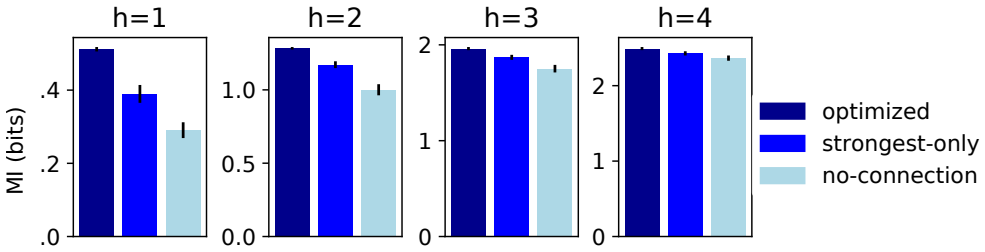

Fig. S. 2.17. Strongest couplings only. After optimizing the connections W (as in Fig. 4), the MI of the fully optimized networks was compared to null couplings and the "strongest only" case, i.e., where every connection $|W_{ij}| < 1$ was set to 0. Error bars: $99th$ CI for the mean.

3

Hippocampus-neocortex interactions

3.1 Aims

The hippocampus and different neocortical areas interact during the encoding, consolidation, and retrieval of memories (Wang and Morris, 2010). A neocortical area that received much scrutiny is the prefrontal cortex (PFC), especially in the context of encoding (Takehara-Nishiuchi and McNaughton, 2008; Barker and Warburton, 2008) and retrieving (Farovik et al., 2008) memories, and its interactions with the hippocampus (Eichenbaum, 2017a). These two areas are directly and indirectly connected (Hoover and Vertes, 2007) and engage in coordinated activity both during awake and sleep periods (Jadhav et al., 2016; Tang and Jadhav, 2019; Zielinski et al., 2019). The importance of PFC and medial temporal interaction has long been established for memory processes such as encoding and retrieval (Simons and Spiers, 2003). Moreover, it has been suggested that PFC may be important for reducing interference during hippocampal-dependent tasks (Guise and Shapiro, 2017), while on the other side inactivating direct hippocampus to PFC projections hinders the encoding of spatial cues and the execution of a spatial working memory task (Spellman et al., 2015). An outflow of spatial and contextual information from the hippocampus to the PFC (Esteves et al., 2021) might also act as a cueing signal required to retrieve the associated cortical trace (Winocur et al., 2007) or to update existing knowledge in a pre-existing schema (Tse et al., 2007; Alonso et al., 2020).

The aim of this chapter is to study the mechanisms that allow the transfer of spatial and contextual information from the hippocampus to the PFC. This will begin by ascertaining whether the flow of spatial information among the two areas is unidirectional or bidirectional. Afterward, using GLM null models as statistical tests, I will detect functional couplings among units across areas, and study the spiking activity of the cells involved in the two areas during synchronized and independent activity. This will allow us to propose a mechanism by which this functional coupling might be implemented, and how assemblies across brain areas are recruited during behavior.

This is a study in collaboration with Karola Kaefer, Federico Stella, and Jozsef Csicsvari.

This study has been submitted for consideration and peer revision to a journal on May 25th, 2022. A previous version of this manuscript, containing some of the results presented here, was previously posted on Biorxiv and can be retrieved here: https://doi.org/10.1101/2021.09.30.462269

Author contributions: M.N., K.K., J.C. conceptualized the study, K.K. conducted the experiments, M.N., K.K., F.S. analyzed the data, M.N. and J.C. wrote the initial version of the

paper, and all authors contributed to manuscript preparation

Theta oscillations as a substrate for medial prefrontal-hippocampal assembly interactions

Michele Nardin[1,@], Karola Kaefer[1,*], Federico Stella[1,*], Jozsef Csicsvari[1,@]

[1] *Institute of Science and Technology Austria (ISTA), Klosterneuburg, 3400, Austria*
[@] *corresponding authors: michele.nardin@ist.ac.at, jozsef.csicsvari@ist.ac.at*
[*] *current address: Donders Institute, Radboud University, Nijmegen 10 6500GL, Netherlands*

The execution of cognitive functions requires coordinated circuit activity across different brain areas that involves the associated firing of neuronal assemblies. Here, we tested the circuit mechanism behind assembly interactions between the hippocampus and medial prefrontal cortex (mPFC) of adult rats by recording neuronal populations during a rule switching task. We identified functionally coupled CA1-mPFC cells that synchronized their activity beyond that expected from common spatial coding or oscillatory influence. During coincident firing, mPFC cells strongly phase locked to CA1 theta oscillations and maintained consistent theta firing phases, independent of the theta timing of their CA1 counterpart. Our results suggest that upregulated theta oscillatory firing of mPFC cells can signal transient interactions with specific CA1 assemblies without mirroring CA1 sequences, thus enabling distributed spatial computations.

3.2 Introduction

Hebb's postulate (Hebb, 1949) that coordinated activity enables neurons to associate with each other and ultimately establish cell assemblies influenced neuroscience from the cellular to system levels for decades. Recent developments in multi-neuronal recordings have provided the means to test the cell assembly hypothesis over larger neuronal populations (Buzsáki, 2010). While neurons can exhibit highly correlated activity, similar firing responses can also arise from common sensory inputs. Reactivation of population firing patterns during sleep provided a potential testing ground for assemblies, considering the diminished sensory input to cortical areas in sleep (Buzsaki et al., 1988). It is also possible to investigate assemblies outside sleep by applying statistical methods that take into account the common sensory inputs (Newsome et al., 1989), or even other factors including common oscillatory drive (Harris et al., 2003).

To understand the contribution of cell assemblies to cognitive functions at the systems level, one needs to determine how assemblies in different brain areas work together and how the activity of assemblies propagates across regions. Synfire chain theory provided a potential mechanism in which the synchronized activity of associated assemblies is part of a synfire chain that propagates across multiple brain regions (Abeles, 1982). Indeed, synchronized network patterns such as hippocampal sharp-wave ripples induce large-scale, brain-wide synchronization (Buzsáki, 2015). Assembly activity in one region can also recruit assemblies with similar coding properties in another region; e.g., during reactivation (Ólafsdóttir et al., 2016; Shin et al., 2019; Ji and Wilson, 2007). However, it is less clear whether information encoded in the form of temporal firing patterns can be efficiently transferred. Hippocampal reactivated firing sequences often do not align with reactivated sequences generated in the medial entorhinal cortex (MEC) or medial prefrontal cortex (mPFC), indicating that temporal coding can show independence across brain areas(O'Neill et al., 2017; Kaefer et al., 2020). In addition, as it

is the case within a region, common inputs can lead to coherent activity without one area influencing the other (Fig. S3.1). Here we simultaneously recorded neuronal population activity in the dorsal hippocampal CA1 area and the prelimbic mPFC during a task requiring switch between spatial or cue-guided rules on a plus maze (Kaefer et al., 2020)(Fig. 3.1A-B). We identified CA1-mPFC assemblies that coordinated their activity independently from common sensory or oscillatory inputs and examined the circuit mechanism behind these assembly interactions.

3.3 Results

Inactivation studies suggest that CA1-mPFC spatial representations mutually influence each other (Patai and Spiers, 2021), even though they code for spatial information in different forms (Kaefer et al., 2020). A large proportion of mPFC cells had symmetric firing fields (Fig 1D), and, unlike CA1, mPFC populations encode relative maze distance more accurately than actual 2D positions (Kaefer et al., 2020). However, spatial representations showed only marginal differences across the two rules (Fig. 3.1C). To quantify the expression dynamics of spatial representations, we tested whether spatial coding in one region lags behind the other. Moreover, to infer whether spatial information in one region is used by the other, we assessed the correlation of decoding errors (Zielinski et al., 2019). The best alignment of decoded positions was seen when mPFC lagged behind CA1 by ≈114ms (Fig. 3.1E, Fig. S3.2), and decoding errors also correlated the strongest at ≈125 ms mPFC delay (Fig. 3.1F). A transfer entropy measure (Bossomaier et al., 2016) that assesses information exchange without assuming spatial coding also indicated a preference for CA1-to-PFC interactions (Binomial test, $p<0.0001$) (Fig. 3.1G). These findings all point to CA1-to-PFC network interactions.

Next, we focused on identifying the individual components of the interregional assembly interactions. Neurons in both regions code for place while also engaging in coherent oscillations. Therefore, we calculated functional cross-correlations of CA1-mPFC cell pairs that measured the strength of correlated firing that occurred independently of spatial firing, oscillations, or other possible common drivers. For each cell, we fitted a generalized linear model (GLM) (Hardcastle et al., 2017) that included spatial position, trajectory, speed, theta oscillatory firing, spiking history, and the within-area spiking of other cells (see Methods). The cross-correlogram calculated from the data was compared (z-scored) to distribution of correlograms generated from simulated spiking (10000 times) using the GML model (Fig. 3.2A,B). In testing different temporal offsets, we found that the maximal percentage of cell pairs with strong functional correlation ($>$4.5 z) occurred with a ≈50ms mPFC delay relative to CA1, and strong functional correlations occurred mostly within a 125 ms time window (Fig. 3.2C). We, therefore, selected functionally-coupled (FC) cell pairs that had $>$4.5 z peak functional cross-correlation within [0,125] ms. Although FC cells synchronized their firing independently of common spatial drive, CA1 FC cells exhibited a tendency for symmetrical place fields (Fig. 3.2D) or fired in a complementary fashion to the symmetrical firing field of their mPFC counterpart (Fig. 3.2E). Moreover, the CA1 place fields predicted the firing field of the FC mPFC cell (Fig. 3.2F). A computational model of spatial information transfer that maximizes mutual information predicted these connectivity rules (Fig. S3.3).

FC mPFC cells exhibited stronger phase locking to CA1 theta oscillations than non-FC cells (Fig. S3.4A). It has been previously reported that mPFC cells show the best phase locking to CA1 theta when a fixed time delay is used (Siapas et al., 2005) and that mPFC theta oscillatory coupling to CA1 changes in strength depending on the cognitive demand (Benchenane et al.,

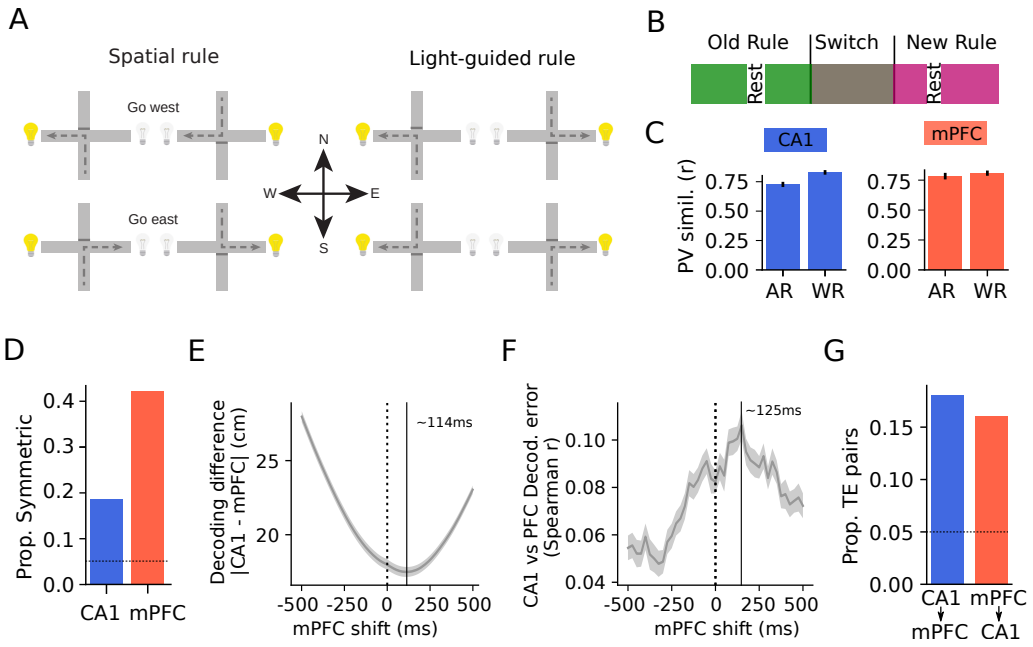

Fig. 3.1. Delayed spatial coding in the mPFC relative to the hippocampus.
(A) Behavioral rules applied during the task. **(B)** Experimental paradigm: three sessions separated by rest periods and the rule is switched in the middle session. **(C)** Similarity of average firing rate population vectors (PV) across rules (AR) vs. within rules (WR) for CA1 and mPFC putative principal cells. **(D)** Proportion of cells exhibiting significant symmetric place fields. Binomial test, CA1: $p<0.001$, mPFC: $p<0.0001$. **(E, F)** Correlation of predicted locations and decoding errors as a function of mPFC time shift. Shaded area: 95% CI, $n=13$ sessions. **(G)** A larger proportion of pyramidal cell pairs exhibiting significant (i.e., higher than 99th percentile of 200 random time shifts) transfer entropy (TE) in the CA1→mPFC direction. Binomial test, $p<0.0001$.

2010; Jones and Wilson, 2005). Thus, for each mPFC cell, we established the optimal delay for CA1 theta phase locking separately at different task phases in the start and goal arms. The median delay was 23 ms in the start arm and 47ms in the goal arm (Fig. S3.4B). The histogram of significant interactions showed similar delays of 25 ms (start arm) vs. 75 ms (goal arm) (Fig. S3.4C).

Individual functional cross-correlations also exhibited theta-modulated peaks (Fig. 3.3A, Fig. S3.5). Although theta-modulated cross-correlations can be seen when both cells phase lock to theta oscillations but otherwise fire independently, in our analyses this common modulation is accounted for by the GLM (Fig. 3.2A, 3.3A). This indicates a transient increase in phase locking when FC cells fire together. Furthermore, the activity of CA1 cell may signal the enhanced phase locking of their mPFC counterparts during synchronized firing. To directly test this, we examined whether mPFC cells exhibited better phase locking to theta oscillations when they fired together with their FC CA1 counterparts (Fig 3B-C). Indeed, when FC cell pairs fired in the same theta cycle, mPFC theta phase locking was stronger than for randomly sampled spikes from the same mPFC cell (Fig. 3.3D, top row). Moreover, such a difference was not significant when non-FC pairs were tested (Fig. 3.3D, bottom row). Although we observed enhanced phase locking of mPFC cells with their FC CA1 counterpart, this does not imply that it is required to detect a functional correlation. To verify that the strengthened mPFC theta phase locking was a critical factor, we randomized the theta spike timing of the mPFC cells during synchronized firing with the FC CA1 counterparts. The cross-correlation peak was reduced by ≈50%, and the number of significant cells dropped to a quarter of the original number (Fig. S3.6). These results demonstrate that the strengthening of mPFC phase

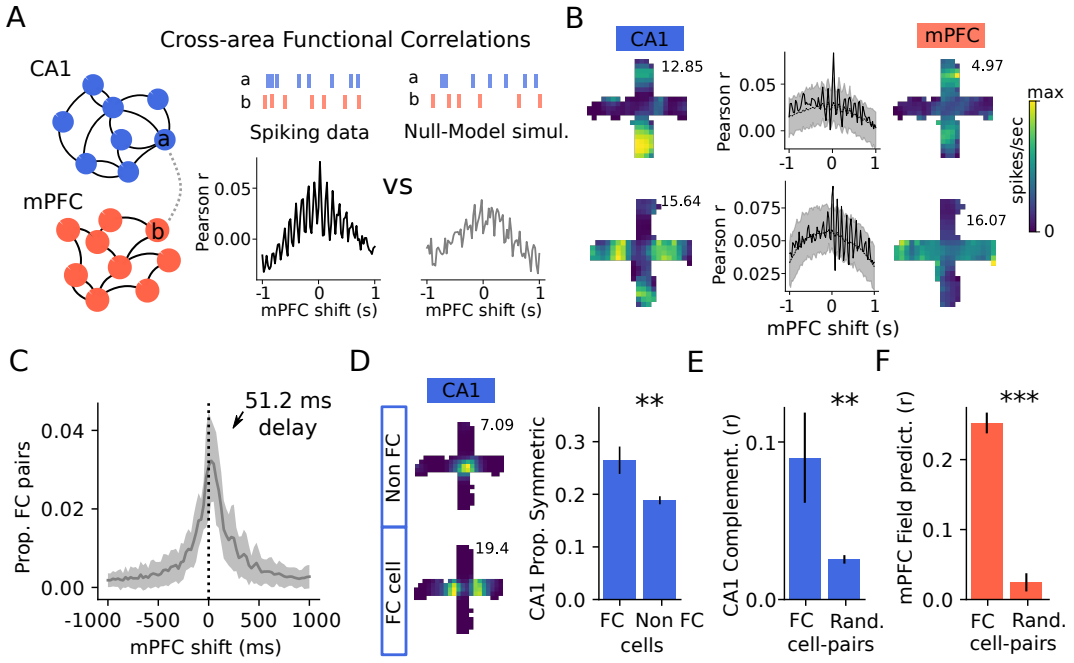

Fig. 3.2. CA1-mPFC functional correlations.
(A) Functional correlations (FC) calculations: real data cross-correlations were compared to a null distribution from 10000 GLM simulated cross-correlations **(B)** Example cross-correlations with significant interaction (z>4.5 relative to GLM). Left and right: CA1 and mPFC firing rate map. Center: cell-pair cross correlograms, real data (solid black line); null model mean (dotted back) ± 4.5 STD (shaded area) **(C)** Proportion of significant FC pairs (z>4.5) as a function of mPFC delay. Shaded area: 95% CI N=13 sessions. **(D)** Left: example place fields of non-FC and FC cells. Right: proportion of symmetric place fields. **(E)** Complementarity (i.e., firing at opposite arms, see Methods) of rate maps of the two CA1 units with the strongest functional correlation to the same mPFC unit (left) vs. random (right). **(F)** Prediction of mPFC firing map by FC CA1 firing maps weighted (left) vs. random (right).(D,E,F) Mann–Whitney U test: ** = p<0.01, *** = p<0.001. Error bars: ± 95% CI.

locking to hippocampal theta is a critical mechanism for CA1-mPFC functional coupling.

Although CA1 cells exhibit, on average, a phase preference for theta oscillations, they can also systematically change their theta phases during theta phase precession (O'Keefe and Recce, 1993). However, mPFC cells strongly phase lock to theta during synchronized firing with FC CA1 counterparts. Thus, mPFC cells may either maintain a consistent theta phase or mirror theta phase changes of their CA1 counterparts. We calculated functional CA1-mPFC cross-correlations separately for cases when the CA1 cell fired at different theta phases (Fig. 3.3E). The plot of significant interactions now exhibited theta modulated peaks, and the peaks shifted in the opposite direction relative to the CA1 cells' theta phase (Fig. 3.3F). This indicates that the correlation delay compensated for the phase shift of the CA1 cell. To verify this, we calculated the average phase difference of FC mPFC-CA1 pairs when CA1 cells fired at different theta phases; the phase difference opposed the CA1 phase changes (Fig. 3.3G). Furthermore, the mPFC theta phase at the goal-arm lagged behind with ≈60° relative to that on the start-arm (Fig. 3.3G), which explains the different time delays for functional correlations (Fig. S3.4C). These findings show that during periods of synchronized firing, each mPFC cell maintained a consistent theta phase and do not follow the theta phase change of their CA1 counterpart.

Next, we tested whether gamma (or other) oscillations contributed to the functional coupling. Gamma oscillations can synchronize distributed sensory representations (Fries et al., 2007) or spatial representations during memory encoding (Colgin et al., 2009). We compared the

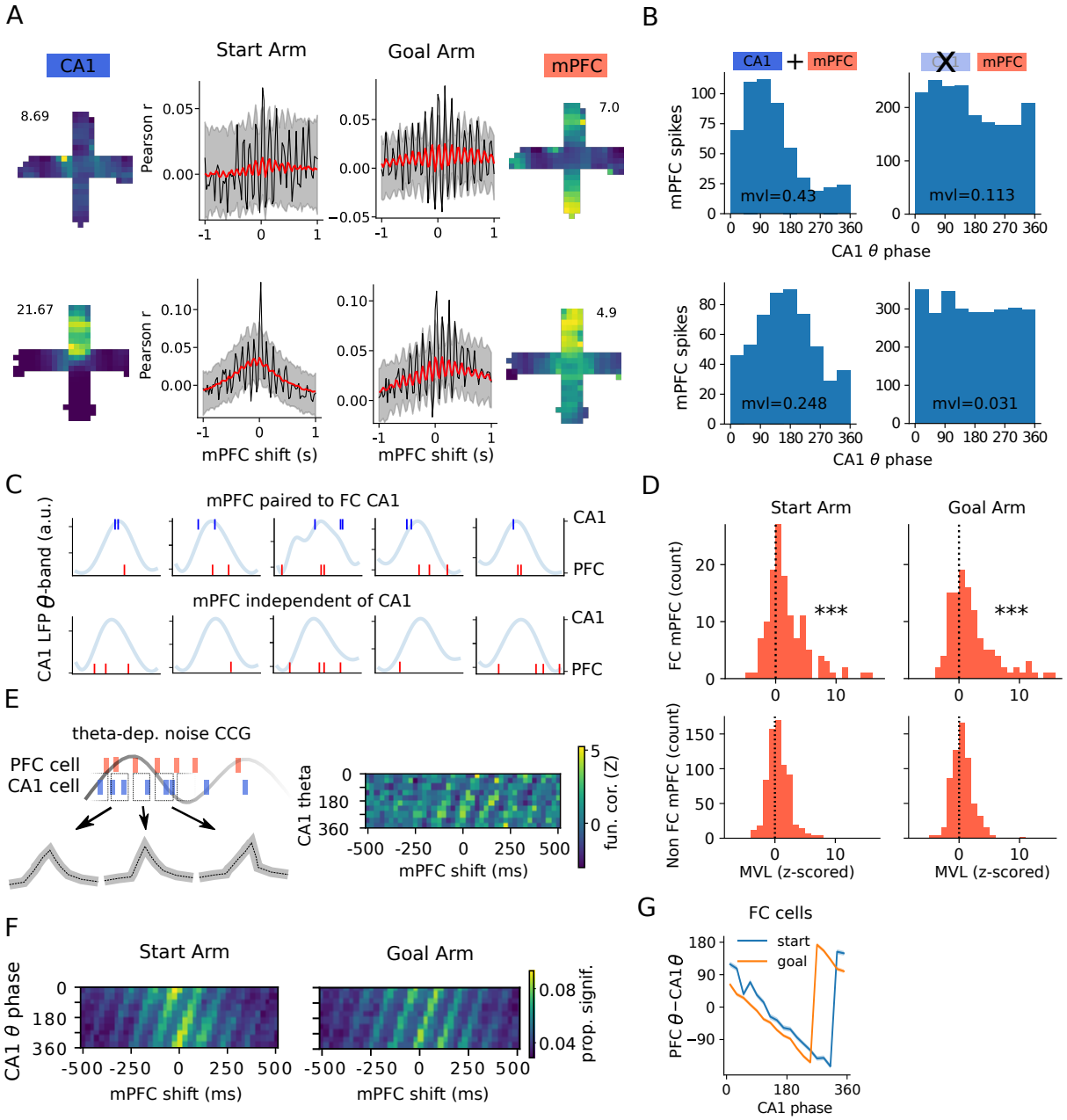

Fig. 3.3. mPFC cells strengthen their phase locking to theta oscillations when cofiring with their FC CA1 counterpart.
(A) Examples of FC cell pairs. Left and right: CA1 and mPFC firing rate map. Center: real cross-correlogram (solid black line) vs. GLM simulation (solid red: average, shaded area: ± 4.5 STD) for start and goal arms. **(B)** Example theta phase histograms when mPFC cells fired together at the same theta cycle with FC CA1 counterparts (left) vs. fired independently (right). Trough of oscillation: 0° **(C)** Example theta cycles with paired CA1 / mPFC spikes vs. independent mPFC firing. **(D)** Top: theta phase locking strength of mPFC cells during paired firing with FC CA1 counterparts. Mean vector length (MVL) during paired activity was z-scored against 200 bootstrapped measurements during independent activity Binomial tests: $p<0.001$. Bottom: same as Top, for randomly selected non-FC pairs. Binomial tests: $p>0.01$. **(E)** Left: schema of theta-dependent functional correlation calculation. Right: example theta-dependent functional correlation. **(F)** Proportion of $z>3$ FC pairs as a function of CA1 theta phase and mPFC time shift. **(G)** Average mPFC - CA1 theta phase difference at different CA1 firing phases. Shaded area: 95% CI

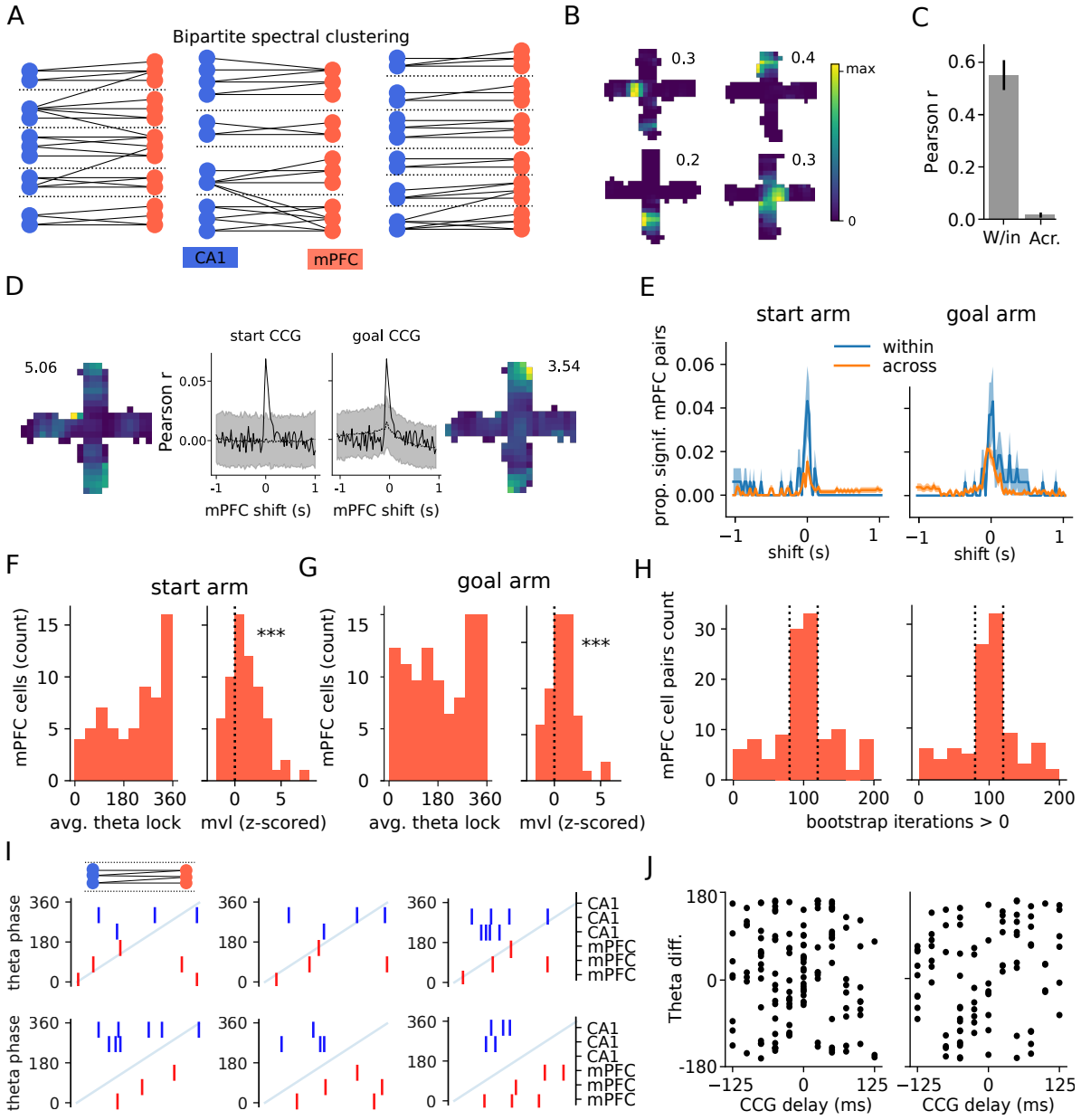

Fig. 3.4. Graph-assembly activity is signaled by mPFC theta sequences.
(A) Graph-assembly examples. Solid lines: functional correlations, dotted lines: assembly boundaries. **(B)** Example graph-assembly firing rate maps. **(C)** Graph-assembly rate map similarity between first vs. second half of recordings (W/in), across different assemblies (Acr.). Mann–Whitney U test: $p<0.001$. **(D)** Example of FC mPFC pair within the same assembly. **(E)** Proportion of significant ($z>4,5$) FC mPFC pairs within vs. across assemblies. Shaded area: 95% CI N=13 sessions. Mann–Whitney U test: $p<0.01$ for shifts = -25, 0, +25 ms. **(F,G)** Histogram of average theta phase (left) and z-scored phase locking strength (right) of mPFC cells during assembly activity. Phase locking strength was z-scored relative to bootstrapped non-assembly activity. Binomial tests: all $p<0.001$. **(H)** Bootstrapped iterations count where mPFC average theta selectivity was higher than the one of its mPFC peers within the same assembly during graph-assembly activation. Dotted lines mark the boundary for the pairs in which the preferred theta phases of the pair were significantly different: 45% start arm and 42% goal arm (Binomial test $p<0.05$). **(I)** Example mPFC sequences during graph-assembly activation. **(J)** The preferred theta phase difference of mPFC cells during graph-assembly activation vs. long-time scale cross-correlation (CCG) delay. Left: start arm (Spearman r=0.01, p=0.9), Right: goal arm (r=0.19, p=0.058).

power spectrum and coherence of the CA1 and mPFC local field potential at times when FC CA1-mPFC cells fired together or independently (Fig. S3.7). No differences were seen in any of the frequency bands, suggesting that selective mPFC cell theta phase locking is the primary mechanism behind the CA1-mPFC functional coupling, and these coupling events do not influence the global oscillatory state of the circuits per se.

So far, we have identified functional correlations between individual CA1-mPFC pairs. Can this pairwise connectivity reveal organized interregional assemblies? When we plotted the graph of FC cells, a grouping tendency emerged (Fig. 3.4A). We applied a bipartite spectral graph-clustering algorithm (Pothen et al., 1990; Newman, 2006) that divided cells into sub-graphs and examined whether these subgraphs met the criteria of functional assemblies. First, we confirmed that mPFC cells within a graph-assembly exhibited stronger functional connections with each other than with mPFC cells of different assemblies (Fig. 3.4D,E). We then identified times when graph-assemblies were active by selecting theta cycles in which at least two mPFC and two CA1 cells from the same graph-assembly fired together. When such a graph-assembly was active, we found that mPFC theta phase locking was stronger as compared to out of assembly activity (Fig. 3.4 F-G). Finally, we showed that assembly activity encoded spatial locations consistently over time and the firing fields of different graph-assemblies were different (Fig. 3.4B,C).

FC mPFC cells tend to fire at their preferred theta phases during graph-assembly activation. In doing so, they may all fire at similar theta phases or in a sequential pattern through firing at different theta phases. First, we calculated the preferred theta phase of mPFC cells during assembly activation. Preferred theta phases spanned the entire theta cycle (Fig. 3.4F-G). During graph-assembly activation, mPFC cell members may therefore exhibit a consistent temporal bias in their spike timing to theta oscillations. For each mPFC cell pair from the same graph-assembly we tested whether they exhibited a significant theta temporal bias during assembly activation, i.e., one cell consistently fired at an earlier theta phase (i.e., $<180^{\circ}$) than the other. Indeed, in total, 45% (start arm) and 42% (goal arm) of FC mPFC pairs exhibited a significant bias (Fig. 3.4H, binomial tests: all $p<0.001$), confirming that at the population level, a significant proportion of cells showed a theta phase bias. Therefore, during assembly activation, a significant portion of mPFC cell pairs show a preferred temporal order in their activity, leading to leading to sequentially-biased firing patterns (Fig. 3.4I). We expect such sequences to be independent of CA1 firing patterns considering that mPFC cells decouple from the theta spike timing of CA1 cells (Fig. 3.3E-G). Theta sequences in the hippocampus reflect the compressed time order at which place cells fire on a linear maze (Dragoi and Buzsáki, 2006). If mPFC cells of a graph-assembly exhibit similar maze-related activation sequences, the temporal activation order of mPFC cells (seen on their long-range cross-correlograms) should predict their theta phase bias during graph-assembly activation (Fig. 3.4J). However, this was not the case, suggesting that during assembly activation, mPFC cells can exhibit unique sequences which are independent of their longer-time scale activation patterns driven by maze trajectories. Hence these sequences could represent unique non-spatial information and signal functional interactions with specific CA1 assemblies.

