

Chapter 16: Computational Models of Hippocampal Cognitive Function

Daniel Bush, MEng., DPhil.

Neil Burgess, BSc., PhD.

UCL Institute of Cognitive Neuroscience & Queen Square Institute of Neurology

University College London

17 Queen Square, London WC1N 3AZ

n.burgess@ucl.ac.uk

drdanielbush@gmail.com

Abstract / Overview

Some of the most striking data relating cognitive behaviour to neuronal firing, damage or metabolic activity in the brain concerns the hippocampus. Computational models have been invaluable in exploring the link between neurons and behaviour, enabling hypothetical mechanisms to be defined precisely and examined quantitatively. Here we review many of these, including models of spatial and mnemonic function, models that stress feed-forward processing through the hippocampal system and those stressing recurrent processing within it. We review the spatial models first, as these are most firmly rooted in the electrophysiology data from this region. Next, we describe models of mnemonic function, specifically associative or episodic memory, following Marr's seminal theory in 1971 and subsequent developments. Finally, we review recent proposals to unify the spatial and mnemonic functions of the hippocampus based, to varying degrees, on two related considerations: first, the correspondence between the place or index-like representations in spatial or mnemonic models and low dimensional latent variables that efficiently describe, generate or predict sensory states; and second, the benefit for prediction and rapid generalisation of finding representations that capture the common structure among transitions or relationships between states across multiple tasks or environments. These models can be seen as computational instantiations of the original idea of a cognitive map that describes states or concepts and the relationships between them.

16.1 Introduction

There have been many attempts to understand and quantify the contribution of the hippocampus to cognition. In this chapter we focus on models of the link between the cognitive ability of the animal and the action of individual cells and synapses. As reviewed in Chapter 14, lesion studies in a variety of mammals (including humans) have implicated the hippocampus in spatial navigation, while human neuropsychology has most notably implicated this region in episodic or declarative memory function (see Chapter 13). In addition to these data a vast body of knowledge has been collected regarding neural representations of the spatial location and orientation of freely moving rodents (see Chapters 11 and 12). Hypotheses regarding the function of the hippocampus have traditionally been expressed in words. However, it is often possible to interpret verbal descriptions in more than one way, or to retrospectively change their interpretation to suit the facts. In addition, it can be difficult to tell whether the proposed explanation would actually work as described, and if so, difficult to make unambiguous quantitative predictions that can be used to test it. These problems become more acute where hypotheses address the question of how a putative function arises from the co-operative behaviour of large numbers of neurons and synapses. One way around this is to express such a hypothesis in terms of equations or computer simulations, referred to as a computational model. An advantage of this approach is that all of the parameter values and assumptions necessary to generate the behaviour concerned are made explicit; another is that the operation of the model is unambiguously specified. These advantages mean that computational modelling has an important role to play in the progress of scientific understanding, most importantly in its interaction with experimental investigation: by predicting critical experiments, undergoing revision to reflect their results, and then predicting further experiments. They are not a panacea for all ills, and the interpretation of a model, the way it works, and the values of its parameters can still be changed or disputed. At the most basic level a computational model can serve as an existence proof of the behaviour that could result from a proposed mechanism, but more generally it can serve to properly define a theoretical understanding and provide a powerful framework in which the nature of a theory relating brain to behaviour can be understood.

Computational modelling of the hippocampus initially followed two largely independent streams, one seeking to explain a general role in associative memory and the other focusing on its role in spatial cognition. More recently, however, broader theoretical frameworks have been proposed that attempt to account for both types of data. Models will usually start from as detailed a biophysical level as is useful for the level of the hypothesis they seek to investigate. While many models reviewed here involve detailed simulation of cellular and synaptic electrophysiology, the aim of this chapter is to explain the neural bases of spatial and mnemonic behaviour – something that is made easier by focusing on the simplest level of description capturing the likely functional consequences of cellular and synaptic events: e.g. whether or not an action potential was fired. Thus, the activity (firing rate) of a neuron is often simply viewed as a monotonic function of the amount by which the net input to it exceeds some threshold value. The net input to a neuron is the sum of the activity of each neuron connected to it weighted by the strength of the connection (occasionally inhibitory inputs are modelled as a divisive term in the net input rather than a subtractive term). ‘Learning’ corresponds to modification of the connection strengths. Most commonly learning is of a ‘Hebbian’ nature (Hebb, 1949) such that simultaneous pre- and post- synaptic activity leads to increased connection strength, and is often used in explicit analogy to synaptic processes like long-term

potentiation (LTP, see Chapter 10; Box 16.1). Other concepts are explained as and where necessary. Readers interested in neural computation more generally should see e.g. (Rumelhart and McClelland, 1986; McClelland and Rumelhart, 1986; Hertz et al., 1990; Anderson, 1995; Gurney, 1997; Dayan and Abbott, 2002; Trappenberg, 2002).

Since the anatomy of the hippocampal formation is similar in rodents, bats, primates and humans it seems sensible to start with neural models of the vast and well-established body of data regarding place and grid cells and the neural representation of space, drawing mostly on experimental data collected in rats. We next consider models that make use of these spatial representations to guide behaviour. Together, models of the representation of location and orientation from sensory input and their service in spatial navigation provide one of the best quantitative accounts of the link between perception, cognition and action and between cells, systems and behaviour. The next part of the chapter concerns attempts to model the more general role of the human hippocampus in memory for personal experience. As we shall see, the role of the recurrent collaterals in area CA3 of the hippocampus maintains a common point of contact between these models: in both the spatial and episodic memory frameworks they are assumed to perform an associative memory function. The chapter concludes with a discussion of theoretical attempts to reconcile these two streams of research (spatial and mnemonic). In particular, this unifying framework proposes that the hippocampus encodes efficient, low-dimensional representations of variables that are useful for planning and prediction, alongside representations that capture common structure in transitions or relationships across multiple task domains to support generalisation. This framework goes some way to explaining experimental data across species in both the spatial and mnemonic domains, and can be seen as implementing a cognitive map (Tolman, 1948; O'Keefe and Nadel, 1978; Eichenbaum and Cohen, 2014; Behrens et al., 2018).

16.2 The Hippocampus and Spatial Representation

This section addresses the representation of spatial location and orientation at the level of single neurons in the hippocampal formation. These models take two structural forms: those relying predominantly on feed-forward connections to capture the data, and those relying predominantly on recurrent connections. Models of the representation of location embodied by the firing of hippocampal place cells are considered first. Next, we consider the complementary representation provided by the firing of grid cells in medial entorhinal cortex (MEC). Importantly, it is not only the firing rates of place and grid cells that encode location, but also the time of firing relative to the ongoing theta rhythm in the local field potential (LFP). Finally, we turn to the equally striking representation of the animal's orientation provided by head direction cells, the nature of which has also been investigated by computational modelling.

16.2.1 Representing Spatial Location and Orientation: Data

A rich set of experimental data have been gathered on the neural representations of spatial behaviour found in and around the hippocampus. Here we briefly summarize those results with the greatest relevance to the models described below (see Chapters 11 and 12 for more details). The firing of place cells in the hippocampi of freely moving rats encodes the location of the

animal, each cell firing when the animal is within a particular portion of its environment (the corresponding ‘place field’). In smaller environments, place cells typically exhibit a single place field, but in larger environments they may exhibit several place fields with no apparent relationship between the location of each (Fenton et al., 2008; Alme et al., 2014; Rich et al., 2014). Cells with similar responses have also been observed in mice and gerbils (McHugh et al., 1996; Mankin et al., 2019); birds (Payne et al., 2021); bats (Ulanovsky et al., 2007; Yartsev et al., 2013); human (Ekstrom et al., 2003) and non-human primates (Hori et al., 2003; Ludvig et al., 2004; Courellis et al., 2019; Mao et al., 2021). In open environments through which the rat can move freely, firing rates are not influenced by the animal’s orientation, while in environments in which movement direction is constrained (e.g. linear tracks, 8-arm mazes) firing is strongly modulated by the rat’s direction of motion.

The location of the place cell representation is controlled by ‘distal’ cues at or beyond the edge of the environment (O’Keefe and Conway, 1978; Muller and Kubie, 1987) more than by those within it (Cressant et al., 1997). A place cell’s spatially localized firing appears to be robust to the removal of subsets of cues, and indeed removal of all of the controlling visual cues while the rat remains in the environment (Muller and Kubie, 1987; O’Keefe and Speakman, 1987), although remaining uncontrolled cues may be important in these cases (Save et al., 2000). In addition, the peak firing rate of the place cell may change significantly when features of the environment are changed, a process known as ‘rate remapping’; while the location of a place field may change dramatically, the cell may stop firing altogether, or previously silent cells may begin to exhibit a place field when the environmental features are changed more significantly, a process known as ‘global remapping’ (Muller and Kubie, 1987; Bostock et al., 1991). Finally, sequences of place cell firing observed during active behaviour are ‘replayed’ on a compressed timescale during subsequent slow-wave sleep and quiescent waking periods (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996; Foster and Wilson, 2006).

In contrast, grid cells recorded in freely moving rodents fire action potentials at multiple spatial locations. These firing fields are typically arranged at the vertices of a regular triangular array covering the whole environment (Hafting et al., 2005). Grid cells were initially discovered in the superficial layers of rodent MEC (Hafting et al., 2005; Fyhn et al., 2008), but have since been identified in the deeper layers (Sargolini et al., 2006) and in pre- and para-subiculum (Boccarda et al., 2010). Moreover, grid-like responses have been recorded in the parahippocampal cortices of the bat (Yartsev et al. 2011), human (Doeller et al., 2010; Jacobs et al., 2013) and non-human primate (Killian et al., 2012). Grid cell firing patterns can be characterised by their scale (i.e. the distance between adjacent firing fields), orientation (of one principal grid axis relative to an external cue), and the phase or spatial offset of their firing fields. Grid scale has been shown to increase in discrete steps along the dorsoventral axis of MEC (Barry et al., 2007; Stensola et al., 2012), and evidence suggests that grid cells which share a common scale form a single functional module (Stensola et al., 2012; Yoon et al., 2013). The scale, relative orientation and offset of grid firing patterns within each module are generally conserved across environments (Fyhn et al., 2007), aside from a transient expansion of grid scale in novel environments that returns to baseline with experience (Barry et al., 2012). The spatial phases of individual grid cells are uniformly distributed across the environment but, importantly, the relative spatial phase of any two simultaneously recorded grid cells from the same module is conserved across all environments visited by the animal (Fyhn et al., 2007; Yoon et al., 2013).

The complementary representation of orientation independent of location is found in head direction cells in the mammillary bodies, anterior thalamic nuclei, dorsal presubiculum and MEC (Taube et al., 1990; Sargolini et al., 2006). These cells code for head direction within an environment, each firing whenever the animal's head points in a specific direction, independently of the animal's location. The orientation of the head direction representation is controlled by distal visual cues in the same way as the place cell representation. The overall orientation of place, grid and head-direction representations may be disrupted by disorientation (rotating the rat in a covered container) and, when recorded simultaneously, all representations have remained in register with each other (Knierim et al., 1995; Sargolini et al., 2006). Interestingly, the firing rate of 'conjunctive' grid cells in the deeper layers of MEC is also modulated by heading direction (Sargolini et al., 2006). In addition, the firing rate of place, grid and head direction cells (McNaughton et al., 1983; Sargolini et al., 2006; Hardcastle et al., 2017), as well as some MEC neurons that do not appear to encode any spatial variables (Kropff et al., 2015), are modulated by running speed.

Interestingly, while the firing rates of both place cells (e.g. Wilson and McNaughton, 1993) and grid cells (e.g. Fiete et al., 2008; Mathis et al., 2012) provide a population vector (Georgopoulos et al., 1986; see Box 16.2) that encodes the animal's location within a given environment, the times at which they fire relative to the LFP theta rhythm encodes additional information (see Chapter 11; O'Keefe and Recce, 1993; Skaggs et al., 1996; Jensen and Lisman, 2000; Hafting et al., 2008; Climer et al., 2013; Jeewajee et al., 2014). Specifically, these cells exhibit theta 'phase precession', firing at a progressively earlier phase of each theta cycle as the firing field is traversed. Importantly, the initial phase of firing upon entry to a place or grid field is typically consistent across cells and, as a result, phase precession generates 'theta sequences' of place and grid cell firing across the population, whereby cells with receptive fields behind the animal fire early, and those ahead of the animal fire late, in each oscillatory cycle (Burgess et al., 1994; Skaggs et al., 1996; Johnson and Redish, 2007). Intriguingly, this phase code for location is conserved across species (Eliav et al., 2018; Qasim et al., 2021) but does not appear to rely on sustained rhythmicity in the theta band: place and grid cells in bats exhibit a phase code for location relative to LFP fluctuations that vary dynamically over a wide range of frequencies (Eliav et al., 2018).

16.2.2 Representing Spatial Location: Feed-forward Models

Feedforward neural networks, in which activity propagates unidirectionally between successive layers of simulated neurons, have had great success in solving pattern recognition and classification problems – producing specific patterns of output activity (e.g. place or grid cell firing fields) given specific patterns of input (e.g. particular constellations of sensory features). Due to their limited internal dynamics, however, feedforward networks are typically unable to integrate the recent history of their inputs – to perform path integration by combining previous estimates of location with self-motion inputs, for example.

Place Cells

Computational modelling of place cell firing began with Zipser (1985). In this model, sensory details of the environment feed-forward to landmark detectors, and thence to place cells. Landmark detectors are neurons specific to a unique place cell and aspect of the sensory scene

(a 'location parameter'). The output of these detectors is proportional to the match between the stored state of a location parameter and its currently perceived state. A place cell's activity corresponds to a thresholded sum of the strengths of the matches it receives from several landmark detectors. Interestingly, the most obvious location parameter – distance from a landmark – was rejected by the author, in favour of measures that scale with environmental size such as the retinal angle between two landmarks. As such, the model captures some of the motility of place fields in the presence of manipulations of environmental cues and some of their robustness to removal of subsets of cues, but (incorrectly) produces place fields that scale up proportionately with environmental expansion (Muller and Kubie, 1987; O'Keefe and Burgess, 1996).

Sharp (1991) followed in the same vein of feed-forward modelling of the response of place cells to sensory input from the environment, but with the incorporation of an element of 'competitive learning' (Rumelhart and Zipser, 1986). Briefly, this involves neurons arranged into groups dominated by lateral inhibition such that only the neuron with the greatest input can fire. Normalized Hebbian learning is then applied (i.e. increasing the strengths of connections between simultaneously active neurons while decreasing the others so that the overall strength of connections to a neuron does not change, see Box 16.1). This learning results in specific neurons coming to represent specific patterns of sensory input: each neuron responding to a particular pattern, or to patterns similar to it. Her model envisaged two types of sensory input regarding each distal cue, one representing its distance from the rat and the other representing both its distance from the rat and its direction relative to the rat's heading. This sensory input passed forwards to a layer of entorhinal cells and thence to a layer of place cells. Competitive learning at each layer causes the entorhinal cells and place cells to respond selectively to the pattern of sensory input present in a particular portion of the environment and produces reasonable robustness to cue removal. The successive layers of competition produce sharper tuning to position and greater robustness to cue removal in place cells than entorhinal cells.

Interestingly, place cell firing in this model is initially directionally modulated, due to the partially directional sensory inputs. During random exploration in an open environment, competitive learning allows a given place cell to learn to respond to the sensory inputs occurring for different orientations at the place field, producing non-directional firing. By contrast, this does not occur during constrained motion (i.e. back and forth in a single direction). This provides a simple account of the directionality of place cell firing, although a more detailed look at the experimental data indicates that, if anything, place fields in open environments are initially non-directional and become directional as a result of experience (Markus et al., 1995; Navratilova et al., 2012a).

In a related model, Franzius et al. (2007) demonstrated that specific assumptions about the form of sensory inputs to place cells could be avoided by simply identifying the most slowly changing features of that input, motivated by the observation that sensory information typically varies much more quickly than behaviourally relevant features of the environment (Wiskott and Sejnowski, 2002). This 'slow feature analysis' (SFA) can produce both place and head direction cell firing patterns from raw visual input, depending on the relative speed of translational movement and head rotation, and could be achieved by biologically plausible learning rules (Sprekeler et al., 2007). As in Sharp's model, place cell responses in the output

layer of the feedforward network are most realistic when SFA is combined with competitive learning.

More recently, it has been demonstrated that spatially modulated, feedforward excitatory and inhibitory inputs governed by different learning rules can also produce a variety of spatial firing patterns in an output neuron (Weber and Sprekeler, 2018). Specifically, if excitatory inputs are subject to a Hebbian learning rule while the strength of inhibitory inputs changes according to the product of the pre-synaptic firing rate and the difference between post-synaptic firing rate and a single, global target value (see Box 16.1), then the output neuron can learn to produce either place or grid cell firing patterns, depending on the relative spatial smoothness of those inputs. If inhibition is uniform across the environment, the output neuron produces a single place field; if inhibitory inputs are smoother than excitatory inputs, the output neuron produces a grid firing pattern by learning a centre-surround input profile; and if inhibitory inputs are less smooth than excitatory inputs, the output neuron produces weakly spatially modulated firing.

In an attempt to derive the specific form of the sensory input to place cells, O'Keefe and Burgess (1996) systematically varied the shape and size of the rat's environment while recording from the same cells. The patterns of firing across environments included place fields that stretched or became bimodal when the environment expanded. These patterns were not consistent with previous models of place fields depending on the relative locations of discrete landmarks from the rat (e.g. Zipser, 1985; Sharp, 1991), but rather indicated continuous dependence on environmental boundaries. Specifically, place fields were viewed as a thresholded linear sum of inputs tuned to respond to the presence of a boundary at a given distance along a given allocentric direction (i.e. independent of the orientation of the rat, and probably determined relative to the head-direction system, see below; Fig. 16.1). These hypothetical inputs were termed 'boundary vector cells'.