3.4 Discussion

Here we identified a subgroup of mPFC and CA1 cells that synchronized their activity beyond that expected from similar spatial coding, oscillations, and background network activity. When such CA1-mPFC pairs fired together, mPFC cells enhanced their phase locking to hippocampal

theta oscillations and maintained a consistent theta phase, even when the phase of their CA1 counterpart varied. These findings suggest a novel mechanism for interregional assembly interactions that goes beyond coherent oscillatory coupling or temporal synchronization(21). While theta oscillations provided a reference frame for synchronization, the enhanced phase locking of mPFC cells alone signaled interaction with their CA1 counterparts. The critical role of this enhanced phase locking was demonstrated by the observation that randomizing mPFC theta spike timing during synchronization abolished about 75% of the detected functional correlations and halved the correlation peaks. Furthermore, the relative spike timing of CA1-mPFC pairs could vary within a theta cycle; therefore, coherent theta oscillatory firing of the pair was not required. Hence, interactions here were not signaled by changes in the firing rate or coherent oscillatory coupling across the regions. Rather transient increase of mPFC phase locking to theta oscillations alone was the primary driver that marked the functional coupling.

This assembly interaction scheme also allows different temporal coding formulas to operate independently in the two cortical regions. While the theta spike timing of CA1 cells can code for place because of theta phase precession (O'Keefe and Recce, 1993), mPFC theta phases can represent other information. The theta phase at which an mPFC cell fired signaled the activity of its FC CA1 counterparts, thus, indicating times when a specific CA1 assembly was active. Different mPFC cells fired at different theta phases enabling them to jointly fire in theta sequences when their CA1 assembly counterparts were active. This suggests that specific theta sequences in the mPFC can signal interactions with specific CA1 assemblies. The importance of single theta oscillatory cycles as units of assembly coding has been demonstrated before, primarily in the hippocampus and MEC, where theta sequences can encode trajectories and even entire spatial maps can flicker across theta cycles (Jezek et al., 2011; Boccara et al., 2019). Similar coding may apply to other brain areas as well, considering that theta oscillations are seen in widespread brain regions beyond the hippocampal formation (Buzsáki, 2002). Theta-related assembly synchronization can enable multiple regions to function and interact with each other in a parallel manner in the limbic system and beyond. Our data also indicate that assembly interactions through theta synchronization can allow each region to use its own unique temporal coding scheme during the time of interactions. Temporal coding may not be simply transferred but often transformed across different cortical processing stages, necessitating further experimental and theoretical work to understand how temporal codes in one region are decoded and transformed downstream.

3.5 Materials and Methods

Experimental methods

The data used in this study is the same as used in (Kaefer et al., 2020). We will report the experimental methods here for completeness.

Subjects and Surgery Four male Long-Evans rats (300-350 g, 2-4 months of age; Janvier, France) were used in this study. The animals were housed in a separate room on a 12 hour light/dark cycle and were taken to the recording room each day prior to the experiments. Animals shared a cage with littermates before surgery. All procedures involving experimental animals were carried out in accordance with Austrian animal law (Austrian federal law for experiments with live animals) under a project license approved by the Austrian Federal Science Ministry (License number: BMWFW-66.018/0015-WF/V3b/2014). Rats were implanted with microdrives housing 32 individually-movable tetrodes, arranged into three bundles targeting the right dorsal hippocampus (specifically dorsal CA1, HPC) and left and right medial prefrontal cortex (specifically prelimbic area, mPFC). The HPC bundle consisted of 16 tetrodes and the two mPFC bundles of 8 tetrodes each. Tetrodes were fabricated out of four 12 mm tungsten wires (California Fine Wire Company, Grover Beach, CA) that were twisted and then heated to bind into a single bundle. Tetrode bundle lengths were cut so that the two mPFC bundles were 1-1.5 mm

longer than the HPC bundle. The tips of the tetrodes were gold-plated to reduce the impedance to around 300 kU. Before surgery, the animal was put under deep anesthesia using isoflurane (0.5%–3%), oxygen (1–2 L/min), and an initial injection of buprenorphine (0.1 mg/kg) and ketamine/xylazine (7:3 ketamine (10%) and xylazine (2%), 0.05ml/100 g). Craniotomies were drilled above the HPC (AP: 2.50 to 4.50, ML: 1.2 to 3.6) and above the mPFC across the sinus (AP: 4.60 to 2.50, ML: 0 to $\pm$ 0.8). Six anchoring screws were fixed onto the skull and two ground screws were positioned above the cerebellum. After dura removal, the tetrode bundles were centered above their respective craniotomies and lowered into the brain at a depth of 2 mm for the mPFC and 1 mm for the HPC. The exact depth of mPFC tetrode implantation was noted to ensure later lowering into the target area. Tetrodes and craniotomies were coated in paraffin wax and the microdrive was anchored to the skull and screws with dental cement. The analgesic meloxicam (5 mg/kg) was given up to three days after surgery and the animal was allowed one week of recovery. Thereafter, tetrodes were gradually moved in 50-200 mm steps into the HPC pyramidal cell layer and mPFC.

Plus maze apparatus and task (Fig. 3.1A,B) Following the recovery period, animals were food-restricted with ad libitum access to water and accustomed to the plus maze and rest box. The plus maze was elevated (80 cm) and consisted of four arms (85 cm long and 12 cm wide), referred to as north, east, south, and west, and a connecting center. The animal was placed in one of the two start arms (north or south) and had to collect a food reward (MLab rodent tablet 20mg, TestDiet, Richmod, USA) in one of the two goal arms (east or west), depending on the rule employed. Access to the arm not chosen as the start was restricted, so that the maze became T-shaped. A small light at the end of one of the two goal arms was switched on. Which arm was chosen as the start and light-on arm was chosen pseudorandomly for every trial, ensuring that an arm was not chosen more than three consecutive times. Once the animal reached a goal arm and 5s passed, the animal was manually picked up and placed in the rest box before commencing to the next trial after a delay of 10s. The animal had to retrieve the reward based on a spatial or response (light) rule. During the spatial rule the reward was always placed in either the east or west arm, while during the response rule the reward was placed in the light-on arm. Importantly, also during the spatial rule one of the two arms was lit, but did not necessarily indicate the location of reward. To prevent an odor-guided strategy pellet dust was scattered along the maze and pellet-filled cups invisible to the animal placed under both goal arms. On each recording day, the animal underwent behavioral blocks as follows: rest, rule 1 (previous day's old rule), rest, pre-switch, rule switching, post-switch, rest, rule 2 (new rule), rest. After the first rest, the animal started by performing trials based on the previous day's old rule. After reaching performance criterion (see below), the animal rested again and afterwards the rule switching phase began. During the pre-switch block the animal had to collect reward based on the last rule of the previous day until reaching the performance criterion (see below). Then the rule was changed and reward had to be collected based on the new rule. The change in rule was not announced to the animal, which had to switch to the new rule through trial-and-error until performing to criterion. Trials performed after the rule change, but before the animal reached good performance comprised the rule switching block, while the post-switch block comprised all trials from the beginning of good performance (defined in the next sub-subsection). The animal had to perform cross-modal switches, i.e., switches from spatial to light or light to spatial rule, never between the two spatial rules. While correct performance of a spatial rule involves two trajectories (e.g., go-east rule: north to east and south to east), correct performance of the light rule can involve any of the four trajectories. Therefore, the performance criterion for the spatial rule was set to 12/15 and for the light rule to 24/30 correct trials, ensuring a similar number of light rule trials where the animal performed trajectories that matched those of the spatial rule. After another rest session, the animal performed a final 20 trials of the newly acquired rule.

Beginning of good performance All trials before the rule change comprised the pre-switch block. Trials performed after the rule change, but before the animal reached good performance comprised the switching block. The beginning of good performance (bgp) was defined as the center index after rule change where the error rate over five consecutive trials dropped to zero.

Histology and reconstruction of recording positions After the final recording day tetrodes were not moved. Animals were administered ketamine/xylazine (7:3 ketamine (10%) and xylazine (2%), 0.1ml/100 g) and overdosed with pentobarbital (300mg/ml) before being transcardially perfused with 0.9% saline followed by 4% formaldehyde. Brains were extracted and stored in 4% formaldehyde. On the same day brains were transferred into 30% sucrose solution until sinking for cryoprotection. Finally, brains were quickly frozen, cut into coronal subsections with a cryostat (50-60 mm), mounted on glass slides and stained with cresyl violet. The positions of tetrode tips were determined from stained subsections and cells recorded from tetrodes outside mPFC were excluded from analysis. For cells recorded from HPC tetrodes the presence of SWRs in the field recordings served as inclusion criteria.

Data Acquisition The extracellular electric signals from tetrodes were pre-amplified using a headstage (4 x 32 channels, Axona Ltd, St. Albans, Hertfordshire,

UK). The amplified local field potential and multiple-unit activity were continuously digitized at 24 kHz using a 128- channel data acquisition system (Axona Ltd). Two red LED bundles mounted on the preamplifier head-stage were used to track the x, y location of the animal. Every day before recording, HPC tetrodes were moved optimizing the yield of recorded cells. Additionally, mPFC tetrodes were lowered every day by 30-50 mm to ensure recording of a new population of cells.

Spike sorting and unit classification Clustering of spikes and unit isolation procedures were described previously (Csicsvari et al., 1998). Briefly, the raw data was resampled to 20 kHz and the power in the 800-9000 Hz range was computed for sliding windows (12.8 ms). Action potentials with a power of > 5 standard deviations (SD) from the baseline mean were selected and their spike features extracted with principal components analysis. Action potentials were then grouped into multiple putative units based on their spike features using an automatic clustering software (http://klustakwik.sourceforge.net; (Harris et al., 2000)). The generated clusters were then manually refined using a graphical cluster-cutting program and only units with clear refractory periods in their autocorrelation, well-defined cluster boundaries and stability over time were used for further analysis. An isolation distance (based on Mahalanobis distance) was calculated to ensure that spike clusters did not overlap (Harris et al., 2000). Putative excitatory principal cells and inhibitory interneurons were discriminated using their auto-correlograms, firing rates and waveforms (Csicsvari et al., 1999).

Dataset details Across the 4 animals, we recorded, processed, and analyzed data from a total of 13 experimental recording days (4,4,3,2 respectively). In all our analysis we included only cells with an average firing rate > 0.25spikes per second in each of the three experimental phases. This comprised a total of 530 hippocampal (putative) principal cells and 160 interneurons, and 477 prefrontal (putative) principal cells and 105 interneurons.

Statistical analysis

Linearized Position To linearize the behavior of the animal, we calculated the distance to the center from the 2D spatial position of the animal (Kaefer et al., 2020). This way a "V-shaped" positive function for each trial was obtained. For each position before the center (i.e., before the global minimum) we subtracted the minimum and then changed the sign. Then, 100 was added to every position to obtain a positive measure of the relative position of the animal between start (0 cm) and goal (200 cm). The center corresponded to 100 cm.

Firing rate maps We inferred the average firing of each cell at each given location of the environment separately for three cases: 2D maps, 1D trajectory dependent, and 1D trajectory independent. In every case, we utilized data from periods when the animal was moving faster than 7cm/s to avoid potential nonlocal population activity (Kaefer et al., 2020).

- **2D maps** We binned the x, y locations of the animal in 5cm square bins. As explained in the experimental setup, spatial coordinates were acquired every $dt = 25.6ms$. We counted how much time (in seconds) the animal spent in each location, which corresponded to the occupancy map. Denoting with $\hat{x}(t), \hat{y}(t)$ the coordinates of the animal at time t, we have

$$occ(x, y) = \sum_t (\hat{x}(t) == x) \cdot (\hat{y}(t) == y)/dt$$

where $a == b$ is 1 if and only if $a = b$ and is 0 otherwise. Afterwards, for each cell i, we counted the number of spikes emitted in each location, and divided by the time spent there. In formulas, denoting by $t_1^i, \ldots, t_K^i$ the times at which K spikes were emitted, we define the rate

$$\hat{\lambda}(x, y) = \sum_{t=t_1^i,\ldots,t_K^i} (\hat{x}(t) == x) \cdot (\hat{y}(t) == y)/occ(x, y)$$

Finally, we regularized $\hat{\lambda}(x, y)$ by convolving with a symmetric 2D Gaussian kernel $\kappa(h, k)$ with $\sigma = 2$bins and truncated at a maximum width $w = 10$ bins:

$$\lambda(x, y) = \sum_{(h=-w,\ldots,w,k=-w,\ldots,w)} \hat{\lambda}(x+h, y+k)\kappa(h, k)$$

Locations x, y with zero occupancy were excluded from the convolution. Locations near the border used only the portion of the kernel that would not exceed the edges (portion appropriately rescaled so as to sum up to 1).

- **Trajectory-dependent 1D maps** Since the arms of the maze are relatively narrow, we computed a linearized version of the firing rate maps, effectively yielding an average firing as a function of the distance to the goal, separately for the 4 trajectories: North to East(NE), North to West (NW), South to East (SE), South to West (SW) (Fig. 3.1A). In words, we selected the trials where the animal followed one trajectory, and with that data binned the linearized position in 10 cm bins and counted the time spent at each discrete location. Afterward, for each cell and each trajectory, we counted the number of spikes emitted in each binned location and divided it by the time spent there. Finally, we convolved the rate maps of each cell for each trajectory with a 1D Gaussian kernel with $\sigma = 1$bin.

- **Trajectory independent 1D maps** We computed the same quantity as above, without separating the trials into 4 trajectory groups.

Population vector similarity (Fig. 3.1C) Population vector (PV) similarity is a measure that allows us to quantify the change in average population activity at any given location across contexts. The PV at a particular location x represents the vector of the average activities of all the cells in the population under study:

$$PV(x) = \{\lambda_1(x), \ldots, \lambda_N(x)\}.$$

We constructed them starting from the previously computed firing rate maps, separately for the two populations and the two rules. We employed a Pearson correlation to quantify the similarity of the PV for each location across rules.

Symmetry of firing rate maps (Fig. 3.1D, Fig. 3.2D) We measured the "mirrordness" of firing rates to check whether single cells fire similarly at opposite sides of the track. To do that, we measured the Pearson correlation of the linearized firing rate maps of opposite arms (S vs N and E vs W) and computed the average of the two measures. Afterward, we computed a null-distribution of symmetry measures obtained from 200 random wrapping of the linearized rate maps (for each cell individually). We deemed a rate map symmetric if its symmetry score was higher than the 95th percentile of the corresponding null-distribution. The random wrapping of rate maps worked as follows. Consider a rate map $\lambda = \{\lambda(0), \ldots, \lambda(M)\}$, ordered as usual from location 0 to M. We randomly selected an integer $k \in \{1, \ldots, M-1\}$, and considered the shuffled rate map

$$\lambda_{shuf} = \{\lambda(k), \ldots \lambda(M), \lambda(0), \ldots, \lambda(M-k)\}$$

Decoding of distance to the goal (Fig. 3.1E, Fig. S3.2) For each decoding scheme, we randomly selected 75% of the trials for computing the firing rate maps, and used the remaining 25% for decoding and assessing decoding quality. We employed a maximum a posteriori (MAP) Bayesian decoding under the assumption that cells are independent and fire according to a Poisson distribution (Zhang et al., 1998). Denoting the firing rate (measured in spikes per second) of cell i at location x with $\lambda_i(x)$, and the same cell emitted n_i spikes in a period dt of time (measured in seconds), the probability that the animal is at location x given the spiking activity is:

$$P(x|n_i) \propto P(x)P(n_i \mid x)$$

where

$$P(x)P(n_i \mid x) = P(x)\exp(-dt\lambda_i(x))(dt\lambda_i(x))^{n_i}/n_i!$$

Combining the information from multiple cells $i = 1, \ldots, N$, and assuming independence, we obtain:

$$P(x \mid \{n_1, \ldots, n_N\}) \propto \prod_{i=1}^{N} P(x|n_i)$$

Assuming that $P(l)$ is uniform, we defined our decoded location x_{MAP} as

$$x_{MAP} = arg\max_x P(x \mid \{n_1, \ldots, n_N\})$$

Correlation of decoding errors (Fig. 3.1F) To check whether the errors of one area agree with the errors of the other area and whether this effect increases at a delay, we employed a delayed version of an analysis proposed by Zielinski and colleagues (Zielinski et al., 2019). If the spatial encoding of PFC is delayed compared to CA1, in that either PFC receives information from CA1, or that information takes longer to reach PFC, then spikes in PFC should represent spatial information that is older than CA1's. Hence, we took all mPFC spiking times and added a time lag $\tau \in [-500, 500]$ms, and for each lag we computed 1D firing rate maps on a fraction of data (random 75%) and decoded the position from the remaining part of the data. Afterward, for each delay, we computed the Spearman correlation between 1D decoding errors from CA1 (without lag) and the lagged PFC activity.

Transfer Entropy (Fig. 3.1G) Transfer entropy is a non-parametric measure of directed (time-asymmetric) transfer of information between two random processes (Wibral et al., 2014). Transfer entropy from a process X to another process Y is the amount of uncertainty reduced in future values of Y by knowing the past values of X, conditioned on past values of Y. More specifically, if X_t and Y_t denote two random processes, transfer entropy from X_t to Y_t is defined as the conditional mutual information between Y_t and the history of X_t, denoted by $\mathbf{X}_{t-1,t-2,\ldots}$, conditioned on the history of the influenced variable $\mathbf{Y}_{t-1,t-2,\ldots}$:

$$T_{X\to Y} = I(Y_t; \mathbf{X}_{t-1,t-2,\ldots} \mid \mathbf{Y}_{t-1,t-2,\ldots}).$$

We used the JIDT package (Lizier, 2014) to estimate this quantity, which was measured across binned (25.6ms) and binarized spike trains of each pair of CA1-PFC cells. Significance was measured by comparing the actual value against 1000 values obtained by randomly shifting the spike trains (uniform random from 1 to 100 seconds)(Boccara et al., 2019).

Detection of cross-area couplings (Fig. 3.2 A,B,C) We employed a statistical modeling approach to detect pairs of CA1-PFC cells that are significantly functionally correlated. With "functionally correlated" we denote cell pairs across brain areas whose noise correlation (Cohen and Kohn, 2011) is much higher than expected from our null-models. For each cell, we fitted a generalized linear model (GLM) (Agresti, 2015) that included all possible covariates measured which could influence and explain the cross-area correlations. These covariates were: linearized spatial position, trajectory, theta selectivity, speed selectivity, spiking history, and within-area spiking of other cells (i.e. PFC cells were fitted with the spiking of other

PFC cells only and, separately, CA1 cells with the spiking of the other CA1 cells only; for further details see subsection "Modelling" below).

These models were used to compute a statistical test, the null hypothesis being that the cross-correlogram among cell pairs is completely explained by external covariates. The alternative hypothesis is that external covariates cannot explain the amount of co-variability; in this case, we considered the correlation as being "functional". This approach is similar to the one introduced in (Nardin et al., 2021a).

With these GLM null-models, we simulated the activity of each cell 10000 times and, for each CA1-PFC cell pair, a cross-correlogram of the responses was computed. Those surrogate cross-correlograms were used to measure how much the actual pairwise cross-correlation measured on real data differed from the simulated ones. We did this for each possible PFC delay in the range of $\pm$1sec.

We considered a cell pair significantly coupled if the peak within $-125, \ldots, +125$ ms of the actual cross-correlogram exceeded the mean plus 4.5 standard deviations of the peaks of the 10000 surrogate cross-correlograms for that pair. This threshold is very conservative and was chosen to compensate for multiple comparisons, and hence represents a Bonferroni correction. Cells in one area that showed significant functional correlation to at least one cell in the other area were termed functionally coupled, often abbreviated to "F.C." or "FC" in our figures.

Complementarity measure (Fig. 3.2E, Fig. S3.3) To measure if two cells are "complementary", i.e., if their joint activity could help to form a symmetric pattern, we computed the Pearson correlation of the linearized firing rate maps for opposite trajectories (for example, SW for cell 1 vs NE for cell 2) and then reported the average across the possible comparisons (SW vs NE and SE vs NW).

mPFC rate map prediction (Fig. 3.2F) We predicted the firing rate maps of mPFC cells starting from the firing rate maps of (simultaneously recorded) CA1 cells. To do so we utilized only functionally coupled cell pairs (see definition above). We summed CA1 maps and weighted them by using the strength of their peak functional correlation (i.e. peak cross-correlation z-scored against GLM simulation). In formulas, denoting with i the mPFC cell for which the prediction has to be made, and with $\{\lambda_1, \ldots, \lambda_N\}$ the firing rate of N functionally coupled CA1 rate maps, and w_{ij} their functional correlation strength, then we defined the predicted mPFC map as

$$\bar{\lambda_i} = \sum_{j=1}^{N} w_{ij}\ \lambda_j$$

We then reported the Pearson correlation between real λ_i and predicted firing rate map $\hat{\lambda}_i$.

Spatial information measure (Fig. S3.3) We computed the spatial information per spike (Skaggs et al., 1993; Souza et al., 2018). In a nuthshell, this corresponds to the first order approximation of the mutual information between position-dependent spiking probability and location, divided by the average firing rate. Denoting with x the location, and $\lambda(x)$ the average firing of a cell at location x, and with $\lambda =< \lambda(x) >_x$ its average firing rate across locations, the spatial information is defined as

$$< \frac{\lambda(x)}{\lambda} \log \frac{\lambda(x)}{\lambda} >_x$$

where $< \cdot >_x$ denotes an average over positions.

CA1 theta oscillation (Fig. 3.3, Fig. S3.4) We averaged the downsampled (5KHz) LFP signal for each tetrode that was implanted in CA1. We then extracted the theta signal ($5 - 15$ Hz) by means of a $3rd-$order Butterworth filter. We detected the local minima of the theta oscillation, and linearly interpolated between each consecutive pair of local minima from 0 to 360 degrees.

Theta locking strength (Fig. 3.3B, Fig. S3.4) We computed the $\theta-$phase of each spike by detecting the angle $(0 - 360)$ of the underlying theta oscillation at the time of activity. We computed phase histograms by binning angles into 30 degree bins. We quantified theta locking strength by means of the mean vector length (MVL): this is defined as the absolute value of the average angle (appropriately converted in radians) in the complex plane:

$$MVL = \text{abs}\left(\frac{1}{N}\sum_{x=1}^{K}\exp(i\theta_x)\right)$$

Theta locking and optimal delay (Fig. S3.4) We employed a measure introduced by Siapas and colleagues in (Siapas et al., 2005) to obtain the optimal delay of mPFC cells theta-locking relative to hippocampal theta. We employed a Rayleigh test for circular uniformity and selected all mPFC cells that yielded a $p < 0.05$. For each one of those cells, we introduced a delay $\tau \in \{-125, -124, \ldots, 124, 125\}ms$ to all spikes, again computed the phase as described above and computed the MVL of these angles. We then selected the delay that yielded the highest MVL for each cell.

Z-scored theta locking strength (Fig. 3.3D) Consider a mPFC-CA1 functionally correlated cell pair. We defined a mPFC spike as being "paired" if it happened in a theta cycle when also the functionally correlated CA1 emitted at least one spike. We tested whether paired spikes had better theta phase locking than independent spikes. To do so, we measured the MVL of the theta-phases of paired spikes, and z-scored it against 200 MVLs measured on randomly subsampled independent spikes. In the (rare) case

where paired spikes were more than independent ones, we randomly subsampled paired spikes instead, and reported the z-score with sign flipped. Excluding these rare cases did not alter the results.

Theta-dependent functional correlation (Fig. 3.3E,F) We employed the same method described above, with the difference that we considered across area cross-correlations only for CA1 spikes that happened at a particular theta phase. In detail, we binned angles into 12 equally sized bins (36 degrees). For each bin $[\theta_m, \theta_M]$, we set to zero all the CA1 activity that happened outside $[\theta_m, \theta_M]$. We then proceeded with the quantification of functional correlations as described above.

Average theta distance between pairs of cells (Fig. 3.3G) For each pair of functionally correlated CA1-mPFC cells, we measured the difference in theta phase among each pair of spikes that happened within the same theta cycles. We then measured the average angular difference as a function of CA1 phase, and reported angular standard error of the mean in the figure. The difference between radian angles θ_1 and θ_2 is defined as:

$$\theta_1 - \theta_2 = \arg\big(\exp(i\,\theta_1 - i\,\theta_2)\big)$$

The average of a set of k radian angles $\theta_1, \ldots, \theta_k$

$$\bar{\theta} = \arg\left(\sum_{j=1}^{k} \exp(i\,\theta_j)\right)$$

The circular variance (for radian angles) is defined as $Var(\theta_1, \ldots, \theta_N) = 1 - MVL(\theta_1, \ldots, \theta_N)$, and the standard deviation of the circular mean as $\sqrt{Var/N}$. In the figures, we reported all these values transformed into angles $(0, \ldots, 360)$ by using:

$$(\theta + \pi)/\pi * 360$$

Randomization of theta phase and drop of correlations (Fig. S3.6) To test whether precise theta timing was required for mPFC cells to be functionally correlated with CA1 cells, we randomized the theta locking. To do so, for each mPFC cell independently we considered different levels of random spike jittering. In detail, to each spike we added a random time uniformly distributed within $[-M, M]$, where M is the maximum jitter allowed. Afterwards, we computed CA1-mPFC functional correlations using mPFC-jittered spike trains, while leaving CA1 untouched. We then measured the reduction of functional correlation strength by considering the relative decrease in magnitude of the peak:

$$\frac{peak_{orig} - peak_{jittered}}{peak_{orig}}$$

Power spectral density (Fig. S3.7) We employed Welch's method to compute averaged power spectral density (Welch, 1967). In a nutshell, this method works by dividing the signal into overlapping segments and averaging the periodograms obtained in each. In detail, we performed short-time Fourier transform centered at each theta cycle, with a Hann window of 400ms, separately for all CA1 and PFC tetrodes. For each CA1-mPFC cell pair, we computed the power spectrum for theta cycles where both cells are active (paired) vs theta cycles where either or both cells are silent (independent). Finally, we averaged the periodograms across cells, separately for CA1 and mPFC.

Bipartite spectral clustering (Fig. S3.4A) We considered a graph were each node represents a cell, and edges are present only among CA1-mPFC cell pairs that were functionally correlated. We then asked if it was possible to subdivide this (bipartite) graph into smaller clusters, containing both CA1 and mPFC cells, so as to maximize the across-area connections within each cluster and minimize the ones outside. We employed the spectral optimization of modularity (SOM) algorithm, first proposed in (Newman, 2006). The algorithm iteratively finds the best split by considering the eigenspectrum of the modularity matrix M of the bipartite graph. We refer the reader to (Newman, 2006) for details.

Assembly firing rate map (Fig. 3.4B, C) We deemed an assembly active whenever at least 2 CA1 and 2 mPFC cells were active in the same theta cycle. Afterwards, we followed the same prescription described above for computing firing rate maps, with the difference that instead of using spike numbers we considered only 1 (active) or 0 (not active) during a given theta time window.

Within area functional correlations (Fig. 3.4D, E) We employed the same statistical procedure presented above for within area functional correlations.

Bootstrapped theta locking difference (Fig. 3.4H) For each pair of mPFC cells belonging to the same assembly, we randomly selected 50 spikes from each cell and measured their average phase. We did so 200 times for each cell pair. Afterwards, we counted how many times the average phase of the first cell was larger than the average phase of the second cell. If the two cells had no phase relationship among each other, the probability of one phase being larger than the other would be $p = 0.5$. If that were the case, 99% of the cases one would have a count between ~ 75 and ~ 125. We quantified the percentage of cell pairs that were outside this confidence interval, and reported the Binomial test with $p = 0.01$ baseline.

Modelling

GLM model of cells response We utilized a GLM model to describe each cell's response propensity as a function of all measured covariates during foraging activity. These detailed models served as null models

for the statistical test we employed to detect functional correlations among cells (see section "Detection of across-area couplings"). We will describe in detail here the covariates used to fit the model and the parameters used in fitting and simulation routines, and refer the reader to other references for the details regarding the theoretical background of GLMs (Agresti, 2015).
The model described the inhomogeneus Poisson activation rate λ_t of cells in 25.6ms time windows, $\tau = 0.0256$s,

$$P(x_t = n|\lambda_t) = e^{-\lambda_t}\frac{\lambda_t^n}{n!}$$

The expected firing rate λ_t takes form

$$\lambda_t = \exp\left(\boldsymbol{\beta}\cdot\boldsymbol{\theta_t}\right)$$

where $\boldsymbol{\beta}$ represents the model coefficients, found by maximum likelihood (below) and $\boldsymbol{\theta}_t$ represent the covariates, which are:

- trajectory-dependent spatial position: we binned the linearized position in 10cm bins, as described above, in 25.6ms time bins. We allowed each cell to have a different encoding for each trajectory separately to allow for maximum flexibility: this resulted in a $80-$dim vector, i.e., 20 location bins for each possible trajectory. At each time point, only the entry of this $80-$dim vector that represented the location of the animal, and the trajectory taken, was set to 1, and all the others to 0 (one-hot encoding variable)
- speed: we binned the speed, which was measured from the behavioral recording in 25.6ms time bins, in 7 non-overlapping and equally populated speed bins, starting from 7cm/s (one hot variable)
- theta phase: we computed the theta phase at the center of each 25.6ms time bin. The computation of the theta phase is detailed in a previous section. We binned the angles in 10 non-overlapping angular bins, and encoded it as a one-hot variable.
- spiking history: the spiking of the last three time bins was used
- whithin-area spiking activity: the spiking activity of all the other cells in the same area, together with the spiking history of each of those cells in the previous time bins were used as additional covariates

The number of parameters of such models ranged from 100 to 400, depending on the number of cells recorded simultaneously in the same area. We utilized the routine *GLM* offered by the package *statsmodels* v0.12.2 in Python 3.7 (Seabold and Perktold, 2010). We fitted the models by using an $L2$ regularization, whose parameter was found by grid search on $(10^{-10}, 10^{-9}, \ldots, 10^{0})$ and cross-validation (train=75%, test=25% of data) independently for each cell.

Normative model of (spatial) information transfer and generalization (Fig. S3.3) Consider two populations of neurons, exemplifying N PFC cells and M hippocampal cells. We will denote with $\vec{x} = (x_1, \ldots, x_N)$ the stochastic (binary) activation of PFC cells, and with $\vec{y} = (y_1, \ldots, y_M)$ the stochastic (binary) activity of CA1 cells. Let us denote with s the distance to the goal, and with k the trajectory. We consider s to be a random variable that take values in $\{0, 1, \ldots, 10\}$, each with equal probability, and k a Bernoulli($p = 0.5$) $\in \{0, 1\}$, independent of s.
We assume that $\vec{x}$ is driven by hippocampal input and internal connectivity, but has no initial selectivity for location. We formalize this request with a stochastic model, which is similar in its formalization to a restricted Boltzmann machine (Smolensky, 1986):

$$P(\vec{x}|\vec{y}) \propto \exp\left(\sum_{i=1}^{N}\sum_{h=1}^{M} C_{ih}x_i y_h + \sum_{i,j=1}^{N} W_{ij}x_i x_j\right) \tag{3.1}$$

where $C \in \mathbb{R}^{N\times M}$ and $W \in \mathbb{R}^{N\times N}$.

We assume that the activity of CA1 cells is both trajectory and position dependent. We have that

$$P(\vec{y}|s) = \sum_k P(\vec{y}|s,\, k)P(k) \tag{3.2}$$

We also have that PFC population activity is position dependent through CA1, i.e. :

$$P(\vec{x}|s) = \sum_{\vec{y}} P(\vec{x}|\vec{y})P(\vec{y}|s) \tag{3.3}$$

For fixed W and $P(\vec{y}|s,\, k)$, we want to find the best C that maximizes the mutual information between $\vec{x}$ and s.

$$I(\vec{x}, s) = \sum_{\vec{x},s} P(\vec{x}, s)\log\frac{P(\vec{x}, s)}{P(\vec{x})P(s)} \tag{3.4}$$

$$\propto \sum_{\vec{x},s} P(\vec{x}|s)\log\frac{P(\vec{x}|s)}{P(\vec{x})} \tag{3.5}$$

We maximize this quantity via Sequential Least SQuares Programming (SLSQP) routine in SciPy. We constrain each C_{ij} to lay in $[-1, 1]$.

Our simulations use $N, M = 10$ neurons, which allows the mutual information to be computed without the need for approximations (by enumerating all possible patterns). Reported estimates are obtained by averaging across 100 randomly initialized networks;

for each simulation, W is a symmetric matrix with entries randomly samples from a $N(0,1)$ distribution, and $P(\vec{y}|s,k)$ was initialized by considering each cell independent of each other, with a gaussian place fields per trajectory and simulated in such a way so as to resemble CA1 single cell statistics measured in the data.

Programming languages and packages used

All scripts were written in Python 3.7 (Van Rossum and Drake Jr, 1995). We used the packages NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), StatsModels (Seabold and Perktold, 2010), AstroPy (Robitaille et al., 2013).

Supplementary Figures

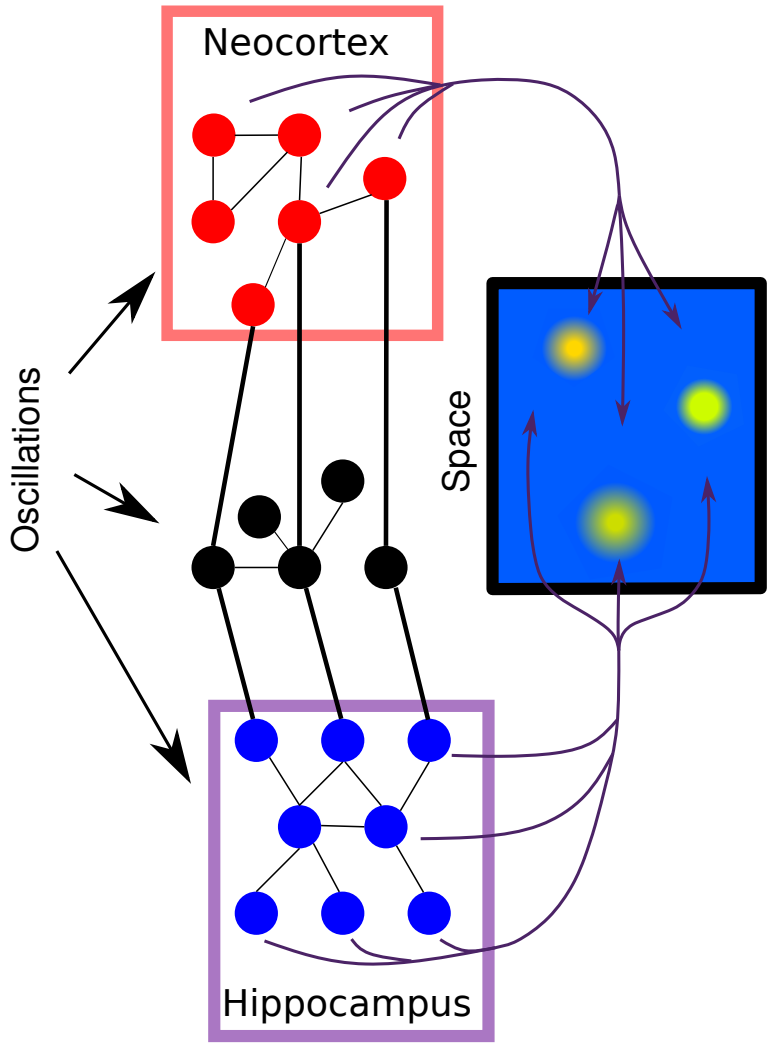

Fig. S3.1. Schematic multi-area assembly synchronization.

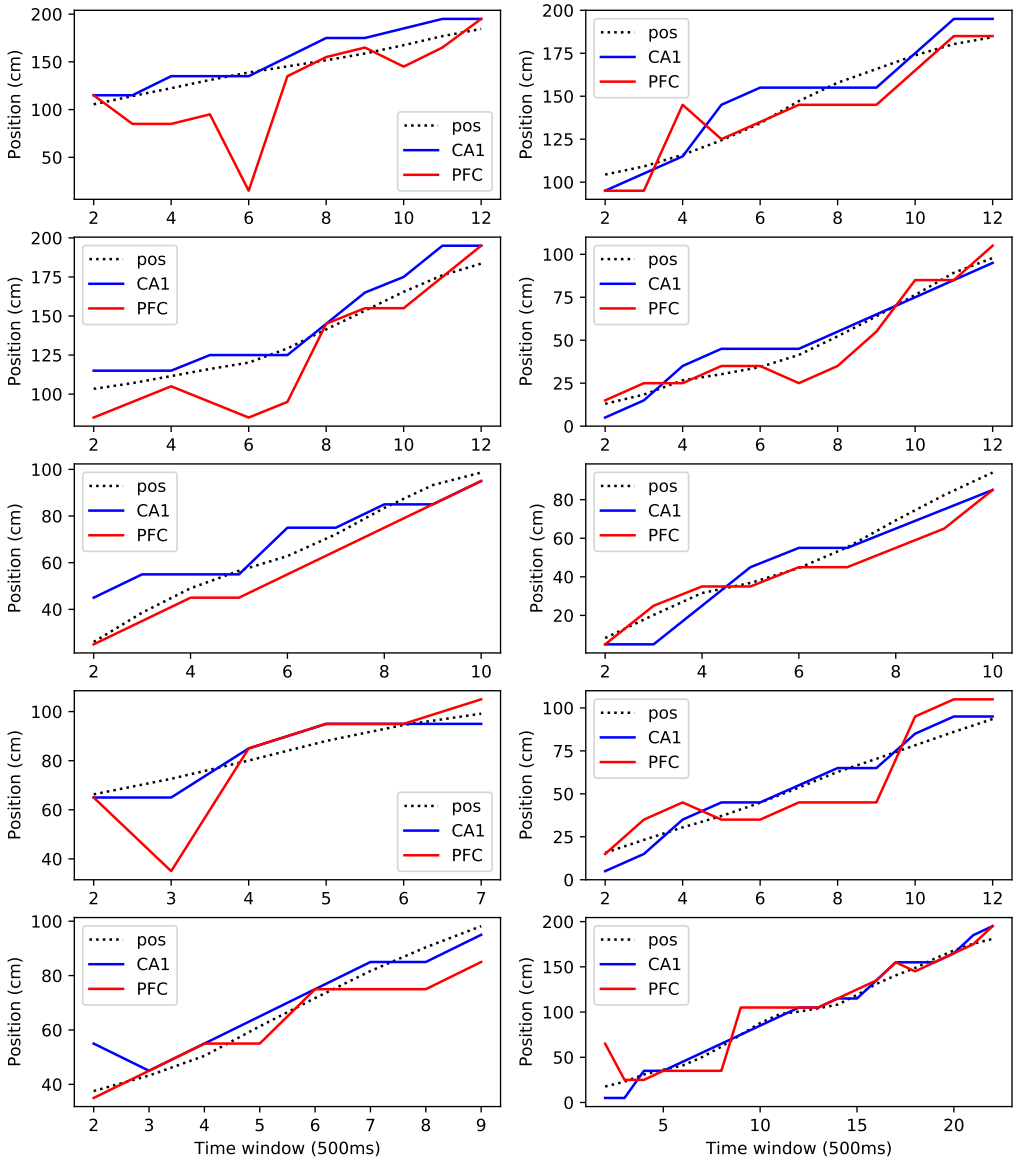

Fig. S3.2. mPFC decoding lags behind HPC.
Decoded position from CA1 (blue) or mPFC (red) population activity in 500ms time windows vs real position of the animal (dotted line).

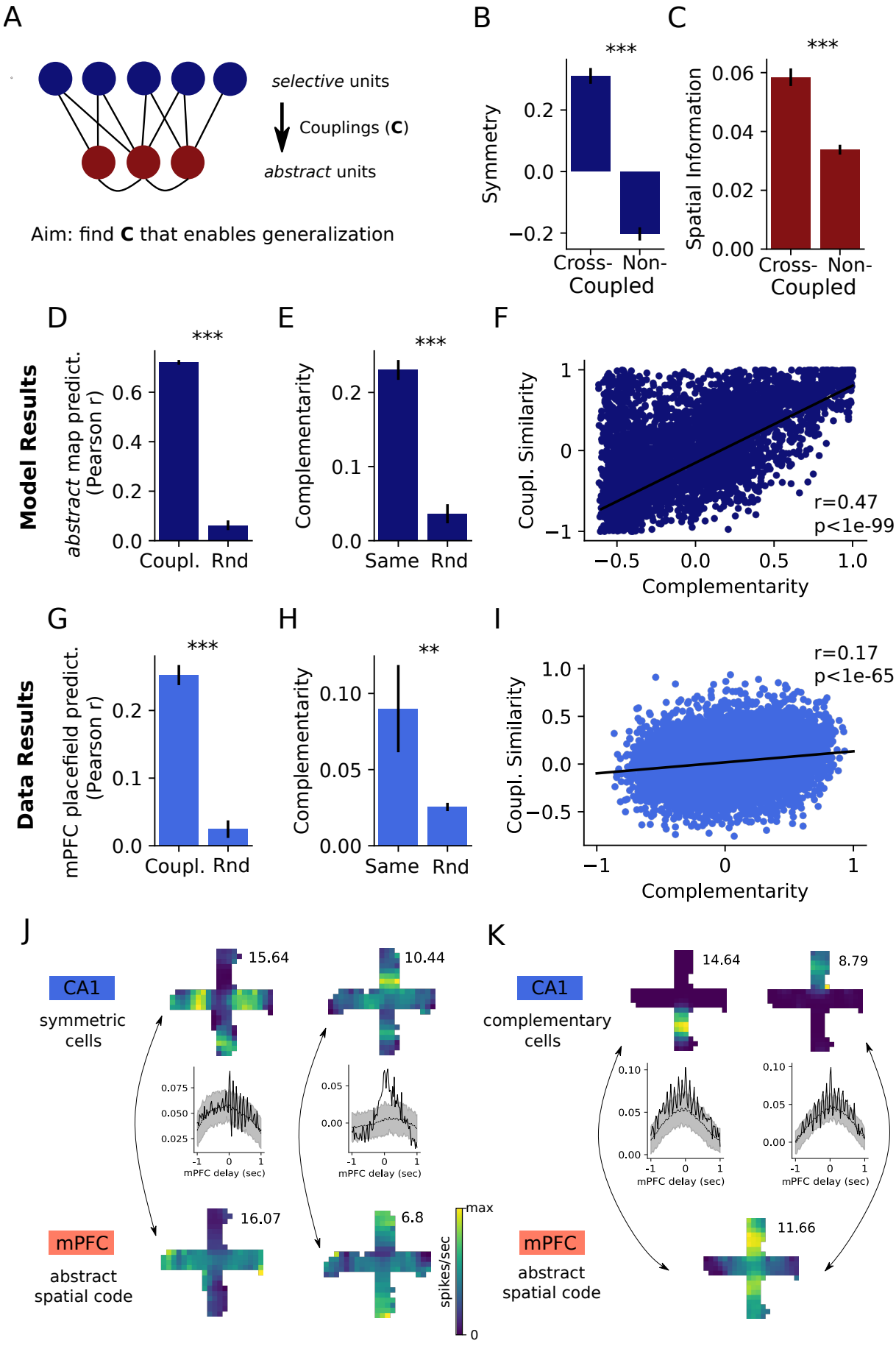
A
selective units
Couplings (C)
abstract units
Aim: find C that enables generalization
B
***
Symmetry
Cross-Coupled
Non-Coupled
C
***
Spatial Information
Cross-Coupled
Non-Coupled
Model Results
D
***
abstract map predict. (Pearson r)
Coupl.
Rnd
E
***
Complementarity
Same
Rnd
F
Coupl. Similarity
r=0.47
p<1e-99
Complementarity
Data Results
G
***
mPFC placefield predict. (Pearson r)
Coupl.
Rnd
H
**
Complementarity
Same
Rnd
I
r=0.17
p<1e-65
Coupl. Similarity
Complementarity
J
CA1
symmetric cells
15.64
10.44
mPFC delay (sec)
mPFC
abstract spatial code
16.07
6.8
spikes/sec
max
0
K
CA1
complementary cells
14.64
8.79
mPFC delay (sec)
mPFC
abstract spatial code
11.66

Fig. S3.3. Cross-area couplings enable generalization. (A) Schematics of information transfer and generalization stochastic model. The input layer is composed of selective units (i.e., CA1-like) exhibiting trajectory dependent spatial firing. The second layer is composed of abstract (i.e., mPFC-like) units, whose activity is influenced by the first layer and internal connectivity. We found the matrix C by maximizing the mutual information between the abstract population firing and distance to the goal. **(B)** Selective units with strong cross-couplings had higher symmetry (i.e., similar firing across trajectories). Here and in the following, error bars represent the 95th confidence interval for the mean. Mann–Whitney U test, $p < 0.001$. **(C)** Abstract units with strong cross-couplings exhibited higher spatial information. Mann–Whitney U test, $p < 0.001$. **(D)** Prediction of abstract units firing map by using selective units firing maps weighted by optimized coupling matrix (left) or random couplings (right). Mann–Whitney U test, $p < 0.001$. **(E)** Complementarity of the two selective units with strongest coupling to the same abstract unit (left) vs random (right). Mann–Whitney U test, $p < 0.001$. **(F)** Complementarity of all selective pair vs coupling similarity to abstract units. Pearson correlation test, p<0.001, r=0.47. **(G)** Same as D), for real data mPFC place fields. Mann–Whitney U test, $p < 0.001$. **(H)** Same ad E), for real data cross-couplings. Mann–Whitney U test, $p < 0.01$. **(I)** Same as F), for real data CA1 cell pairs. Pearson correlation test, $p < 0.001$, r=0.17. **(J)** Generalization mechanism 1. CA1 cells with symmetric firing fields (top row, two example firing rate maps, peak firing rate in top-right corner) show significant cross-couplings with mPFC cells (bottom row, two examples). Middle row shows the cross correlograms between top and bottom cells. (black line = real cross-correlation, black-dotted = average cross-correlation from null-model simulations, shaded ares = ±3STD of the mean from null-model simulations). **(K)** Generalization mechanism 2. CA1 cells with complementary firing rate maps (top row) both significantly cross-coupled to the same mPFC cell (bottom row). Cross-correlograms as in J). Throughout the figure, **= $p < 0.01$, *** = $p < 0.001$.

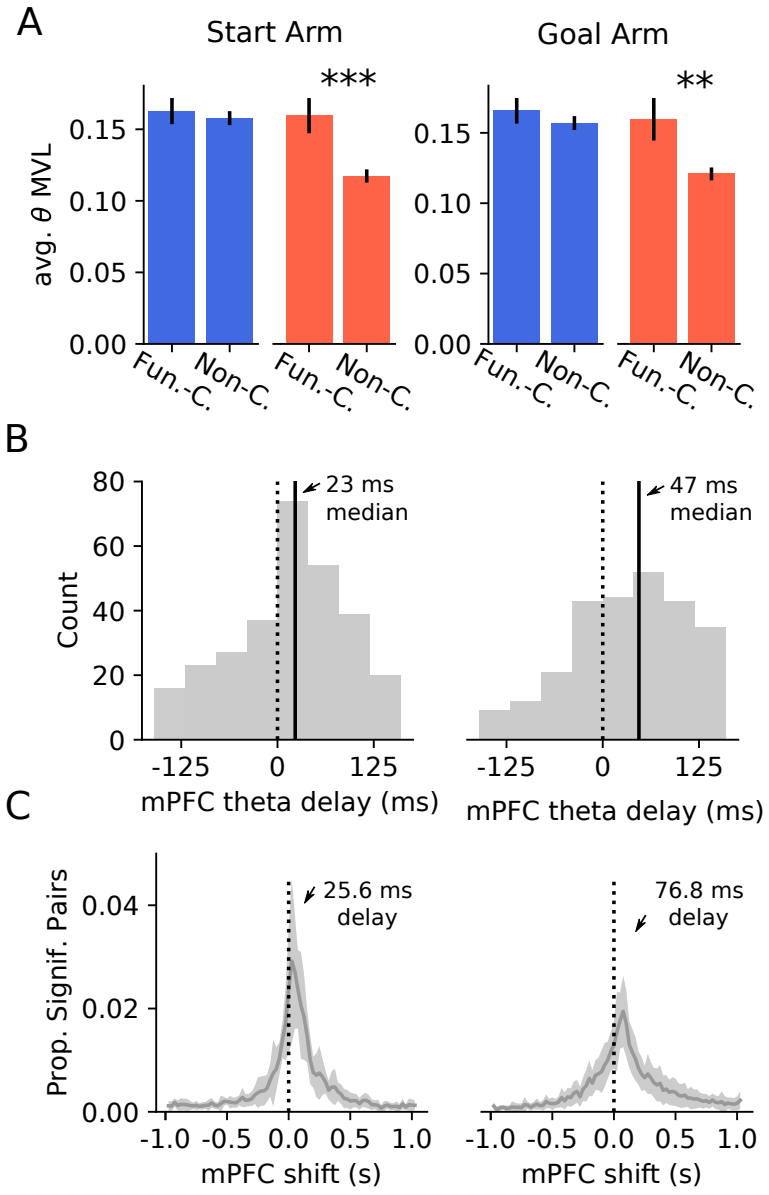

Fig. S3.4. Start vs Goal arms delayed couplings.
(A) average theta phase locking strength of CA1 (blue) or mPFC (red) cells, comparing FC vs non-FC. Left: start arm (Mann–Whitney U test, CA1: $p > 0.05$, mPFC: $p < 0.001$). Right: goal arm (Mann–Whitney U test, CA1: $p > 0.05$, mPFC: $p < 0.01$). **(B)** Proportion of significant cell pairs ($z > 4.5$) in start arm (left) vs goal arm (right) as a function of mPFC delay. Shaded area represents the 99% CI for the mean. N=13 sessions. **(C)** Optimal delay that produced the best theta locking of each mPFC cell, found by computing a phase histogram for each delay and measuring the mean vector length (Siapas et al., 2005). Mann–Whitney U test for start vs goal arm optimal theta delay, p<0.001. KS test for the distributions, $p < 0.001$.

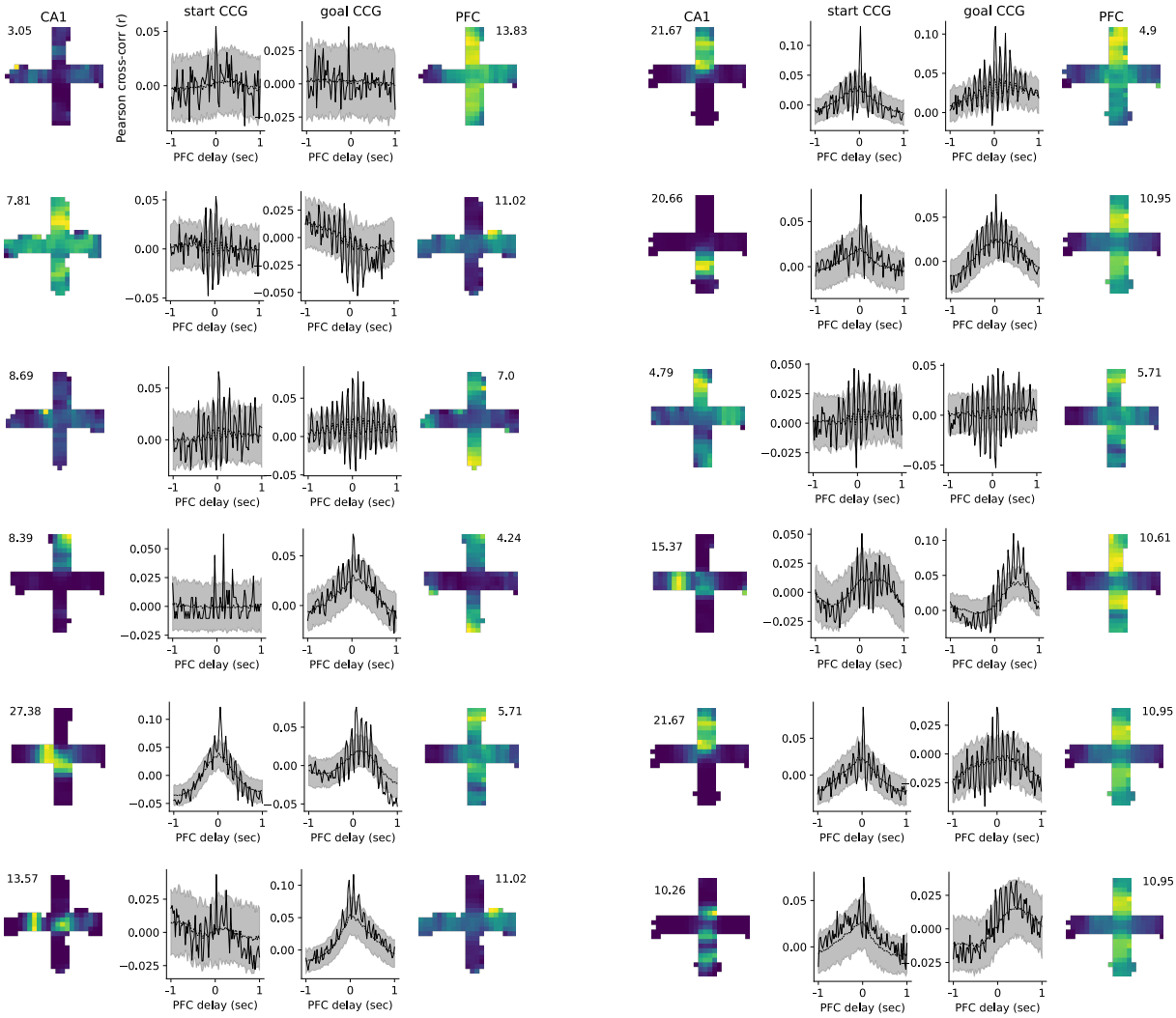

Fig. S3.5. Start vs Goal Functional Cross-correlations.
Additional examples of cross-correlations for cell pairs with significant interaction ($z > 4.5$ relative to GLM). Left and right: CA1 and mPFC firing rate map. Center: cell-pair cross correlograms, real data (solid black line); null model mean (dotted back) $\pm$ 3STD (shaded area). **TODO**: decrease space, increase legibility

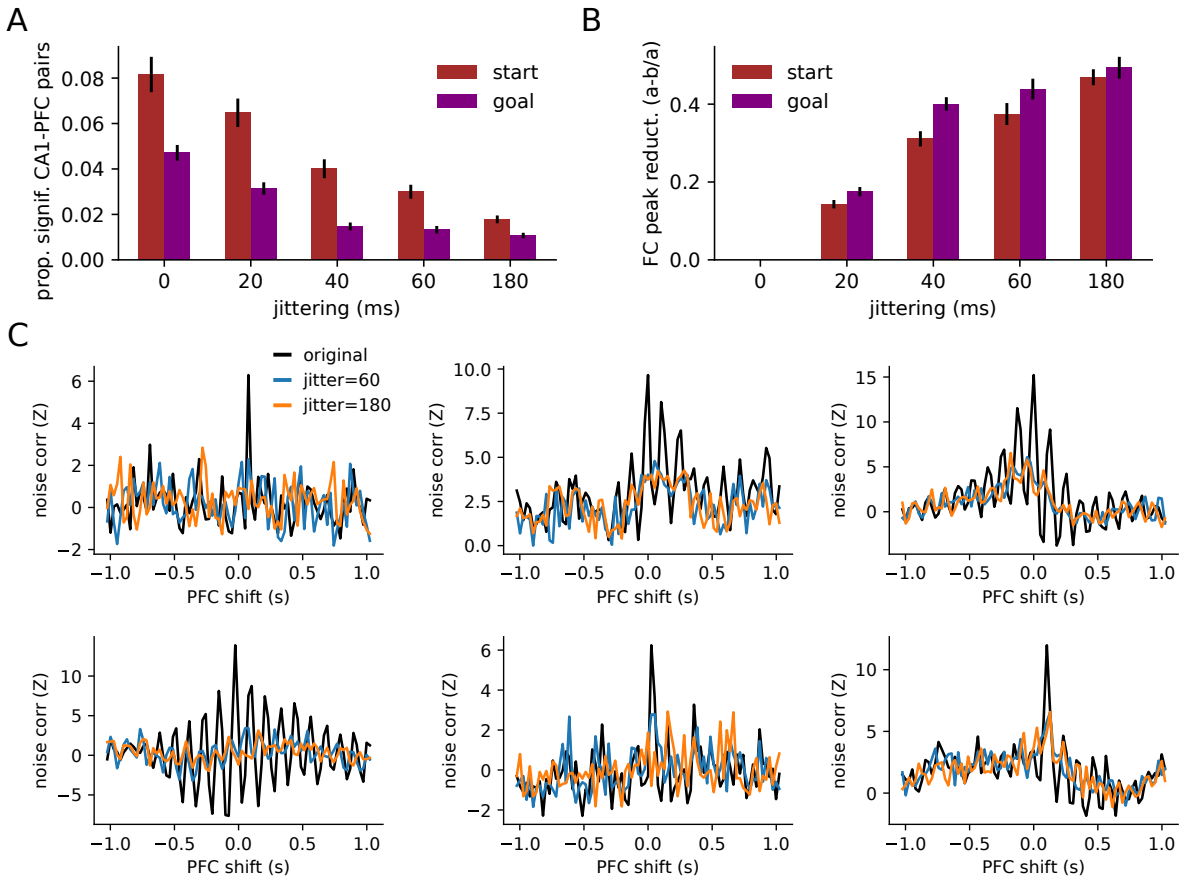

Fig. S3.6. CA1-mPFC FC firing depends on precise theta-timing.
(A) Proportion of significant FC pairs ($z > 4.5$ within theta-cycle delays) for start (brown) vs goal (purple) as a function of mPFC spike jittering (uniform +- 0, 20, 40, 60, 180 ms). **(B)** Relative reduction of strongest FC peak of measured as (peak_orig – peak_jitt)/peak_orig for different levels of mPFC spike jittering. **(C)** Example functional cross-correlations showing the effect of jittering (black: no jittering, blue: 60ms, orange: 180ms).

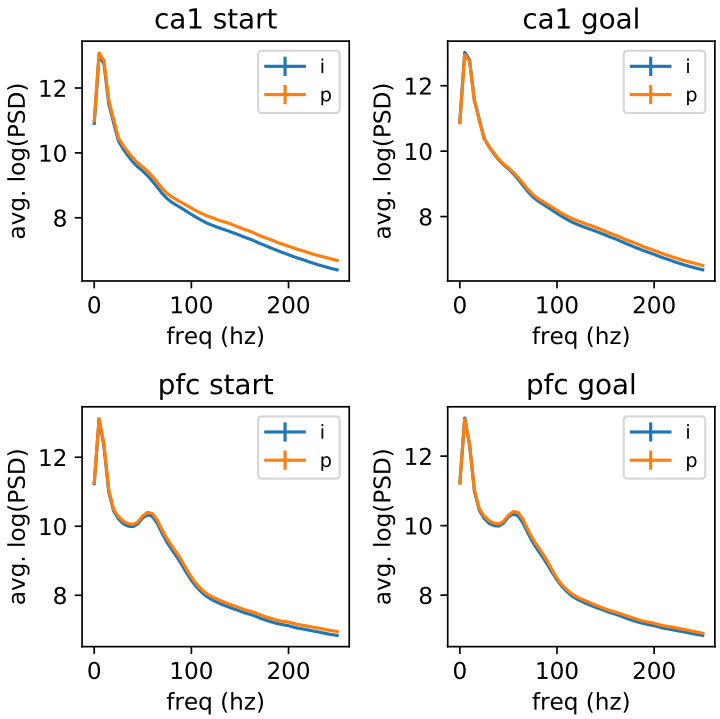

Fig. S3.7. Paired firing of mPFC-CA1 FC pairs is not accompanied by significant changes in underlying oscillations.
Average PSD for paired (orange) vs independent (blue) activity of CA1 (top) and mPFC (bottom) FC cell pairs. PSD computed using Welch's method on 400ms time segments centered on mid-theta cycles using a Hanning window and constant detrending.

4

Extra-hippocampal memory consolidation

4.1 Aims

Most theories agree on the fact that the responsibility for long-term memory storage resides outside the hippocampus (Squire et al., 2015). In particular, it is believed that, over time, there are repeated interactions between the hippocampus and the neocortex which lead to a gradual transfer of memories (Frankland et al., 2004). This process is mediated mostly through the entorhinal cortex, which is, anatomically speaking, an interface between the hippocampus and most of the neocortex (Witter, 2011). Therefore, it makes sense to study how spatially tuned entorhinal cells (Fyhn et al., 2004; Hafting et al., 2005; Diehl et al., 2017) encode spatial locations during and after learning, and how these representations change over time.

Here, I will study neural activity that correlates with the learning of novel spatial memories. One of the clearest correlates of learning in hippocampal CA1 cells is the accumulation of firing around salient locations (Hollup et al., 2001; Dupret et al., 2010; Gauthier and Tank, 2018; Lee et al., 2020). This accumulation has been shown to emerge during learning and to be predictive of future memory retention (Dupret et al., 2010). Interestingly, while the CA1 area accumulates firing around goal locations, CA3 hippocampal subregion does not (Dupret et al., 2010). It is unknown how the MEC would encode goal learning, if at all, nor how quickly and for how long.

The aim of this chapter is to study the responses of simultaneously recorded single cells in CA1 and MEC during and after goal learning. The study will focus on the accumulation of firing of different subsets of spatially-tuned cells, the speed of such accumulation, and the fine details of the movement of firing fields towards goal locations.

This is a study in collaboration with Charlotte Boccara, Federico Stella, Joseph O'Neill, and Jozsef Csicsvari.

This is the author's version of the work. It is included in this dissertation by permission of the AAAS. The definitive version was published in the journal Science on March 29th, 2019, volume 363, pages 1443–1447. DOI: https://doi.org/10.1126/science.aav4837

Author contribution: C.N.B and J.C. designed and implemented the study and wrote the manuscript. C.N.B performed the experiments. J.O'N helped with the implementation of the study, C.N.B and M.N. performed cluster cutting, C.N.B., M.N. and F.S. planned the analyses, M.N. and F.S. analyzed the data, M.N. wrote the methods, all authors discussed the results and contributed to the manuscript.

The Entorhinal Cognitive Map is Attracted to Goals

Charlotte N Boccara[&@+], Michele Nardin[&], Federico Stella, Joseph O'Neill[#], Jozsef Csicsvari[@]

Institute of Science and Technology Austria (ISTA), Klosterneuburg, 3400, Austria
[&] These authors contributed equally to this work
[@] corresponding authors: charlotte.boccara@medisin.uio.no, jozsef.csicsvari@ist.ac.at
[+]current address: University of Oslo, Sognsvannsveien 9 Domus Medica, 0372 Oslo, Norway.
[#]current address: Cardiff University, 70 Park Place, Cardiff, CF10 3AT, United Kingdom.

Grid cells with their rigid hexagonal firing fields are thought to provide an invariant metric to the hippocampal cognitive map. Yet, environmental geometrical features have recently been shown to distort the grid structure. Given that the hippocampal role goes beyond space, we tested the influence of non-spatial information on the grid organization. We trained rats to daily learn three new reward locations on the cheeseboard maze, while recording from the medial entorhinal cortex and the hippocampal CA1 region. Many grid fields moved towards goal location, leading to long-lasting deformations of the entorhinal map. Therefore, distortions in the grid structure contributes to goal representation during both learning and recall, which demonstrates that grid cells participate in mnemonic coding, beyond providing a simple metric of space.