By fitting a place cell's firing pattern across several different environmental shapes, the model can predict its firing pattern in an environment of novel shape (Hartley et al., 2000). In addition, boundary vector cells with allocentric firing patterns predicted by this model were later identified in the medial entorhinal cortex (Solstad et al., 2008) and subiculum (Lever et al., 2009). Importantly, however, the recent observation that place cell firing is modulated by heading direction relative to some fixed reference point indicates that some egocentric influence remains (Jercog et al., 2019). It is also important to mention that the subiculum, where boundary vector cells appear to be most numerous, has traditionally been considered an output of the hippocampus, and so it is not clear how firing patterns in this region might give rise to hippocampal place cells (although more recent studies suggest there may be projections to CA1, if not CA3, e.g. Sun et al., 2019). In addition, place cell activity must be at least partially determined by non-boundary related inputs, or firing patterns would be conserved across all geometrically identical environments. More generally, the feedforward models described above cannot account for the persistence of place cell firing in darkness, when visual inputs are presumed to be absent and the updating of spatial representations is mediated by self-motion (i.e. 'path integration') inputs.

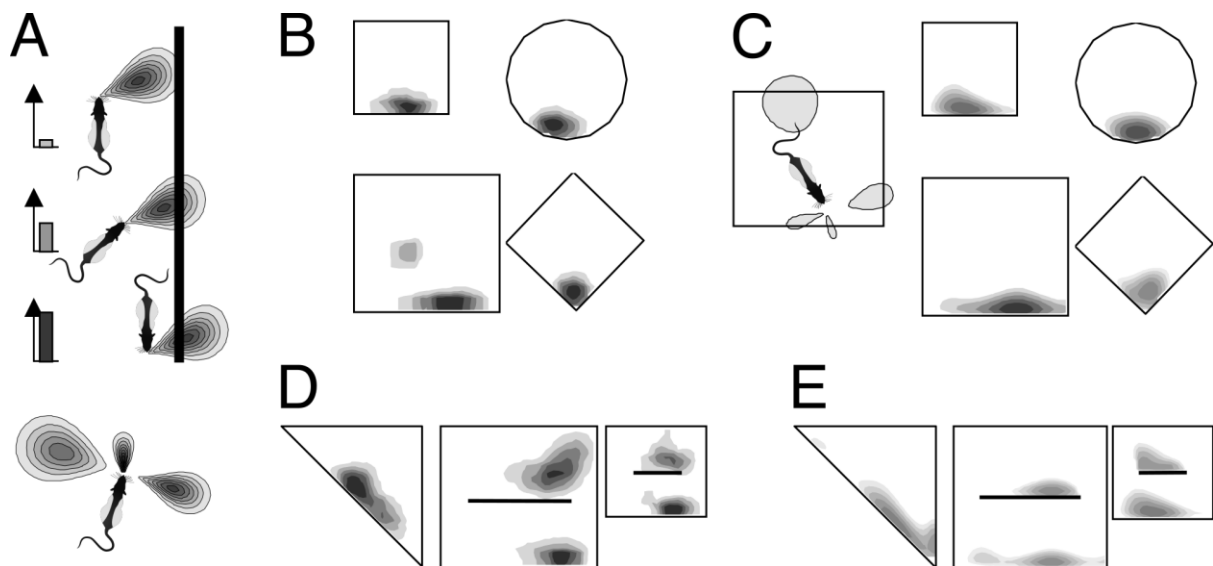


Figure 16.1 Model of the influence of environmental boundaries on place fields, assuming a stable directional reference frame. Place fields are composed from thresholded linear sums of the firing rates of ‘boundary vector cells’ (BVCs). **A:** Above: Each BVC has a Gaussian tuned response to the presence of a boundary at a given distance and bearing from the rat (independent of its orientation). Below: The sharpness of tuning of a BVC decreases as the distance to which it is tuned increases. The only free parameters of a BVC are the distance and direction of peak response. **B:** Place fields recorded from the same cell in four environments of different shape or orientation relative to distal cues. **C:** Simulation of the place fields in **B** by the best fitting set of 4 BVCs constrained to be in orthogonal directions (BVCs shown on the left, simulated fields on the right). The simulated cell can now be used to predict firing in novel geometrical configurations of boundaries. Real and predicted data from three novel configurations are shown in **D** and **E** respectively, showing good qualitative agreement. Adapted from (Burgess and Hartley, 2002).

Another phenomenon not addressed by the above models is the ‘remapping’ of place cell representations across different environments. This ‘remapping’ can be both partial and incremental over time (e.g. Bostock et al., 1991; Skaggs and McNaughton, 1998; Lever et al., 2002), with the eventual creation of stable but distinct patterns of firing in the two environments. The factors influencing the speed and extent of remapping are currently not well understood (Sanders et al., 2020; see Section 16.5.2), but one common change in an individual place cell is to continue to fire in the environment in which it fires most strongly, and to stop firing in the other. This aspect of remapping was addressed by Fuhs and Touretzky (2000) in a model of learning in the perforant path projection from entorhinal cortex to place cells in CA3. They found that the usual learning rules relating synaptic modification to the product of the pre- and post-synaptic activity (i.e. ‘Hebbian’ learning), or to its covariance, were unable to reproduce this behaviour. In the case of Hebbian learning, a place cell with strong firing in one environment and weak firing in the other will strengthen its firing in both environments. In the case of covariance learning, exposure to the second environment will tend to lead to loss of the place cell representation in the first environment. By contrast, the BCM learning rule (Bienenstock et al., 1982; see Box 16.1), which explicitly makes the direction of synaptic modification dependent on the strength of post synaptic activity, did produce the desired result: strong firing remaining stable and weak firing reducing with experience. This type of learning also captures the way place fields become more coherent with time, and the dynamics of their response to the introduction of a barrier into the environment (Barry and Burgess, 2007).

Evidence of experience-dependent change in place cell firing also comes from experiments by Mehta et al. (1997; 2000). They found that, over the first few runs through a CA1 place field on a linear track, the spatial distribution of firing changes from roughly symmetrical to take on a slight asymmetry caused by additional firing at low rates earlier on the track. They suggest that this results from the known temporal asymmetry of LTP (which is greater when pre-synaptic activity precedes post-synaptic activity than vice versa, see e.g. Bi and Poo, 1998) acting on the CA3 to CA1 pathway (see Chapter 10). Other models have implicated the recurrent connections within CA3 as responsible for this effect (see Section 16.2.3). Interestingly, the theta phase precession of place cell firing (see Section 16.3.1 and Chapter 11 for data and Section 16.3.5 for models) acts to increase the effect of temporal asymmetry in LTP: causing place cells with fields early on the path to fire before those with fields later along the path within each theta cycle.

Finally, following the discovery of grid cells, numerous theoretical models have demonstrated how input from grid modules of two or more spatial scales could be combined to generate place fields through an effective Fourier synthesis (e.g. Rolls et al., 2006; Solstad et al., 2006). Using either hardwired synaptic weights or some form of Hebbian learning rule, these models set the effective strength of grid cell inputs to decline with their spatial offset from the output place field (Cheng and Frank, 2011). Grid cell to place cell models can produce either single or multiple place fields, although the secondary fields often exhibit six-fold symmetry - particularly when all grid inputs share a single orientation - in apparent contrast with empirical data. More restricted place field firing can be generated by introducing some variation in firing rate between the receptive fields of each grid cell, in line with experimental data (Ismakov et al., 2017). Finally, making independent changes to the orientation and / or spatial phase of each grid module (Fyhn et al., 2007), and / or incorporating a ‘gating’ input representing abstract contextual signals (Hayman and Jeffery, 2008), can account for the remapping of output place field responses, while remapped field locations may reflect movements between the vertices of an underlying grid (Whittington et al., 2020). Several predictions of these models appear to be at odds with empirical data, however (Bush et al., 2014). Most notably, place cell firing patterns appear to precede those of grid cells in the developmental timeline (Langston et al., 2010; Wills et al., 2010); and grid firing patterns are eliminated by inactivation of medial septum, with little effect on place cell responses in either novel or familiar environments (Koenig et al., 2011; Brandon et al., 2014).

Grid Cells

In addition to modelling place cell responses, two main classes of grid cell model suggest that their regular, periodic firing fields can be generated by feed-forward input from other regions. Interestingly, the first class reverses the logic of grid-to-place cell models by suggesting that grid cell firing patterns might be generated using feed-forward input from the hippocampus through a process analogous to principal component analysis (PCA) of place cell firing covariance (Castro and Aguiar, 2014; Dordek et al., 2016) or, relatedly, eigen decomposition of the transition matrix between places (Stachenfeld et al., 2017). These models build on the observations that a modified Hebbian learning rule acting on feed-forward projections can approximate a process of PCA (Oja, 1982; see Box 16.1); that feed-forward projections exist from CA1 to the deeper layers of MEC; and that grid cell firing patterns appear to rely on stable

place cell activity (Bonnevie et al., 2013). Hence, employing this modified learning rule at feed-forward projections from a population of simulated place cells – or approximating that process mathematically using non-negative PCA - creates output units with grid-like firing patterns. This account bears some resemblance to earlier models which suggested that feed-forward input from place cells to a network of neurons with spike-frequency adaptation was also sufficient to generate grid-like firing patterns (Kropff and Treves, 2008; Si and Treves, 2013; D’Albis and Kempter, 2017).

In contrast, most theoretical models of grid cell firing assume that their principal input is a self-motion signal, with periodic firing patterns resulting from the integration of that velocity signal over time, consistent with a proposed role in path integration (McNaughton et al., 2006; further discussion in the following section). In particular, following accounts of theta phase precession in place cells (O’Keefe and Recce, 1993; Lengyel et al., 2003; see Section 16.2.4), the oscillatory interference (OI) model proposes that grid firing patterns can be accounted for at the single cell level by constructive interference between two or more oscillatory inputs (Burgess et al., 2005; Burgess et al., 2007; Blair et al., 2008; Burgess, 2008; Hasselmo, 2008). In its simplest one-dimensional (1-D) form, one oscillation has a baseline frequency and the other ‘velocity controlled oscillator’ (VCO) has a frequency that varies linearly from that baseline with the speed of movement (Burgess, 2008). In rodents, the baseline frequency is generally assumed to be the 5-12Hz movement related theta oscillation (Vanderwolf, 1969; O’Keefe and Nadel, 1978; Burgess et al., 2007).

These two signals generate grid cell membrane potential oscillations (MPOs) modulated by an ‘envelope’ frequency that is equal to the difference in baseline and VCO frequencies; and a ‘carrier’ frequency between the baseline and active frequencies (see Fig 16.2A). The envelope corresponds to the grid cell rate code – being spatially periodic and approximately Gaussian or cosine tuned; while the carrier corresponds to the temporal code - being higher in frequency than the baseline oscillation, and thus causing the grid cell to fire at progressively earlier phases of that baseline oscillation as the firing field is traversed (i.e. generating phase precession; see Section 16.2.4). The scale of the resultant grid firing pattern is controlled by the slope of the VCO movement speed / burst firing frequency relationship, which determines how quickly the VCO and baseline oscillation move in and out of phase during movement.

The OI model accounts for two-dimensional (2-D) grid firing patterns by incorporating input from multiple VCOs whose burst firing frequencies vary linearly with movement speed along different preferred directions. Because distance is the time integral of velocity, and phase is the time integral of frequency, the phase of each VCO – if sampled at fixed intervals (i.e. at the peak or trough of the baseline oscillation) - encodes (periodic) displacement in its preferred direction. A grid cell that receives input from two or more VCOs whose preferred directions differ by multiples of 60° will exhibit a triangular array of firing fields at locations where those VCO inputs are in phase. The specific location or offset of those firing fields can be manipulated by adding a constant phase shift to one or more VCO inputs. Hence, the OI model proposes that each VCO performs path integration along different 1-D axes, while grid cells simply ‘read-out’ the activity of multiple VCO inputs by firing whenever they are in phase.

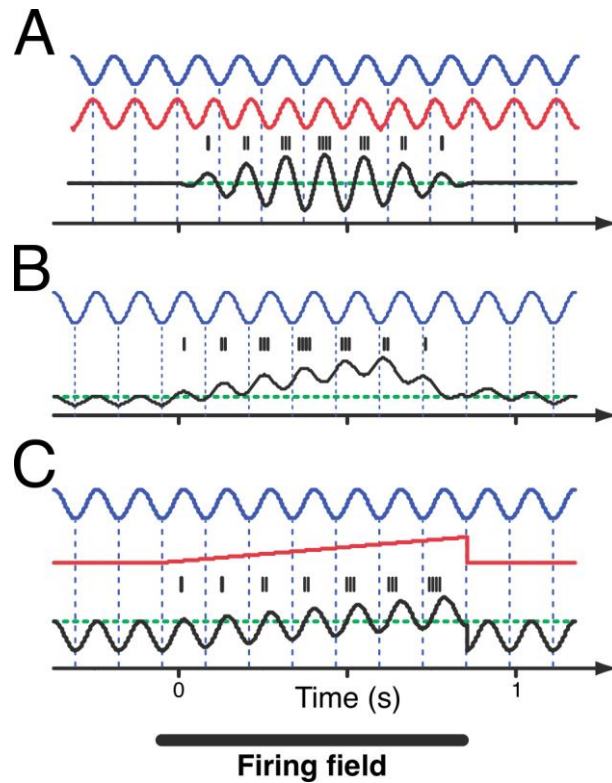


Figure 16.2 Models of phase coding in place and grid cells. **A:** The oscillatory interference (OI) model of phase precession and grid cell firing in 1D environments. A baseline oscillation with frequency f_{base} (blue line) and a velocity controlled oscillation (VCO) with frequency f_{VCO} that varies from f_{base} linearly with movement speed. Constructive interference between these two oscillations generates a spatially periodic activity pattern with a ‘carrier’ frequency (black line) equal to their mean frequency $(f_{base} + f_{VCO})/2$ and an ‘envelope’ equal to their difference in frequency $(f_{VCO} - f_{base})$. This activity pattern corresponds to spatially periodic firing fields within which spikes (black dashes) are fired at the peak of the membrane potential oscillation (MPO, black line) but progressively earlier phases of the baseline oscillation, as observed in grid cells. **B:** Schematic representation of intracellular recording data from place (Harvey et al., 2010) and grid (Dominsoru et al., 2013; Schmidt-Hieber and Hausser, 2013) cells. The membrane potential (black line) shows a ramping depolarisation but no increase in MPO amplitude within the firing field. Spikes (black dashes) are fired at the peak of MPOs but progressively earlier phases of the LFP theta oscillation (blue line) **C:** The depolarising ramp model of phase precession in place and grid cells. Ramp-like depolarisation (red line) causes the membrane potential (black line) to exceed firing threshold (dashed green line) progressively earlier in each cycle of the LFP oscillation (blue line), such that spikes (black dashes) are fired at progressively earlier MPO phases (cf. panel B). Adapted from (Burgess and O’Keefe, 2010).

The OI model accounts for both periodic firing patterns and theta phase precession in grid cells, and is supported by the observation that a reduction in theta band activity caused either by passive transport of the animal (Winter et al., 2015) or inactivation of medial septum (Brandon et al., 2011; Koenig et al., 2011) leads to the loss of grid firing patterns. In addition, cells with VCO-like properties have been identified in the hippocampal formation (Welday et al., 2011). The OI model has been challenged by the observation that oscillatory activity with a relatively constant frequency is absent from the hippocampal formation of bats (Yartsev et al., 2011) and humans (Watrous et al., 2013). Importantly, however, it is the phase difference between VCO and baseline oscillations that encodes location, and the baseline frequency can therefore vary so long as this phase relationship is preserved (Burgess, 2008; Blair et al., 2014; Orchard, 2015;

Bush and Burgess, 2020), e.g. if it is itself derived from the active frequencies (Burgess and Burgess, 2014). Indeed, models forming grid cells from path-integrating ‘stripe’ or ‘band’ cells (Mhatre et al. 2012; Horiuchi and Moss 2015) are equivalent to oscillatory interference with a baseline frequency of 0 Hz. Nonetheless, the OI model alone cannot account for the strong interactions between grid cells from the same functional module that are clearly indicated by experimental data (Yoon et al., 2013), nor for the subthreshold membrane dynamics observed during movement through the grid field, which exhibit no change in the amplitude of theta band MPOs inside grid firing fields (see Fig 16.2B; Domnisoru et al., 2013; Schmidt-Hieber and Hausser, 2013).

16.2.3 Representing Spatial Location and Orientation: Feed-back Models

The long-range recurrent connections between pyramidal cells in area CA3 of the hippocampus have long been interpreted as enabling this region to work as an auto-associative neural network, see e.g. (Marr, 1971; Hopfield, 1982; Amit, 1989). This type of network is most often used to provide a content-addressable memory, a subject explored in Section 16.4. Elsewhere, it has been demonstrated that di-synaptic recurrent inhibitory connections couple grid cells in MEC (Couey et al., 2013). In this section we consider the role played by recurrent collaterals in the spatial representations of place, grid and head-direction cells. Recurrent networks can take advantage of self-generated internal dynamics to imbue spatial representations with a dependence on the recent history of inputs. As such, these networks can readily account for path integration – that is, using self-motion signals to update previous estimates of self-location; as well as leveraging attractor dynamics (see Box 16.3) to ensure that spatial representations are robust to the withdrawal of a subset of sensory cues. Interestingly, direct experimental evidence for an associative function for CA3 has emerged, with indications that the NMDA receptors in this region are involved in making both the place fields and the rat's spatial memory robust to cue removal (Nakazawa et al., 2002). In parallel, attractor dynamics have been found in the place cell representation of two environments of different shape after fast remapping caused by exposure to the two environmental shapes made of different materials (Wills et al., 2005), and analogous results have been obtained from the human hippocampus using fMRI (Stemmers et al., 2016). In these representations, in contrast to those that have not fast-remapped (Leutgeb et al., 2005), the two shapes act as point attractors: all place cells in intermediate shaped environments coherently returning to one or other representation. Substantial evidence has also emerged for continuous attractor network dynamics in grid cell firing patterns (Yoon et al., 2013; Gardner et al., 2022).