4.2 Introduction

To decide upon relevant behavior, individuals rely on dynamic neural representations of their world, computed from current and past experiences. The cognitive map formed by an extended network of specialized cell types coding for defined spatial features is essential for accurate navigation (O'Keefe and Nadel, 1978; Moser et al., 2017). While hippocampal place cell activity is restricted to discrete, sparse place fields in specific environments, parahippocampal grid cells present multiple firing fields, arranged in regular hexagonal arrays that densely tessellate all environments (Hafting et al., 2005; O'Keefe and Dostrovsky, 1971). This led to the hypothesis that grid cells provide a universal invariant metric for spatial cognition (Moser and Moser, 2008). As such, they were originally considered to have a narrower role than hippocampal place cells, which code for multimodal information beyond simple spatial representations (Moser et al., 2017; Eichenbaum et al., 1999). New data, however, suggest a more complex grid code (Marozzi et al., 2015; Lipton et al., 2007; Diehl et al., 2017). Non-spatial factors modulate local field firing rates without affecting the grid structure (Ismakov et al., 2017) while topographically organized auditory stimulus can drive grid-like structure(Aronov et al., 2017). Furthermore, geometrical environmental features can influence the rigid grid structure, thus challenging the role of grid cells to provide invariant metrics(Stensola et al., 2012; Krupic et al., 2018; Barry et al., 2007). However, grid distortions could also encode more complex behavioral information. We therefore tested the influence of behaviorally-relevant information on the entorhinal cognitive map.

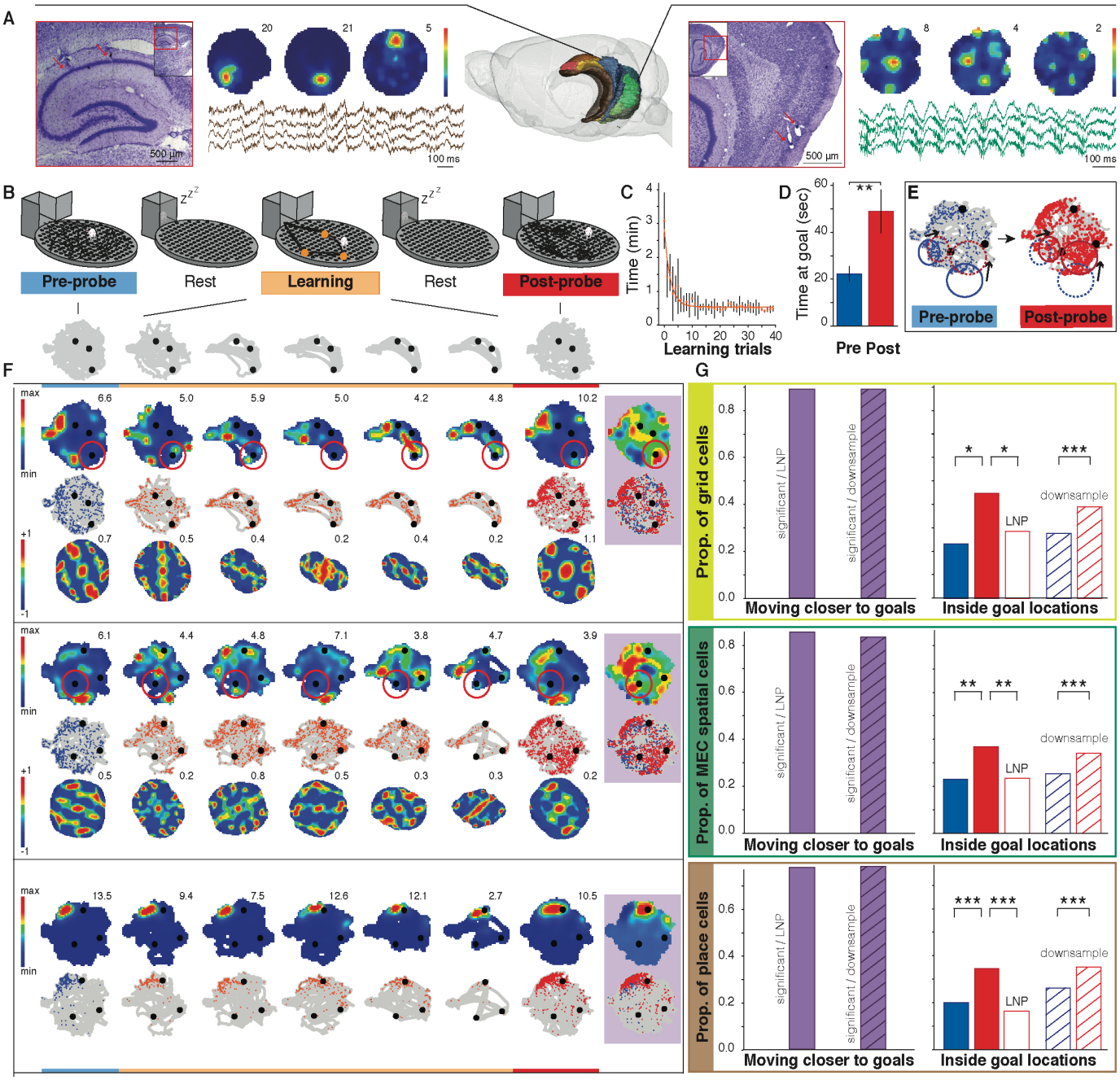

Fig. 4.1. MEC and CA1 spatial cells move towards newly-learnt goals. (A) Dual-recordings in CA1 (left-half) and MEC (right-half). Nissl-stained sections (red arrows: electrode tracks), rate maps of simultaneously-recorded place and grid cells and LFP traces. **(B)** Behavioral sequence: pre-probe, pre-rest, learning, post-rest and post-probe. Bottom: animal's path (grey) with goals (dots). **(C)** Average normalized learning-curve (orange). **(D)** Memory retention test. Average time at goals: pre-learning (blue) and post-learning (red). T-test, P=0.0053. Black error-bars: SEM across sessions. **(E)** Example grid cell showing spikes (colored dots) and field (circles) movement towards goals (black dots). **(F)** Example maps of grid (up), spatial MEC (middle) and CA1 place (bottom) cells across paradigm. **(G)** Left-half: Proportions of spatial cells with fields moving significantly closer to goals. Significance calculated against LNP data [purple] or downsampled [dashed purple, binomial test, all Ps<0.0001, grids (up): 89%; MEC (middle): 84%; CA1 (bottom): 79%]. Right-half: Proportion of cells with their strongest fields at goals in pre-learning (blue) and post-learning (red). Fisher's exact test, grids (up): P=0.027, N=56; MEC (middle): P=0.0047, N=157; CA1 (bottom): P=0.00018, N=245). Empty bars: control post-probe LNP data (all Ps<0.018 against post-probe). Dashed bars: control down-sampled data (all Ps<0.00001). See supplementary figures for detailed legends.

4.3 Results

We trained rats to learn daily three new hidden reward locations on a cheeseboard maze, while recording simultaneously from the medial entorhinal cortex (MEC) and the hippocampal CA1 region (Fig. 4.1A, Fig. S4.1). This hippocampus-dependent task (Dupret et al., 2010) consisted of three phases: pre-probe, learning and post-probe, where the probes verified memory retention in the absence of food rewards (methods and Fig. 4.1B–D). This paradigm changed daily the cognitive valence of local points in an otherwise familiar environment, leading to the accumulation of CA1 place fields around reward locations (i.e. goal remapping, (Dupret et al., 2010)). This allowed us to test how rewards can be dynamically encoded in MEC neural representation during goal learning.

The majority of grid cells (80–90%) had at least one of their firing fields significantly moving towards a goal (Fig. 4.1E–G, binomial test, $P<0.00001$, see methods and Fig. S4.2–3). MEC non-periodic spatial and CA1 place cells showed a similar behavior (Fig. 4.1F–G and S4.3, binomial test, $P<0.0001$). This led to the accumulation of entorhinal and hippocampal strongest firing fields at goal locations (Fig. 4.1G). Because successful learning led to a higher number of visits of reward locations (Fig. 4.1D), we used a Linear-Nonlinear-Poisson (LNP) spiking model to verify that the reorganization of firing fields was independent of variations in trajectories, speed or heading between pre- and post-probe by comparing real data to maps generated with this method (Hardcastle et al., 2017). We also performed an additional control by downsampling the pre- and post-probe map to match occupancy in each spatial bin (see methods and Fig. 4.1G). Conjunctive head-direction coding did not influence field movement to goal (Fig. S4.4).

The proportion of cells with their strongest fields within goal locations increased progressively during learning in both entorhinal and CA1 spatial cells (Fig. S4.5A–C). To test whether these changes were long-lasting, we performed the same analyses on pre-probe, this time using the previous day's goal locations. MEC cells retained the accumulation of firing fields around previous goal locations (Fig. S4.6A–B). Unexpectedly, CA1 changes were more transient (Fig. S4.6C), suggesting differences in memory trace lability between those two regions, in line with models arguing for faster plasticity in CA1 than in MEC (Rennó-Costa and Tort, 2017). We also observed faster CA1 plasticity during learning by scoring the development of goal representation as a function of firing related to goal vicinity (see methods and Fig. S4.5D–G).

Given that pre-probe grid map maintained an accumulation of firing fields at the previous day's goals, we examined the effect of the local change in cognitive valence on the grid structure itself. For three animals we added the exploration of a familiar open field, which acted as a control environment without valence bias (see methods). Grid scores were significantly higher in the open field compared to both probes (Fig. 4.2A–B; 1-way analysis of variance (ANOVA), $P<0.00001$), independently of differences in spatial sampling, trajectories, speed or heading (Fig. 4.2B and methods; 1-way ANOVA, $P<0.00001$).

To test whether the grid score drop resulted from local map distortions around goals, we used a Laplacian of Gaussian (LoG) filter to detect individual fields independently of their peak firing rate (see methods and Fig. S4.7; (9, 10)). While there were no significant differences in the number of fields, their size, spacing or ellipticity after learning (Fig. S4.8), we found a significant decrease in the mean distance between field center and closest reward (Fig. 4.3A–B; Kolmogorov-Smirnov (KS) test, $P=0.01472$) and an increase in the number of grid fields near goals (Fig. 4.3C; Fisher's exact test, $P=0.0145$).

To determine which parameters contributed to grid field movement, we calculated the "strength

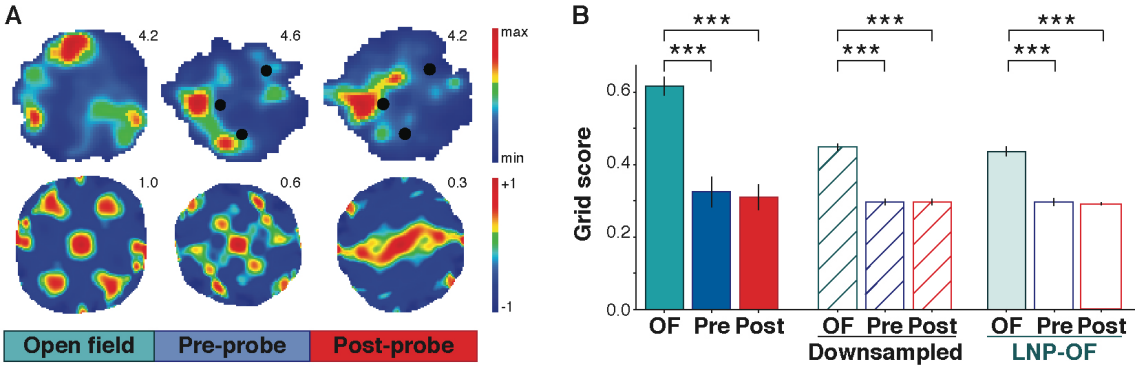

Fig. 4.2. Grid score degradation on the cheeseboard. (A) Example of MEC grid cell, which exhibits its highest grid score on the open field and a degraded score on pre- and post-probes. Top: rate maps with maximum firing rate in red and lowest in blue with peak firing rate (Hz) on the upper right corner. Bottom: corresponding spatial autocorrelogram maps, range from +1 (red) to -1 (blue), with the grid score noted on the upper right corner. **(B)** Average grid score (±SEM) across MEC cells in open field (OF: solid cyan), pre-probe (Pre: solid blue) and post-probe (Post: solid red) (1- way ANOVA, $P<0.00001$). Dashed bars: downsampled control data (1-way ANOVA, $P<0.00001$). Light cyan and empty bar: control data obtained with LNP spiking model in open field (LNP-OF) (1-way ANOVA, $P<0.00001$).

of attraction" of each field towards a goal and correlated it against spatial parameters and behavioral performance (Fig. 4.3D–E, S4.9). The strongest correlation was found with the pre-probe distance to goal location: the closest fields being generally subjected to a strong attraction while fields over a distance of ca. 30 cm showing little detectable attraction (Fig. 4.3D–E). Most attracted fields moved towards the closest goals and the most visited post-probe goal was pulling the strongest attraction (Fig. S4.9). Different fields of a given cell could be attracted to different goals, depending on their relative position to the goals (Fig. S4.7H–K). We observed a weak increase in the peak grid field firing rate in post-probe, however this increase was not associated with goal locations and there was no correlation with strength of attraction (Fig. S4.10). Finally, the increase of activity at goals correlated with memory retention (correlation analysis, $r=0.77$, $P=0.002$, Fig. 4.3F). There was no significant difference of in the strength of attraction between CA1 and MEC fields, however pre-probe field distance influenced only MEC fields (Fig. S4.11).

To test whether the local movement of firing fields towards the closest goals could explain the grid score drop in probe sessions, we applied a movement towards imaginary goals on grid fields recorded in the open field, following the distribution of movements determined after learning. This resulted in a grid score drop was analogous to that observed between open fields and cheeseboard environments (Fig. 4.3G; 1-way ANOVA, $P=0.0003$).

We subsequently examined the reorganization of place-related assemblies at the population level (Fig. 4.4 and S4.12–13). Both MEC and CA1 cell assemblies showed a significant reduction in vector similarity between pre- and post-probe as compared to the intrinsic variability in comparing the two halves of the pre-probe (Fig. S4.12B; t-test, all $Ps<0.0001$). This was independent of changes in spatial sampling (Fig. S4.12B; t-test, all $Ps<0.0001$). MEC population vector similarity across pre/post probe sessions were weaker around goal locations than away from them (Fig. S4.12C; t-test, $P=0.0003$) and positively correlated with distance from the goal locations (Fig. 4.4E; $P<0.0001$). In contrast, the reorganization of CA1 population vectors between pre- and post-probes did not exhibit significant positive correlation with goal distance (Fig. 4.4E; $P=1$). Given that goal remapping of individual MEC and CA1 spatial cells occurred incrementally during learning (Fig. S4.5), we next examined

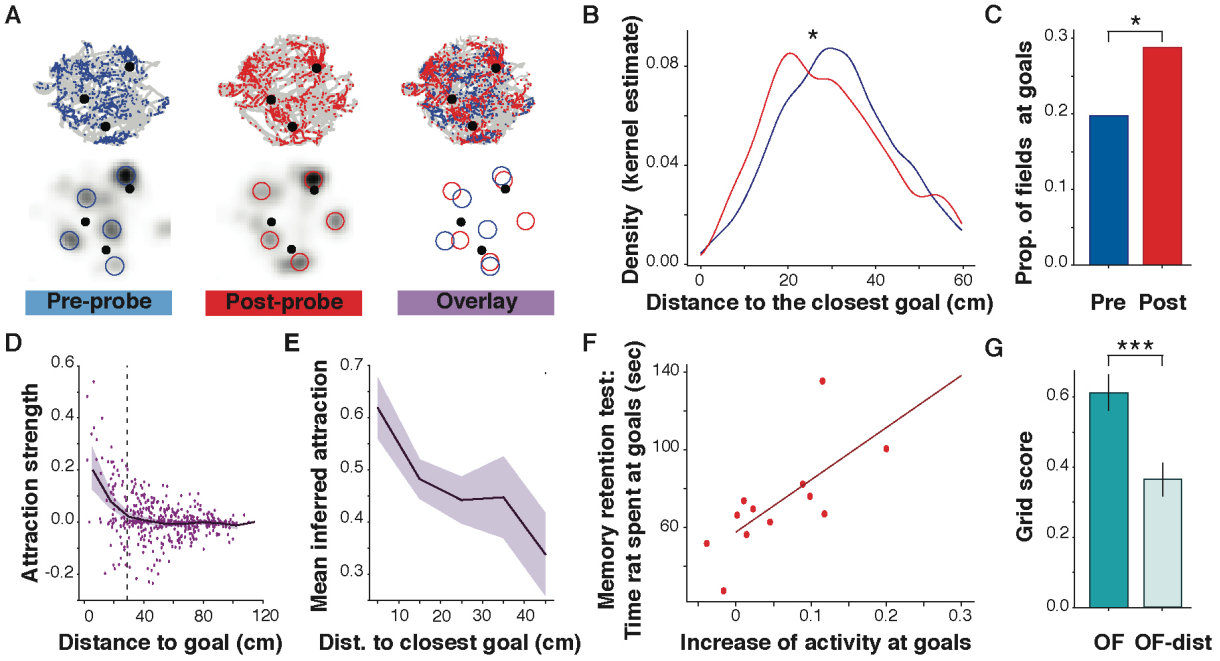

Fig. 4.3. Stronger attraction of MEC fields closest to goal locations. (A) Example of grid fields movement from pre-learning (blue) to post-learning (red); overlay (right). Top: animal path (grey), spikes (colored dots), goals (black dots). Bottom: field detected (colored circles) detected by LoG filter; grey-scale rate maps. **(B)** Density of fields as a function of distance to closest goal in pre- and post-probe; KS test, P=0.01472. **(C)** Proportion of fields at goals in pre- and post-probes; Fisher's exact test, P=0.0145. **(D)** Scatter plot of attraction strength against pre-probe goal distance. Running average (purple line, shaded SEM). Significance threshold: dashed line (t-test on each 10 cm window below 29 cm: P<0.05). **(E)** Running average of mean inferred attraction as a function of distance to closest goal. Spearman correlation analysis, r=-0.192, P=0.01225; standard-deviation (shadow). **(F)** Memory retention (time at goals) against normalized mean firing rate increase at goals. One dot per session. Line: regression analysis, r=0.7, P=0.011. **(G)** Grid score in open field (left) and distorted open field (right, generated following real-data field movement distribution); 1-way ANOVA, P=0.0003.

how the expression of assemblies dynamically shifted toward post-probe goal representation. We computed the Fisher's z-scored correlation coefficients of the population activity in 125 millisecond time bins, with the population vectors representing the current location of the animal in the pre- and post-probes (see example session on Fig. 4.4B and S4.13). As learning progressed, similarity to post-probe representation increased both in MEC and CA1 cell assemblies. Yet, when examining the fine temporal structure of assembly expression during learning, we observed a fast-paced flickering between pre- and post-probe representations. The distribution of z-scores were then compared with a control distribution obtained by a cell-id shuffling procedure (Fig. 4.4C). Significant differences were detected for both CA1 and MEC (KS test, P<0.00001 for all combinations). Moreover, we observed a significantly heavier-tailed distribution of real z-scores compared to the control shuffled ones (binomial test, P<0.0001). Real data, therefore, showed a stronger tendency to have extreme values, suggesting that there were no intermediate representations but rather flicker between the two competitive representations of pre- and post-probes. The MEC flickering we observed was reminiscent of previously reported CA1 flickering (Dupret et al., 2013; Jezek et al., 2011). Given that MEC vector correlation increased with goal distance (Fig. 4.4D), we restricted flickering analyses to goal locations. The distribution of flickering scores was significantly different and generally shifted towards more positive values (Fig. S4.12–13, KS and Mann-Withney U tests, P<0.0001). We also observed inter-regional differences in goal flickering dynamics with CA1

reaching a plateau faster than MEC during learning (Fig. S4.12–14).

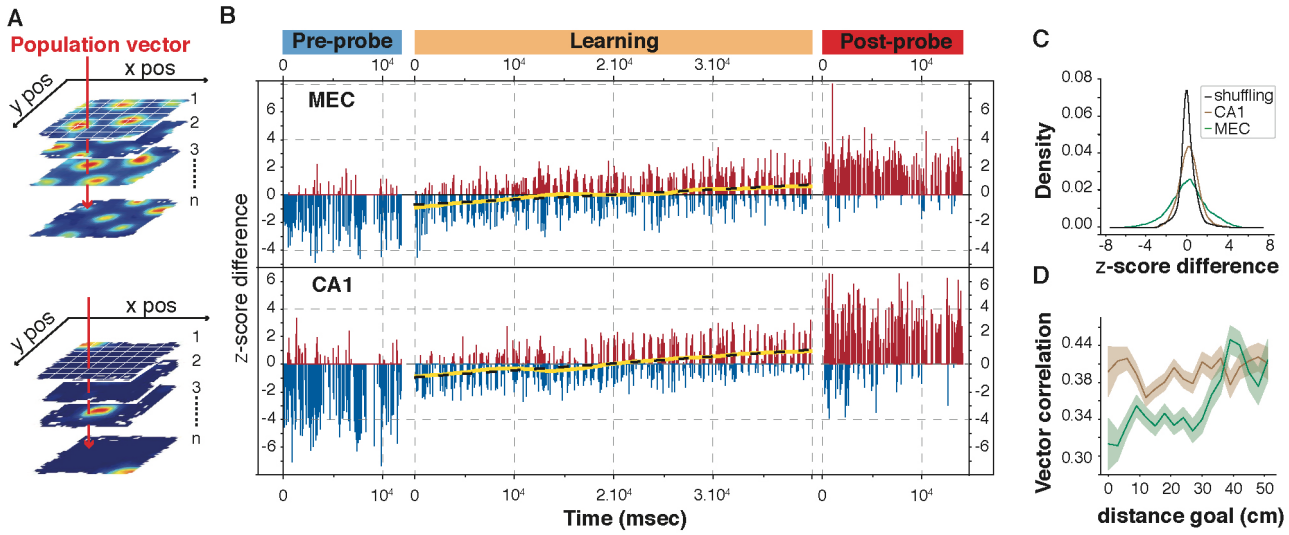

Fig. 4.4. Flickering in MEC and CA1 between co-existing representations. (A) Schemas illustrating population-vector computation. **(B)** Example of progression of z-scored correlation coefficients of ensemble activity across paradigm with pre-probe (negative, blue) and post-probe (positive, red) population vectors in MEC (up) and CA1 (down). Yellow line: smoothed score progression; black dashed line: regression analysis. **(C)** Distribution of z-scores in MEC (green), CA1 (brown) and cell- ID shuffling (black); KS-test: MEC vs. shuffling: $P<0.00001$, CA1 vs. shuffling: $P<0.00001$. The shuffling distributions did not differ: KS-test, $P>0.1$. **(D)** Running average of population vector similarity as a function of distance to goal. MEC: green; CA1: brown. 1-sided-t-test regression analysis: MEC: $r=0.1181$, $P<0.00001$; CA1: $r=-0.00227$, $P=1$. Spearman correlation analysis: MEC: $r=0.10415$, $P<0.00001$, CA1: r=- 0.02331, $P=0.02369$. Shadows: standard-deviation.

4.4 Discussion

Goal learning can lead to the local and long-lasting distortion of the entorhinal spatial maps. This demonstrates the influence of non-geometrical cognitive factors onto the grid structure itself. These findings support emerging theories proposing that the grid pattern carries a broader organizational role for both spatial and non-spatial information in more complex and naturalistic behaviors (Bush et al., 2015; Garvert et al., 2017). Grid structure distortions have been recently linked to the geometrical features of the recording environment (Stensola et al., 2015; Krupic et al., 2018; Barry et al., 2007), which may reflect distorted perception of space. Here, we found evidence for a grid code at the structural level that goes beyond simple metrics: individual grid fields moved towards newly-learned goal locations, leading to the deformation of the grid map, independently from variations of spatial sampling, trajectories, speed or heading inherent to our behavioral paradigm. Field attraction strength to a goal was proportional to the original goal-field distance, locally constraining the deformation of the entorhinal spatial representation. While local remapping took place in MEC, CA1 reorganized through global remapping. Moreover, field reorganization towards goal location was maintained overnight for MEC but not for CA1 although both maintained reorganized fields after learning (up to 2h). This argues for a higher lability of CA1 spatial memory traces compared to MEC cells (Hainmueller and Bartos, 2018). The role of the hippocampus in goal encoding was recently highlighted by reports of a subpopulation of CA1 neurons with an angular tuning for goal direction (Sarel et al., 2017; Gauthier and Tank, 2018). While our results are consistent with the role assigned to CA1 in computational models of goal directed navigation (Bush et al.,

2015; Hinman et al., 2018), differences in goal coding between CA1 and MEC require updates in the current models. Finally, we showed that assembly expression of different goal-related context rapidly alternated – flickered – in the MEC during learning, similar to CA1 flickering (Dupret et al., 2013; Jezek et al., 2011). The simultaneous expression of the old and new MEC representations in the same trial periods, suggests that multiple maps can be stored in the MEC. The apparent absence of intermediate representations in MEC has implications as to how new/modified maps could be dynamically encoded and (re)organized in the CA1-MEC circuits during learning. This may open new avenues of computational research as to its role in inferential reasoning associative memory.

4.5 Materials and methods

Subjects and Surgery Four male long-Evans rats (3–5 months old, 300–400 g at implantation, housed and food deprived as described previously (O'Neill et al., 2017) were implanted with 16 independently movable tetrodes under deep anesthesia using isoflurane (0.5%–3%), oxygen (1–2 l/min), and an initial dose of buprenorphine (0.1 mg/kg). Our tetrodes were arranged in a drive with two bundles: one targeting the CA1 region of the hippocampus and the other the medial entorhinal cortex (MEC). All procedures involving experimental animals were carried out in accordance with Austrian animal law (Austrian federal Law for experiments with live animals) under a project license approved by the Austrian Federal Science Ministry. Each tetrode consisted of four 12 µm tungsten wires twisted together with their tip gold plated to reduce electrode impedance to 200–600 kΩ. Surgical implantations of electrodes were conducted following procedures previously described (26). In brief, MEC and CA1 bundles consisted of 8 tetrodes each. MEC bundles were positioned on the medio-lateral (ML) axis at [4.0 mm to 5.7 mm] lateral to the midline and, on the antero-posterior (AP) axis, [-7.4 mm to -8.8 mm] posterior to bregma. CA1 coordinates were ML [2.7 mm to 3.7 mm] and AP [-2.7 mm to -3.7 mm]. Electrodes were implanted 0.9 mm below dura at a 10-degree angle, relative to the skull.

Data acquisition General data acquisition procedures have been described previously (O'Neill et al., 2017). In brief, after a recovery period of 1 week, tetrodes were progressively lower in maximum steps of 200 µm per day until reaching CA1 pyramidal layer and MEC superficial layers while the animals were trained on the cheeseboard maze (see behavioral training procedures below).

Apparatus The cheeseboard maze is similar to what has been described previously (Dupret et al., 2010). It consisted of a PVC circular board (120 cm in diameter, 2 cm in thickness) with a total of 177 food wells (2.5 cm in diameter, 1.5 cm in depth) drilled into the surface of the maze in evenly spaced parallel rows and columns (8 cm between the centers of each well). A PVC grey start-box (27 cm long, 19 cm wide and 59 cm high) was equipped with a door (35 cm high and placed along the edge of the board perpendicular to the rows of food wells. The top of the box was open to allow tracking the animal inside. A small glass-made cup (2.5 cm in diameter) was placed inside the start-box. Three out of four animals were also trained on a familiar open field (OF) arena very similar to the cheeseboard maze and, in most sessions, consisting of a plain circular environment of the same diameter and the same material, without food wells drilled in it. In two sessions, we used a rectangular environment. The same start box was used for both environments. All environments were surrounded by black curtains and polarized by a 30 cm wide white cue card attached to the curtain.

Training procedures Training procedures were similar to what has been described previously (Dupret et al., 2010). In brief, rats were pre-handled before surgical procedures. Following the postoperative recovery period, they were food-deprived so that their weight was reduced and maintained at 85% of their age-matched preoperative weight. They were first habituated to retrieve pellets in the open field while electrodes were lowered to the region of interest. Once coverage of the open field was satisfactory, animals were trained to retrieve hidden pellets on the cheeseboard. First, the rat was allowed to freely explore the whole cheeseboard for at least 30 min for 3 days. Then, the rat was trained to chase for food-rewards and come back to the start-box. Three groups of visible food pellets (MLab rodent tablet 45 mg, TestDiet) were spread out on the surface of the cheeseboard maze while the rat was inside the start box. For each trial, the door was temporarily opened, the animal was allowed to exit the box and retrieve all the rewards while another additional reward was placed in the glass-made cup situated within the start- box. Once all the rewards had been collected, the door was re-opened, and the rat was gently conducted back to the start box to find and consume the additional food reward within the

start box. That procedure was repeated until the rat started to return back consistently on its own after having collected all the rewards within the board (3–4 days). A similar procedure was applied over the following days, this time with three hidden rewards (i.e., one food pellet per baited location) within the cheeseboard maze (2–3 days). The same baited locations were used from one day to the other. To prevent the use of an odor-guided search strategy during these experiments, food pellet dust was scattered across the maze before each experiment, the board was periodically wiped (using the towel used to handle the rat daily) and the board was rotated relative to the start- box between learning trials and between rest and probe sessions. This initial phase of the experiment ended when the rat was familiar with the whole procedure.

Behavioral paradigm: cheeseboard spatial memory test The animals were housed in a separate holding room and were taken to the recording room each day prior to the experiments. Each daily experiment consisted of a sequence of five recording sessions in the following order: a probe test ("pre-probe"), an immobility/sleep rest session ("pre-rest"), a learning session, an immobility/sleep rest session ("post-rest") and a probe test ("post-probe"). The two probe tests (15-20 min) were never rewarded. After both the pre-probe and the learning sessions, rats were allowed to settle down within the start box for the rest sessions ($\sim$25 min). During the learning session, rats were given successive trials ($\sim$40 trials) to find the three hidden rewards placed in randomly selected food wells. Three out of four animals were also exposed daily to a familiar open field (OF) arena. In that case the rat was let free to explore (15-20 min) the familiar open field environment before the beginning or after the end of the above described experimental procedure. Pellet dust and pellet crumbs were scattered on the disk. In all, 13 sequences of probe–rest–learning–rest–probe in four animals were analyzed, with a total of 490 CA1 pyramidal cells and 262 MEC principal cells. 9 out of 13 sessions (3 rats) also had the OF exploration before the pre-probe or after the post-probe exploration, with a total of 370 CA1 pyramidal cells and 209 MEC principal cells.

Spike sorting The spike detection in the local field potential and sorting was performed as previously described (Csicsvari et al., 1999). Action potentials were extracted by first computing power in the 800–9000 Hz range within a in a sliding window (12.8 ms). Action potentials with a power of $>$5 SD from the baseline mean were selected and spike features were then extracted by using principal components analyses. The detected action potentials were then segregated into putative multiple single units by using the Klustakwik automatic clustering software (Harris et al., 2000) (http://klustakwik.sourceforge.net/). These clusters were then manually refined by a graphical cluster cutting program (Csicsvari et al., 1999). Only units with clear refractory periods in their autocorrelation and well-defined cluster boundaries were used for further analysis. We further confirmed the quality of cluster separation by calculating the Mahalanobis distance (Harris et al., 2000) between each pair of clusters. Periods of waking spatial exploration, immobility, and sleep were clustered together and the stability of the isolated clusters was examined by visual inspection of the extracted features of the clusters over time. Pyramidal cells and interneurons in the CA1 region were discriminated by their autocorrelations, firing rate (average 0.1 5 Hz) and waveforms, as previously described. Putative MEC principal cells were identified by their firing rate: they had to keep a stable average firing over the entire recording day between 0.1 and 7 Hz. In this way we were able to identify the activity of 262 putative excitatory MEC neurons (209 with OF), as well as 490 CA1 pyramidal units (370 with OF).

Histology and reconstruction of recording positions Electrodes were not moved after the final recording session. The rats were killed with an overdose of pentobarbital and were transcardially perfused with 0.9% saline (wt/vol) followed by 4% formaldehyde (wt/vol). The brains were extracted and stored in 4% formaldehyde. At least 24 h later, the brains were quickly frozen, cut in sagittal sections (30 μm) using a cryostat, mounted and stained with cresyl violet (Nissl). Every section in the area of the tetrode trace was retained. The positions of the tips of the recording electrodes were determined from digital pictures of the brain sections. The laminar locations of the recording electrodes in MEC and CA1 were determined on the basis of cytoarchitectonic criteria (29), as well local field potential sharp-wave response during rest (30).

Behavioral performance Behavioral performance was calculated offline using the animal's position records from the tracking data. Learning performance was assessed by constructing learning curves based on the time and the distance travelled to retrieve all three rewards for each trial. Since the baseline time and distance changed daily, we normalized the learning curves by each day's baseline performance (last 20 trials) and then averaged them pointwise in order to compute an overall mean learning curve. Memory retention performance was assessed during the first 5 min of each probe by scoring the time spent in the goal areas, defined as 15 cm in diameter centered on the learned bait locations. In subsequent analysis, the learning was subdivided in learning blocks, each consisting of seven trials and an overlap of two with the neighboring one (Fig. S4.5H).

Occupancy maps Position estimates were based on tracking the middle positions between LEDs on the head stage. The x–y plane of the cheeseboard was divided into bins of 3 cm $\times$ 3 cm and occupancy- maps were calculated during exploratory epochs (speed $>$3 cm/s) measuring the amount of time spent in each

spatial bin by the animal, based on the tracking data. The number of periods ($\tau = 25.6ms$) spent in each bin were counted using a Triweight Kernel (Dunn et al., 2017) with a bandwidth of σ = 3cms and centered at the center of each spatial bin x_b :

$$occ(x_b) = \sum_x \tau K(x_b, x)$$

where

$$K(x_b, x) = \frac{4}{9\pi\sigma^2}\left[1 - \frac{||x_b - x||^2}{9\sigma^2}\right]$$

Spatial firing rate maps The x–y plane of the cheeseboard was divided into bins of 3 cm × 3 cm and rate-maps were calculated during exploratory epochs (speed >3 cm/s) by dividing the number of spikes recorded in each bin by the occupancy and then smoothed with a Gaussian filter with a standard deviation (SD) of two bins. Bins with less than 250 ms occupancy time were not considered. The number of spikes in each bin were counted using a Triweight Kernel with a bandwidth of 3 cm. The peak rate was defined as the rate in the bin with the highest rate in the firing rate map. Normalized firing rate maps were the original firing rate maps divided by the peak firing rate (unless otherwise specified). Z-score normalized maps were the original firing rate maps to which the mean firing rate was subtracted and divided by the map SD.

Sparsity measure and spatially selective cells labelling Sparsity measure represents the proportion of the environment in which a cell fires, corrected for occupancy time. It is formally defined as

$$\frac{(\sum_i P_i R_i)^2}{\sum_i P_i R_i^2}$$

where P_i is the probability of the rat occupying bin i and R_i is the firing rate in bin i. Hippocampal place cells were screened for their spatial tuning using a sparsity value of no more than 0.3, whereas entorhinal spatially modulated cells were selected using a sparsity threshold of 0.6. This discrepancy is due to the fact that entorhinal cells usually present multi- field firing patterns, giving rise to a higher sparsity score even when spatially modulated.

Shuffling procedure for cell labelling To keep the firing structure of each cell while disrupting the link between firing and spatial position we used a spike list wrapping shuffling procedure (32). In brief, considering a spike train with time span [0, END], a random time between 20 sec and END -20 sec was picked and added to all the spiking times. The spiking times which resulted above END were brought to the beginning by subtracting END, and the new 'wrapped' spike train was used to compute a new rate map. This was repeated 200 times for each cell in each environment independently.

Spatial autocorrelograms The autocorrelogram represents the map of shifted self-coherence and was calculated as in (Boccara et al., 2010). For any regularly spaced grid of gaussian bumps one expects to obtain again the same regular grid as autocorrelogram, because the coherence will be high when the shift lets the peaks overlap again.

Grid score The grid score represents the amount of rotational coherence of a certain map. It was calculated similarly to (Stensola et al., 2015): From a spatial rate autocorrelogram whose center was excluded, we considered the Pearson correlation of the autocorrelogram rotated by 30, 60, 90, 120 and 150 degrees (±3 degrees offsets). Only bins closer to the center than a radius s were considered. The grid score using this particular radius s, was defined as the difference between the average of the maximum correlations around 60 and 120 degrees (±3 degrees offsets) and the average of the minimum correlations around 30, 90 and 150 degrees (±3 degrees offsets). Eventually, the final grid score of the cell was defined as the maximum grid score over values of s ranging from twenty to forty bins, computed at intervals of one bin width. A small amount of ellipticity (among the axes ratios 1, 1.1 and 1.2) was also allowed in order to correct for possible deformations in the grid structure (Stensola et al., 2015).

Grid cells labelling A cell was labelled as grid cell only when its grid score was higher than the 95 th percentile of the distribution of scores coming from a spike list wrapping-shuffling procedure (see above). The threshold was calculated separately for each cell in each single environment. The cells were classified as grid cells based on their score in the open field familiar and stable environment (3 out of 4 animals). The scores in pre-probe were used only when the open field data was not present (1 out of 4 animals). With this procedure a total of 56 cells were identified.

Head-direction modulation To determine whether cells had their activity modulated by head-direction, we computed the mean vector length of the angular spiking distribution (Boccara et al., 2010). A cell was labelled as directional only when its mean vector length was higher than the 95 th percentile of the distribution of scores coming from a spike list wrapping-shuffling procedure (see above). The threshold was calculated separately for each cell in each single environment.

Accumulation of firing at goal locations This analysis tested for high spiking activity close to the goal locations. The bins with high firing rate were detected as bins with an intensity above 80% of the peak. The cells with high firing within 15 cm from any goal location were counted in pre-probe rate maps and post- probe rate maps. Maps were computed using the activity of the first 10 minutes of each probe, in order to standardize the time considered across animals and sessions. The increase between pre-probe

and post-probe was tested using a Fisher's exact test on contingency tables. The same procedure was used during learning on rate maps computed during trial blocks (as described above, 7 trials with an overlap of 2) and averaged across cells.

Downsampling procedure To verify that the observed effects were not due to the fact that the rat spent more time close to goal locations in post-probe, we used a random downsampling procedure. Each cell was associated with a list ($position_t$, $spike_t$), where $position_t$ represents the (x,y) coordinate of the animal at a particular time point and $spike_t$ the number of spikes the cell emitted in that particular time bin. Each point was then associated to its own spatial bin in a map with 10 x 10 spatial bins (15 cm size). To correct for uneven sampling between pre and post-probe, we randomly sampled while allowing repetitions, for each map, a number of events corresponding to the minimum occupancy of the two. Using the downsampled data, we calculated new rate maps using the same procedures as described above. This was repeated 200 times for each cell.

Linear–nonlinear model and Poisson simulations To assure that the change observed in the rate maps were not due to behavioral differences, we used a linear–nonlinear (LN) model for the entorhinal cells that takes into consideration position, heading and speed of the animal. The model has been used for entorhinal cells by (Hardcastle et al., 2017) and is described there in detail. In brief, this model has been used to quantify the dependence of spiking on a combination of variables (position, heading and speed): it allows to estimate the firing rate of a cell during a time bin as an exponential function of the sum of the relevant value of each variable projected onto a corresponding set of parameters. Once the LN model was fitted to the data observed in pre-probe, the expected firing rate were computed based on the behavior of post-probe. Using the instantaneous expected firing rate, we simulated the firing of a cell in post-probe as a non-homogeneous Poisson process (31). This procedure was repeated 200 times for each cell and is referred through the text as linear- nonlinear Poisson spiking (LNP) model.

Firing-by-vicinity score The firing-by-vicinity score correspond to the sum of the firing field rates of cells near goals during learning. To obtain this score we first constructed a two-dimensional kernel function by summing three 2D Gaussian kernels with a SD of 30 cm and centered at each goal. We then multiplied the normalized firing rate map of each cell with the two-dimensional kernel function. Finally, we summed up the kernel-multiplied spatial firing rates. Thus, we calculated, for each cell, their firing rates near the goals, taking into consideration those rates in the immediate vicinity and less those further away. During learning, we computed the score for each cell on a sliding window of seven trials in order to have a better temporal resolution and compared it across cell types.

Firing field detection It has been recently shown that firing fields of grid cells present different firing intensities, so a thresholding procedure is not suitable to detect grid fields (Dunn et al., 2017). To detect grid fields independently of their firing intensities, we employed a Laplacian-of-Gaussian (LoG) filter blob detection algorithm (31). This algorithm consists of convolving the image with a LoG kernel. The Laplacian filter measures the local curvature of a surface and is defined as the sum of the unmixed second derivatives of a (possibly multidimensional) real function and can be thought in 2D as a measure of local curvature of a surface. The convolved image emphasizes areas of high curvature of a gaussian-smoothed version of the image and will have negative peaks where the peaks of the fields are (Fig. S4.7). The gaussian kernel, used before the Laplacian filter, had the standard deviation of a gaussian bell fit on the central peak of the autocorrelogram. This measure was also used as the average dimension of the fields of a map. The detected peaks were then filtered using the following criteria: -) Intensity: the intensity had to be higher than the 75 th percentile of the distribution of intensities of the map. -) Overlap: if two fields overlapped, only the one with the strongest intensity was considered. -) Distance to the border: in order to avoid "half fields" or distorted fields close to the borders, all the fields detected less than 10 cm from the border were discarded.

Firing fields matching To study the movement of individual grid fields between pre and post-probe, we paired them based on their spatial overlap and distance between centers. Once the fields were detected in pre- and post-probe, the pairing worked as follows: -) We calculated the distance of each field in pre-probe to each field in prost-probe. -) Fields were matched starting from the pair with the biggest overlap (smallest distance). Once matched they were then excluded. -) The fields were paired up until reaching a maximum distance of 20 cm. The threshold of 20 cm is given by the fact that the average spacing between fields in our dataset was 35 cm and the average field size was about 15 cm.

Significant movement towards a goal We used two methods to determine the significance of movement towards a goal of each field against LNP data with method 1 or downsampled data with method 2 (Fig. 4.1G). Method 1: We first simulated 100 times the pre-probe firing fields with the LNP spiking model using the real post-probe behavior. We then evaluated the number of times a firing field moved more in the LNP-generated maps than in the real maps. Next, we compared each field movement against chance on a cell-by-cell basis using a 95% significance level. Finally, we compensated for multiple comparison with a Holm-Bonferroni correction. Method 2: This method is very similar to method 1. We randomly downsampled 100

times the pre- and post-probe rate maps in order for them to reach the same occupancy. We then evaluated the number of times a firing field moved more in the downsampled maps than in the original maps. Finally, we compared each field movement against chance on a cell-by- cell basis using a 95% significance level with a Holm-Bonferroni correction. After we assessed whether a field moved significantly or not, we checked whether it moved towards or further away from goals.

Attraction strength In order to quantify the movement of the fields towards the goal locations, we measured the relative movement of each field and weighted it by a directionality score. If we denote with d_pre the distance of the field from a goal in pre-probe and d_post the same in post probe, the attraction strength was defined as:

$$att = (d_{pre} - d_{post})/(d_{pre} + d_{post}) * |cos(theta)|$$

where theta is the angle between the vector of movement and the vector pointing from the field to the goal location.

Inferred attraction strength To verify the hypothesis that attraction is stronger on fields located close to goals, we used an algorithm that tests for the best fit between pre- and post-probe when moving the firing fields towards a goal location. For each pre-probe map, we detected the firing fields and then moved them towards one (or more) goal location(s). Each field was free to move towards any goal (or combination of goals) by any percentage of the distance that separated it from the goal it was moving towards. We then measured the correlation between the post probe map and the modified pre-probe maps, using any combination of movements towards any of the wells, and selected the one that correlated best. The inferred attraction for each field was the percentage of movement of the map that fitted best the post-probe map, and we scatter- plotted it against the distance of the fields in pre-probe in order to carry out a correlation test.

Field deformation, size and spacing To check whether the fields were more deformed on the cheeseboard, we fitted a (diagonal) 2D gaussian bell on each field detected in the open field, in pre- and post-probe; and computed a deformation score. The score was computed as the absolute value of the difference of the two standard deviations divided by the sum of the two. As such we had a score ranging from 0 to 1, where 0 means perfectly round and 1 means perfectly flat. The mean size of the fields was computed as the standard deviation of a perfectly symmetric 2D gaussian bell fitted on the central blob of the autocorrelogram, whereas the mean spacing between fields (i.e. grid scale) as the distance between the central blob and the one outside the center with the highest intensity of the autocorrelogram.

Grid score degradation To test whether the movement of fields observed in probes could explain the drop of grid score, we took the firing rate maps of the grid cells recorded in the open field and applied a movement on each field following the real distribution of movements from pre- to post- probe. We repeated this procedure 100 times for each cell and then compared the grid scores of the distorted maps with the original ones.

Rate remapping To compare the extent of rate remapping between pre- and post-probes, we employed a rate remapping score computed as:

$$remap_{score} = (rate_{post} - rate_{pre})/(rate_{post} + rate_{pre})$$

We compared both the peak firing rate of each cell and also the firing rate of each detected firing field. We then compared the distribution of scores with downsampled data, as well as with the maps generated using the LNP spiking model in order to simulate the spiking using the rate modulation fitted on pre-probe and the real behavior of post-probe.

Correlation with behavioral performance We computed the average increase of firing at goals between pre- and post-probe using the firing-by-vicinity method described above. We then correlated it with the time the animal spent around the most visited reward area in post-probe during the first 5 or 10 minutes.

Population vector The population vector represents the simultaneous average activity of all the selected cells in a particular spatial bin, computed in a certain experimental session. In brief, all the spatial rate maps of the cells considered (here: CA1 place cells or MEC spatially selective cells) in a given session were stacked along the z-axis. Thus, the population vector in a given spatial bin is the 'vertical' z vector for that particular (x,y) set of coordinates (Dupret et al., 2013).

Population vector similarity against distance from goals To analyze the dependence of population vector similarity on the distance from the goals, the spatial rate maps were z-scored and then stacked along the z-axis for pre and post-probe as described above. For each spatial bin (with an occupancy of at least 250 ms in both pre- and post-probe), the Pearson correlation between the two firing rate vectors was calculated and then scatter plotted against the distance from the goal location. We then tested whether the firing rate similarity was positively correlated with the distance from the closest goal using both a one-sided (positive) correlation analysis, as well as with a Spearman correlation analysis.

Flickering To test the effectiveness of learning from a population point of view, we compared the instantaneous firing during learning with the average activity that we observed in pre- and post-probe. The activity

of each cell of the selected ensemble (CA1 pyramidal or MEC principal cells) was binned into 125 ms windows and smoothed with a 1D gaussian filter with 250 ms SD. Each time-window overlapped 100 ms with the previous. We then measured the correlation between the instantaneous population activity in each time-window during learning with the population vector activity of pre- and post-probes. In this way, we obtained two time series expressing the similarity of the instantaneous population activity with the representation observed during pre- and post-probes. We employed a Fisher's z-transform of the correlation coefficients and took their difference. We next compared the distribution of the z-scores against the scores coming from a cell-ID shuffling procedure, and then verified whether the two distributions were significantly different using a Kolmogorov-Smirnov test. Finally, we used a binomial test to check whether the number of scores outside the 95 th percentile threshold (1.645) was higher than chance. This procedure was executed only on sessions that had at least 10 CA1 and 10 MEC units (10 out of 13 sessions).

Average flickering during learning trials To study the temporal behavior of flickering during learning, we averaged the difference of Fisher z-scored correlation coefficients during each learning trial, pooling together all the sessions.

Supplementary figures

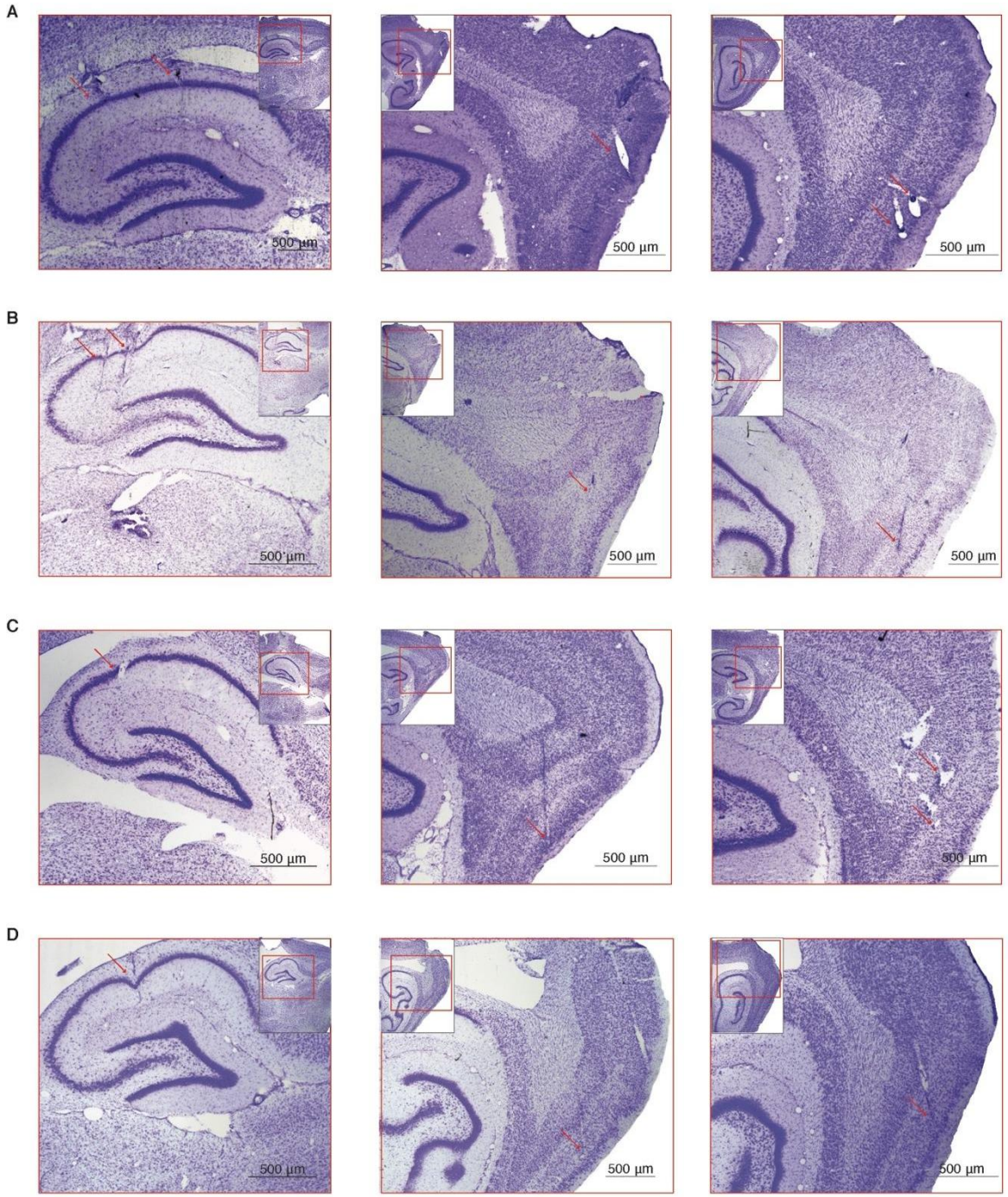

Fig. S4.1. Histological verification of electrode positioning. Nissl-stained sagittal sections showing the medial entorhinal cortex and the CA1 region of the hippocampus with electrode tracks (red arrows) terminating respectively in the superficial layers (II/III) and in the pyramidal cell layer. Inserts show the entire region and the location of the delineation of the magnified regions (red boxes). **(A)** animal 1, **(B)** animal 2, **(C)** animal 3, **(D)** animal 4. The vast majority of the entorhinal cells were recorded in the superficial layers (II/III). A minority of cells were recorded in LV (animal 3, one tetrode).

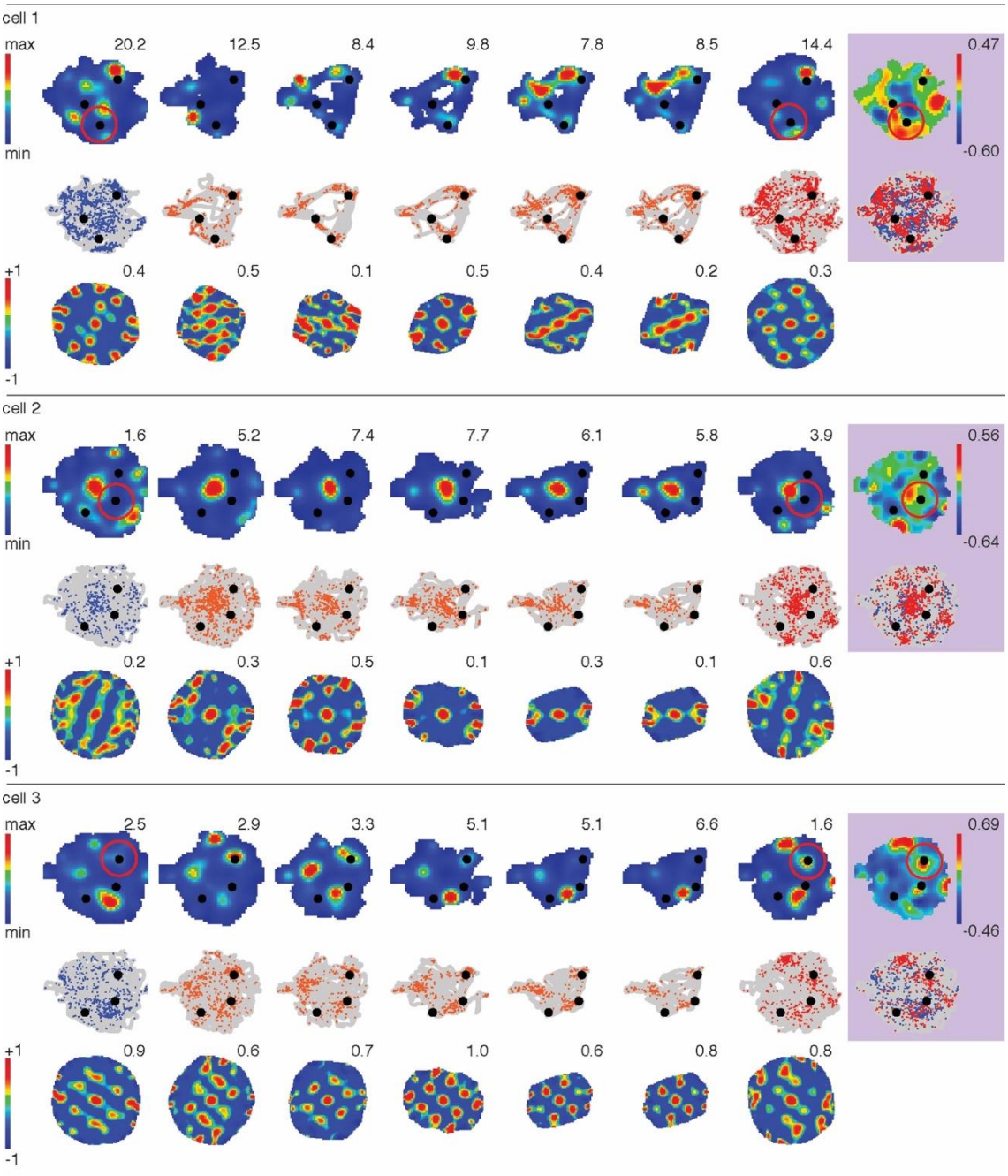

Fig. S4.2. (Part 1/2) Additional examples of grid cell spatial maps during pre-probe, learning and post-probe.

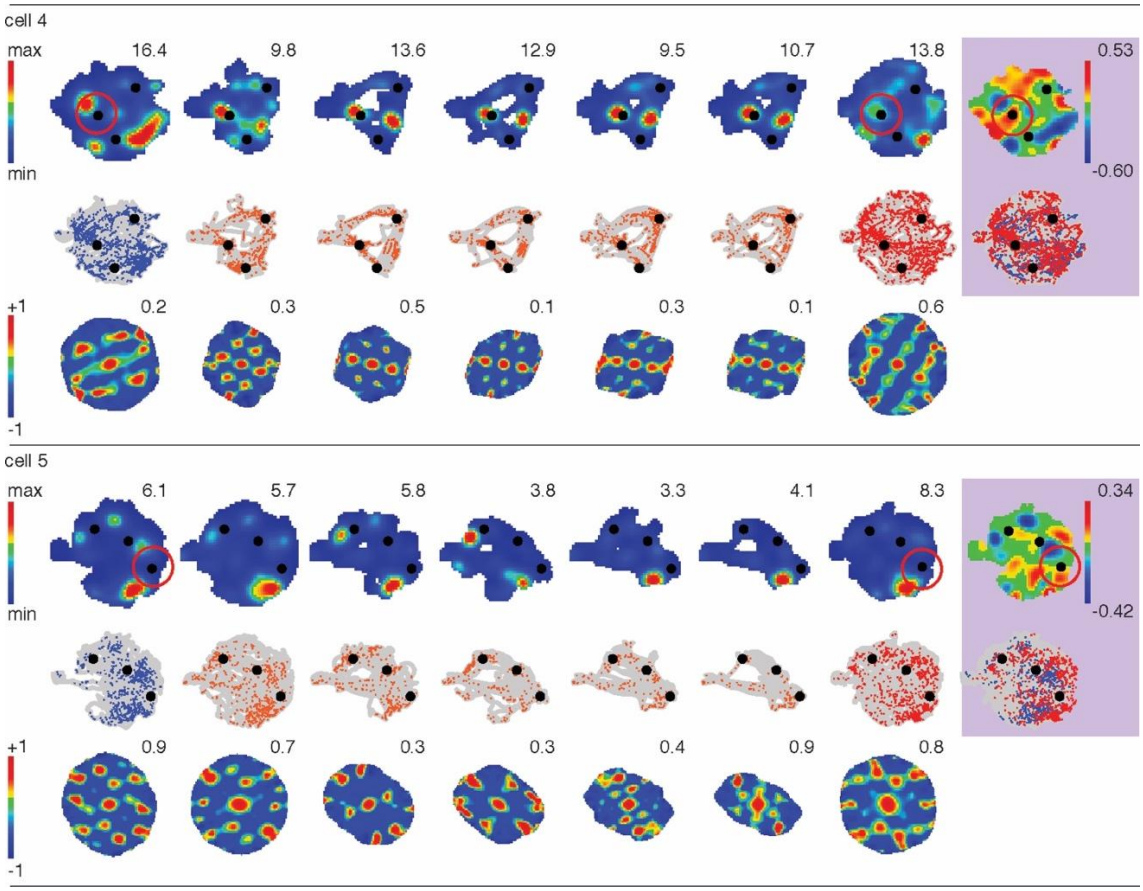

Fig. S4.2. (Part 2/2)Additional examples of grid cell spatial maps during pre-probe, learning and post-probe. Spatial maps of five example MEC cells showing the accumulation of activity around goal location after learning. For each cell there are three panels of maps from top to bottom. Top panels show the firing rate maps, scale on the left of the raw and peak firing rate (Hz) in the upper right corner of each map. Middle panels represent the path of the animal (grey) with colored dots (blue: pre-probe; orange; learning; red: post-probe) representing the spikes of the cells. Bottom panels show the spatial autocorrelograms, scale is on the left of the raw going from -1 to +1. Purple areas on the right are divided in two panels: the upper panels show the differential heat rate maps with scale bar on the right and pre-(blue); the lower panels show the spike overlays of pre-probe (blue) and post-probe (red). The red circles highlight goals pulling strong attraction.

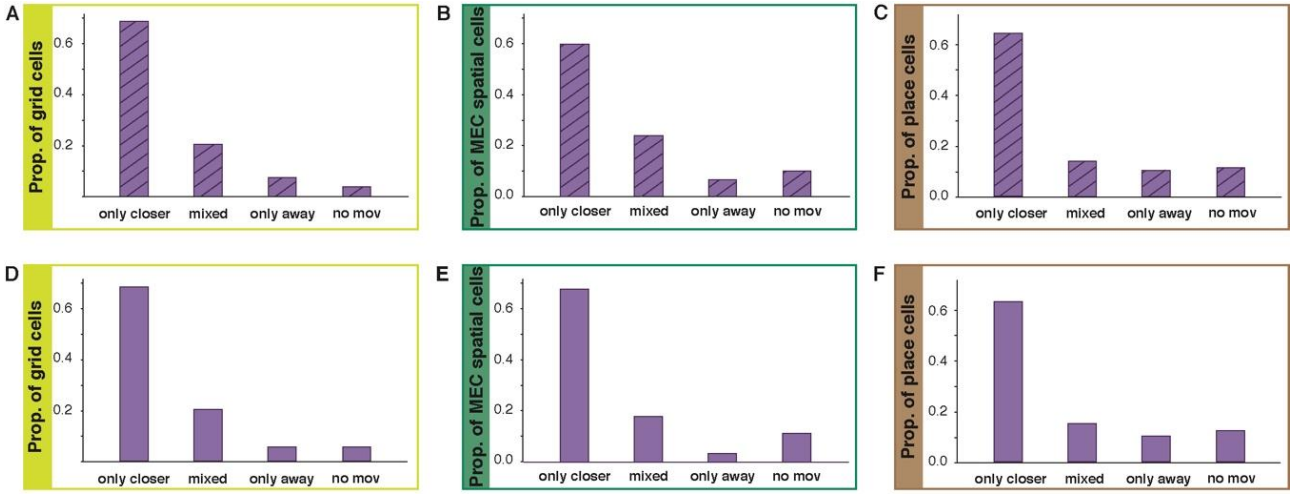

Fig. S4.3. Distribution of movement relative to goal among all spatial cells recorded. (A) Distribution of grid cells according to the significant movement of their grid fields relative to goals. Significance calculated against downsampled data (see methods). Left "only closer": all detected moving fields significantly moved closer to goals (69%). Middle left "mixed": detected fields presented a mixed behavior with some moving towards the goal and some not (20%), thus 89% grid cell show goal-attraction of at least one field (Fig. 4.1). Middle right "only away": all detected fields significantly moved away from goals (7%). Right "no mov": none of the detected fields moved significantly (4%). **(B)** Same as (A) but for entorhinal spatial cells. In the same order as above: 60%, 24%, 6%, 10%. **(C)** Same as (A) but for hippocampal place cells. In the same order as above: 64%, 15%, 10%, 11%. **(D)** Same as (A) but with significance calculated against LNP data. In the same order as above: 66%, 20%, 5%, 5%. **(E)** Same as (B) but with significance calculated against LNP data. In the same order as above: 68%, 18%, 3%, 11%. **(F)** Same as (C) but with significance calculated against LNP data. In the same order as above: 63%, 15%, 10%, 12%.

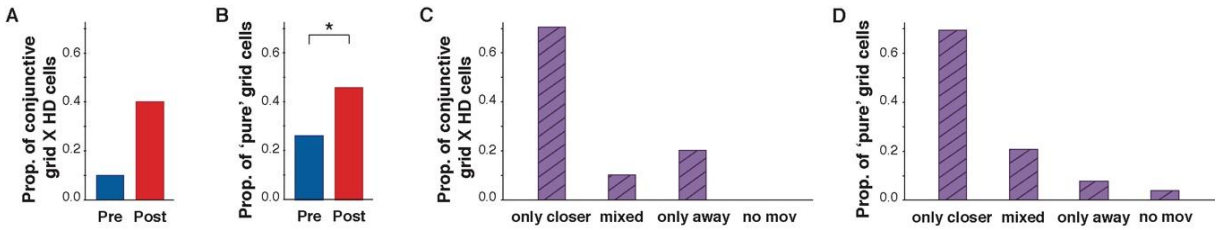

Fig. S4.4. Influence of head-direction (HD) modulation on field movement. (A) Conjunctive grid-by-head-direction cell proportions at goals pre-learning (blue) and postlearning (red). Fisher's exact test: P=0.3, N=10. Binomial test: P=0.0127, N=10. As in Fig. 4.1. Note that the non-significance for the Fisher's exact test is due to the very small number of conjunctive cells. **(B)** As in (A) for non-directional 'pure' grid cells. Fisher's exact test: P<0.001, N=46. Binomial test: $P < 0.0001$. **(C)** Proportions of conjunctive grid-by-headdirection cells (as in Fig. S4.3). Left "only closer": all detected fields significantly move closer to goals (70%). Middle left "mixed": detected fields present a mixed behavior with some moving towards the goal and some not (10%). Middle right "only away": all detected fields significantly move away from goals (20%). Right "no mov": none of the detected fields move significantly (10%). Significance calculated against downsampled data. **(D)** As in (C) for non-directional 'pure' grid cells. In the same order: 67%, 17%, 4%, 2%.