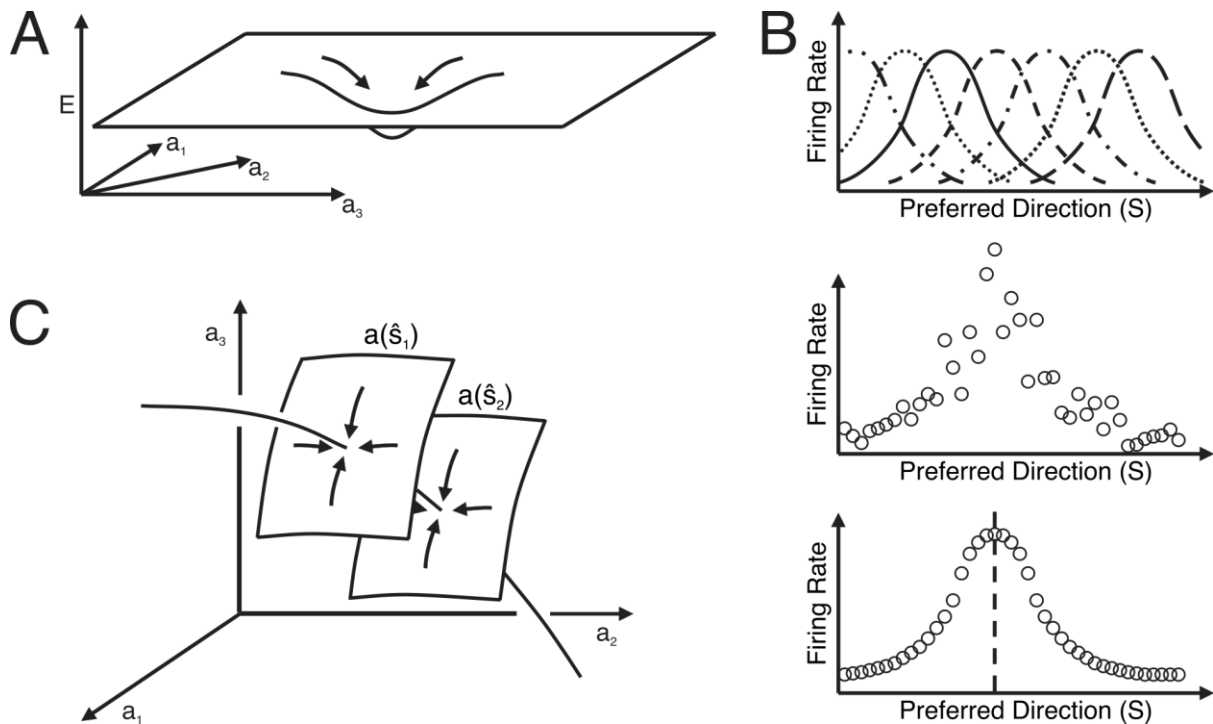


Figure 16.3 Point attractors and line attractors in neural systems (see Box 16.3). **A:** A point attractor is a pattern of activation $\mathbf{a} = (a_1, a_2, a_3, \dots)$ into which other nearby patterns will evolve under the dynamics of the network (determined by the pattern of connection strengths, update rule etc). In some cases a function can be defined that can only decrease under the dynamics (a 'Lyapunov' function, or the energy E of a physical system), so that attractor states lie at local minima of this function. **B:** Population encoding and decoding (see Box 16.2). Neural populations can encode the current value of a variable in the pattern of activation across neurons, each of which is tuned to respond to a preferred value. These are often imagined to be laid-out so that the position of each neuron on the ordinate corresponds to its preferred value. For example, a set of head-direction cells might each be tuned to a different 'preferred' direction (top). If firing rates are noisy it may be difficult to estimate the actual value of the variable (middle). Recurrent connections can be organized so that all other patterns of activation evolve into a smooth bump-shaped patterns of activation (below). This process can provide an optimal way of decoding the value of the variable from the population (the peak of the bump – dashed line). **C:** The set of smooth bump-shaped patterns of activation form a 'line attractor': a continuous set of patterns of activity onto which other nearby patterns will evolve, but along which movement is unimpeded. Locations along the line attractor can be thought of as estimates of the variable (\hat{S}), all of the patterns of activity that end up at a given estimate (\hat{S}_1 , say) form a subspace $\mathbf{a}(\hat{S}_1)$ within which the intersection with the line is a point attractor. Adapted from (Latham et al., 2003).

Continuous attractor models of head direction cells

The simplest examples of the use of continuous attractors (see Box 16.3) to model spatial representations come from models of the representation of head-direction rather than location. In many respects the literature on head-direction cells (HDCs) is much more straightforward than that on place or grid cells. The overall orientation of the head direction representation can be controlled by sensory cues in a similar way to the place and grid cell representations. Unlike place cells, however, there have been no reports to date of HDCs changing their preferred orientations relative to each other (i.e. remapping). Even when the rat is disoriented, asleep, in a symmetrical environment without polarising cues, or if angular head velocity inputs are inhibited, the preferred directions of simultaneously recorded HDCs remain consistent – if they rotate, all rotate together (Peyrache et al., 2015; Butler et al., 2017; Bassett et al., 2018;

Chaudhuri et al., 2019; but see Kornienko et al., 2018). This is analogous to grid cell firing patterns, which preserve the spatial offset of their firing fields (or relative phase) across all environments visited by the animal (Fyhn et al., 2007; Yoon et al., 2013) and during sleep (Trettel et al., 2019; Gardner et al., 2019). For this reason all models of the head-direction system follow a similar basic mechanism of a 1-D continuous attractor or ‘line attractor’ (Skaggs et al., 1995; Redish et al., 1996; Zhang, 1996) from which the 2-D continuous attractor models of place and grid cells developed (see Figs. 16.3-5 and Box 16.3).

If HDCs are imagined laid out in a ring with each cell’s location corresponding to its preferred direction, and each is connected to its neighbours (see Box 16.3), then activity will be smoothly peaked at the current heading direction. Skaggs et al.’s model contains two further rings of cells, with each cell receiving connections from the corresponding HDC. One ring is composed of ‘left rotation’ cells which project back to the HDCs to the left (anticlockwise) of their location, while the other is composed of ‘right rotation’ cells which project back to HDCs to the right (clockwise) of their location. The left rotation cells corresponding to the current heading direction are activated when the rat is turning left due to inputs from the vestibular system as well as the HDCs, causing the HDC activation to move leftwards. Alternatively, a similar system can be formed of just two turn-modulated rings with offset connections (Redish et al., 1996). In addition to these cells, the HDCs receive input from ‘sensory cells’ (‘visual cells’ in Fig. 16.5A) to become associated with those sensory inputs appearing at a stable bearing during exploration of a new environment. These sensory inputs subsequently prevent the cumulative errors that would otherwise occur in the integration of angular velocity.

This basic model has been implemented and extended in different ways, developing in hand with our knowledge of the operation of the head-direction system. This is now thought to involve a circuit from the mammillary bodies (MB) to anterior thalamic nuclei (ATN) to dorsal presubiculum (PS), with visual inputs arriving in PS via retrosplenial cortex (RSc). In this circuit, cells in the MB code for head direction further in the future (60 to 70 ms) than those in the ATN (20 to 30 ms), while those in the PS code for current or past head direction (0 to -10ms; see Blair and Sharp, 1995; Taube and Muller, 1998; Taube, 1998; Blair et al., 1998). Various of the additional detailed properties of these systems, such as time advances and asymmetric responses during turning, have been incorporated, e.g. (Touretzky and Redish, 1996; Blair et al., 1997; Goodridge and Touretzky, 2000). In addition, some models have addressed the problem of parallax, whereby proximal sensory cues that present different orientations depending on location within the environment introduce errors into the head direction circuit (Bicanski et al., 2016); and head-direction coding in three dimensions (Page et al., 2018; Laurens and Angelaki, 2018). Finally, it is worth noting the incredible correspondence between the ring attractor model of head direction coding and activity in the ellipsoid body of the fly brain (Seelig and Jayaraman, 2015; see Fig. 16.5B). Here we focus on the hippocampus, however, and return to models of location.

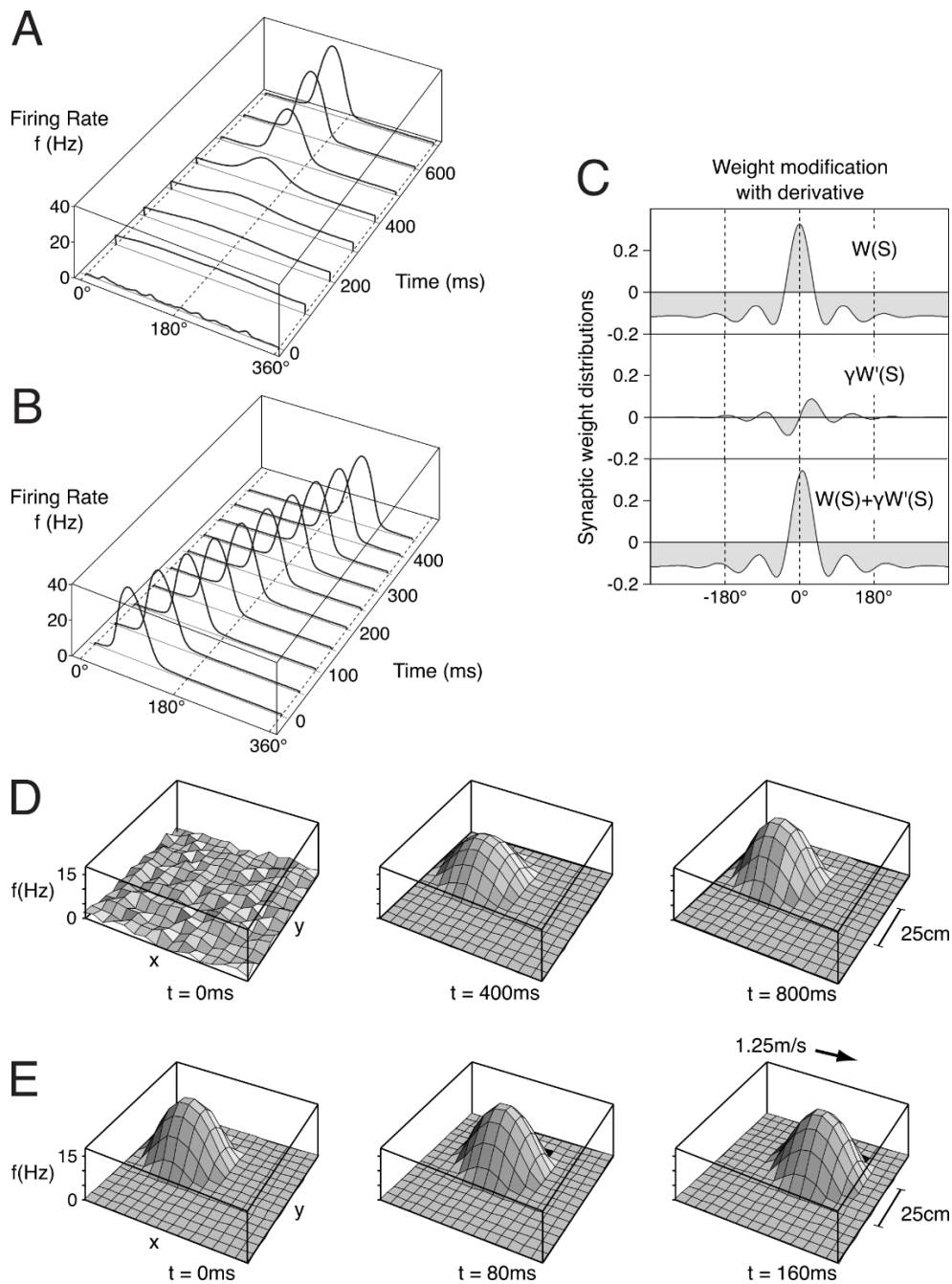


Figure 16.4 Continuous attractor networks of place and head-direction cells. **A:** Emergence of a stable firing profile from an arbitrary initial state in a network of head direction cells arranged as a 1-D continuous attractor. The cells are indexed by their preferred firing direction and connected by weights with a symmetrical distribution, i.e. an even function of the difference between cell's tuning directions (see C, top row). **B:** Movement of the peak caused by an asymmetrical component in the weight distribution (see C, middle row). **C:** Distribution of connection weights in a 1-D attractor network (bottom row), showing symmetrical component (top row) and additional asymmetric component (middle row). Note the slight asymmetry in the combined connection weights. The asymmetric component is the spatial derivative of the symmetric component along the direction of drift and its size (γ) determines the speed of drift of the represented head-direction. **D:** A 2-D place cell network similar to the 1-D head-direction cell network, showing emergence of a stereotyped stable firing profile from an arbitrary initial state, using a symmetric weight distribution (a Gaussian with constant inhibitory background). **E:** As with the 1-D network, the addition of an asymmetric component to the connection weights causes the represented location to drift (again, the asymmetric component is the spatial derivative along the direction of drift and its size determines the speed of drift). Adapted from (Zhang, 1996).

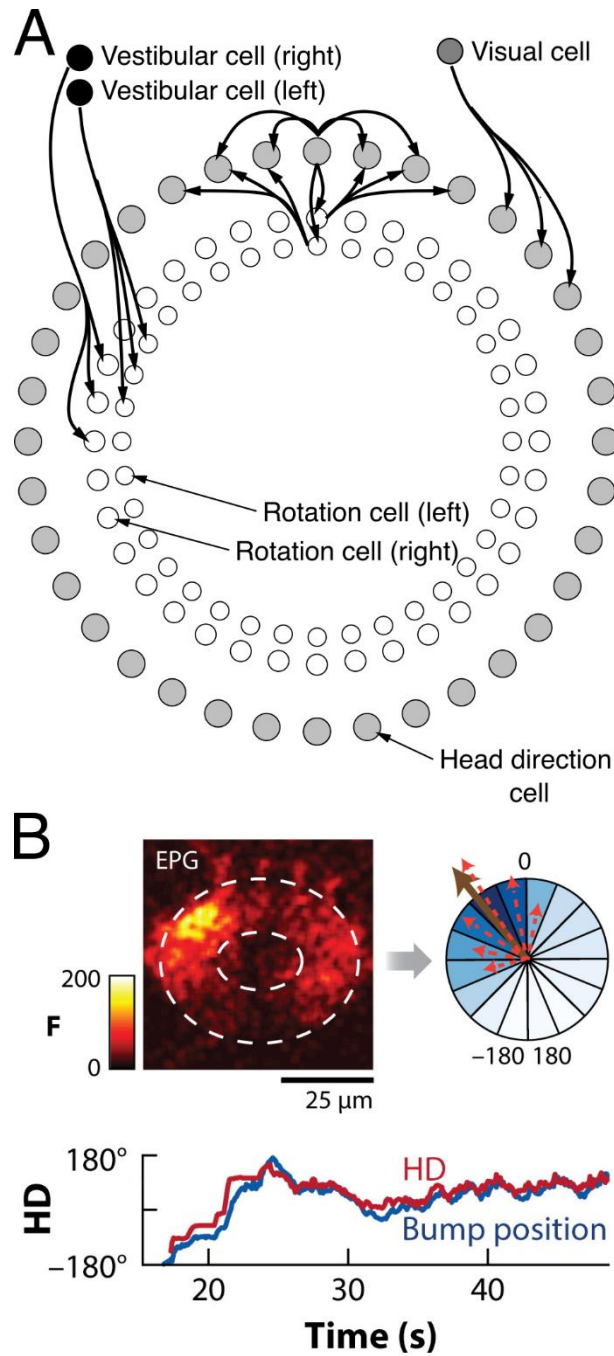


Figure 16.5 Ring attractor network model of head-direction cells and empirical data. **A:** Skaggs et al.'s (1995) model of head direction cells, showing the lateral connections among head direction cells providing a continuous attractor, and connections from left or right rotation cells, visual inputs and vestibular inputs. The input from rotation cells serves the same purpose as the asymmetric component of lateral connections in Fig. 16.4C. Adapted from Skaggs et al. (1995). **B:** Empirical data from *Drosophila*. Snapshot of calcium activity (F) in EPG neurons of the ellipsoid body, showing a single activity bump (top left panel). This can be used to compute a resultant vector that tracks the angular position of the activity bump (top right panel). This angular position (blue line) maintains close correspondence with the fly's heading direction (HD, red line) while walking on a spherical treadmill. Adapted from Hulse and Jayaraman (2020).

Continuous attractor models of place cells

Samsonovich and McNaughton (1997) produced a detailed model of the place cell representation as a continuous attractor, following (Zhang, 1996). In this model, the recurrent connections in CA3 are pre-configured to provide several different continuous attractor representations of location (termed ‘charts’). Each chart involves a different set of place cells, the relative positions of whose place fields are predetermined. The strength of the recurrent connection between two cells in a chart is set as a Gaussian function of the proximity of their place fields. The place cells connect with a ‘path integration’ (PI) system, originally hypothesized to reside in the subiculum, in which neurons respond to combinations of the rat’s location and orientation (Sharp, 1996; Cacucci et al., 2004). Specifically, the place cells connect to PI neurons representing similar locations while the return projections connect back to place cells representing slightly different locations shifted along the rat’s direction of orientation. The gain of this return projection is modulated by information relating to the rat’s speed of self-motion (presumably carried by motor efference signals). This system, while demanding a highly specific set of hard-wired connections, provides a self-consistent continuous attractor representation of location that moves automatically with self-motion. Finally, the hippocampus also receives sensory input so that, when the rat is placed in an environment for the first time, associations between the sensory scene and the internal representation of location can be formed which can then be used to periodically reset the system.

Overall, the model can be seen as a possible implementation of O’Keefe and Nadel’s (1978; pp 220-230) view of the role of path integration in supporting short-term continuity in a cognitive map. Subsequent developments eliminated the need for a separate path integration system by incorporating direction and velocity modulation of place cells within the continuous attractor network (Conklin and Eliasmith, 2005). Samsonovich and McNaughton (1997) also suggest that remapping reflects the system switching between uncorrelated charts. This is a reasonable model of the situation after global remapping, which is consistent with each chart acting as an attractor (Wills et al., 2005). However, the model is not consistent with situations in which individual place fields can move relative to each other in response to environmental change e.g. (O’Keefe and Burgess, 1996; Fenton et al., 2000), nor with slow (Lever et al., 2002) or partial remapping (e.g. Skaggs and McNaughton, 1998). To fit these data requires feed-forward inputs to dominate, replacing the model’s main feature.

Conversely, the Samsonovich and McNaughton (1997) model is consistent with data showing that the stability of the place cell representation is dependent on NMDA receptors (Kentros et al., 1998); that place field locations remain consistent with each other but slowly drift in the absence of anchoring sensory cues or if the rat is consistently disoriented before each trial (Knierim et al., 1995); and that the membrane potential of place cells exhibits a ramp like depolarisation as the firing field is traversed, presumably reflecting a release from recurrent inhibition (see Fig 16.2B; Harvey et al., 2010). In addition, the notion that synaptic connections between place cells are preconfigured is consistent with the ‘preplay’ of place cell sequences before the first visit to a new environment (Dragoi and Tonegawa, 2011). The involvement of some form of path integration is also suggested by the increased influence of the boundary the rat is running from compared to the one it is running towards (O’Keefe and Burgess, 1996; Gothard et al., 1996; Redish et al., 2000). We note that the important role played by path integration in updating the representation of spatial location does not necessarily imply that the

hippocampus is required for behaviour in all tests of path integration, as other brain regions may be sufficient to support this function. For example, in Alyan and McNaughton's (1999) experiment, hippocampal lesions did not prevent rats returning to the start of an outbound path in complete darkness.

Elsewhere, recurrent networks have also been used to examine place field directionality, as modelled in a feed-forward manner by Sharp (1991). In these models (Brunel and Trullier, 1998; Kali and Dayan, 2000) place cell firing is initially derived from orientation-specific sensory input at each location (referred to as the 'local view' from that location), while the dynamics of the network are strongly dependent on the recurrent connections in CA3. As with Sharp (1991), in these models, directionally constrained exploration results in orientation-dependent place fields. If exploration is unconstrained and random, however, then Hebbian learning in the recurrent collaterals of these models will result in a continuous attractor of the sort hard-wired by Samsonovich and McNaughton (1997), and an orientation-independent place cell representation of space. One caveat to this is that the Hebbian learning must be modulated by novelty to prevent inhomogeneity in exploration causing non-uniform weight profiles and unrealistic firing patterns (Kali and Dayan, 2000). Such novelty information has been suggested as a function for the cholinergic septal inputs to the hippocampus (Hasselmo et al., 1996; see Section 16.4.4; but see Hasselmo and Fehlau, 2001). In addition, Kali and Dayan (2000) demonstrated how novelty-modulated learning could be used to control the formation of independent place cell maps in environments that are sufficiently distinct – although this may not be sufficient to capture the complexities of all data concerning remapping. Like Sharp's (1991) model, these recurrent models are also inconsistent with data showing that place cell firing in open fields is initially non-directional but can become directional as a result of experience (Markus et al., 1995; Navratilova et al., 2012a).

Continuous attractor models of grid cells

Although continuous attractor network models of place cell firing are consistent with several aspects of the experimental data, they are generally unable to account for the heterogeneous changes in place cell firing that follow various environmental manipulations (i.e. slow and partial remapping), and the conjunctive location by movement direction modulated 'shifter' cells required to update place cell firing patterns have remained elusive. Conversely, continuous attractor network models appear much better suited to account for grid cell firing patterns (McNaughton et al., 2006). As direct recurrent excitatory connections between grid cell candidate neurons in MEC are sparse, however, recurrent connectivity in most implementations is mediated by disynaptic inhibition from interneurons, which have been shown to densely innervate MEC principal neurons (Dhillon and Jones, 2000; Couey et al., 2013; Pastoll et al., 2013; Fuchs et al., 2016). Uniform excitatory input to such a network will generate one or more stable activity packets or 'bumps', and self-motion information can then be used to translate the position of this activity packet across the neural sheet in accordance with the animal's movement in the real world (Fuhs and Touretzky, 2006; Guanella et al., 2007; Burak and Fiete, 2009). As with the continuous attractor models of place cells described above, most of these models suggest that the activity bump is shifted by asymmetric interactions between grid cells in the neural sheet. This can be achieved by rate-coded input from conjunctive grid x movement direction cells, which bear some relation to the conjunctive grid

x head direction cells identified in the deeper layers of MEC (Sargolini et al., 2006). If the recurrent inhibitory input from these conjunctive cells to other cells in the network is shifted along the axis of their preferred firing direction, then their firing will shift the activity bump in the movement direction.

In the case of a single activity bump, a continuous attractor network model of grid cell firing must exhibit a twisted torus topology, such that movement of a set distance in a direction corresponding to any multiple of 60° across the neural sheet will return it to its original position, thus accounting for the hexagonal symmetry of the grid firing pattern in the real world (Guanella et al., 2007; Pastoll et al., 2013). In the case of multiple bumps, the circular weight profile dictates that the location of activity bumps on the neural sheet exhibit six-fold symmetry through circular close packing. To ensure that activity bumps smoothly appear and disappear at the edges of the neural sheet, either periodic boundary conditions are imposed (which places constraints on the dimensions of the neural sheet), or alternatively the synaptic weights (Fuhs and Touretzky, 2006) or feedforward synaptic inputs (Burak and Fiete, 2009) are smoothly modulated to zero towards the edges of the neural sheet. Importantly, population activity is constrained by the synaptic connections between neurons such that grid cell firing patterns can only ever encode a single location at any time. Hence, grid cells in the continuous attractor network effectively perform path integration, tracking the animal's location by integrating self-motion signals. Like any path integration system, this representation will accumulate noise over time, but this can be ameliorated by learned associations with sensory inputs that can 'reset' the location estimate (as with models of place and head direction firing patterns, described above, and analogous to 'pose cells' in robotics models of simultaneous localisation and mapping, e.g. Milford and Wyeth, 2008). Incorporating associations with cells encoding the location of different landmarks, formed by slow Hebbian learning, also produces shifts and deformations of the grid firing pattern consistent with those observed during navigation in familiar environments (Ocko et al., 2018).

Continuous attractor network models of grid cell firing readily account for the modular organisation of grid cells (Barry et al., 2007; Stensola et al., 2012), for the consistent offset of firing patterns from co-recorded grid cells across environments (Fyhn et al., 2007; Yoon et al., 2013), and for the apparently coherent drifts in grid firing patterns relative to the environment during active movement (Hardcastle et al., 2015; Chen et al., 2016; Perez-Escobar et al., 2016; Almog et al., 2019). In addition, subthreshold membrane potential recordings during the traversal of grid firing fields revealed a ramp depolarisation (i.e. release from inhibition) predicted by continuous attractor network models (see Fig 16.2B; Domnisoru et al., 2013; Schmidt-Hieber and Hausser, 2013). However, these models generally predict that local interneurons should exhibit grid firing patterns, for which experimental evidence is lacking (Buetfering et al., 2014; but see Solanka et al., 2015; Shipston-Sharma et al., 2016). These models also struggle to account for distortions of the grid firing pattern (e.g. Krupic et al., 2015); the rate modulation of different firing fields (e.g. Ismakov et al., 2017); and for the theta modulation and phase code for location exhibited by grid cells (Hafting et al., 2008) without relying on unrealistic subthreshold currents (Navratilova et al., 2012b; Pastoll et al., 2013), as discussed in the next section.

16.2.4 Modelling Phase Coding in Place and Grid Cells

Most of the place and grid cell models described above focus on the rate code for location exhibited by these cells, while ignoring the concurrently expressed theta phase code for location within the firing field. Nonetheless, the origin of the phase coding of place and grid cell firing with respect to the concurrent theta rhythm of the EEG (O'Keefe and Recce, 1993; Hafting et al., 2008) has been the subject of several computational models. Before considering these models, we very briefly review the relevant experimental findings (see Section 16.2.1 and Chapter 11 for more details).

The theta rhythm is a large amplitude LFP oscillation of around 6 to 10 Hz that is present whenever the rat is actively moving its head through the environment. As the rat traverses a place or grid field, the corresponding cell tends to fire spikes with a systematic phase relationship to the theta rhythm. On entering the field spikes are fired at a 'late' phase, and as the rat passes through the field, spikes are fired at successively earlier phases so that, on exiting the field, the phase of firing may have 'precessed' by up to 360 degrees (corresponding to an 'early' phase). Interestingly, the phase of firing correlates better with the location of the rat within the place or grid firing field than with other variables such as the time spent within the field, or the instantaneous firing rate of the cell (Huxter et al., 2003; Climer et al., 2013; Jeewajee et al., 2014).

An appealingly simple feed-forward model of the place cell phase code assumes that excitatory synaptic input might increase as the rat runs through the place field, while the theta rhythm might reflect a saw-tooth shaped inhibitory input (i.e. inhibition decreasing through each cycle; see Fig 16.2C; Harris et al., 2002; Mehta et al., 2000). In this model, firing phase would advance simply because the increasing excitatory input manages to overcome the inhibitory input successively earlier in each cycle. The cause of the increasing excitatory input might reflect an exaggerated form of the asymmetry reported by (Mehta et al., 1997), or an increasing then decreasing input but with lack of firing on the decreasing portion due to effects such as habituation (Harris et al., 2002). Interestingly, this subthreshold 'ramp' depolarisation appears to be consistent with whole cell recordings from place cells during navigation in virtual reality environments (Harvey et al., 2010; see Fig 16.2B). These models also capture the observation that the phase shift becomes more reliable over the first few runs of a trial, as does the asymmetry of place fields (Mehta et al., 2002; Feng et al., 2015), and allow phase to be analysed in terms of firing rate during non-translational behaviours such as dreaming or wheel running (Harris et al., 2002). However, they predict a relationship between firing rate and phase – both driven by the amplitude of subthreshold depolarisation – which does not appear to exist in empirical data (Huxter et al., 2003), and go against the finding that, while the development of asymmetry in place fields over the first few runs of a trial is prevented by blockade of NMDA receptors, the phase shift phenomenon is unaffected by this manipulation (Ekstrom et al., 2001). In addition, the correlation between phase and location is stronger than that between phase and rate, and, on the linear track at least, the weaker correlation is a side-effect of the stronger one (O'Keefe and Burgess, 2005).

As with models of place and grid cell firing, an alternative formulation stresses the role of recurrent connectivity as opposed to feed forward connections. Specifically, simulations of the Samsonovich and McNaughton (1997) model in which net activation is made to oscillate at theta frequency show something qualitatively similar to the phase shift, due to path integration

occurring within each cycle. That is, the initially active set of place cells settles to those with fields centred on the rat and then expands to include those with fields centred ahead of the rat. The first quantitative model of this phase shift was proposed by Tsodyks et al., (1996). In this model, the recurrent connections between place cells in CA3 are asymmetrically arranged so that each place cell projects to place cells further along a learned path (see also Blum and Abbott, 1996). External input to a CA3 place cell arrives at a fixed (early) phase of theta, causing place cell activity at this phase which in turn propagates through the recurrent connections to place cells with fields further along the path and causes them to fire. Overall activity is inhibited at the end of each theta cycle, preventing the further propagation of activity into the next cycle. Thus, when the rat enters a place field the corresponding cell starts to fire at a late phase due to propagated activity from cells with fields earlier on the path, and fires earlier within each cycle as the rat advances due to activity having to propagate through fewer cells, until finally firing at the early phase due solely to external inputs (see Kang and De Weese, 2019 for a related model in grid cells).

Several similar mechanisms, depending on the association of place cells firing earlier along a learned path to those firing later along it, have been proposed (Touretzky and Redish, 1996; Wallenstein and Hasselmo, 1997; Jensen and Lisman, 1996). The Jensen and Lisman (1996) model also makes interesting suggestions for the gamma rhythm, in separating the firing of cells corresponding to the current and successively further advanced locations, and for the dynamics of NMDA channels, in separating each route retrieval into successive cycles of the theta rhythm. Wallenstein and Hasselmo (1997) emphasize the role of GABA_B receptors in varying the relative influence of the inputs to CA1 from CA3 compared to those directly from EC over the theta cycle: allowing sensory (EC) input to dominate early and predictive input from CA3 to dominate late in the cycle (see also Chance, 2012). It is important to note, however, that the apparent lack of direct excitatory connections between principal cells in superficial MEC suggest that these models could not explain phase precession in grid cell firing. Instead, Navratilova et al. (2012b) suggest that after-spike dynamics could account for phase precession in an attractor network model of grid cell firing, although it is difficult to see how these ionic currents might be coordinated by movement speed on the requisite timescale (i.e. to give a faster change in firing phase during faster runs through the place or grid field).

These models do, however, produce a phase shift that is limited to 360 degrees; that is more strongly correlated with position than time; and greater for well-learned paths than random exploration. Other aspects of these models are less consistent with empirical data. Firstly, since the initial firing of a place cell depends on both the externally driven activity of other cells and its propagation through the network, it seems likely that, on a cell-by-cell and on a run-by-run basis, the initial phase of firing should be more variable than the (externally driven) final phase of firing, but this is not the case in the data (Skaggs et al., 1996; Huxter et al., 2003). Secondly, the observation that the theta phase preference of place cell firing is maintained after transient perturbation of the hippocampus suggests that phase precession might be generated by external, rather than internal, mechanisms (Zugaro et al., 2005).

A third type of model stresses the inherent oscillatory nature of some cellular processes, as did Jensen and Lisman (1996), but for different reasons. Specifically, O'Keefe and Recce (1993) pointed out that the phase and amplitude characteristics of place cell firing could be modelled as the interference pattern between an 11Hz external input to the cell (perhaps the sensory

input) and a 9Hz external or internal oscillation corresponding to LFP theta (perhaps driven by the septal input). This produces an oscillation of 10Hz corresponding to firing that shifts in phase relative to LFP theta and a 1Hz envelope, one half cycle of which corresponds to the place field. This model was subsequently extended (Lengyel et al., 2003), identifying the first input as a voltage-controlled oscillation of the membrane potential (see e.g. Hoppensteadt, 1986) in the dendrites, and the second as an inhibitory input to the soma of fixed frequency. The frequency of the dendritic oscillation was assumed to increase above that of the somatic oscillation proportionally to the strength of the dendritic input, which is assumed to be zero outside the place field and proportional to the rat's running speed within it (McNaughton et al., 1983; Ekstrom et al., 2001; Czurko et al., 1999; Huxter et al., 2003). Thus the two oscillations destructively interfere outside of the place field, while phase of firing relative to the somatic input within the field can shift more rapidly as the rat runs faster, preserving the relationship between phase and location. In addition, the dendritic oscillation must be weakly driven in anti-phase to the somatic input so that, in the absence of any dendritic input, it ensures complete destructive interference.

Corroborative evidence for interference models comes from the observation that the increase in place field size along the dorsoventral axis of the hippocampus parallels a corresponding decrease in the intrinsic firing frequency of place cells – reducing towards LFP theta frequency in more ventral regions (Maurer et al., 2005). Similarly, the dorsoventral increase in grid field size is matched by a corresponding gradient in h-current that reduces the resonant frequency of stellate cells in MEC (Giocomo et al., 2011). Issues with the original model include why only one half cycle of the interference pattern is observed, and the demonstration that somatic and dendritic oscillatory processes cannot remain independent, but quickly phase lock in real neurons (Remme et al., 2010). As discussed above (Section 16.2.2), however, it is possible that the full interference pattern is expressed by the periodic firing fields of entorhinal grid cells; and that the interference pattern is generated entirely from feed-forward inputs, rather than independent processes in a single cell (Burgess et al., 2007; Burgess, 2008; Hasselmo, 2008). This raises the possibility that theta phase precession in hippocampal place cells might be inherited from grid cell inputs (consistent with some empirical observations, e.g. Bonnevie et al., 2013; Schlesiger et al., 2015), and several models of such inheritance exist (e.g. Jaramillo et al., 2014). Another important point is that, to ensure that the moving representation of location generated by the entire hippocampus within each theta cycle is coherent, phase precession must be coordinated across multiple grid and place field scales expressed along the dorsoventral axis, wherein theta oscillations act as a travelling wave (e.g. Lubenov and Siapas, 2007; Leibold and Monsalve-Mercado, 2017). Finally, it is important to reiterate that interference models do not rely on any specific assumptions about either the frequency or stationarity of the baseline (or ‘somatic’) oscillation (which is generally equivalent to the dominant LFP frequency, see e.g. Geisler et al., 2010): phase coding of location within the firing field arises from the difference between this and the active (i.e. velocity-controlled) oscillation. Hence, the baseline frequency does not need to occupy any particular value, nor remain constant over time (Burgess, 2008; Blair et al., 2014; Orchard, 2015; Bush and Burgess, 2020).