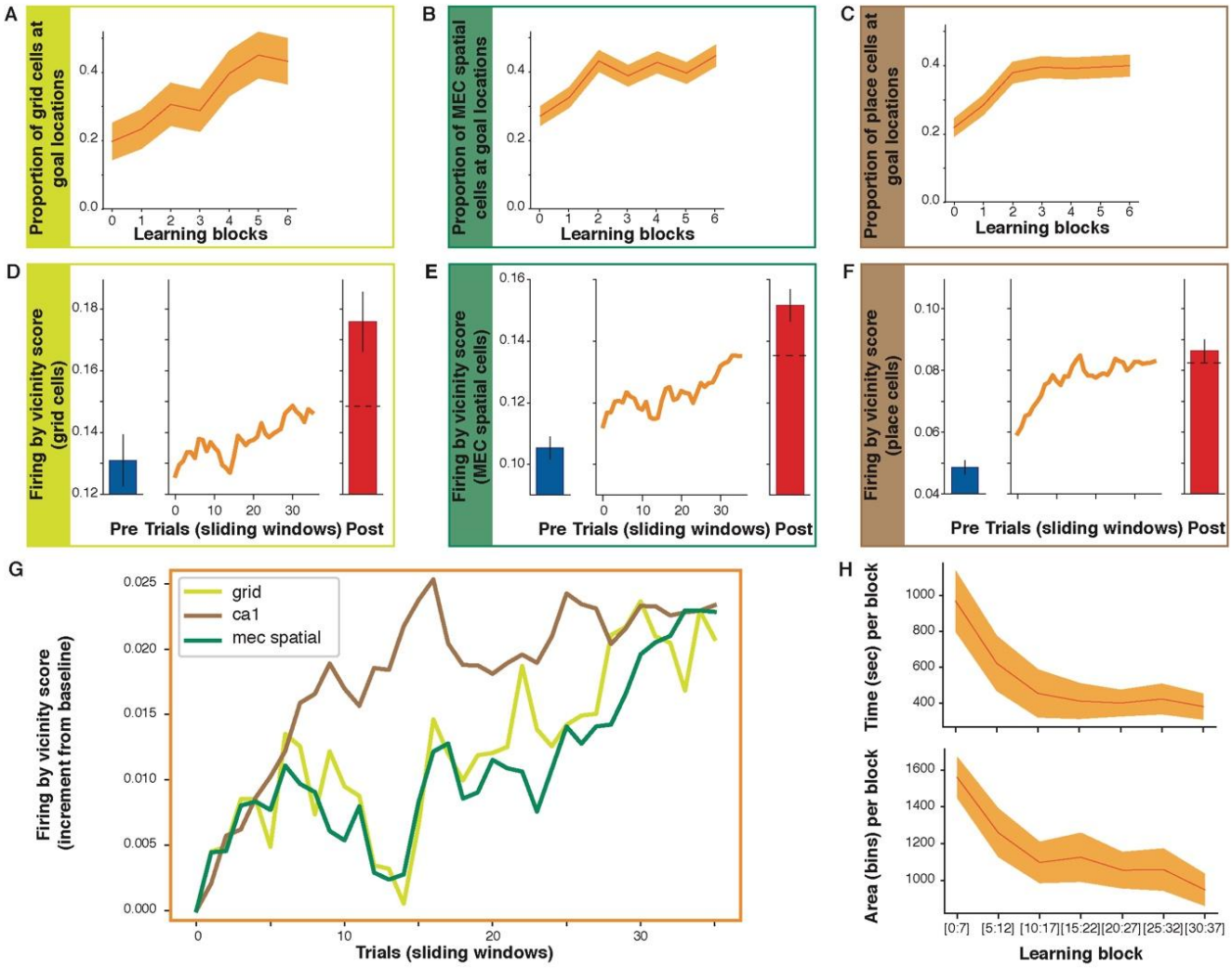

Fig. S4.5. Comparative dynamics of goal-encoding emergence during learning. (A) Proportions of grid cells with strongest firing at goals across learning-blocks. Spearman correlation, r=0.928, P=0.0025. **(B)** Same as (A) for entorhinal spatial cells, r=0.835, P=0.0193. **(C)** Same as (A) for hippocampal place cells, r=0.53, P=0.2193, first 4 blocks: r=0.99, P<0.00001. Orange shadows: standard-errors. The analyses used in Fig. 4.1G (right) and S4.5A–C are based on the selection of the strongest firing fields (top 20% strongest firing bins) moving from outside to inside goal locations. Yet, it did not include the firing fields with low firing rate or movement occurring outside or within goal locations. Thus, we computed a firing-by-vicinity score where we summed the firing field rates of cells near goals by multiplying the firing rate with three 2D kernel functions centered at the goal locations, see fig 4.1G (left) and S4.5D-G. **(D)** Evolution of firing-by-vicinity score in grid cells across behavioral paradigm. Left: average in pre-probe (blue bar +/− SEM). Middle: scores per sliding windows of seven trials (orange). Right: average in post-probe (red bar +/− SEM). Dashed line: level measured at the end of learning. Mann-Withney U test, pre- vs. post-probe: $P < 0.001$. Regression test, trials 1–15: $P > 0.98$, trials 16–30: $P < 0.0001$. **(E)** Same as (D) for entorhinal spatial cells. Mann-Withney U test, pre- vs. post-probe: $P < 0.001$. Regression test, trials 1–15: $P > 0.9$, trials 16–30: $P < 0.001$. **(F)** Same as (D) for hippocampal place cells. Mann-Withney U test, pre- vs. post-probe: $P < 0.001$. Regression test, trials 1–15: $P < 0.0001$, trials 16–30: $P > 0.81$. **(G)** Comparison of firing-by-vicinity score (after baseline subtraction) across learning (seven trials sliding windows): CA1 place cells (brown), MEC spatial cells (dark green), MEC grid cells (light green). **(H)** Average distribution of time (top) and area (bottom) per learning block (as presented in Fig. S4.5A–C).

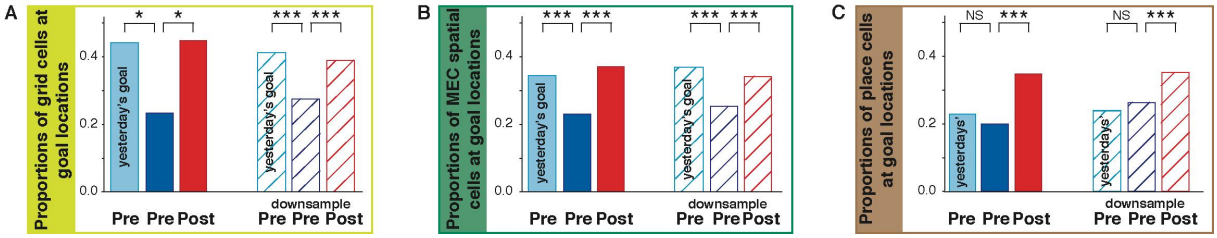

Fig. S4.6. Longer-lasting map deformation in the MEC compared to CA1. (A) Proportion of grid cells around goal locations from previous day before learning ("yesterday's goal", light blue). To allow easy comparison, proportion of grid cells around today's goal are shown in bright blue (before learning) and red (after learning). Fisher's exact tests: pre-probe, old vs. new wells: P=0.01756, N=43. Dashed bars show control data obtained with downsampled data following the same color code: light blue (before learning, old goal locations); bright blue (before learning, new goal locations); bright red (after learning, new goal locations). Fisher's exact tests: pre-probe, old vs. new wells: P<0.00001, N=4300. **(B)** Same as (A) but for entorhinal spatial cells. Fisher's exact tests: pre-probe, old vs. new wells: P=0.0047, N=122; pre-probe, old vs. new wells downsampled: P<0.00001, N=12200. **(C)** Same as (A) but for hippocampal place cells. Fisher's exact tests: pre-probe, old vs. new wells: P=0.432, N=183; pre-probe, old vs. new wells downsampled: P=0.07823, N=18300.

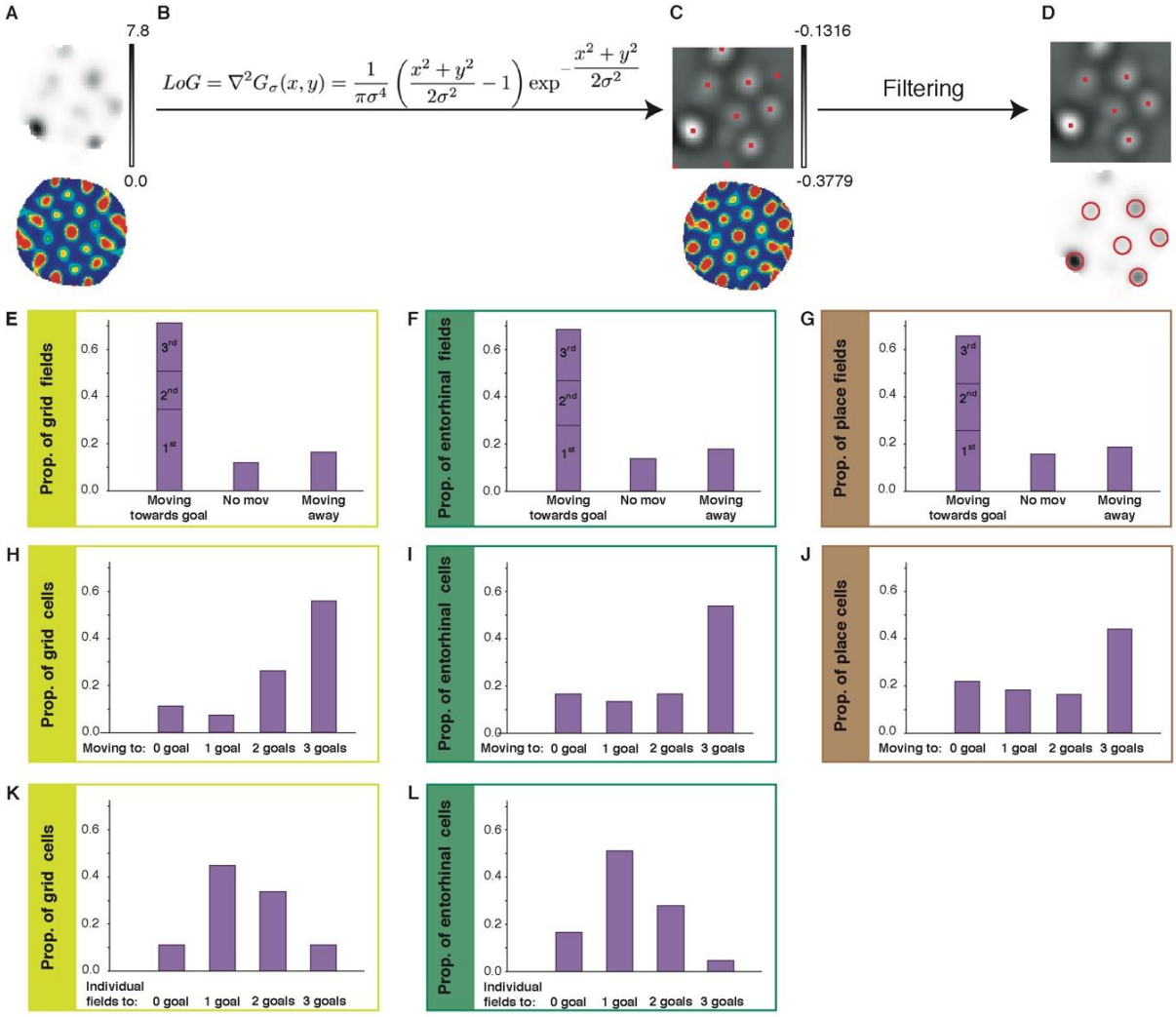

Fig. S4.7. Laplacian of Gaussian (LoG) based field detection and field-to-goal movement. A–D: Schematic of the field detection algorithm. **(A)** Example grid cell with uneven field rate. Up: rate map with peak rate in black. Bottom: spatial autocorrelogram with peak correlation in red. **(B)** LoG kernel to convolve the image. **(C)** Up: same rate map as in (A) convolved with the LoG kernel, where has been estimated as the standard deviation of a symmetric Gaussian bell fitted on the central peak of the autocorrelogram. Red dots: detected local minima (minimum distance of 5 bins). Bottom: Corresponding autocorrelogram of the convolved rate map. **(D)** Up: convolved rate map with the filtered field centers (fields too close to each other, with too low intensity or too close to the border were excluded). Bottom: original rate map with circles on top of the detected fields. E–G: Proportions of spatial fields moving towards goals (left), not moving (middle) or away from goals (right) for grid cells **(E)**, spatial MEC cells **(F)** and place cells **(G)**. Left bars stratification represents proportions moving towards closest (1st), second closest (2nd) or furthest (3rd) goal. H–J: Proportion of cells with fields presenting none (0 goal, left) or significant movement towards one (middle left), two (middle right) or three goals (right) between pre- and post-learning for grid **(H)**, spatial MEC **(I)** and place cells **(J)**. Note that for this analysis, one field may be attracted to several goals. K–M: Proportion of cells whose individual fields showed strongest attraction towards one, two or three goals for grid cells **(K)** and spatial MEC cells **(L)**. Note that for this analysis, each individual field could only be attracted to one goal. Given that the vast majority of place cells presented only one field, this analysis was not performed for them.

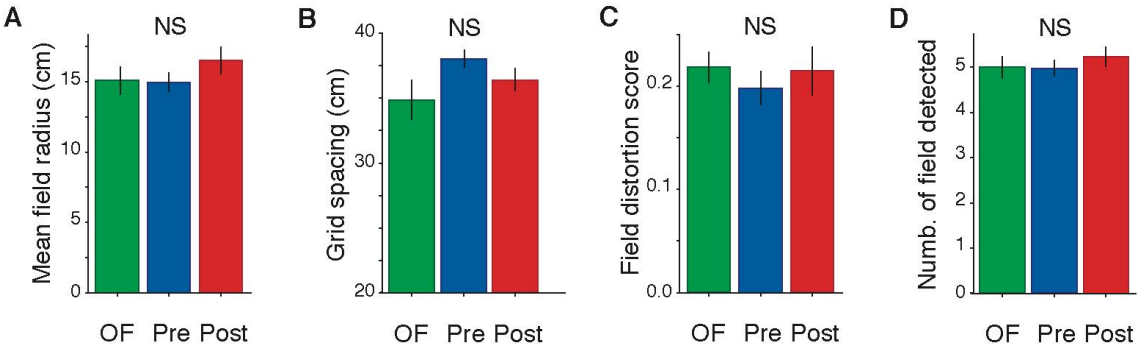

Fig. S4.8. Comparison grid properties across open field, pre- and post-probe. (A) Mean field radius across environments; one-way ANOVA, P=0.38638. **(B)** Mean grid spacing across environments; one-way ANOVA, P=0.12. **(C)** Mean field distortion-score across environments; one-way ANOVA, P=0.31242429. **(D)** Mean number of fields per grid across environments; one-way ANOVA, P=0.59.

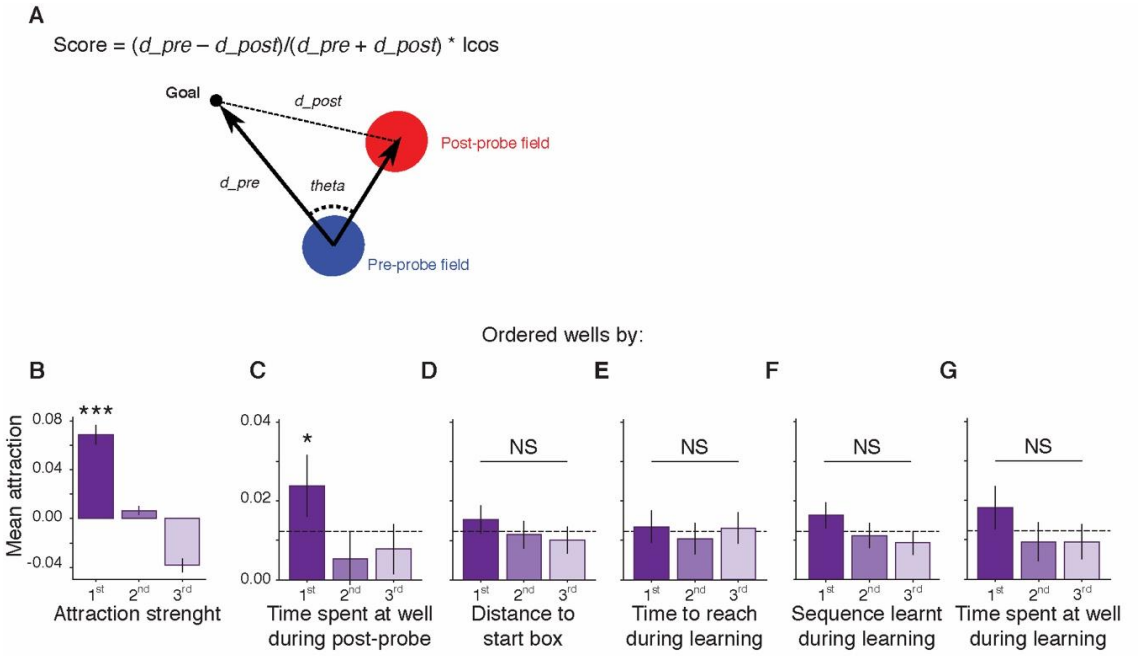

Fig. S4.9. Attraction score and influence of various factors on attraction pulled by each reward well. (A) Schematic illustrating attraction score computation (see methods for details). Grid fields mean attraction strength to wells ordered by: **(B)** attraction strength, **(C)** time spent at the well during post-probe, **(D)** distance of the well to the start box, **(E)** time to reach a well during learning, **(F)** place in the sequence of learning and **(G)** time spent at the well during learning. First ordered well is dark purple (right bar), second well medium purple (middle bar) and third well light purple (left bar). Significance given by one-way ANOVA, B: P<0.00001 (***), P=0.0363. C: P=0.036 (*), D: P=0.225 (NS: Non-significant), E: P=0.301 (NS), F: P=0.124 (NS), G: P=0.119 (NS).

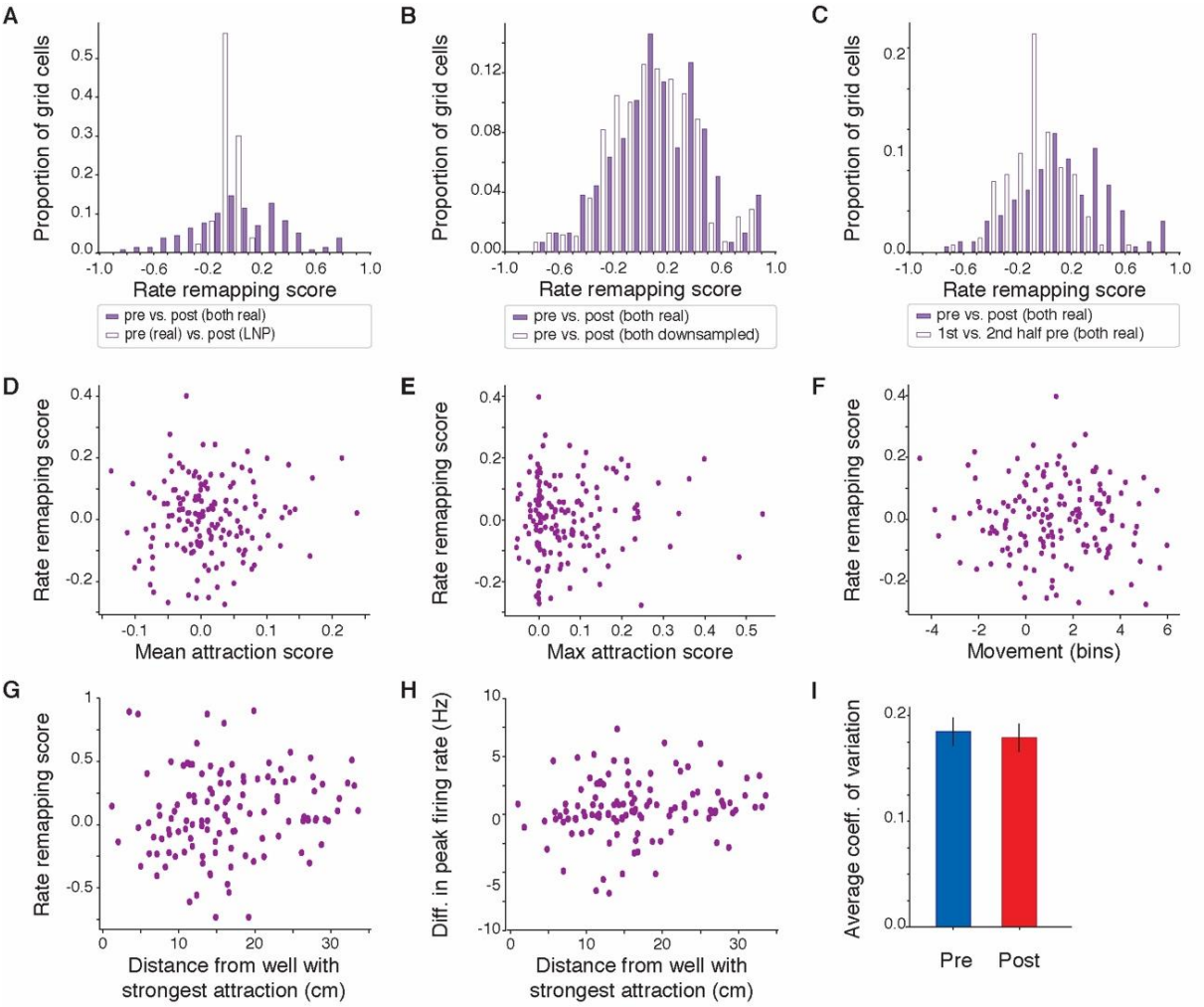

Fig. S4.10. Rate remapping absence of influence on goal remapping. Rate remapping scores were computed for each firing field independently as the difference of firing rate in post- and pre-probe divided by the sum. **(A)** Rate remapping scores between pre- and post-probe in real data (purple bars) or in LNP post-probe maps (blank bars) generated using post-probe behavior and LNP spiking model based on pre-probe activity. Note a significant positive tail showing rate remapping for real data through an increase of rates in post-probe. Rank-sum test: $P<0.0001$, Levene test for Variance: $P<0.0001$. **(B)** Same as in (A) for [pre- vs. postprobe] real data (purple bars) compared to control [pre- vs. post-probe] downsampled data (blank bars). Note the absence of significant differences between the two distributions showing the independence of rate remapping from spatial sampling. Rank-sum test: $P=0.87$, Levene test: $P=0.39$, KS test: $P=0.76$. **(C)** Same as in (A) for [pre- vs. post-probe] real data (purple bars) compared to intrinsic variability in first vs. second half of pre-probe (blank bars). Note a significant positive tail showing positive rate remapping from pre- to postprobes, larger than the intrinsic rate variability within a session. Rank-sum test: $P<0.023$, Levene test: $P=0.0021$. (D-H) Regression analyses showing the absence of correlation between rate remapping and various movement measures on a field-by-field basis. Regression analysis, all $Ps>0.1$. **(D)** Mean attraction score as a function of rate remapping score. **(E)** Maximum attraction score as a function of rate remapping score. **(F)** Field movement as a function of rate remapping score over bins of 3 cm each. **(G)** Distance from well with strongest attraction as a function of rate remapping score. **(H)** Distance from well with strongest attraction as a function of difference in peak firing rate. **(I)** Average coefficient of variation of firing field intensities across cells. one-way ANOVA, $P>0.2$.

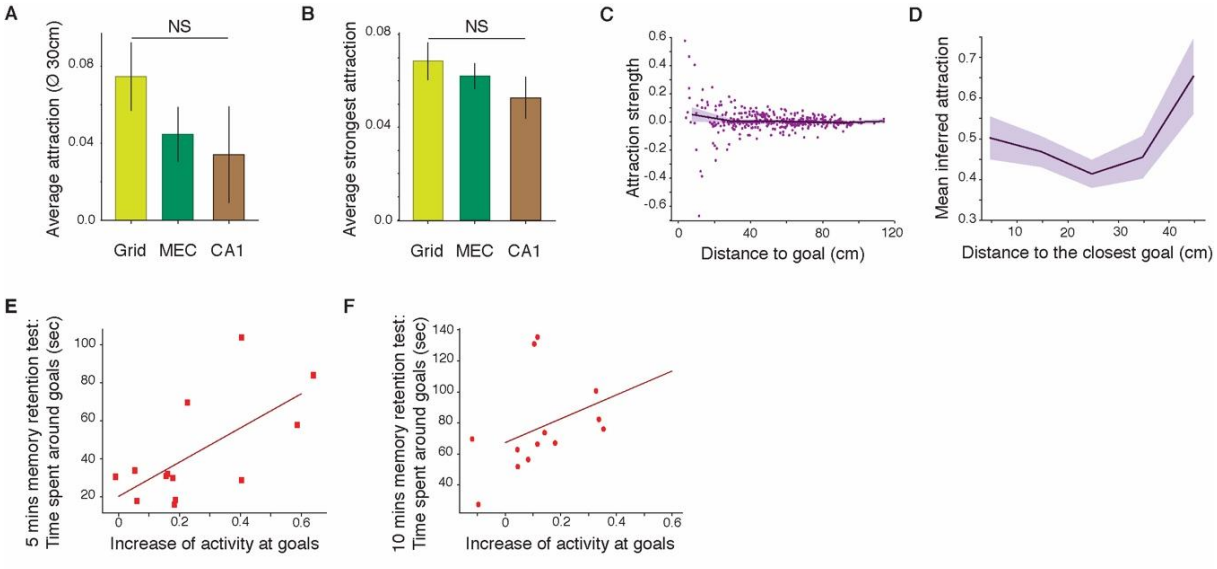

Fig. S4.11. CA1 goal attraction strength compared to MEC: absence of influence of goal distance and weaker correlation with memory retention. (A) Average strength of attraction calculated for fields within a diameter of 30 cm around goals for grid (light green), MEC spatial (dark green) and CA1 place (brown) fields. one-way ANOVA, P=0.3873. **(B)** Same as (A) but for average strongest attraction. one-way ANOVA, P=0.4357. **(C)** Scatter plot of CA1 attraction strength against pre-probe's distance to goal. Running average (purple line, shaded SEM). No significant difference found from zero (t-test on each 10 cm window: all Ps>0.05). **(D)** Running average of mean CA1 inferred attraction as a function of distance to closest goal was not significantly different. Spearman correlation analysis, r=0.0207, P=0.748; shadow: standard deviation. **(E)** Memory retention (time at goals) against normalized mean CA1 firing rate increase at goals during the first 5 mins in post-probe. One dot per session. Red line: regression analysis: r=0.66, P=0.015. **(F)** Same as (D) except that the mean firing rate increase was calculated over 10 mins. Regression analysis: r=0.38, P=0.201. Note that the correlation presented on figure 3F for MEC showed the normalized mean firing rate increase at goals during the first 10 minutes in post-probe. Entorhinal increase of activity at goals also significantly correlated with the 5 minutes memory retention test (data not shown).

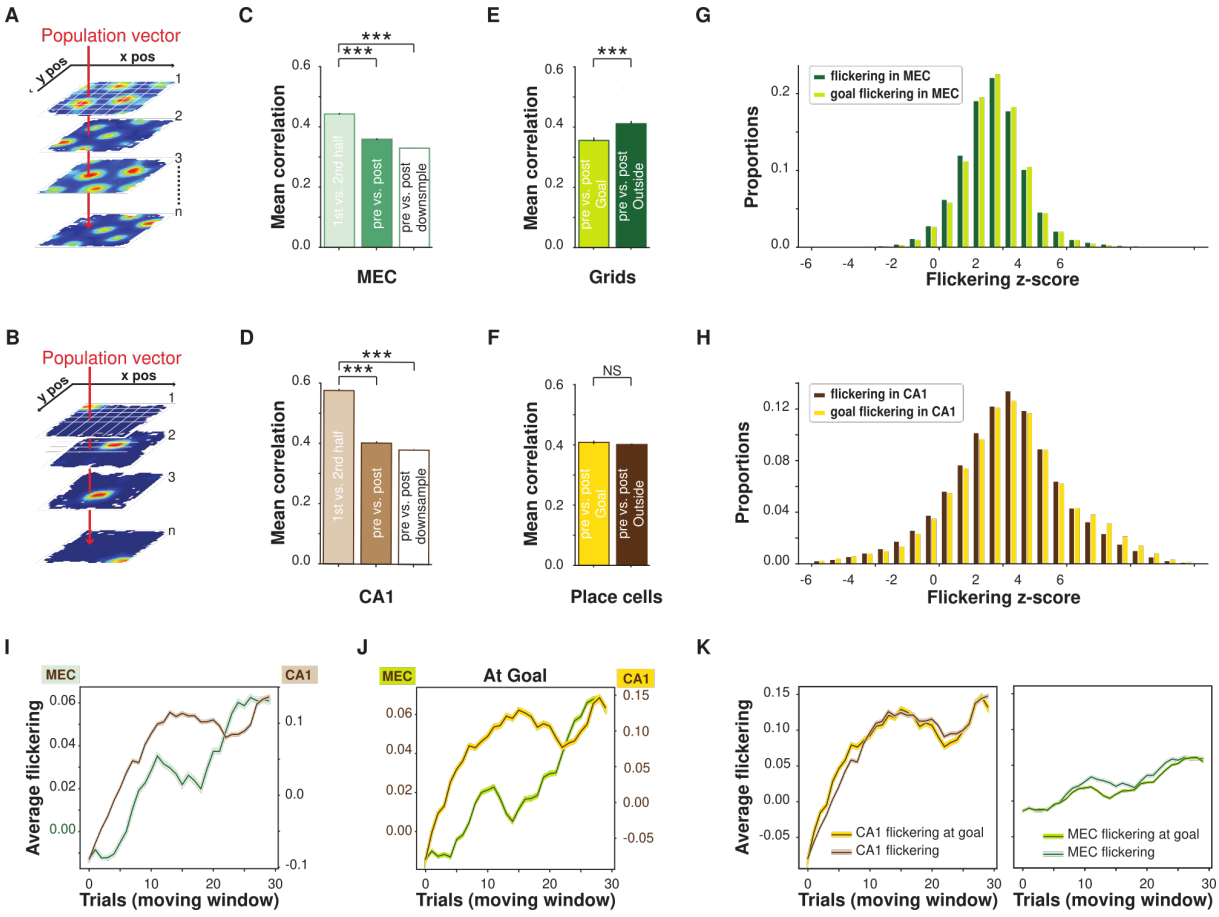

Fig. S4.12. Population vector and flickering analyses: focus on goal areas. (A–B) Schemas illustrating population-vector computation for grid cells (A) and place cells (B). **(C)** Population vector similarity in MEC. Left: first vs. second half of pre-probe (light green bar). Middle: pre-probe vs. post-probe (green bar). Right: downsampled pre-probe vs. post-probe (blank bar); t-tests: $P<0.00001$ for all. **(D)** Population vector similarity in CA1. Left: first vs. second half of pre-probe (light brown bar). Middle: pre-probe vs. post-probe (brown bar). Right: downsampled pre-probe vs. post-probe (blank bar); t-tests: $P<0.00001$ for all. **(E)** Local population vector similarity in MEC restricted around or outside goal locations. Preprobe vs. post-probe for grid cells inside (left yellow-green bars) and outside (right dark green bars) goal area; t-tests: $P=0.00033$. **(F)** Same as (E) for CA1 populations. Pre-probe vs. postprobe for place cells inside (left yellow bars) and outside (right dark brown bars) goal area; t-tests: $P=0.2001$. **(G)** Comparison of distributions of MEC flickering z-scores in the entire environment and restricted to goal areas. KS test: $P<0.0001$. Rank-sum test: $P<0.0001$. **(H)** Same as (G) for CA1 populations. KS test: $P<0.0001$. Rank-sum test: $P<0.0001$. Note an increase flickering in goal areas. **(I)** Average of the z-scores values during learning per trial across sessions in MEC (green) and CA1 (brown). Shades represent SEM. Note scales are different for MEC and CA1. Regression test, trials 1-16: MEC: $r=0.91$, $P<0.0001$, CA1: $r=0.98$, $P < 0.00001$; trials 15-30: MEC: $r=0.93$, $P<0.00001$, CA1: $r=0.19$, $P=0.46$. **(J)** Same as in **(I)** but restricted around goal locations. Regression test, trials 1-16: MEC: $r=0.79$, $P=0.002$, CA1: $r=0.95$, $P<0.00001$; trials 15-30: MEC: $r=0.97$, $P<0.00001$, CA1: $r=0.09$, $P=0.76$. **(K)** Same data as in (I) and **(J)** but scaled to allow different comparison between CA1 and MEC. Note that: though the passage from positive to negative flickering may be informative, we believe one should focus on the latency to reach a plateau.