16.2.5 Hybrid Models of Place and Grid Cell Firing

In recent years, several groups have proposed hybrid models of place and grid cell firing that incorporate contributions from both feedforward and recurrent inputs to account for a greater body of experimental data; as well as models that emphasise the interactions between complementary spatial representations provided by place and grid cells, respectively. This is consistent with empirical data which indicates that place cell firing patterns tend to be more strongly dictated by sensory inputs, and grid cell firing patterns by self-motion (Chen et al., 2019). As such, it is possible that the path integration input to place cells, which accounts for stable activity patterns when sensory cues are diminished, arises from grid cell inputs; while the sensory input to grid cells, revealed by the ‘resetting’ of accumulated error in grid cell firing patterns during periods of running away from environmental boundaries or other prominent sensory features, arises reciprocally from place cell inputs. These models, building on earlier models of place (e.g. Wan et al., 1993; Touretzky and Redish, 1996; Arleo and Gerstner, 2000) and grid (e.g. Ocko et al., 2018) cell firing, emphasise the importance of interactions between place and grid cells to establish robust representations of location (e.g. Renno-Costa and Tort, 2017; Agmon and Burak, 2020), and also account for coherent remapping in the grid and place cell populations (Fyhn et al., 2007) and the heterogeneity of grid cell in-field firing rates (Ismakov et al., 2017).

Similarly, to account for a more complete body of empirical data relating to grid cells, several hybrid models of grid firing patterns that incorporate contributions of both feedforward and recurrent inputs have been proposed (Hasselmo and Brandon, 2012; Schmidt-Hieber and Haussler, 2013; Bush and Burgess, 2014). Specifically, these models make use of continuous attractor dynamics to ensure relative stability among the firing patterns of grid cells from within the same module; and oscillatory interference to shift the activity bump. As such, path integration is performed by VCO inputs to grid cells, rather than by conjunctive cells within the grid cell network (Welday et al., 2011). This solves the problematic issue for continuous attractor network models that conjunctive cells in MEC are modulated by head, rather than movement, direction, which is not sufficient to support accurate path integration (Raudies et al., 2015). These hybrid models can therefore account for a greater body of experimental data, including both the rate and temporal firing pattern of grid cells, the relative stability of grid cell firing patterns from the same module, and the subthreshold ramp depolarisation of grid cells inside the firing field. Nonetheless, like all continuous attractor network models of grid cell firing, they predict the existence of interneurons with grid firing patterns in MEC, for which strong evidence has not yet been found (Buetfering et al., 2014; but see Solanka et al., 2015; Shipston-Sharma et al., 2016).

16.3 The Hippocampus and Spatial Navigation

In this section we consider the contribution of hippocampal spatial representations to guiding behaviour. We focus on large-scale navigation, the spatial behaviour most commonly associated with the hippocampus and medial temporal lobes (see Chapter 14). This compares to planning movements in smaller scale spaces and over shorter durations, such as visually guided reaching, which is most commonly associated with the posterior parietal lobe, see e.g. (Burgess et al., 1999). As with models of spatial representation, these models can be

approximately divided into those stressing the role of feed-forward connections, and those stressing the role of recurrent connections.

16.3.1 Spatial Navigation: Data

Behavioural data indicate that rats learn about the spatial layout of their environment during exploration in the absence of explicit goals or rewards (e.g. Tolman, 1948) and can profit from being placed at the goal location without having explored the rest of the environment (Keith and McVety, 1988). These processes are referred to as ‘latent learning’. Rats also appear to be able to perform short cuts and detours. These abilities contributed to the idea that rats form a cognitive map of their environment as opposed to simply learning to associate individual stimuli with responses, see (Tolman, 1948; O’Keefe and Nadel, 1978; but see Grieves and Dudchenko, 2013). In the framework of reinforcement learning (RL), this corresponds to a ‘model-based’ approach, whereby a model of the world is used to predict the outcomes of different actions. This world model may require extensive learning, but that can proceed in the absence of reward and subsequently support flexible planning. In contrast, ‘model-free’ RL, whereby a value function that maps actions in each state to long-term cumulative reward is learned by trial and error, may be less computationally expensive but is more rigid: when the goal or the optimal route to the goal changes, learning must begin again from scratch. Stimulus-response associations undoubtedly play an important role in spatial navigation, such as when the goal is directly visible or a well-learned turn or sequence of turns is to be performed. However, there seems to be good evidence that these types of behaviour are less dependent on the hippocampus than those associated with cognitive mapping (Morris et al., 1982; Packard and McGaugh, 1996; O’Keefe and Nadel, 1978; Doeller et al., 2008; Vikbladh et al., 2019). These issues are discussed in more detail in Chapter 14. Of note, many models of navigation combine contributions from parallel model-based and model-free strategies alongside dynamic arbitration between the (sometimes conflicting) output of each strategy, to account for more behavioural data (Guazzelli et al., 1998; Arleo and Gerstner, 2000; Chavarriga et al., 2005; Dolle et al., 2010; Geerts et al., 2020). However, we focus specifically on models of hippocampal navigation here.

The initial spur to the association of the hippocampus with a cognitive map of the rat’s environment was the discovery of place cells, whose activity is not easily described in terms of a simple response to a single stimulus (like that of concept cells in the human hippocampal formation; Quiroga et al., 2005). Indeed, some recent empirical studies have begun to demonstrate a causal role for rodent place cells in spatial behaviour (de Lavilleon et al., 2015; Robinson et al., 2020). Nonetheless, an explanatory gap remains between the properties of place cells and the properties required of a system for spatial navigation. Three features of place cell firing are particularly problematic. Firstly, information about a place in an environment (i.e. the firing of the corresponding place cells) can only be accessed locally (by actually visiting that place). Although hippocampal replay events may allow non-local place cell activity during quiescent waking or rest, it is not yet clear if these events are utilised during active navigation or can represent novel trajectories through known environments, rather than simply recapitulating previous experience (Gillespie et al., 2021; but see Gupta et al., 2010; Olafsdottir et al., 2015). An alternative solution to navigating with place cells may be provided

by ‘spatial view cells’ in the macaque hippocampus (Rolls et al., 1997) which fire as a function of where the monkey is looking rather than where it is physically located.

Secondly, there is limited evidence that place field activity is modulated by the location of the current goal more than by the location of any other cue (e.g. Speakman and O’Keefe, 1990; Hok et al., 2007; Duvelle et al., 2019) – although there is some evidence that place (Hollup et al., 2001; Lee et al., 2006; Dupret et al., 2010; Mamad et al., 2017; Kaufman et al., 2020) and grid (Boccaro et al., 2019; Butler et al., 2019) fields shift slightly towards persistently rewarded locations. That is, place cells appear to tell you where you currently are rather than where your goal is (particularly if that location is not where reward is delivered). Nonetheless, several recent empirical studies have described the modulation of place and non-place cell firing rates in the mammalian hippocampus by goal direction (e.g. Sarel et al., 2017; Kunz et al., 2021; Ormond and O’Keefe, 2021), which could be used to support navigation (by selecting a movement direction that maximises those firing rates), although the origin of this goal direction signal and manner in which it is learned are not yet clear. Similarly, place cell theta sequences (see Section 16.2.1) appear to dynamically explore different movement trajectories away from the current location (e.g. Johnson and Redish, 2007; Wikenheiser et al., 2015; Kay et al., 2020), akin to a process of deliberation or ‘vicarious trial and error’ (VTE; Redish, 2016), but their potential role in navigation has not yet been explicitly modelled.

Thirdly, the phenomenon of global remapping indicates that place cells do not provide any metric information about the relative location of their place fields, and therefore cannot be used to generalise information across environments (i.e. learning the relation between the place field locations of two cells in one environment provides no information about their relative location in a second environment). The compact and efficient representation of large-scale space offered by grid cells appear better suited to support long-range navigation by computing direct vectors between start and goal locations (Fiete et al., 2008; Bush et al., 2015; Stemmler et al., 2015) or supporting generalisation across relational structures, in a broader sense (Behrens et al., 2018). Nonetheless, these vectors may need further refinement in light of the sensory and affective properties of intermediate locations, which may be furnished by place cells or those encoding the presence of boundaries (e.g. Edvardson et al., 2019).

16.3.2 Spatial Navigation: Feed-forward Models

Zipser’s (1986) ‘view field’ model built on the observation that, in some circumstances, place cell firing is modulated by the orientation of the rat. In this model, a set of orientation dependent place cells or ‘view-field units’ become associated to a set of ‘goal units’ which encode the direction to the goal relative to the current heading direction. So long as the appropriate cells become associated, the population vector of directions represented by the goal units guide the rat to the goal, as goal units driven by place cells representing the current location will fire the most strongly. However, this model requires the direction towards the goal to be continuously maintained during initial exploration of the environment, perhaps as a path integration vector, in order for that direction to be associated with active place cells in each location. In a second (‘beta coefficient’) model, Zipser suggested that this could be avoided by calculating and storing the location of the goal relative to subsets of landmarks (as the coefficients of the linear sum of landmark locations that is equal to the goal location). In this model, learning at the goal location is sufficient to support navigation back to that location, although the neural

mechanisms required to implement the desired calculations are not explained. A similar model was proposed by Wilkie and Palfrey (1987) but, again, this model was primarily heuristic and no biologically plausible implementation of the landmark distance matching procedure was provided. More importantly, it is not clear if these models require an explicit representation of place, as they act directly on sensory input (analogous to models of insect navigation, e.g. Cartwright and Collett, 1983). Nonetheless, each of these models can account for the latent learning of goal locations, as well as behavioural search patterns following the movement or removal of prominent visual cues.

Several models have followed Zipser's view field model in associating places or local views to movements (see Trullier et al. (1997) for a wider review of biologically based artificial navigation systems). McNaughton and Nadel (1990), suggested that routes might be learned as a chain of associations from a local view to an action and thence to the next local view, and so on. This model was not actually simulated, and simply storing routes is insufficient to enable spatial navigation. Even if a given route can be correctly selected in terms of the locations to which it leads, navigational abilities such as generating novel shortcuts and detours will be beyond a simple route-based system. The task of accumulating route-independent spatial information faces several issues, including the 'credit assignment' problem: deciding which actions along a route are critical in determining whether it eventually leads to the goal.

Brown and Sharp (1995) provided a more sophisticated model for associating locations with actions (see also Sharp et al., 1996). In their model the possible actions in a place are represented by left turn and right turn cells (in nucleus accumbens) driven by each place cell. These 'turn cells' receive modifiable connections from head-direction cells (HDCs) which support the rat's spatial learning. When the rat reaches the goal, connections between head-direction and turn cells are modified according to a recency-weighted index of their simultaneous activity. Thus, if turning left in a particular place when facing north leads immediately to the goal, then the HDC representing north becomes more strongly associated to the left turn cell driven by the corresponding place cell. The recency weighting of connection modification is designed to provide an approximate solution to the credit assignment problem (when applied over many trials) by effectively dividing credit according to the number of steps within which an action leads to finding the goal. This model successfully simulates learning in the Morris water maze but does not show latent learning: performance would not be affected by whether the rat can look around from the goal location, and navigation to the goal would be strongly affected if stereotyped routes were used during learning.

A related way to think about spatial navigation is to imagine defining a surface over the environment on which gradient ascent leads to the goal, like the value function in RL (Dayan, 1991; Foster et al., 2000, see below). The simplest model of this sort has place cells connected to a goal cell via reward-modulated Hebb-modifiable connections such that encountering a goal causes the strengthening of its input connections from concurrently active place cells, see (Burgess and O'Keefe, 1996; Fig. 16.6A). The activity of the goal cell will subsequently increase with proximity to the goal since the net activity of those place cells with strengthened connections will increase with the proximity to the goal. The task for the rat is then to move in the direction that increases the firing rate of the cell representing the desired goal (Fig. 16.6B). This type of model qualitatively captures the rapid nature of learning a goal location once place cell firing has become established and the ability to learn simply by being at the goal location

as opposed to having to find it many times. However, finding the goal would involve the rat hunting around to determine the best direction in which to move. This VTE behaviour is often observed at choice points but is less common in the open field (Redish, 2016). A second problem raised by this model is the range over which spatial information is accessible. If there are no place cells that fire at both the goal location and the current location of the rat, then there will be no gradient in the firing rate of the goal cell (being locally zero). Hence, this type of model requires the place cell population to include some firing fields that have non-zero firing rates at any two points in the environment, however far apart. Although larger place fields have been observed in ventral hippocampus (e.g. Jung et al., 1994; Kjelstrup et al., 2008), this places strong constraints on the spatial range over which goal directed navigation might be feasible.

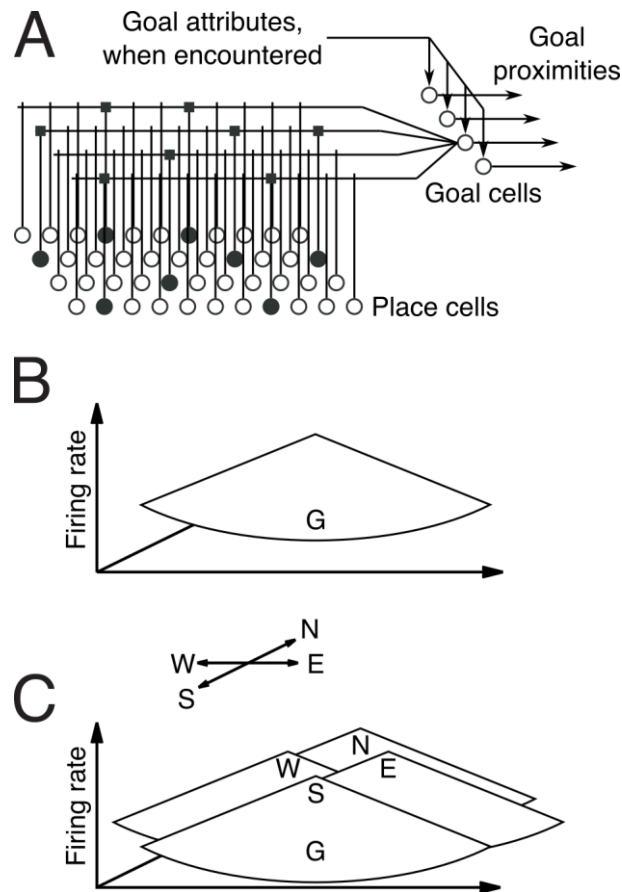


Figure 16.6 Simple model of navigation with place cell. **A:** A ‘goal’ cell stores a goal’s location by taking a snapshot of place cell activity via Hebbian synaptic modification when the goal cell is excited by the attributes of a particular goal location. Solid circles are active place cells; open circles are inactive place cells and solid squares mark potentiated synapses between place cell axons and goal cell dendrites. **B:** The firing rate map of the goal cell (roughly an inverted cone) during subsequent movements of the rat codes for the proximity of the goal (G). **C:** The firing rate maps of 4 goal cells whose population vector codes for the goal location (G). Each is associated with the allocentric direction u_i in which the location of its peak firing rate is displaced from G; thus the vector sum of the directions u_i weighted by the instantaneous firing rates f_i of the goal cells (i.e. $\sum_i f_i u_i / \sum_i f_i$) codes for the direction of the rat from the goal, and the net firing rate of the goal cells (i.e. $\sum_i f_i$) codes for the goal’s proximity. Adapted from (Burgess and O’Keefe, 1996).

Burgess et al. (1994) proposed a biologically inspired navigation model that aimed to address both issues. First, to calculate the direction to the goal from any location after a single visit to

that goal, this model made use of the empirical observation that place cells which fire at a late phase of the theta rhythm have fields peaked ahead of the rat (see Section 16.3.5). Specifically, the model posits a set of goal cells that are each associated with a different head-direction, and assumes that synaptic connections between place cells and each directional goal cell are only modified at late phases of the theta cycle (consistent with observations of theta phase dependent synaptic plasticity, e.g. Pavlides et al., 1988; Huerta and Lisman, 1995; Holscher et al., 1997). Hence, when the rat is at the goal and facing north, connections are formed between place cells slightly north of the goal (i.e. ahead of the rat) and the 'north' goal cell; and turning to face each direction while located at the goal allows connections between place cells offset in each different direction from the goal and the corresponding goal cell to be learned (Fig. 16.6C). This produces a set of directional goal cells with firing fields that are offset in the corresponding direction from the goal, such that the population vector of goal cell firing rates at any location within their firing fields indicates the direction of the goal. In addition, different sets of directional goal cells can be used to encode the location of (and thus support navigation to) different salient locations. Second, the range over which spatial information is available from place cell firing was increased by interposing a set of subicular cells between the place and goal cells. Weaker inhibitory competition between subicular cells coupled with competitive learning in the inputs from place cells during exploration causes the subicular cells to build up larger firing fields, each one effectively composed of several place fields. This learning is goal independent and corresponds to latent learning in preparing the ground for effective one-shot learning of the location of any goals, should they be encountered.

As described by Foster et al. (2000), another way to ameliorate the issue of navigational range being limited to the diameter of the largest place field is to make use of a temporal difference (TD) learning rule (Sutton and Barto, 1988). Under this model-free RL formulation, place cells provide a representation of the task 'state' (i.e. the rat's current location; Dayan, 1991) and its value reflects the expected number of steps needed to reach the goal (one unit of reward being received on reaching the goal). TD learning can be implemented by connecting place cells to 'actor' and 'critic' units with connections that are adjusted according to a modified Hebbian rule (see Arleo and Gerstner, 2000 for an alternative implementation that makes use of the Q-learning algorithm). The activation of the critic unit is equal to the expected future reward from the current state, discounted by distance into the future (see e.g. Dayan and Abbott, 2002), a more principled analogue of the simple goal cell formulation above. The set of actor units, only one of which can be active at a given time, represent movement in different directions. At each step the connection weight from a place cell to the critic unit or to the active actor unit is adjusted by an amount proportional to the product of the place cell's firing rate and the amount by which the reward exceeds that expected from the change in the activity of the critic unit. This type of learning with reward prediction errors is consistent with a role for dopaminergic modulation of LTP (e.g. Montague et al., 1996; Schultz et al., 1997). Over many routes to the goal, ideally involving performing all actions at all locations many times, this rule causes both the critic to provide an accurate estimation of value, and the appropriate actions to be associated with each state (see Fig. 16.7). For a given environmental configuration and goal location, this can provide the optimal strategy, which is not the case with the approximate recency weighting implemented by Brown and Sharp (1995) above (and Blum and Abbott, 1996, below) or, if obstacles are present, with goal cells simply indicating the physical proximity of the goal (Burgess et al., 1994). As mentioned earlier, however, learning is both experience-dependent

and goal-dependent: knowing to head north from a given place results from this route having previously led to the goal, and there is no transfer of learning when the goal is moved. Like the Brown and Sharp (1995) model, therefore, navigation can be strongly influenced by taking stereotyped routes during learning and is inflexible to changes in reward location.

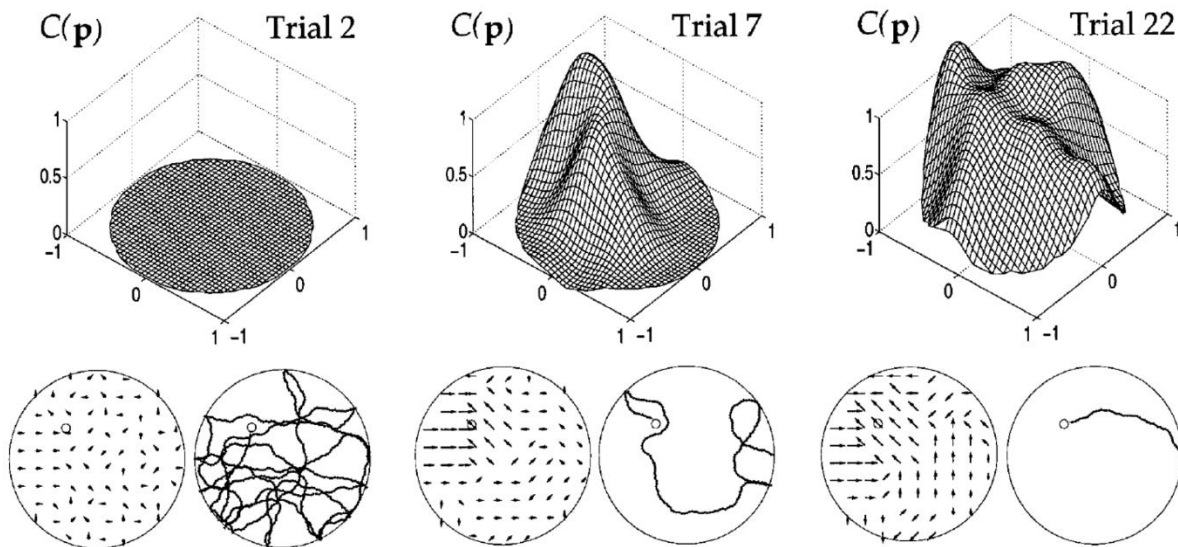


Figure 16.7 Learning in the actor-critic system in a water maze. The plots for each trial show the critic's value function $C(p)$ (above), the preferred actions at various locations (below left, the length of each arrow is related to the probability that the particular action shown is taken by a logarithmic scale) and a sample path (below right). Trial 2: After a timed-out first trial, the critic's value function remains zero everywhere, the actions point randomly in different directions, and a long and tortuous path is taken to the platform. Trial 7: The critic's value function having peaked in the northeast quadrant of the pool, the preferred actions are correct for locations close to the platform, but not for locations further away. Trial 22: The critic's value function has spread across the whole pool and the preferred actions are close to correct in most locations, and so the actor takes a direct route to the platform. Adapted from (Foster et al., 2000).

To simulate goal independent learning over many trials in which the location of the goal changes, Foster et al. (2000) proposed that a second system learns to form a coordinate representation of the rat's position by using the rat's locally accurate ability to estimate self-motion (potentially foreshadowing the computational benefit and generative mechanisms of grid cell firing patterns, see Sections 16.3.4 and 16.2, respectively). The place cells are connected to two units which learn to estimate the x and y coordinate of the rat, again using TD learning to adjust connection weights. In this system, the change in a connection weight to the x unit is proportional to the place cell activation times the amount by which the change in x , as estimated by self-motion, exceeds that estimated by the change in the activity of the x unit. The explicit representation of x and y coordinates enables accurate navigation after one exposure to the goal and so corresponds well with latent learning. No neural implementation of this vector navigation procedure was described, however.

As described by Dayan (1993), an alternative approach to making TD learning more flexible and support more efficient RL is to decompose the learned value function (i.e. the discounted sum of expected rewards at future states, given a specific action or policy) into a separate reward function and discounted expectancy of occupying each future state, or 'successor

representation' (see Section 16.5.2). It has been hypothesised that the SR is the true variable encoded by place cell activity (Stachenfeld et al., 2017; see also Gustafson and Daw, 2011). Under this formulation, place cells do not encode current location per se, but a predictive representation of future locations given the current location, which is learned as a function of the navigational history of the animal. The SR combines some advantages of both model-free and model-based approaches to navigation: it can be used to efficiently identify the optimal trajectory from current to goal locations, like model-free RL; but is flexible to changes in the goal location, like model-based RL, provided that the transition structure of the environment and policy remain unchanged. In addition, the notion that place cells may encode the SR accounts for various empirical observations, such as the increase in place field asymmetry during repeated trajectories in the same direction of motion (Mehta et al., 2000); the local remapping of place fields close to a novel barrier in a linear maze (Alvernhe et al., 2011); and the skewing of place fields towards a repeatedly visited reward location (Markus et al., 1995; Hollup et al., 2001; see Stachenfeld et al., 2017 for further details). The SR is also consistent with some aspects of latent learning, as it can be learned during exploration of an environment prior to the introduction of any reward; and recent analyses suggest that some aspects of human and rodent navigation behaviour can be best predicted by an SR formulation (de Cothi et al., 2021). Nonetheless, like many of the place cell navigation models described above, navigation using the SR is limited to the diameter of the largest place field (beyond which the product of the reward function and SR will be zero); and will be biased by stereotyped movement patterns during exploration (which dictate the structure of the SR).

16.3.3 Spatial Navigation: Feed-back Models

As well as being linked with spatial representation and associative memory (see Box 16.3), the CA3 recurrent collaterals have been proposed to play a role in spatial navigation. The first model to formalize such a role was suggested by Muller et al., (1991; 1996) and focused on the Hebb-associative effects of LTP on the recurrent collaterals. If pre- and post-synaptic firing within a short time interval leads to a small increase in synaptic strength, then the firing of place cells as the rat moves around an environment will lead to the strength of a connection between two place cells depending on the proximity of their place fields. This occurs simply because the greater the overlap between place fields the more often they will fire near-coincidentally during random exploration. Muller et al., (1996) show that, after extensive exploration, the synaptic strengths represent a 'cognitive graph': each approximately representing the minimum path-length between the centres of the place fields of the cells it connects. Their model proposes that the rat navigates by moving through the place fields of the cells most strongly connected to the cells with fields at the current and destination positions. This mechanism, reminiscent of a resistive grid (Connelly et al., 1990), works well but relies on a graph search. It is not easy to imagine how such a process could be implemented in a biologically plausible fashion given the apparent lack of influence of the goal location on the firing of place cells. One early suggestion was that activation corresponding to the goal location occurs in entorhinal cortex while activation corresponding to the current location occurs in CA3, and activation in each region spreads along the available paths (although more slowly in CA3 than entorhinal cortex) until a commonly activated location is detected in CA1 (Gorchetnikov and Hasselmo, 2002). This location represents the next immediate destination for the rat, although how this information is interpreted in terms of whether to turn left or right

is not described. More recent formulations have suggested that activation may spread backwards from the goal location through cortical columns in prefrontal cortex that represent each action taken and state visited, thus providing the shortest sequence of actions needed to reach the goal (Hasselmo et al., 2005; see also Martinet et al., 2011).

In a related model, Blum and Abbott (1996) make use of the temporal asymmetry of LTP to strengthen recurrent connections from CA3 place cells that fire early on the rat's path to those that fire later along it (see also Gerstner and Abbott, 1997). This causes the activation of place cells to spread backwards along the path, see also (Mehta et al., 1997; Mehta et al., 2000) and models of spatial representation above. If the firing of place cells is interpreted by systems downstream of CA3 as representing the location of the rat, this shift in firing (backwards along the path) would be interpreted as a shift in the location of the rat forward along the path. Blum and Abbott suggested that navigation along a previously performed route could be performed by moving from the current location (e.g. read from CA1) to the shifted location represented in CA3, although how this could be implemented was not described. To enable navigation to a goal location the rules for synaptic change were modified to be proportional to the amount of pre- and post-synaptic activation weighted by how recently it occurred prior encountering the goal, i.e. a similar modification to that suggested by Brown and Sharp (1995), above.

It is interesting to note that the symmetric pattern of connection strengths learned in Muller et al.'s model resembles that used in continuous attractor models of place cell firing, and so also serves to produce a consistent pattern of activity to represent each location. By contrast, the additional asymmetry in connection strengths learned in Blum and Abbott's model serve to shift the represented location along the learned route. Indeed this latter property has been shown to allow the represented location to move smoothly along the route over time, suggested as a model for mental replay during sleep (Redish and Touretzky, 1998; Bush et al., 2012). The goal-independent encoding of spatial proximity (as in the symmetric connections of Muller et al.'s cognitive graph) or of previously travelled routes (as in the asymmetric connections of the later models) correspond to latent learning. However, many visits are required to learn the synaptic weights that will incorporate any new goal location, which probably falls short of a rat's abilities. A further drawback associated with these models is that none make clear the details of how the rat's brain might deduce the direction it should move in, or if it would be able to generate a short-cut or detour. They also assume place fields of fixed relative location but might still make some interesting predictions regarding the locus of search in environments that had changed in shape or size. To build up a true distance metric in complex environments would take a long time, in common with the RL approaches (see Foster et al., 2000; Section 16.3.2).

These models can be viewed in the context of a more general set of higher-level algorithms for navigation based on directed graphs. Lieblein and Arbib (1982) describe a 'world graph' in which locations are represented as nodes connected by (asymmetric) edges that represent the movement necessary to get from one node to the next. In terms of the navigation of autonomous agents, Scholkopf and Mallot (1995) describe a 'view graph' in which the local view or sensory perception at given locations form the nodes, and the actions required to get from one view to the next form the edges (see also McNaughton and Nadel (1990) for a description of how a view graph might be implemented in the hippocampus). Mallot and Gillner (2000) argue that such view graphs are a good model for human navigation despite their simplicity. One key

requirement for building a world graph is to be able to decide whether to assign a new node to a location. This can be dictated on the basis of its familiarity, see (Touretzky and Redish, 1996; Kali and Dayan, 2000) for models relating to this, or possibly on the basis of the sequence of actions that lead back to a location. Lieblisch and Arbib (1982) further suggest that the nodes of a world-graph might also represent a location's motivational valence and thus become a general model for goal-directed behaviour even though its creation might correspond to latent learning. Finally, George et al. (2021) describe how 'clone structured cognitive graphs', in which nodes that correspond to a specific constellation of sensory inputs are cloned to disambiguate the recent history of inputs that led to that state, can account for a wide variety of data regarding hippocampal spatial representations and navigation behaviour. Specifically, by separating behavioural states that are associated with equivalent sensory inputs but differ in their temporal context, the directed graphs that are learned by this framework during exploration can account for the existence of 'splitter cells' (Frank et al., 2000; Wood et al., 2000; Grieves et al., 2016) and various remapping phenomena.

16.3.4 Models of Navigation with Grid Cells

In contrast to place cells, grid cells exhibit several properties that naturally afford large-scale vector navigation and address many of the issues relating to place cell navigation models described above. The regular periodic firing patterns of grid cells potentially provide a compact code that resembles a residue number system, encoding locations over a very large range that approaches the lowest common multiple of the spatial scales of all grid modules (Gorchetchnikov and Grossberg, 2007; Fiete et al., 2008; Sreenivasan and Fiete, 2011; Mathis et al., 2012). Importantly, grid cells generally fire in all environments visited by an animal and do not exhibit global remapping but tend to maintain a constant phase relationship across all environments (Fyhn et al., 2007; Yoon et al., 2013). Hence, the periodic firing patterns of grid cells appear to provide a framework with which to infer the vector between two locations, even when those locations are much farther apart than the largest grid scale and the intervening space has not been explored (Erdem and Hasselmo, 2012; Kubie and Fenton, 2012; Bush et al., 2015). The grid cell population effectively provides an efficient coordinate system for large scale space, which should theoretically allow the direction and distance between any two previously visited points (at which the grid cell population activity is known) to be computed (Stemmler et al., 2015; Behrens et al., 2018). Consistent with this view, units with grid-like firing patterns emerge in some recurrent networks trained to perform vector navigation (Banino et al., 2018; Cuevas and Wei, 2018).

Several possible methods of computing navigational vectors from grid cell population activity using realistic neural networks have been proposed. One possible solution is to perform 'linear look ahead' by propagating sweeps of activity sequentially through the grid cell population in different directions, beginning at the start location. The time taken for grid cells encoding the goal location to become active, or the activity of a neuron that integrates the total output of the grid cell population during that time, subsequently provides an indication of the distance to the goal in that direction which can be combined across any two non-collinear axes to produce a direct movement vector (Erdem and Hasselmo, 2012; Kubie and Fenton, 2012; Bush et al., 2015). Such activity sweeps are consistent with observations of grid cell 'replay' in MEC (Olafsdottir et al., 2016; O'Neill et al., 2017), and this model has the advantage of requiring no

additional circuitry or mechanisms to solve the navigation problem, beyond those already in place to update the grid representation and its association with environmental sensory-driven place cell representations during exploration. In addition, this model predicts that more time and greater metabolic activity within the hippocampal formation should be associated with the construction of longer vectors, consistent with some experimental data (Kosslyn et al., 1978; Sherrill et al., 2013; Howard et al., 2014a).

Alternatively, the distance between different locations encoded in grid cell population activity can be directly linearly decoded using a ‘distance cell’ model, analogous to neural network models of the mental number line (Dehaene, 1997). Briefly, separate arrays of distance cells code for each direction of travel along two non-collinear axes and receive input from grid cells in each module with synaptic weights proportional to their mean firing rate at that location on that axis. Winner-take-all dynamics ensure that only a single distance cell in each array is active, and all distance cells provide input to a readout neuron with synaptic weights that increase in strength linearly with distance along the axis. Hence, the firing rate of that readout neuron signals the distance from the origin to that location along that directional axis, and the combined output of readout neurons stimulated, via the distance cells, by grid cell population activity encoding the start and goal location is then directly proportional to the distance between those locations along that directional axis (Bush et al., 2015). Because distance and direction are decoded linearly, the distance cell model is rapid, and predicts no scaling of computational time or effort when decoding increasingly large vectors. However, it does require a significant amount of additional neural circuitry, and it is not clear how the fine-tuned weights of those connections may be learned during development or exploration.

In sum, although the properties of grid cell firing fields appear to be perfectly suited to support vector navigation in large-scale space, and despite behavioural evidence that navigation in real (Chen et al., 2015) and virtual (Bellmund et al., 2020) environments is consistent with grid firing patterns, the exact mechanism by which this might be achieved has yet to be elucidated. Of note, this mechanism might not be restricted to support the navigation of an agent through the world – it has also been proposed to support the planning of saccades during visual exploration (Bicanski et al., 2019), consistent with reports of grid-like responses in MEC during visual search (e.g. Killian et al., 2012; Nau et al., 2018; Julian et al., 2018). Indeed, it has been suggested that grid cells might be used throughout neocortex to encode the location of sensory features in different modalities using a consistent coordinate system (Hawkins et al., 2019). In spatial navigation, however, it seems likely that vector navigation would have to be combined with some other method for evaluating the sensory and affective properties of intervening locations during the planning of feasible routes through cluttered environments that prevent movement in a straight line between current and goal locations. One such method could involve generating potential trajectories by the firing of place cells influenced by grid vectors and the presence of boundaries (Edvardsen et al., 2019). This mechanism would be consistent with the recent observation that the replay of goal-directed place cell sequences circumnavigates the current configuration of barriers (Widłowski and Foster, 2022). Overall, several plausible computational mechanisms have been proposed that utilise place cells to label specific locations and grid cells to encode the spatial relationship between those locations, but no one specific mechanism has so far have been conclusively endorsed by experimental data.

16.4 The Hippocampus and Associative Memory

In contrast to the vast amount of animal work linking hippocampal damage to deficits in spatial cognition, the major impairment noted in humans following bilateral damage to the hippocampus is amnesia: a much more general impairment in memory. The extent of this impairment into various subdivisions of memory and into information acquired prior to the damage is a contentious issue. Here we briefly review the data on human hippocampal function in memory, introduce the canonical ‘complementary learning systems’ model (McClelland et al., 1995) derived from Marr’s seminal paper in 1971 and discuss various developments made to this framework over the years.