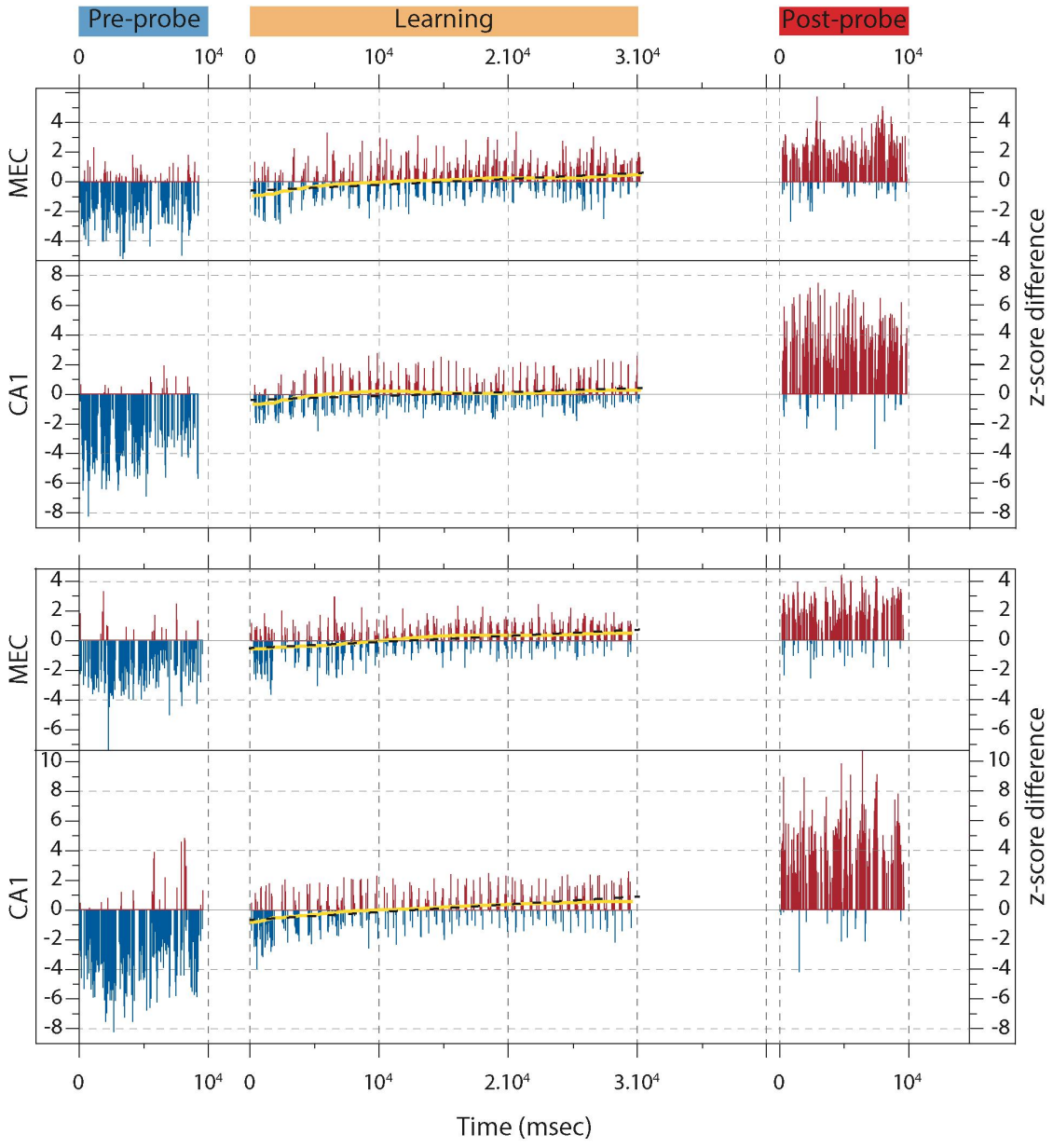

Fig. S4.13. Additional examples of z-scores regression slopes. Examples of progression of z-score difference of MEC (up) and CA1 (down) populations activity similarity to pre- and post-probe representations during two example days spanning (from left to right): pre-probe, learning and post-probe. Z-score difference more similar to pre-probe are represented by blue bars and to post-probe by red bars. Yellow line: smoothed score progression; black dashed line: regression analysis.

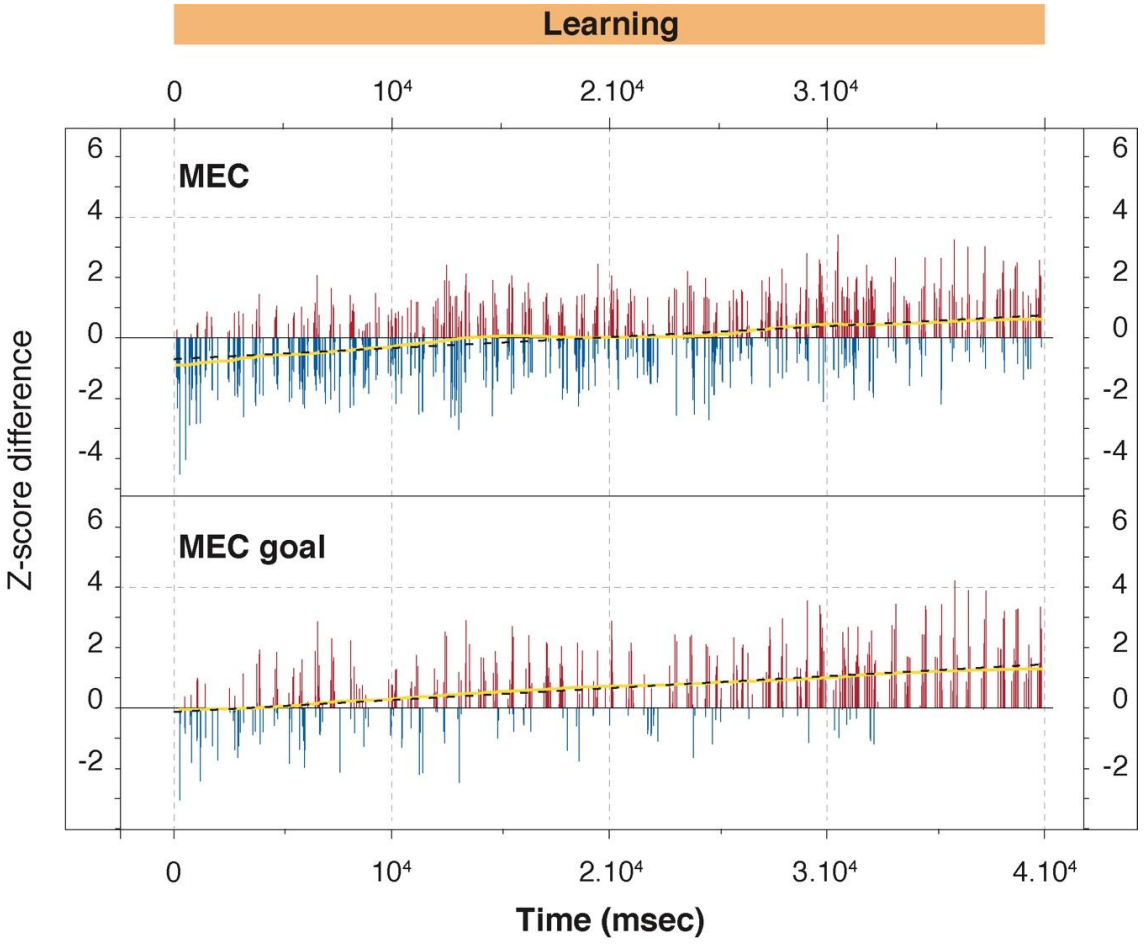

Fig. S4.14. Example of z-scores regression slopes at goal location during learning. Example from Fig. 4.4. Top: analyses performed on whole cheeseboard. Bottom: analyses restricted to goal locations only

5

Conclusion

In this thesis, I have presented 3 projects to which I have substantially contributed. These projects describe different aspects of spatial coding; they involve the statistical analysis and modeling of single cell activity in different brain areas of freely behaving rats. Here, I will briefly summarize the findings in order to discuss the implications they might have on the existing theories of memory consolidation and the potential underlying mechanisms. Finally, I will conclude by describing future quests and research directions.

In chapter 1 I have introduced concepts related to memory, described the hippocampal role in the processing of memories and concurrent coding of spatiotemporal coordinates, and summarized several memory theories. These theories, although substantially different, have similar fundamental building blocks: hippocampal-dependent quick acquisition of novel memories, hippocampal-neocortical interactions, and extra-hippocampal long-term memory assimilation (or transformation) and storage.

In chapter 2, I have described a mechanism that allows the hippocampal area CA1 to precisely encode novel cognitive maps in the presence of input noise. My collaborators and I developed a method to detect CA1 cell-cell interactions while rats freely explored familiar and novel environments. Through data analysis and theoretical modeling of population spatial representations, one principle emerged: a nonlinear dependence of the pairwise interaction probability with place field similarity. This relation improved spatial coding and decodability, especially during the exploration of a novel environment. In the following section, I will argue that the same mechanism could be employed to quickly and efficiently encode other types of novel information.

In chapter 3, I presented a project where my collaborators and I described a mechanism for hippocampal and the medial prefrontal cortex (mPFC) cell assemblies interactions. The underlying mechanism involved a transient increase in the phase locking of mPFC cells to hippocampal theta, selectively during synchronization periods. This gave rise to functional connections among cell pairs across regions: these connectivity patterns were similar to those observed in a model of information transfer and generalization. In the following section, I will argue that this mechanism might be used by the hippocampus to cue neocortical memories or allow for the transfer of information.

In chapter 4, I showed results from the analyses of hippocampal CA1 and MEC cells activity during and after the learning of behaviorally-relevant locations. Among the other findings, my collaborators and I observed that, similarly to CA1, albeit at different speeds, MEC spatially tuned cells increased their spiking activity at goal locations. Moreover, firing reorganization at

goal locations was maintained overnight in the MEC, but not in CA1. In the following section, I will argue that the MEC might have a larger-than-thought role in memory processing during complex behaviors.

5.1 Impact of results on existing theories of memory

During the learning of salient reward locations, we observed that both CA1 hippocampal and medial entorhinal cells increase their firing around novel goal locations (Fig. 4.1, S4.5). Another research group found a similar effect in entorhinal cells (Butler et al., 2019), and they suggested that increasing firing at goals is used to improve spatial precision around salient locations. This could reconcile the findings with a purely spatial view of the MTL (McNaughton et al., 2006), although not with a rigid metric view of grid cells (Moser and Moser, 2008). Nonetheless, the amount of firing accumulation around goals is predictive of future memory retention (Fig. 4.3), suggesting that spatial precision is only a co-objective rather than the main objective of MEC; this is in line with reports of field-by-field variability in grid cells (Stemmler and Herz, 2017). These results suggest that not only the hippocampus but also the MEC play a broader role in mnemonic processing by encoding both spatial and non-spatial information.

The dynamics of CA1 hippocampal learning-related activity argue against a recall-centered view of the hippocampus (Miller and Matzel, 2000): it is not clear why CA1 would rapidly (Fig. S4.5) and transiently (Fig. S4.6) accumulate firing around goals if its role was limited to memory recall. These findings seem to also contradict the multiple trace theory (Rosenbaum et al., 2001; Nadel and Moscovitch, 1997), which states that the hippocampus maintains its own long-term accounts of each episodic memory. Nonetheless, it is possible that it is not the hippocampus, but other MTL areas that maintain their own account of the episodic memory: the overnight maintenance of goal-firing in MEC cells argues for a longer-lasting trace in non-hippocampal MTL areas. This also corroborates the idea that the MEC, to a first approximation, acts as a memory buffer (Buzsáki, 1989), and would agree with studies showing that MTL lesions that extend beyond the hippocampus cause larger amnesia (Squire, 2004).

The reason why the hippocampus is so good at storing novel information has been studied from cellular and synaptic perspectives (Silva et al., 2009). Here, we studied the statistical relations between CA1 hippocampus pairs of neurons (Fig. 2.1). We found that an increase of positive interactions among similarly tuned cells, especially during the exploration of a novel environment, improves spatial information, decodability, and linear separability of population responses to different stimuli. There is no reason to believe that this mechanism is used only for encoding spatial information. It is possible that the hippocampus employs the same method to combat noise when encoding any sort of variables or relations among elements of a cognitive map (Buzsáki and Moser, 2013; Garvert et al., 2017); this could be a high-level mechanism that helps the hippocampus to quickly store novel information, aiding learning.

Considering that the MEC is the main gateway between the hippocampus and the neocortex, its slower emergence of goal-related firing suggests that information is not accumulated at the same time in all neocortical areas (Takashima et al., 2009). Nonetheless, the presence of previous knowledge, or a mental schema, has been reported to speed up neocortical learning (Tse et al., 2007). This is especially true for the task studied in this thesis since animals are heavily pre-trained and then repeat the task over several days (Dupret et al., 2010; Boccara et al., 2019; Kaefer et al., 2020). It is possible that a quick formation of a new memory in the

hippocampus happens concurrently in the relevant neocortical areas (Wang and Morris, 2010), e.g. the prefrontal cortex. We proposed here a mechanism that could be used to transfer spatial and/or contextual memories, together with a generalization method (Fig. S3.3)(Nardin et al., 2021b). It is possible that the same mechanism also acts as a cueing system, i.e., the hippocampus cues or tags the prefrontal cortex with spatiotemporal and contextual information for consolidation or later retrieval.

To conclude, the findings presented here shed light on aspects of various consolidation theories, clarifying details and providing clues as to how memories are stored throughout the brain.

5.2 Possible underlying mechanisms

Connections among neurons can be studied from a synaptic point of view (Perin et al., 2011) or, as discussed in chapters 2 and 4, as high-level statistical second-order descriptions (Averbeck et al., 2006; Cohen and Kohn, 2011). These statistical relations, across brain areas, allowed us to describe a novel mechanism that binds hippocampal and prefrontal assemblies during behavior (Fig. 3.2). This mechanism relies on a transient increase in theta locking of mPFC cells when co-active with CA1 functionally coupled counterparts (Fig. 3.3). It has been reported that mPFC cells are phase-locked to hippocampal theta oscillations (Jones and Wilson, 2005), although at a delay (Siapas et al., 2005). A mechanism based on theta oscillations would resonate with human studies, where it has been observed that recall rate of previously learned words increased as a function of prestimulus medial temporal theta amplitude (Guderian et al., 2009). Transient theta synchrony between the hippocampus and prefrontal cortex has been previously described as a potential mechanism to form prefrontal assemblies and tag memories for long-term storage (Benchenane et al., 2010). Here we further showed how this mechanism can bind pairs of CA1-mPFC cells, transiently synchronize them, and give rise to excess correlations that we picked up with statistical methods (Cohen and Kohn, 2011; Savin and Tkačik, 2017).

Within the hippocampus, we studied statistical pairwise interactions on a time scale of 25ms, which has been shown to organize the coactivity of hippocampal assemblies (Harris et al., 2003). This time window allows for a multitude of synaptic patterns within local circuits, including direct dendritic excitation, lateral inhibition, multisynaptic patterns, and global inhibitory signals. This approach disregards fine synaptic details but allows one to approach the hippocampus from an information-theory perspective, unlocking different methods employed in studies of optimal coding (Tkačik et al., 2010). We found that CA1 hippocampal cells increased positive interactions among similarly tuned cells during the exploration of a novel environment. A possible explanation for a novelty-induced increase in interactions is a reduction in overall inhibition (Fig. S2.3H) (Frank et al., 2004; Nitz and McNaughton, 2004; Wilson and McNaughton, 1993) and inhibitory synchrony (Fig. S2.3J), likely controlled by extra-hippocampal signals (Arriaga and Han, 2019). This novelty-induced effect could potentially allow stronger cell-cell excitatory interactions that receive similar inputs by decreasing lateral inhibition (Klausberger and Somogyi, 2008) or allowing polysynaptic activity of pyramidal neurons (Crepel et al., 1997) while promoting place field formation (Sheffield et al., 2017; Cohen et al., 2017). Moreover, it is known that the hippocampus contains a variety of inhibitory cells (Klausberger and Somogyi, 2008; Dudok et al., 2021) that respond differently to novelty (Arriaga and Han, 2019). A recent computational approach showed how the interplay between different types of inhibition, and their varying activity levels during familiarization, can lead to the emergence and stabilization of novel place fields (Pedrosa and Clopath, 2020). It is

possible that the same mechanism is responsible for the varying degrees of positive interactions in the hippocampus (Fig. 2.2).

Inhibitory cells also play a big role in the learning of novel goal locations: in fact, learning dynamically changes interneurons' activity during learning (43% decreasing vs 25% increasing) (Dupret et al., 2013). Moreover, changes in monosynaptic excitatory-inhibitory cells lead to the dynamic emergence of novel assemblies (Dupret et al., 2013). It has been suggested that these effects are due to dopaminergic terminals in the hippocampus and that these can improve learning and memory retention (McNamara et al., 2014; Kempadoo et al., 2016). It is not clear whether an external reward signal to the hippocampus (Schultz et al., 1997; Takeuchi et al., 2016) could cause such a localized increase in firing, although it has been suggested that dopamine could be a mechanism to induce novelty-dependent decreases in hippocampal inhibition(Li et al., 2003; Arriaga and Han, 2019). It is possible that exposure to a novel environment shares mechanisms that allow the hippocampus to quickly learn goal-related information.

5.3 Different modeling approaches: MaxEnt vs. Likelihood based

The study of the statistical properties of single-cell neural responses to external stimuli, generally called "receptive fields", has a long history (Sherrington, 1906; Hartline, 1938; Hubel and Wiesel, 1962). One class of models that is widely used to describe the receptive fields of many types of stimulus-responsive neurons is the generalized linear model (GLM) (Agresti, 2015). This type of model includes a linear regression part (that captures the linear dependence on external stimuli), a nonlinearity (that is thought of as a transfer function) and a noise class (generally assumed to be Binomial or Poisson) (Dayan and Abbott, 2005; Pillow, 2007). This class of models has been used to capture properties of both sensory and non-sensory areas, e.g. retinal neurons (Pillow et al., 2008) and entorhinal cells (Hardcastle et al., 2017). GLMs are very flexible, easily allow for the inclusion of many covariates (both external, e.g. visual stimuli, or internal, e.g. oscillations or spiking history), and can accomodate different types of regularization. Moreover, convergence to a single global maximum is guaranteed if one is to fit the model by maximizing the likelihood while using an exponential link function Agresti (2015); Pillow (2007).

On the other side, in the context of statistical hypothesis testing, one is interested in comparing data measurements against a null model that captures certain aspects of the data (Savin and Tkačik, 2017). Among the possible choices of mathematical models that can reproduce measured data moments (e.g., mean, variance, ...), one is interested in the one that does not incorporate additional structure or assumptions. This is formalized by requesting the model to have maximal entropy (MaxEnt) while reproducing certain single-cell or population moments (Jaynes, 1982). This guarantees that statistical hypothesis testing is not biased by (potentially unreasonable) assumptions hard-wired into the null model. This is the case of the study presented in chapter 2, where we analyzed hippocampal neurons during free exploration. There, a null model that reproduced exactly the firing of each cell given 2D position and global synchrony was ideal to rigorously measure the extent of within-area interactions.Although close to a MaxEnt in spirit, this model is inferred by maximizing the marginal likelihood of the data given parameters and is technically based on the assumptions of neural Poisson noise, GP priors, and exponential link function, making the claim of maximum entropy only partially correct. Nonetheless, our procedure to propagate uncertainty in the estimation of parameters

throughout the analyses allows for a viable way to increase entropy while keeping the chosen moments fixed.

Although theoretically elegant, it is not always possible to use maximum entropy models. For example, in chapters 3 and 4, GLM models were used to regress neural firing against many covariates (i.e. 2D position, head direction, speed, theta oscillations, firing history, and cofiring of other neurons). This would have been very hard to do in a MaxEnt setting without adding additional assumptions. Although a maximum-likelihood GLM model does not guarantee the perfect reproduction of statistical moments or maximal entropy, it was used to make sure that the results were not influenced by the lack of one or more covariates. This represents a compromise between rigor and viability, and allows to carry out statistical tests that require to take into account many different covariates.

5.4 Open questions and future directions

It is not clear how hippocampal interactions would adapt during learning. It is possible that efficient coding hypotheses could explain the change in interaction structure under the assumption that there is a need for a non-uniform representation of space. On the other hand, it is also possible that mnemonic requirements would lead to higher-dimensional representation, making spatial information alone not sufficient to explain remapping and changes in interaction structures. It would be interesting to study the evolution of pairwise interactions during/after learning and compare data results with the predictions of different efficient coding hypotheses. Furthermore, statistical interactions between CA1 and MEC cells could shed light on the initial steps of memory consolidation in MEC. More generally, the interplay between hippocampal and entorhinal cells might play a big role during sleep after learning. It is possible that the accumulation of firing around goals in MEC is driven by the hippocampus, potentially through transient synchronization periods driven by sharp wave ripples. The fact that MEC accumulates firing slower than the hippocampus does not exclude the possibility that PFC learns aspects of rewards at the same pace as the hippocampus. It would be interesting to study PFC neural responses and hippocampal-prefrontal interactions during goal learning and, more generally, if prefrontal - hippocampal interactions are important to establish novel spatial representations in PFC. One could think of studying them during initial exposures to novel environments, and also see what they would do during initial learning (without a schema) or during learning after being skilled (schema).

Bibliography

Abeles, M. (1982). *Local cortical circuits: Studies of brain function*. Springer, Berlin.

Abraham, W. C., Jones, O. D., and Glanzman, D. L. (2019). Is plasticity of synapses the mechanism of long-term memory storage? *NPJ science of learning*, 4(1):1–10.

Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.

Alonso, A., Meij, J. v. d., Tse, D., and Genzel, L. (2020). Naïve to expert: Considering the role of previous knowledge in memory. *Brain and Neuroscience Advances*, 4:2398212820948686.

Amit, D. J. (1989). Modeling brain function : The world of attractor neural networks.

Aronov, D., Nevers, R., and Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722.

Arriaga, M. and Han, E. B. (2019). Structured inhibitory activity dynamics in new virtual environments. *eLife*, 8:e47611.

Atick, J. J. and Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural computation*, 2(3):308–320.

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–366.

Barak, O., Rigotti, M., and Fusi, S. (2013). The Sparseness of Mixed Selectivity Neurons Controls the Generalization–Discrimination Trade-Off. *The Journal of Neuroscience*, 33(9):3844–3856.

Barker, G. R. I. and Warburton, E. C. (2008). NMDA Receptor Plasticity in the Perirhinal and Prefrontal Cortices Is Crucial for the Acquisition of Long-Term Object-in-Place Associative Memory. *Journal of Neuroscience*, 28(11):2837–2844.

Barlow, H. (2001). Redundancy reduction revisited. *Network: computation in neural systems*, 12(3):241.

Barlow, H. B. et al. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01).

Barry, C., Ginzberg, L. L., O'Keefe, J., and Burgess, N. (2012). Grid cell firing patterns signal environmental novelty by expansion. *Proceedings of the National Academy of Sciences*, 109(43):17687–17692.

Barry, C., Hayman, R., Burgess, N., and Jeffery, K. J. (2007). Experience-dependent rescaling of entorhinal grids. *Nature neuroscience*, 10(6):682–684.

Bartlett, F. C. and Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge university press.

Bassett, D. S. and Sporns, O. (2017). Network neuroscience. *Nature neuroscience*, 20(3):353–364.

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., and Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, 100(2):490—509.

Bellmund, J. L. S., Gardenfors, P., Moser, E. I., and Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415):eaat6766.

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y., Battaglia, F. P., and Wiener, S. I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron*, 66(6):921–936.

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., and Salzman, C. D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell*.

Bliss, T. V. and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, 232(2):331–356.

Boccara, C. N., Kjonigsen, L. J., Hammer, I. M., Bjaalie, J. G., Leergaard, T. B., and Witter, M. P. (2015). A three-plane architectonic atlas of the rat hippocampal region. *Hippocampus*, 25(7):838–857.

Boccara, C. N., Nardin, M., Stella, F., O'Neill, J., and Csicsvari, J. (2019). The entorhinal cognitive map is attracted to goals. *Science*, 363(6434):1443–1447.

Boccara, C. N., Sargolini, F., Thoresen, V. H., Solstad, T., Witter, M. P., Moser, E. I., and Moser, M. B. (2010). Grid cells in pre-and parasubiculum. *Nature neuroscience*, 13(8):987–994.

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*, 16(3):309–317.

Borghuis, B. G., Ratliff, C. P., Smith, R. G., Sterling, P., and Balasubramanian, V. (2008). Design of a neuronal array. *Journal of Neuroscience*, 28(12):3178–3189.

Bossomaier, T., Barnett, L., Harré, M., and Lizier, J. T. (2016). Transfer entropy. In *An introduction to transfer entropy*, pages 65–95. Springer.

Brinkman, B. A., Weber, A. I., Rieke, F., and Shea-Brown, E. (2016). How do efficient coding strategies depend on origins of noise in neural circuits? *PLoS computational biology*, 12(10):e1005150.

Bruce, D. (2001). Fifty years since lashley's in search of the engram: refutations and conjectures. *Journal of the History of the Neurosciences*, 10(3):308–318.

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.

Brun, V. H., Otnæss, M. K., Molden, S., Steffenach, H.-A., Witter, M. P., Moser, M. B., and Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science*, 296(5576):2243–2246.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198.

Bush, D., Barry, C., Manson, D., and Burgess, N. (2015). Using grid cells for navigation. *Neuron*, 87(3):507–520.

Butler, W. N., Hardcastle, K., and Giocomo, L. M. (2019). Remembered reward locations restructure entorhinal spatial maps. *Science (New York, N.Y.)*, 363(6434):1447—1452.

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, 33(3):325–340.

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362–385.

Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, 25(10):1073–1188.

Buzsaki, G., Bickford, R. G., Ponomareff, G., Thal, L. J., Mandel, R., and Gage, F. H. (1988). Nucleus basalis and thalamic control of neocortical activity in the freely moving rat. *Journal of neuroscience*, 8(11):4007–4026.

Buzsáki, G. and Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature neuroscience*, 16(2):130–138.

Buzsáki, G. (1989). Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience*, 31(3):551–570.

Cacucci, F., Wills, T. J., Lever, C., Giese, K. P., and O'Keefe, J. (2007). Experience-dependent increase in ca1 place cell spatial information, but not spatial reproducibility, is dependent on the autophosphorylation of the α-isoform of the calcium/calmodulin-dependent protein kinase ii. *Journal of Neuroscience*, 27(29):7854–7859.

Chalk, M., Marre, O., and Tkacik, G. (2016). Relevant sparse codes with variational information bottleneck. *Advances in Neural Information Processing Systems*, 29:1957–1965.

Chalk, M., Marre, O., and Tkačik, G. (2018). Toward a unified theory of efficient, predictive, and sparse coding. *Proceedings of the National Academy of Sciences*, 115(1):186–191.

Cohen, J. D., Bolstad, M., and Lee, A. K. (2017). Experience-dependent shaping of hippocampal ca1 intracellular activity in novel and familiar environments. *Elife*, 6:e23040.

Cohen, M. R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature neuroscience*, 14(7):811–819.

Colgin, L. L. (2013). Mechanisms and functions of theta rhythms. *Annual review of neuroscience*, 36:295–312.

Colgin, L. L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., Moser, M. B., and Moser, E. I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, 462(7271):353–357.

Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., and Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bulletin & Review*, 28(6):1796–1810.

Crepel, V., Khazipov, R., and Ben-Ari, Y. (1997). Blocking GABAA Inhibition Reveals AMPA- and NMDA-Receptor-Mediated Polysynaptic Responses in the CA1 Region of the Rat Hippocampus. *Journal of Neurophysiology*, 77(4):2071–2082.

Csicsvari, J., Hirase, H., Czurko, A., and Buzsáki, G. (1998). Reliability and state dependence of pyramidal cell–interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron*, 21(1):179–189.

Csicsvari, J., Hirase, H., Czurkó, A., Mamiya, A., and Buzsáki, G. (1999). Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *Journal of Neuroscience*, 19(1):274–287.

Csicsvari, J., Hirase, H., Mamiya, A., and Buzsáki, G. (2000). Ensemble patterns of hippocampal ca3-ca1 neurons during sharp wave–associated population events. *Neuron*, 28(2):585–594.

da Silveira, R. A. and Berry, M. J. (2014). High-fidelity coding with correlated neurons. *PLoS computational biology*, 10(11):e1003970.

Das, A. and Fiete, I. R. (2020). Systematic errors in connectivity inferred from activity in strongly recurrent networks. *Nature Neuroscience*, 23(10):1286–1296.

Dayan, P. and Abbott, L. F. (2005). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press.

de Bono, M. and Maricq, A. V. (2005). Neuronal substrates of complex behaviors in c. elegans. *Annu. Rev. Neurosci.*, 28:451–501.

Delatour, B. and Witter, M. (2002). Projections from the parahippocampal region to the prefrontal cortex in the rat: evidence of multiple pathways. *European Journal of Neuroscience*, 15(8):1400–1407.

Derdikman, D., Whitlock, J. R., Tsao, A., Fyhn, M., Hafting, T., Moser, M. B., and Moser, E. I. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience*, 12(10):1325–1332.

Diehl, G. W., Hon, O. J., Leutgeb, S., and Leutgeb, J. K. (2017). Grid and nongrid cells in medial entorhinal cortex represent spatial location and environmental features with complementary coding schemes. *Neuron*, 94(1):83–92.

Doi, E. and Lewicki, M. S. (2007). A theory of retinal population coding. *Advances in neural information processing systems*, 19:353.

Dragoi, G. and Buzsáki, G. (2006). Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron*, 50(1):145–157.

Dudai, Y. (2012). The Restless Engram: Consolidations Never End. *Annual Review of Neuroscience*, 35(1):227–247. Good overview of consolidation theories Consolidation vs reconsolidation Sleep.

Dudai, Y., Karni, A., and Born, J. (2015). The Consolidation and Transformation of Memory. *Neuron*, 88(1):20–32.

Dudai, Y., Moscovitch, M., Schacter, D. L., and Morris, R. G. (2007). Part 1. Memory. In *Science of Memory*, volume 1, pages 11–35. Oxford University Press, Oxford, England, 1 edition.

Dudchenko, P. A. (2010). *Why people get lost: the psychology and neuroscience of spatial cognition*. Oxford University Press, USA.

Dudok, B., Klein, P. M., Hwaun, E., Lee, B. R., Yao, Z., Fong, O., Bowler, J. C., Terada, S., Sparks, F. T., Szabo, G. G., Farrell, J. S., Berg, J., Daigle, T. L., Tasic, B., Dimidschstein, J., Fishell, G., Losonczy, A., Zeng, H., and Soltesz, I. (2021). Alternating sources of perisomatic inhibition during behavior. *Neuron*.

Dunn, B., Mørreaunet, M., and Roudi, Y. (2015). Correlations and functional connections in a population of grid cells. *PLoS Comput Biol*, 11(2):e1004052.

Dunn, B., Wennberg, D., Huang, Z., and Roudi, Y. (2017). Grid cells show field-to-field variability and this explains the aperiodic response of inhibitory interneurons. *bioRxiv*, page 101899.

Dupret, D., O'neill, J., Pleydell-Bouverie, B., and Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience*, 13(8):995–1002.