16.4.1 Hippocampus and Memory: Data

Substantial bilateral damage to the hippocampus and medial temporal lobes almost invariably leads to amnesia, characterized as a drop in the memory component of the intelligence quotient (MIQ) of at least 20 points relative to full-scale IQ. Since only a relatively small number of cases of damage restricted to the hippocampus have been studied (e.g. Kartsounis et al., 1995), it is difficult to draw general conclusions regarding its role in memory as opposed to the roles of surrounding cortical areas. However, some general points can be made (see Spiers et al., 2001; Squire et al., 2004; Squire and Wixted, 2011 and Chapter 13 for details). These include a ubiquitous deficit in long-term memory for personally experienced events that occur after the lesion (i.e. an ‘anterograde’ deficit in ‘episodic’ memory) alongside spared procedural and working memory. The extent of retrograde amnesia (loss of memory for information acquired prior to the lesion) appears to vary across patients and possibly across types of information (Nadel and Moscovitch, 1997; Winocur et al., 2010). Memory loss can extend over the entire lifetime or be restricted to shorter periods prior to the damage but does seem to be relatively limited in the case of lesions to the fornix (Aggleton and Brown, 1999).

More controversial findings include the relative sparing of semantic memory (memory for facts) and of familiarity-based recognition in some cases of focal hippocampal damage (Vargha-Kardem et al., 1997; Manns et al., 2003). A relative sparing of recognition memory is consistent with findings in monkeys showing that this type of memory is more strongly dependent on nearby cortical areas (e.g. Zola-Morgan et al., 1994; Baxter and Murray, 2001; Gaffan, 1994; Aggleton and Brown, 1999). There is also some evidence for specific impairments in short-term memory for the relations among co-occurring items or features (e.g. Hannula et al., 2006; Olson et al., 2006; Pertzov et al., 2013) and the topography of visual scenes (Lee et al., 2006; Hartley et al., 2007); while in animals, hippocampal lesions are associated with deficits in sequence learning (Fortin et al., 2002) and trace conditioning, in which a response must be made at a fixed delay after the disappearance of a cue (Solomon et al., 1986). Finally, it should be noted that the human hippocampus, particularly in the right hemisphere, is also involved in spatial navigation (Burgess et al., 2002) and model-based decision making (Vikbladh et al., 2019); and that intact hippocampal function appears to be required for imagining new experiences (i.e. ‘scene construction’ or ‘episodic future thinking’; Atance and O’Neill, 2001; Hassabis et al., 2007; Schacter et al., 2012; see Chapter 13).

16.4.2 Marr's (1971) Hippocampo-neocortical Model of Long-term Memory

Much of the modelling work on the role of the hippocampus in episodic or associative memory can be considered part of a long tradition reaching back to Marr (1971). In this section we sketch the main components that provide a common framework for subsequent models, and indicate how these components correspond to various aspects of the data on human memory (see also Willshaw and Buckingham, 1990; Burgess et al., 2001a; Norman et al., 2008). We refer to this as the canonical hippocampo-neocortical model (see Fig. 16.8) and note its extension as ‘complementary learning systems theory’ following (McClelland et al., 1995). We also note the related recent focus on the potential contribution of episodic memory to efficient RL, which is discussed in Section 16.5.4.

In Marr's (1971) model, events in the outside world are represented by patterns of activity in neocortical areas. The role of the hippocampus is to store these representations over the short term so that relevant events can be categorized and stored for the long term in neocortex (see Marr's (1970) model of cerebral neocortex). This is achieved by mapping the neocortical representation of an event into a ‘simple representation’ in hippocampus, with modifiable connections to and from the hippocampus storing the mappings between those (full and simple) representations. The CA3 recurrent collaterals are also modified to store the simple representation as an associative memory so that, if it is incompletely activated, the ‘collateral effect’ will result in the full representation being recovered via a process of ‘pattern completion’. Thus, partial activation of the neocortical representation of an event can lead to complete reactivation of its simple representation in the hippocampus, which can in turn reactivate the entire neocortical representation (e.g. Horner et al., 2015).

In Marr's view, the capacity of the hippocampal system should be enough to store a day's events so that the process of categorization and long-term storage in neocortex can take place during the following night's sleep. We note that this now seems at odds with the much larger extent of retrograde amnesia following hippocampal lesions, and the possibility that some types of information (e.g. episodic as opposed to semantic) remain forever dependent on the hippocampus (Nadel and Moscovitch, 1997; Winocur et al., 2010; see Section 16.4.5). Marr further suggests that the simple representations need only reflect those parts of the event through which it will be addressed, and that they should be sparsely encoded to reduce possible interference between representations – i.e. undergo a process of ‘pattern separation’. The sparsity of a representation refers to the fraction of neurons that are active: if this is very low, the chance of the same neuron being active in the representation of different events is small and interference between representations is minimized.

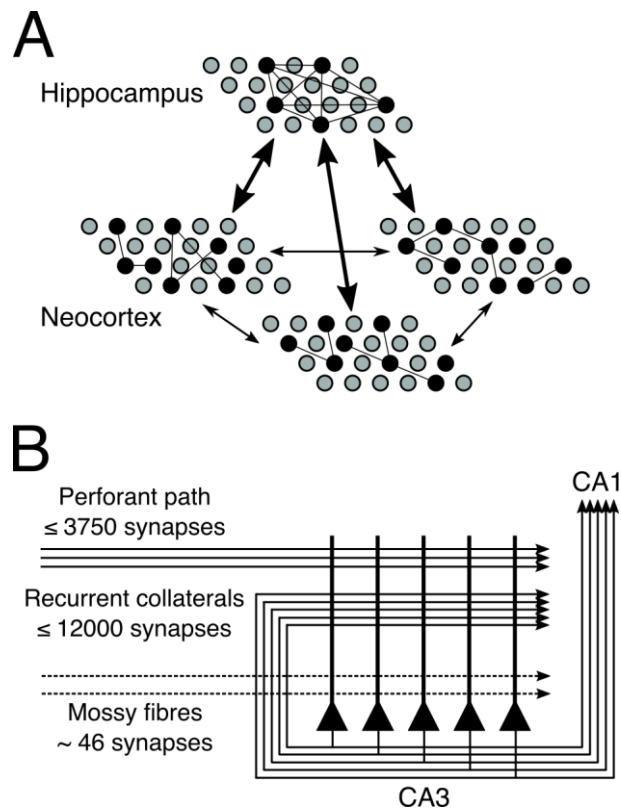


Figure 16.8 The canonical hippocampo-neocortical model of long-term memory. **A:** Illustration of the model (strong connections and active cells shown in black). Relatively dense recurrent connections and sparse representations in the hippocampus enable efficient pattern completion. Connections between neocortex and hippocampus allow the hippocampal representation of an event to be associated with its sensory details, including reactivation of the representations in different neocortical areas dealing with different sensory modalities. Abstracted semantic representations may also be learned over time in neocortex. The recurrent connections within each neocortical area allow uni-modal recognition. Adapted from (Burgess et al., 2001a). **B:** Anatomy of the inputs to CA3 pyramidal cells, showing the approximate number of synapses in the rat. Adapted from (Treves and Rolls, 1992).

16.4.3 Associative Memory and the Hippocampus

Much of the development and analysis of this canonical model has followed from research on the associative properties of feed-forward (e.g. Willshaw et al., 1969) and recurrent networks (Kohonen, 1972; Gardner-Medwin, 1976; Hopfield, 1982) based on Hebbian learning (see Fig. 16.9). Initial developments concentrated on matching the major anatomical properties of the hippocampus (see Chapter 3 and Fig. 16.8B) with constraints on the representations and learning mechanisms indicated by a functional analysis of associative memory (see Fig. 16.10). The first major attempt of this sort (McNaughton and Morris, 1987) highlighted the potential contribution of the dentate gyrus (DG).

Specifically, the much larger number of projection cells in the DG (around one million granule cells in the rat) than in either entorhinal cortex (EC, around 200,000 layer II cells project into the hippocampus) or region CA3 (around 300,000 pyramidal cells) indicate that it could be used for ‘pattern separation’ (Amaral et al., 1990). This means that distinct (i.e. non-overlapping or orthogonal) patterns of activation will be created in DG despite similarity in patterns of EC activation representing similar events. Thus pattern completion in CA3 will not

lead to a novel event causing retrieval of the representation of similar, familiar events, which would lead to interference between old and new memories. Of course, one cannot have perfect pattern completion *and* perfect pattern separation of incoming patterns of activity - there is a balance between recognising new input as a noisy version of a stored pattern versus a new pattern to be stored in its own right. In this respect the nature of the overlapping and distinct content may be as important as its absolute similarity (see Section 16.4.4).

Second, the specific nature of the various synaptic inputs to CA3 pyramidal cells suggest different functions. The input from DG comes from a small number (around 46) of very large synapses proximal to the soma. A much larger number of connections are received further up the dendrites from within CA3 (up to 12000) and on the distal apical dendrites from EC (up to 3750). It was suggested that the powerful input from DG (via ‘detonator synapses’) serve to impose a new pattern of activity to be learned in CA3 in the face of interference due to feedback via the recurrent connections, which will tend to cause the system to return to a previously stored pattern of activity (McNaughton and Morris, 1987). Once the representation of a new event has been imposed, Hebbian modification of both the recurrent connections and the connections from EC can occur. The large number of recurrent synapses per cell allow for a large auto-associative memory capacity, while the large number of synapses in the input from EC allow for a large hetero-associative memory capacity in associating the EC representation to the CA3 representation (Treves and Rolls, 1992; see Fig. 16.8B).

Third, the requirements of Hebbian learning in the CA3 recurrent collaterals and the inputs from EC to CA3, but not those from DG, are consistent with the physiology of these various connections. The synapses in the former two pathways are thought to be capable of NMDA receptor dependent LTP, while the mossy fibre connections from DG show only non-Hebbian modification (see Chapter 10). Equally, the divisive normalisation required by associative networks (Willshaw et al., 1969; see Box 16.3) is consistent with the action of interneurons providing inhibition by opening ion channels near to the soma to shunt input current in the dendrites (see Fig. 16.9). Further analysis of auto-associative memory indicates that ‘progressive recall’ improves performance (Gardner-Medwin, 1976). Under this model, inhibition is slowly reduced during retrieval so that the first few cells that become active are the most likely to be correct and feedback from their activation decreases the chances of subsequent erroneous activation. Such periodic fluctuation of inhibition (or equivalently, the firing threshold) may provide a functional interpretation for the theta rhythm (see also Sections 16.4.4 and 16.4.6). Finally, the canonical model has been elaborated to include separate input and output representations in the entorhinal cortex (in the superficial and deep layers, respectively). Many of these ideas have been reviewed or developed further in (McClelland et al., 1995; Amit, 1989; Rolls and Treves, 1997; Hasselmo and McClelland, 1999; Redish, 1999; Rolls and Kesner, 2006).

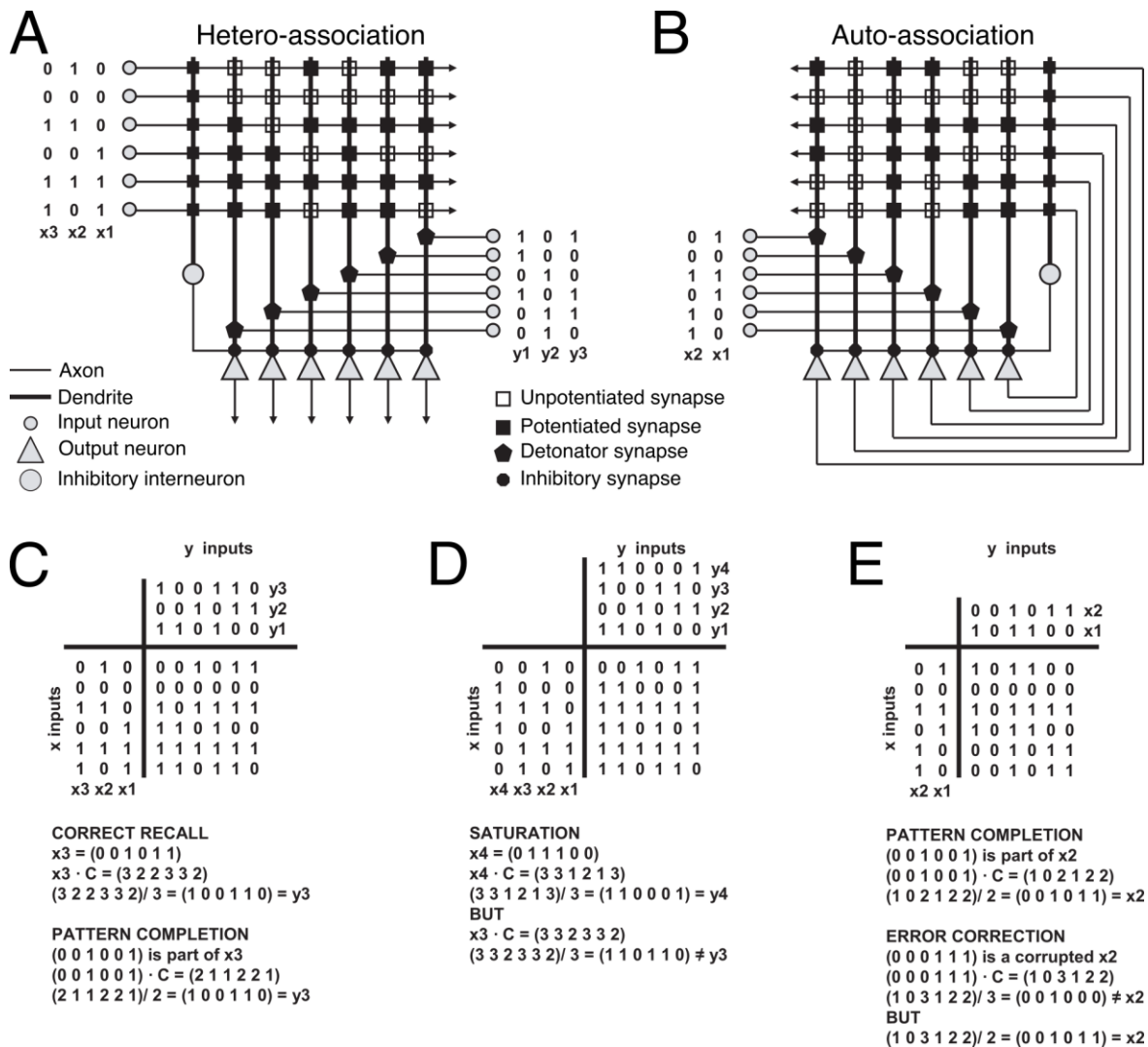


Figure 16.9 The biological implementation of associative memory in the hippocampus. **A:** A hetero-associative network associates pattern of activity Y1 with input X1, and Y2 with X2 etc, to form an associative memory (the matrix of connection weights shown in A is the result of successive presentations of input pattern X1 and output pattern Y1, X2 and Y2, X3 and Y3 to an initially blank matrix). This enables input X1 to reproduce activation Y1. Even an incomplete or corrupted version of X1 can reproduce Y1, see C. Storing too much information leads to ‘saturation’ of the system and retrieval failures, see D. **B:** An auto-associative network associates pattern X1 with itself, and X2 with itself etc, to form an auto-associative memory. This enables an incomplete or corrupted pattern to be denoised, see E. The essential functional components of these networks are: a set of powerful ‘detonator’ synapses that can impose the pattern of activity to be stored; a set of extensively connected inputs with modifiable synapses; and a set of inhibitory interneurons whose role is to set a divisive threshold for the activity of the cells. **C-E:** Details of associative memory using correlation matrix formalism (Willshaw et al., 1969; Kohonen, 1972). Pairs of binary vectors (e.g., X1, Y1) are presented to the system for storage. A Hebbian learning rule is used: a synaptic connection is strengthened (set to 1 from 0) given pre- and post-synaptic activity (i.e. both set to 1). Since the net input to a cell is the sum of inputs multiplied by the strength of the connection mediating it, the net input to a cell corresponds to the number of active inputs arriving via strengthened connections. To fire, the net input to a cell must equal the number of currently active inputs. Thus retrieval of vector given its corresponding paired associate is achieved by matrix-vector multiplication (e.g. pattern Y3 in A is extracted by multiplying the matrix rows by corresponding elements of vector X3 and summing the columns) followed by integer division by the number of active bits in the input vector. Provided not too many patterns have been stored, any unique subset of an X vector will recall the correct Y vector. The auto-associative case (B and E) work as the hetero-associative cases (A,C,D) with output Y1= input X1, Y2=X2, etc. Adapted from (McNaughton and Nadel, 1990).

16.4.4 The Hippocampal Representation and Novelty

The considerations of sparsity and pattern separation regarding the hippocampal representation of an event raise the issue of how these representations relate to the various elements of its content and context. First note that the conflicting processes of pattern separation and pattern completion serve to define the similarity space of retrieval, i.e. which dimensions a retrieval cue can vary along but still reinstate the event representation and which dimensions serve to discriminate different events. In the limit of complete orthogonalisation in the dentate gyrus, the hippocampal representations are independent of the details of the events and can be thought of as simply an index for them (Teyler and DiScenna, 1986). However, as the hippocampal representation must initially be activated by inputs from neocortex the overall memory system is still ‘content addressable’ and its behaviour (e.g., pattern separation or pattern completion) will depend on how different aspects of an event and its context contribute to the activation of its hippocampal representation. This relates to Marr’s observation that the hippocampal representation should include only those aspects of an event used for its retrieval.

In a simple associative memory, all elements of the representation of an event are equally associated with all other elements. As implied by Marr, however, it seems plausible that some aspects of an event are better able to cue associative retrieval than others, while other aspects of an event can be associatively retrieved more easily than others. Thus the ‘simple representations’ envisaged for the hippocampus would reflect some aspects more than others. A related suggestion is that the hippocampal representation reflects efficient compression of the neocortical representations – extracting the distinguishing features of each event (Gluck and Myers, 1996). For example, the name of somebody you only met once is often a good cue to recalling the meeting but can be hard to retrieve, while the location of the meeting is often both a good cue and relatively easy to retrieve. Similarly, the sequential position of an item in a list is easier to use as a cue than it is to retrieve (Jones, 1976). Thus even the simplest associative model of memory should include asymmetric associations between the elements of an event.