Dupret, D., O'Neill, J., and Csicsvari, J. (2013). Dynamic reconfiguration of hippocampal interneuron circuits during spatial learning. *Neuron*, 78(1):166–180.

Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1(1):41–50.

Eichenbaum, H. (2017a). Prefrontal–hippocampal interactions in episodic memory. *Nature Reviews Neuroscience*, 18(9):547—558.

Eichenbaum, H. (2017b). The role of the hippocampus in navigation is memory. *Journal of Neurophysiology*, 117(4):1785–1796.

Eichenbaum, H. and Cohen, N. (2014). Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? *Neuron*, 83(4):764—770.

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.

Eichenbaum, H., Stewart, C., and Morris, R. (1990). Hippocampal representation in place learning. *Journal of Neuroscience*, 10(11):3531–3542.

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, 425(6954):184–188.

El-Gaby, M., Reeve, H. M., Lopes-dos Santos, V., Campo-Urriza, N., Perestenko, P. V., Morley, A., Strickland, L. A., Lukács, I. P., Paulsen, O., and Dupret, D. (2021). An emergent neural coactivity code for dynamic memory. *Nature Neuroscience*, .(0):1–11.

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, 2(10):704–716.

Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D., and Peterson, K. J. (2011). The cambrian conundrum: early divergence and later ecological success in the early history of animals. *science*, 334(6059):1091–1097.

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.

Farovik, A., Dupont, L. M., Arce, M., and Eichenbaum, H. (2008). Medial Prefrontal Cortex Supports Recollection, But Not Familiarity, in the Rat. *Journal of Neuroscience*, 28(50):13428–13434.

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.

Flaxman, S., Wilson, A., Neill, D., Nickisch, H., and Smola, A. (2015). Fast kronecker inference in gaussian processes with non-gaussian likelihoods. In *International Conference on Machine Learning*, pages 607–616. PMLR.

Frank, L. M., Brown, E. N., and Wilson, M. A. (2001). A comparison of the firing properties of putative excitatory and inhibitory neurons from ca1 and the entorhinal cortex. *Journal of neurophysiology*, 86(4):2029–2040.

Frank, L. M., Stanley, G. B., and Brown, E. N. (2004). Hippocampal plasticity across multiple days of exposure to novel environments. *Journal of Neuroscience*, 24(35):7681–7689.

Frankland, P. W., Bontempi, B., Talton, L. E., Kaczmarek, L., and Silva, A. J. (2004). The Involvement of the Anterior Cingulate Cortex in Remote Contextual Fear Memory. *Science*, 304(5672):881–883.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual review of neuroscience*, 32:209–224.

Fries, P., Nikolić, D., and Singer, W. (2007). The gamma cycle. *Trends in neurosciences*, 30(7):309–316.

Frisch, K. v. et al. (1926). Neue versuche uber die bedeutung von duftorgan und pollenduft fur die verstandigung im bienenvolk. *Zeitschrift für Vergleichende Physiologie*, 4:1–21.

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.

Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature neuroscience*, 11(7):823–833.

Funahashi, S. and Kubota, K. (1994). Working memory and prefrontal cortex. *Neuroscience research*, 21(1):1–11.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., and Moser, M.-B. (2004). Spatial Representation in the Entorhinal Cortex. *Science*, 305(5688):1258–1264.

Gallistel, C. and Matzel, L. D. (2013). The Neuroscience of Learning: Beyond the Hebbian Synapse. *Annual Review of Psychology*, 64(1):169–200.

Gallos, L. K., Makse, H. A., and Sigman, M. (2012). A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proceedings of the National Academy of Sciences*, 109(8):2825–2830.

Garvert, M. M., Dolan, R. J., and Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *Elife*, 6:e17086.

Gauthier, J. L. and Tank, D. W. (2018). A Dedicated Population for Reward Coding in the Hippocampus. *Neuron*, 99(1):179—193.e7.

Gava, G. P., McHugh, S. B., Lefèvre, L., Lopes-dos Santos, V., Trouche, S., El-Gaby, M., Schultz, S. R., and Dupret, D. (2021). Integrating new memories into the hippocampal network activity space. *Nature Neuroscience*, 24(3):326–330.

Geiller, T., Vancura, B., Terada, S., Troullinou, E., Chavlis, S., Tsagkatakis, G., Tsakalides, P., Ócsai, K., Poirazi, P., Rózsa, B. J., et al. (2020). Large-scale 3d two-photon imaging of molecularly identified ca1 interneuron dynamics in behaving mice. *Neuron*, 108(5):968–983.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6:721–741.

Ginosar, G., Aljadeff, J., Burak, Y., Sompolinsky, H., Las, L., and Ulanovsky, N. (2021). Locally ordered representation of 3D space in the entorhinal cortex. *Nature*, 596(7872):404–409.

Girardeau, G. and Lopes-dos Santos, V. (2021). Brain neural patterns and the memory function of sleep. *Science (New York, N.Y.)*, 374(6567):560–564.

Giusti, C., Pastalkova, E., Curto, C., and Itskov, V. (2015). Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112(44):13455–13460.

Glanzman, D. L. (2010). Common Mechanisms of Synaptic Plasticity in Vertebrates and Invertebrates. *Current Biology*, 20(1):R31–R36.

Granot-Atedgi, E., Tkačik, G., Segev, R., and Schneidman, E. (2013). Stimulus-dependent maximum entropy models of neural population codes. *PLoS Comput Biol*, 9(3):e1002922.

Green, J., Vijayan, V., Mussells Pires, P., Adachi, A., and Maimon, G. (2019). A neural heading estimate is compared with an internal goal to guide oriented navigation. *Nature neuroscience*, 22(9):1460–1468.

Gridchyn, I., Schoenenberger, P., O'Neill, J., and Csicsvari, J. (2020). Optogenetic inhibition-mediated activity-dependent modification of ca1 pyramidal-interneuron connections during behavior. *Elife*, 9:e61106.

Guderian, S., Schott, B. H., Richardson-Klavehn, A., and Düzel, E. (2009). Medial temporal theta state before an event predicts episodic encoding success in humans. *Proceedings of the National Academy of Sciences*, 106(13):5365–5370.

Guise, K. G. and Shapiro, M. L. (2017). Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron*, 94(1):183–192.

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.

Hainmueller, T. and Bartos, M. (2018). Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature*, 558(7709):292–296.

Hardcastle, K., Ganguli, S., and Giocomo, L. M. (2015). Environmental Boundaries as an Error Correction Mechanism for Grid Cells. *Neuron*, 86(3):827—839.

Hardcastle, K., Maheswaranathan, N., Ganguli, S., and Giocomo, L. M. (2017). A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron*, 94(2):375–387.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. (2020). Array programming with numpy. *Nature*, 585(7825):357–362.

Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nature Reviews Neuroscience*, 6(5):399–407.

Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424(6948):552–556.

Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H., and Buzsaki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of neurophysiology*, 84(1):401–414.

Hartley, T., Lever, C., Burgess, N., and O'Keefe, J. (2013). Space in the brain: how the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120510.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology-Legacy Content*, 121(2):400–415.

Harvey, B. M., Klein, B. P., Petridou, N., and Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, 341(6150):1123–1126.

Hasselmo, M. E. (2011). *How we remember: Brain mechanisms of episodic memory*. MIT press.

Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior*, 4:60–78.

Henze, D. A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K. D., and Buzsaki, G. (2000). Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *Journal of neurophysiology*, 84(1):390–400.

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., and Tkačik, G. (2014). Variance predicts salience in central sensory processing. *Elife*, 3:e03722.

Hill, D. N., Mehta, S. B., and Kleinfeld, D. (2011). Quality metrics to accompany spike sorting of extracellular signals. *Journal of Neuroscience*, 31(24):8699–8705.

Hinman, J. R., Dannenberg, H., Alexander, A. S., and Hasselmo, M. E. (2018). Neural mechanisms of navigation involving interactions of cortical and subcortical structures. *Journal of Neurophysiology*, 119(6):2007–2029.

Hok, V., Chah, E., Save, E., and Poucet, B. (2013). Prefrontal Cortex Focally Modulates Hippocampal Place Cell Firing Patterns. *The Journal of Neuroscience*, 33(8):3443–3451.

Hollup, S. A., Molden, S., Donnett, J. G., Moser, M.-B., and Moser, E. I. (2001). Accumulation of Hippocampal Place Fields at the Goal Location in an Annular Watermaze Task. *The Journal of Neuroscience*, 21(5):1635–1644.

Hoover, W. B. and Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Structure and Function*, 212(2):149–179.

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.

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.

Humphries, M. D. (2017). Dynamical networks: finding, measuring, and tracking neural population activity using network science. *Network Neuroscience*, 1(4):324–338.

Huth, J. E. (2013). *The Lost Art of Finding Our Way:*. Harvard University Press.

Inostroza, M. and Born, J. (2013). Sleep for Preserving and Transforming Episodic Memory. *Annual Review of Neuroscience*, 36(1):79–102.

Ismakov, R., Barak, O., Jeffery, K., and Derdikman, D. (2017). Grid cells encode local positional information. *Current Biology*, 27(15):2337–2343.

Ito, H. T., Zhang, S.-J., Witter, M. P., Moser, E. I., and Moser, M. B. (2015). A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation. *Nature*, 522(7554):50–55.

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.

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952.

Jenkins, J. G. and Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology*, 35(4):605–612.

Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., and Moser, M. B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, 478(7368):246–249.

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

Jones, M. W. and Wilson, M. A. (2005). Theta rhythms coordinate hippocampal–prefrontal interactions in a spatial memory task. *PLoS biology*, 3(12):e402.

Josselyn, S. A., Köhler, S., and Frankland, P. W. (2015). Finding the engram. *Nature Reviews Neuroscience*, 16(9):521–534.

Jung, M. W., Qin, Y., McNaughton, B. L., and Barnes, C. A. (1998). Firing characteristics of deep layer neurons in prefrontal cortex in rats performing spatial working memory tasks. *Cerebral cortex (New York, NY: 1991)*, 8(5):437–450.

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. *Journal of Neuroscience*, 14(12):7347–7356.

Kaefer, K., Malagon-Vina, H., Dickerson, D. D., O'Neill, J., Trossbach, S. V., Korth, C., and Csicsvari, J. (2019). Disrupted-in-schizophrenia 1 overexpression disrupts hippocampal coding and oscillatory synchronization. *Hippocampus*, 29(9):802–816.

Kaefer, K., Nardin, M., Blahna, K., and Csicsvari, J. (2020). Replay of behavioral sequences in the medial prefrontal cortex during rule switching. *Neuron*, 106(1):154–165.

Kandel, E. R., editor (2013). *Principles of neural science*. McGraw-Hill, New York, 5th ed edition.

Kanitscheider, I., Coen-Cagli, R., and Pouget, A. (2015). Origin of information-limiting noise correlations. *Proceedings of the National Academy of Sciences*, 112(50):E6973–E6982.

Kanter, B. R., Lykken, C. M., Avesar, D., Weible, A., Dickinson, J., Dunn, B., Borgesius, N. Z., Roudi, Y., and Kentros, C. G. (2017). A Novel Mechanism for the Grid-to-Place Cell Transformation Revealed by Transgenic Depolarization of Medial Entorhinal Cortex Layer II. *Neuron*, 93(6):1480–1492.e6.

Karklin, Y. and Simoncelli, E. P. (2011). Efficient coding of natural images with a population of noisy linear-nonlinear neurons. *Advances in neural information processing systems*, 24:999.

Karlsson, M. P. and Frank, L. M. (2008). Network dynamics underlying the formation of sparse, informative representations in the hippocampus. *Journal of Neuroscience*, 28(52):14271–14281.

Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D., and Kandel, E. R. (2016). Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proceedings of the National Academy of Sciences*, 113(51):14835–14840.

Kim, S. S., Rouault, H., Druckmann, S., and Jayaraman, V. (2017). Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849–853.

Kitamura, T., Ogawa, S. K., Roy, D. S., Okuyama, T., Morrissey, M. D., Smith, L. M., Redondo, R. L., and Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a memory. *Science*, 356(6333):73–78.

Klausberger, T. and Somogyi, P. (2008). Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations. *Science*, 321(5885):53–57.

Klinzing, J. G., Niethard, N., and Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, 22(10):1—13.

Kraus, B., Robinson, R., White, J., Eichenbaum, H., and Hasselmo, M. (2013). Hippocampal "Time Cells": Time versus Path Integration. *Neuron*, 78(6):1090–1101.

Krupic, J., Bauza, M., Burton, S., and O'Keefe, J. (2018). Local transformations of the hippocampal cognitive map. *Science*, 359(6380):1143–1146.

Lashley, K. (1950). In search of the engram. society of exp. biology symp., no. 4 (1950).

Latora, V. and Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701.

Latuske, P., Kornienko, O., Kohler, L., and Allen, K. (2017). Hippocampal Remapping and Its Entorhinal Origin. *Frontiers in behavioral neuroscience*, 11:253.

Lee, J. S., Briguglio, J. J., Cohen, J. D., Romani, S., and Lee, A. K. (2020). The Statistical Structure of the Hippocampal Code for Space as a Function of Time, Context, and Value. *Cell*, 183(3):620–635.e22.

Leutgeb, J. K., Leutgeb, S., Moser, M. B., and Moser, E. I. (2007). Pattern separation in the dentate gyrus and ca3 of the hippocampus. *science*, 315(5814):961–966.

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B., and Moser, E. I. (2004). Distinct Ensemble Codes in Hippocampal Areas CA3 and CA1. *Science*, 305(5688):1295–1298.

Li, S., Cullen, W. K., Anwyl, R., and Rowan, M. J. (2003). Dopamine-dependent facilitation of ltp induction in hippocampal ca1 by exposure to spatial novelty. *Nature neuroscience*, 6(5):526–531.

Linsker, R. (1989). An application of the principle of maximum information preservation to linear systems. In *Advances in neural information processing systems*, pages 186–194.

Lipton, P. A., White, J. A., and Eichenbaum, H. (2007). Disambiguation of overlapping experiences by neurons in the medial entorhinal cortex. *Journal of Neuroscience*, 27(21):5787–5795.

Lisman, J., Cooper, K., Sehgal, M., and Silva, A. J. (2018). Memory formation depends on both synapse-specific modifications of synaptic strength and cell-specific increases in excitability. *Nature neuroscience*, 21(3):309–314.

Lizier, J. T. (2014). Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1:11.

Lopes-dos Santos, V., Ribeiro, S., and Tort, A. B. (2013). Detecting cell assemblies in large neuronal populations. *Journal of neuroscience methods*, 220(2):149–166.

Low, R. J., Gu, Y., and Tank, D. W. (2014). Cellular resolution optical access to brain regions in fissures: imaging medial prefrontal cortex and grid cells in entorhinal cortex. *Proceedings of the National Academy of Sciences*, 111(52):18739–18744.

Magee, J. C. and Grienberger, C. (2020). Synaptic plasticity forms and functions. *Annual review of neuroscience*, 43:95–117.

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):1–9.

Marozzi, E., Ginzberg, L. L., Alenda, A., and Jeffery, K. J. (2015). Purely translational realignment in grid cell firing patterns following nonmetric context change. *Cerebral Cortex*, 25(11):4619–4627.

Martin, S. J., Grimwood, P. D., and Morris, R. G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual review of neuroscience*, 23(1):649–711.

Mathis, A., Herz, A. V., and Stemmler, M. B. (2013). Multiscale codes in the nervous system: The problem of noise correlations and the ambiguity of periodic scales. *Physical Review E*, 88(2):022713.

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.

McKenzie, S., Huszár, R., English, D. F., Kim, K., Christensen, F., Yoon, E., and Buzsáki, G. (2021). Preexisting hippocampal network dynamics constrain optogenetically induced place fields. *Neuron*, 109:1–15.

McNamara, C. G., Tejero-Cantero, , Trouche, S., Campo-Urriza, N., and Dupret, D. (2014). Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nature Neuroscience*, 17(12):1658–1660.

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. *Experimental brain research*, 52(1):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. and Morris, R. G. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in neurosciences*, 10(10):408–415.

Menzel, R. and Erber, J. (1978). Learning and memory in bees. *Scientific American*, 239(1):102–111.

Meshulam, L., Gauthier, J. L., Brody, C. D., Tank, D. W., and Bialek, W. (2017). Collective behavior of place and non-place neurons in the hippocampal network. *Neuron*, 96(5):1178–1191.

Miller, E. K. (2000). The prefontral cortex and cognitive control. *Nature reviews neuroscience*, 1(1):59–65.

Miller, R. R. (2021). Failures of Memory and the Fate of Forgotten Memories. *Neurobiology of Learning and Memory*, 181:107426.

Miller, R. R. and Matzel, L. D. (2000). Memory involves far more than 'consolidation'. *Nature reviews. Neuroscience*, 1(3):214–6.

Mizuseki, K. and Buzsaki, G. (2014). Theta oscillations decrease spike synchrony in the hippocampus and entorhinal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120530.

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, E. I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, 31(1):69—89.

Moser, E. I. and Moser, M. B. (2008). A metric for space. *Hippocampus*, 18(12):1142–1156.

Moser, E. I., Moser, M. B., and McNaughton, B. L. (2017). Spatial representation in the hippocampal formation: a history. *Nature neuroscience*, 20(11):1448–1464.

Moser, E. I. and Paulsen, O. (2001). New excitement in cognitive space: between place cells and spatial memory. *Current opinion in neurobiology*, 11(6):745–751.

Müller, M. and Wehner, R. (1988). Path integration in desert ants, cataglyphis fortis. *Proceedings of the National Academy of Sciences*, 85(14):5287–5290.

Muller, R. (1996). A Quarter of a Century of Place Cells. *Neuron*, 17(5):813–822.

Nadel, L. (1991). The hippocampus and space revisited. *Hippocampus*, 1(3):221–229.

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

Nagele, J., Herz, A. V., and Stemmler, M. B. (2020). Untethered firing fields and intermittent silences: Why grid-cell discharge is so variable. *Hippocampus*.

Nardin, M., Csicsvari, J., Tkacik, G., and Savin, C. (2021a). The structure of hippocampal ca1 interactions optimizes spatial coding across experience. *bioRxiv*.

Nardin, M., Kaefer, K., and Csicsvari, J. (2021b). The generalized spatial representation in the prefrontal cortex is inherited from the hippocampus. *bioRxiv*, page 2021.09.30.462269.

Navawongse, R. and Eichenbaum, H. (2013). Distinct pathways for rule-based retrieval and spatial mapping of memory representations in hippocampal neurons. *Journal of Neuroscience*, 33(3):1002–1013.

Navratilova, Z. and Mcnaughton, B. L. (2014). Models of path integration in the hippocampal complex. In *Space, time and memory in the hippocampal formation*, pages 191–224. Springer.

Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104.

Newsome, W. T., Britten, K. H., and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341(6237):52–54.

Nitz, D. and McNaughton, B. (2004). Differential modulation of ca1 and dentate gyrus interneurons during exploration of novel environments. *Journal of neurophysiology*, 91(2):863–872.

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(2):171–5.

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

O'Keefe, J. and Recce, M. L. (1993). Phase relationship between hippocampal place units and the eeg theta rhythm. *Hippocampus*, 3(3):317–330.

Ólafsdóttir, H. F., Carpenter, F., and Barry, C. (2016). Coordinated grid and place cell replay during rest. *Nature neuroscience*, 19(6):792–794.

O'Neill, J., Senior, T., and Csicsvari, J. (2006). Place-selective firing of ca1 pyramidal cells during sharp wave/ripple network patterns in exploratory behavior. *Neuron*, 49(1):143–155.

O'Neill, J., Senior, T. J., Allen, K., Huxter, J. R., and Csicsvari, J. (2008). Reactivation of experience-dependent cell assembly patterns in the hippocampus. *Nature neuroscience*, 11(2):209–215.

O'Neill, J., Boccara, C. N., Stella, F., Schönenberger, P., and Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science*, 355(6321):184–188.

Panzeri, S., Schultz, S. R., Treves, A., and Rolls, E. T. (1999). Correlations and the encoding of information in the nervous system. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1423):1001–1012.

Patai, E. Z. and Spiers, H. J. (2021). The versatile wayfinder: Prefrontal contributions to spatial navigation. *Trends in cognitive sciences*.

Pedrosa, V. and Clopath, C. (2020). The interplay between somatic and dendritic inhibition promotes the emergence and stabilization of place fields. *PLoS computational biology*, 16(7):e1007955.

Perin, R., Berger, T. K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, 108(13):5419–5424.

Pillow, J. (2007). *Likelihood-based approaches to modeling the neural code*, volume 70. MIT press Cambridge, Massachusets.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999.

Posani, L., Cocco, S., Ježek, K., and Monasson, R. (2017). Functional connectivity models for decoding of spatial representations from hippocampal ca1 recordings. *Journal of Computational Neuroscience*, 43(1):17–33.

Pothen, A., Simon, H. D., and Liou, K.-P. (1990). Partitioning sparse matrices with eigenvectors of graphs. *SIAM journal on matrix analysis and applications*, 11(3):430–452.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). Numerical recipes in c++. *The art of scientific computing*, 2:1002.

Puchalla, J. L., Schneidman, E., Harris, R. A., and Berry, M. J. (2005). Redundancy in the population code of the retina. *Neuron*, 46(3):493–504.

Rasch, B. and Born, J. (2013). About Sleep's Role in Memory. *Physiological Reviews*, 93(2):681–766.

Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer.

Rennó-Costa, C. and Tort, A. B. (2017). Place and grid cells in a loop: implications for memory function and spatial coding. *Journal of Neuroscience*, 37(34):8062–8076.

Reversat, A., Gaertner, F., Merrin, J., Stopp, J., Tasciyan, S., Aguilera, J., De Vries, I., Hauschild, R., Hons, M., Piel, M., et al. (2020). Cellular locomotion using environmental topography. *Nature*, 582(7813):582–585.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590.

Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., et al. (2013). Astropy: A community python package for astronomy. *Astronomy & Astrophysics*, 558:A33.

Roediger, H. L., Dudai, Y., and Fitzpatrick, S. M., editors (2007). *Science of Memory: Concepts*. Oxford University Press, New York.

Rolls, E. T. (1999). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, 9(4):467–480.

Rolls, E. T., Stringer, S. M., and Elliot, T. (2006). Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning. *Network: Computation in Neural Systems*, 17(4):447–465.

Rosenbaum, R. S., Winocur, G., and Moscovitch, M. (2001). New views on old memories: re-evaluating the role of the hippocampal complex. *Behavioural brain research*, 127(1-2):183–197.

Roy, D. S., Park, Y.-G., Kim, M. E., Zhang, Y., Ogawa, S. K., DiNapoli, N., Gu, X., Cho, J. H., Choi, H., Kamentsky, L., et al. (2022). Brain-wide mapping reveals that engrams for a single memory are distributed across multiple brain regions. *Nature communications*, 13(1):1–16.

Rueckemann, J. W., DiMauro, A. J., Rangel, L. M., Han, X., Boyden, E. S., and Eichenbaum, H. (2016). Transient optogenetic inactivation of the medial entorhinal cortex biases the active population of hippocampal neurons. *Hippocampus*, 26(2):246–260.

Rumyantsev, O. I., Lecoq, J. A., Hernandez, O., Zhang, Y., Savall, J., Chrapkiewicz, R., Li, J., Zeng, H., Ganguli, S., and Schnitzer, M. J. (2020). Fundamental bounds on the fidelity of sensory cortical coding. *Nature*, 580(7801):100–105.

Ryan, T. J., Roy, D. S., Pignatelli, M., Arons, A., and Tonegawa, S. (2015). Engram cells retain memory under retrograde amnesia. *Science*, 348(6238):1007–1013.

Sakurai, Y. (2002). Coding of auditory temporal and pitch information by hippocampal individual cells and cell assemblies in the rat. *Neuroscience*, 115(4):1153–1163.

Sara, S. J. (2000). Retrieval and Reconsolidation: Toward a Neurobiology of Remembering. *Learning & Memory*, 7(2):73–84.

Sarel, A., Finkelstein, A., Las, L., and Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science (New York, N.Y.)*, 355(6321):176—180.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B., and Moser, E. I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762.

Sauer, J.-F., Folschweiller, S., and Bartos, M. (2022). Topographically organized representation of space and context in the medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, 119(6):e2117300119.

Savin, C. and Tkacik, G. (2016). Estimating nonlinear neural response functions using gp priors and kronecker methods. In *Advances in Neural Information Processing Systems*, pages 3603–3611.

Savin, C. and Tkačik, G. (2017). Maximum entropy models as a tool for building precise neural controls. *Current opinion in neurobiology*, 46:120–126.

Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., and Redish, A. D. (2005). Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience*, 131(1):1–11.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306):1593–1599.

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

Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, volume 57, page 61. Austin, TX.

Seelig, J. D. and Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551):186–191.

Sheffield, M. E., Adoff, M. D., and Dombeck, D. A. (2017). Increased prevalence of calcium transients across the dendritic arbor during place field formation. *Neuron*, 96(2):490–504.

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

Shin, J. D., Tang, W., and Jadhav, S. P. (2019). Dynamics of awake hippocampal-prefrontal replay for spatial learning and memory-guided decision making. *Neuron*, 104(6):1110–1125.

Siapas, A. G., Lubenov, E. V., and Wilson, M. A. (2005). Prefrontal phase locking to hippocampal theta oscillations. *Neuron*, 46(1):141–151.

Silva, A. J., Zhou, Y., Rogerson, T., Shobe, J., and Balaji, J. (2009). Molecular and cellular approaches to memory allocation in neural circuits. *Science*, 326(5951):391–395.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216.

Simons, J. S. and Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, 4(8):637–648.

Skaggs, W. E., McNaughton, B. L., and Gothard, K. M. (1993). An information-theoretic approach to deciphering the hippocampal code. In *Advances in neural information processing systems*, pages 1030–1037.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. Technical report, Colorado Univ at Boulder Dept of Computer Science.

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

Souza, B. C., Pavão, R., Belchior, H., and Tort, A. B. (2018). On information metrics for spatial coding. *Neuroscience*, 375:62–73.

Spellman, T., Rigotti, M., Ahmari, S. E., Fusi, S., Gogos, J. A., and Gordon, J. A. (2015). Hippocampal–prefrontal input supports spatial encoding in working memory. *Nature*, 522(7556):309–314.

Spiers, H. J. and Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage*, 31(4):1826–1840.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3):171–177.

Squire, L. R. and Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology*, 5(2):169–177.

Squire, L. R., Genzel, L., Wixted, J. T., and Morris, R. G. (2015). Memory Consolidation. *Cold Spring Harbor Perspectives in Biology*, 7(8):a021766.

Stefanini, F., Kushnir, L., Jimenez, J. C., Jennings, J. H., Woods, N. I., Stuber, G. D., Kheirbek, M. A., Hen, R., and Fusi, S. (2020). A distributed neural code in the dentate gyrus and in ca1. *Neuron*, 107:703–716.

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

Stemmler, M. and Herz, A. V. (2017). Spatial Cognition: Grid Cells Harbour Three Complementary Positional Codes. *Current Biology*, 27(15):R755–R758.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M. B., and Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, 492(7427):72—8.

Stensola, T., Stensola, H., Moser, M. B., and Moser, E. I. (2015). Shearing-induced asymmetry in entorhinal grid cells. *Nature*, 518(7538):207–212.

Suryanarayana, S. M., Pérez-Fernández, J., Robertson, B., and Grillner, S. (2020). The evolutionary origin of visual and somatosensory representation in the vertebrate pallium. *Nature Ecology & Evolution*, 4(4):639–651.

Takashima, A., Nieuwenhuis, I. L., Jensen, O., Talamini, L. M., Rijpkema, M., and Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *Journal of Neuroscience*, 29(32):10087–10093.

Takehara-Nishiuchi, K. and McNaughton, B. L. (2008). Spontaneous Changes of Neocortical Code for Associative Memory During Consolidation. *Science*, 322(5903):960–963.

Takeuchi, T., Duszkiewicz, A. J., Sonneborn, A., Spooner, P. A., Yamasaki, M., Watanabe, M., Smith, C. C., Fernández, G., Deisseroth, K., Greene, R. W., et al. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature*, 537(7620):357–362.

Tang, W. and Jadhav, S. P. (2019). Sharp-wave ripples as a signature of hippocampal-prefrontal reactivation for memory during sleep and waking states. *Neurobiology of Learning and Memory*, 160(Current Neurology and Neuroscience Reports 14 2014):11—20.

Taube, J. S., Muller, R. U., and Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, 10(2):420–435.

Tkačik, G., Marre, O., Amodei, D., Schneidman, E., Bialek, W., and Berry, M. J. (2014). Searching for collective behavior in a large network of sensory neurons. *PLoS computational biology*, 10(1):e1003408.

Tkačik, G., Prentice, J. S., Balasubramanian, V., and Schneidman, E. (2010). Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences*, 107(32):14419–14424.

Tocker, G., Barak, O., and Derdikman, D. (2015). Grid cells correlation structure suggests organized feedforward projections into superficial layers of the medial entorhinal cortex. *Hippocampus*, 25(12):1599–1613.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, 55(4):189.

Tonegawa, S., Morrissey, M. D., and Kitamura, T. (2018). The role of engram cells in the systems consolidation of memory. *Nature Reviews Neuroscience*, 19(8):485–498.

Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global.

Treves, A. and Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal ca3 network. *Hippocampus*, 2(2):189–199.

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(5821):76–82.

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, 53(1):1–25.

Van Hateren, J. H. (1992). A theory of maximizing sensory information. *Biological cybernetics*, 68(1):23–29.

Van Kesteren, M. T., Ruiter, D. J., Fernández, G., and Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in neurosciences*, 35(4):211–219.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python tutorial*, volume 620. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Vazdarjanova, A. and Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of ca3 and ca1 ensembles. *Journal of Neuroscience*, 24(29):6489–6496.

Vertes, R. P. (2006). Interactions among the medial prefrontal cortex, hippocampus and midline thalamus in emotional and cognitive processing in the rat. *Neuroscience*, 142(1):1–20.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Wadhwa, N. and Berg, H. C. (2022). Bacterial motility: machinery and mechanisms. *Nature Reviews Microbiology*, 20(3):161–173.

Wang, S.-H. and Morris, R. G. (2010). Hippocampal-Neocortical Interactions in Memory Formation, Consolidation, and Reconsolidation. *Annual Review of Psychology*, 61(1):49–79.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

Weber, F., Machens, C. K., and Borst, A. (2012). Disentangling the functional consequences of the connectivity between optic-flow processing neurons. *Nature neuroscience*, 15(3):441–448.

Wehner, R., Michel, B., and Antonsen, P. (1996). Visual navigation in insects: coupling of egocentric and geocentric information. *The Journal of experimental biology*, 199(1):129–140.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73.

Whitlock, J. R., Heynen, A. J., Shuler, M. G., and Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *science*, 313(5790):1093–1097.

Whittington, J. C., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T. E. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, 183:1–15.

Wibral, M., Vicente, R., and Lindner, M. (2014). Transfer entropy in neuroscience. In *Directed information measures in neuroscience*, pages 3–36. Springer.

Wilson, M. and McNaughton, B. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124):1055–1058.

Winocur, G. and Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, 17(5):766–780.

Winocur, G., Moscovitch, M., and Sekeres, M. (2007). Memory consolidation or transformation: context manipulation and hippocampal representations of memory. *Nature neuroscience*, 10(5):555–557.

Witter, M. (2011). Entorhinal cortex. *Scholarpedia*, 6(10):4380.

Wolbers, T. and Hegarty, M. (2010). What determines our navigational abilities? *Trends in cognitive sciences*, 14(3):138–146.

Wood, E. R., Dudchenko, P. A., Robitsek, R., and Eichenbaum, H. (2000). Hippocampal Neurons Encode Information about Different Types of Memory Episodes Occurring in the Same Location. *Neuron*, 27(3):623—633. Left versus right trajeltones.

Xu, W. and Südhof, T. C. (2013). A neural circuit for memory specificity and generalization. *Science*, 339(6125):1290–1295.

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.

Zielinski, M. C., Shin, J. D., and Jadhav, S. P. (2019). Coherent coding of spatial position mediated by theta oscillations in the hippocampus and prefrontal cortex. *Journal of Neuroscience*, 39(23):4550–4565.

Zola-Morgan, S., Cohen, N. J., and Squire, L. R. (1983). Recall of remote episodic memory in amnesia. *Neuropsychologia*, 21(5):487–500.