The nature of the hippocampal representation has also been elucidated by recordings of place cells in rodents and concept cells in humans (Quiroga et al., 2005; 2012). These data suggest that specific patterns of neural activity might consistently encode different elements of an event - a specific population of concept cells when perceiving Jennifer Aniston (a well-known US movie star), and a specific population of place cells when one occupies a specific location, for example. These ‘invariant’ responses to elements that may appear in multiple events challenge the idea that hippocampal representations of similar events are completely orthogonalized. However, it is also clear that place cells exhibit remapping (i.e. hippocampal representations change) when individual locations become associated with different task contingencies (e.g. Markus et al., 1995; Frank et al., 2000; Wood et al., 2000; Grieves et al., 2016). Whether the same is true for concept cells in the human hippocampus has yet to be established. As such, pattern separation may be driven by necessity – when new events must be dissociated from previous experiences that share overlapping features to support different behavioural outputs, consistent with the notion that the hippocampus encodes task relevant latent variables (see Section 16.5.1).

Consideration of the different requirements of encoding and retrieval also raises some interesting questions. Notwithstanding the suggestion that ‘detonator synapses’ from the DG

can impose a new pattern of activation on CA3 in spite of the retrieval-related feedback from recurrent connections (McNaughton and Morris, 1987), there must be some mechanism to determine whether the system should be optimised for encoding or retrieval. One proposal is that the supply of acetylcholine (ACh) from the medial septum switches the hippocampus between encoding and retrieval modes (Hasselmo et al., 1995; Wallenstein and Hasselmo, 1997; Murre, 1996). In this model, elevated ACh increases the rate of synaptic modification of the recurrent connections in CA3 and suppresses the synaptic transmission of intrinsic activity (within CA3 and from CA3 to CA1), enabling new patterns of activity to be stored without interference from retrieved memories. Conversely, decreased ACh reduces the rate of synaptic modification and enhances synaptic transmission both within CA3 and from CA3 to CA1, providing favourable conditions for memory retrieval and output to neocortex (Hasselmo, 1999). The delivery of ACh is determined by the novelty of the neocortical input, as represented by direct activation of CA1 from EC, compared to the most similar previously stored event, as represented by the input to CA1 from CA3 after settling to a stored state. Specifically, if both CA3 and EC inputs to CA1 are matching (as with a familiar stimulus), strong activation of CA1 drives interneurons in the medial septum which decrease the activity of cholinergic cells projecting to the hippocampus (see Fig. 16.10). More generally, a medial temporal lobe circuit that can generate a novelty signal by comparing incoming sensory information with previous memories encoded in the recurrent connections of CA3 may be useful to enhance the encoding of salient or unexpected events.

Two types of evidence support this hypothesis. Firstly, by emphasizing the connections between the hippocampus and medial septum, the model begins to address data showing the importance of the fornix, the large fibre bundle connecting the hippocampus with medial septum and other subcortical structures. In the model, sectioning the fornix will prevent the learning of new memories due to lack of ACh which corresponds well to the effects of damage to the fornix in neuropsychological patients (see e.g. Gaffan and Gaffan, 1991; Aggleton and Brown, 1999; Spiers et al., 2001). Secondly, blocking ACh receptors (by injecting scopolamine) impairs encoding – which impairs recollection more strongly than familiarity-based recognition, as the former relies on forming novel associations within the hippocampus (Hasselmo and Wyble, 1997). This will also be the case in the model, assuming that CA3 serves to associate events and their contexts, since recall is more reliant on these associations than recognition. However, disruption of the hippocampus might also impair recall more than recognition in many other models, e.g. due to disrupting associations with context (see Section 16.4.6).

In addition, ACh enhances theta band activity, which has been proposed to dynamically schedule periods of encoding and retrieval within each oscillatory cycle by alternately inhibiting intra-hippocampal connections and entorhinal inputs to the hippocampus, respectively (Hasselmo et al., 2002). Around the trough of the theta cycle (as measured at the hippocampal fissure), strong input from entorhinal cortex is paired with inhibition of synaptic transmission (but not plasticity) at recurrent collaterals within CA3 to produce favourable conditions for the encoding of new associations. Conversely, around the peak of the theta cycle, entorhinal input is inhibited while synaptic currents within hippocampus are enhanced to produce favourable conditions for the retrieval of existing associations. This model is supported by observations of a shift in the preferred firing phase of place cells towards the trough of the theta cycle during exposure to novel environments, consistent with the encoding of new

information, which is disrupted by cholinergic antagonists (Douchamps et al., 2013). The dynamic scheduling of encoding and retrieval by the theta rhythm is also supported by the observation that LTP is preferentially induced at specific phases of the oscillatory cycle (Pavrides et al., 1988; Huerta and Lisman, 1995; Holscher et al., 1997).

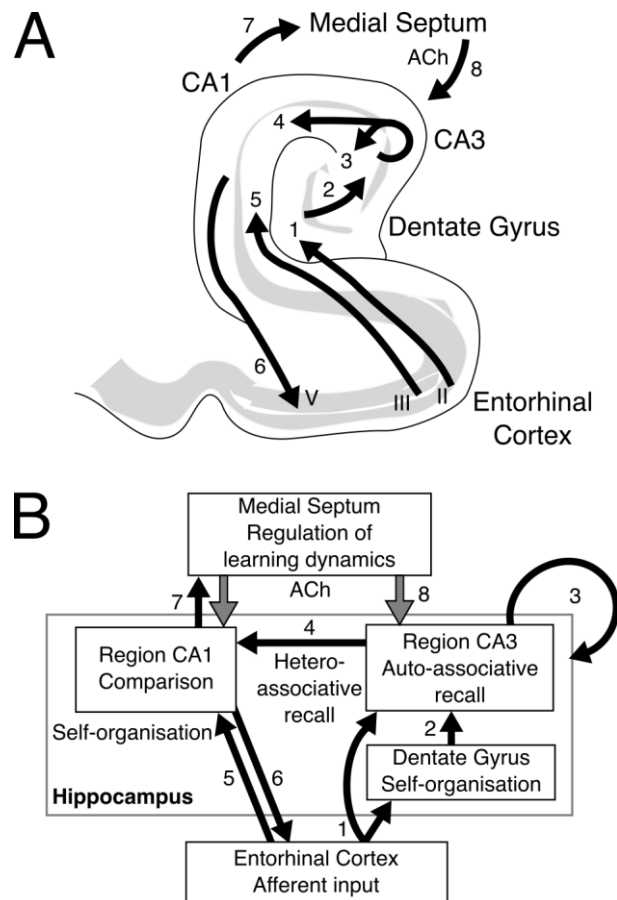


Figure 16.10 Schematic representation of **A**: hippocampal anatomy and **B**: computational models of hippocampal episodic memory function. Numbers label synaptic connections mediating various functions in the model. 1. Synapses of the perforant path fibres projecting from entorhinal cortex layer II to the dentate gyrus undergo sequential self-organization to form sparse, less overlapping representations of entorhinal activity patterns. Direct connections to CA3 may be used to cue recall. 2. Mossy fibres transfer sparse dentate gyrus activity to CA3. 3. Excitatory recurrent connections in region CA3 mediate auto-associative encoding and retrieval of the features of episodic memories. 4. Schaffer collaterals encode and retrieve associations between CA3 activity and activity patterns induced by entorhinal input to CA1. 5. Perforant path input to region CA1 undergoes self-organization, forming new representations of entorhinal cortex input for comparison with recall from CA3. 6. Projections from region CA1 to deep layers of the entorhinal cortex allow representations in CA1 to activate the associated patterns in entorhinal cortex. 7. Output from region CA1 to the medial septum regulates cholinergic modulation. 8. Cholinergic modulation from the medial septum sets appropriate dynamics for encoding in hippocampus. Adapted from (Hasselmo and McClelland, 1999).

16.4.5 Consolidation and Cross Modal Binding of Events in Memory

Ever since the initial reports of dense anterograde amnesia but weaker or temporally graded retrograde amnesia after bilateral medial temporal lobectomy (Scoville and Milner, 1957; see Chapter 13), researchers have considered how the hippocampus contributes to the long-term consolidation of memories. One hypothesis is that information is consolidated elsewhere in the

brain after which it can be recalled without the hippocampus. Marr's (1971) model follows this view, suggesting that the day's events are stored in the hippocampus before information deemed relevant to the animal's future is categorised and incorporated into the long-term store of knowledge in neocortex. However, experimental data regarding the gradient of retrograde amnesia (i.e. the sparing of memories acquired sufficiently long enough before the damage) is inconsistent and remains controversial in both animals (see Chapter 14 and Squire et al., 2015) and humans (see Chapter 13 and Spiers et al., 2001). One problem in the human data is that the amnesia often extends back to childhood, implying that the hippocampus stores several decades worth of information. Nonetheless, several arguments have been put forward for the transfer of information from the hippocampus into a consolidated form in neocortex. In particular, this hypothesis has been developed by complementary learning systems theory, which describes the benefits of combining a fast learning (hippocampal) system that can rapidly acquire the details of specific experiences with a slow learning (neocortical) system that can gradually extract structural or semantic knowledge from those experiences to support generalisation (Marr, 1971; McClelland et al., 1995; McClelland, 2013; Kumaran et al., 2016).

Complementary learning systems theory is consistent with memory for the unique content and context of a specific event (i.e. episodic memory) depending on the hippocampus, but semantic memory depending on other neocortical areas (see e.g. Graham and Hodges, 1997). It is also consistent with the suggestion of Nadel and Moscovich (1997) that semantic memories show a temporal gradient of retrograde amnesia while detailed episodic memories do not. The model is less obviously consistent with the suggestion that semantic information can be acquired despite early bilateral hippocampal pathology (Vargha-Khadem et al., 1997). However, the possibility of partial sparing of the hippocampus (Squire and Zola, 1998), or the use of external rehearsal of information (Baddeley et al., 2001), might provide explanations for these developmental cases within the framework of the model. The process of memory consolidation also has many parallels with the training of deep neural networks - e.g. those that use a fixed feed-forward structure and error driven learning rules to optimally extract the statistics of their inputs following the repeated and interleaved presentation of large numbers of training examples (Rumelhart et al., 1986; LeCun et al., 2015; Kumaran et al., 2016). We note that some consolidation processes might also occur within the hippocampus without transfer of information from one region to another - using protein synthesis to stabilise changes in the strength of individual synapses, for example (Sutton and Schuman, 2006; Redondo and Morris, 2011).

Evidence for the replay of hippocampal activity patterns in the rodent hippocampus during quiescent waking periods and sleep led to the hypothesis that these events might mediate long-term memory consolidation (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996; Carr et al., 2011; Olafsdottir et al., 2018). Consistent with this view, hippocampal replay events are coordinated with increased neocortical activity (Siapas and Wilson, 1998; Sirota et al., 2003; Battaglia et al., 2004; Ji and Wilson, 2007); and disrupting replay events during rest or sleep impairs the acquisition of spatial memory tasks (Girardeau et al., 2009; Nakashiba et al., 2009; Ego-Stengel and Wilson, 2010). In addition, hippocampal replay appears to favour rewarded experiences (Ambrose et al., 2016), which aligns well with Marr's hypothesis that salient or statistically unusual events should be preferentially consolidated in order to direct future behaviour. Indeed, one theoretical account suggests that the replay of different experiences is prioritised in order to optimise future reward (Mattar and Daw, 2018).

Importantly, a number of studies have demonstrated that neural activity patterns observed during learning are also reactivated during rest in the human brain, and that reactivation is correlated with subsequent memory performance (e.g. Staresina et al., 2013; Deuker et al., 2013; Schapiro et al., 2018; Schreiner and Staudigl, 2020).

Early models of memory consolidation focussed on the anatomical convergence of information from different sensory modalities onto the hippocampus. In the absence of dense long-range connections between different sensory cortical areas, associations between the elements of an event such as its sight, sound and smell cannot be formed in the lower-level cortices. Thus, Damasio (1989) suggested that ‘convergence zones’ must exist where these ‘cross-modal bindings’ could be formed. Several models have extended this idea to include rapid learning of a hippocampal, or medial temporal lobe, representation reciprocally connected to all the unimodal cortical representations of the event that allows them to be associated to each other (Alvarez and Squire, 1994; Murre, 1996; Moll and Miikkulainen, 1997). These models also suggest that, after multiple rehearsals of a memory driven by the hippocampus, long-range associations can be learned directly between the unimodal representations, finally making the stored information independent of the hippocampus.

A re-interpretation of the experimental data regarding the gradient of retrograde amnesia associated with medial temporal lobe damage led Nadel and Moscovitch (1997) to a different conclusion regarding consolidation (see also Winocur et al., 2010). Noticing that, in many instances, retrograde amnesia extends back over a much longer time than that envisaged by Marr, they proposed that the hippocampus remains necessary for the retrieval of detailed episodic or spatial information. To account for both the common occurrence of a temporal gradient in memory for other types of information, and the possibility of partial damage to the hippocampus, they proposed a new model. According to this ‘multiple trace theory’, whenever a memory is rehearsed or reactivated, a new hippocampal representation is formed, again connected to all of the neocortical representations. The result of this is that, while a complete lesion of the medial temporal lobe impairs retrieval of all memories, the older the memory, the more robust it will be to partial damage by virtue of being represented in multiple locations within hippocampus (but see Teng and Squire, 1999; Rosenbaum et al., 2000 for evidence that early spatial memories become hippocampus independent).

Both the arguments regarding data abstraction and anatomical convergence were represented in a model by Kali and Dayan (2004). In this model the hippocampus serves to aid the learning of a hierarchical semantic system, by being able to reinstate (i.e. replay) the top-level (i.e. entorhinal / perirhinal / parahippocampal) representation of events during sleep. The semantic system contains reciprocal connections between these top-level cortical areas and those below them in the hierarchy, but no direct connections between areas at the same level. Aided by the hippocampus, the neocortical system forms an associative representation of activity patterns in lower cortical areas. This is achieved by the top-level representations learning a ‘generative model’ of representations lower in the hierarchy (e.g. Hinton and Sejnowski, 1999; see Section 16.4.8). Thus top-level representations can be cued by input from a subset of lower cortical areas and then cause pattern completion in all lower areas via the reciprocal connections. In further simulations, Kali and Dayan (2004) note that regular reactivation of episodic hippocampal representations is required to maintain them in register with the slowly changing semantic representations. As such, the occasional replay or rehearsal of episodic information

from the past may be required to maintain functional connections between the hippocampus and neocortex following the initial stages of memory consolidation.

Another argument for consolidation refers to the effects of interference. In deep neural networks, new information can interfere with previously stored information, sometimes ‘catastrophically’ such that all information is lost (McCloskey and Cohen, 1989). This problem can also occur in associative memories (see Fig 16.9). McClelland et al. (1995) proposed a solution to this in which long-term consolidation involves random interleaved re-presentation of all previous knowledge along with newly acquired knowledge, to produce a single integrated neocortical representation of semantic knowledge. Note, however, that this mechanism requires the temporary store to have capacity for the entire dataset, although more recent studies have demonstrated that ‘generative replay’ (in which the hippocampus constructs replay sequences from a generative model, rather than replaying specific events; see Section 16.4.8) can avoid this issue (Shin et al., 2017; van de Ven et al., 2020; Stoianov et al., 2021). Other solutions include associative memories with continuous but bounded connection weights (Hopfield, 1982), which avoid the issue of catastrophic interference by storing only the most recently presented input patterns; and ‘constructive’ algorithms in which the addition of new information is accompanied by the addition of new processing units (Gallant, 1986; Mezard and Nadal, 1989; Freat, 1990; Fahlman and Lebiere, 1990). These algorithms may relate to the observation of neurogenesis in the rodent dentate gyrus that is related to learning (Deng et al., 2010), although recent data suggests that this is absent in the adult human hippocampus (Franjic et al., 2021). Notably, ‘modern’ Hopfield networks (also known as dense associative memories, e.g. Krotov and Hopfield, 2016; Demircigil et al., 2017) are not subject to the same capacity constraints as the canonical model, exhibit reduced interference between stored patterns and more rapid convergence. However, biologically realistic implementations of these networks have yet to be fully elucidated (but see Krotov and Hopfield, 2020).

Relatedly, experimental data indicate that new memories are consolidated more quickly if they are consistent with previously existing knowledge or ‘schema’ (Tse et al., 2007; Squire et al., 2015). It is relatively straightforward to incorporate this feature into the canonical model, given that slow learning is only required when new information is inconsistent with existing knowledge and may therefore lead to catastrophic interference (McClelland, 2013). Interestingly, recent theoretical work has also suggested that, in order to optimise generalisation from previous experience in novel circumstances, memories of ‘unpredictable’ events (i.e. those that are not consistent with the statistical regularities extracted from a wider corpus of experience) should remain dependent on the hippocampus, while memories of ‘predictable’ events can be consolidated in neocortex (Sun et al., 2021). This addresses an issue known as ‘overfitting’, whereby extensive training to capture the statistics of even the most unpredictable exemplars impairs generalisation. This model also ameliorates the issue of hippocampal capacity, as only memories of unpredictable events need to be stored in the hippocampus over the longer term.

In contrast, some empirical data have presented a challenge to classic models of neocortical memory consolidation. For example, preliminary evidence indicates that the hippocampus may support rapid generalisation or ‘statistical learning’ (e.g. Zeithamova et al., 2012; Schapiro et al., 2014), a role ascribed to neocortex in the canonical model. This is implicitly at odds with a process of pattern separation, however, by which overlapping sensory inputs are

orthogonalized, reducing the representational overlap between related events and impairing generalisation. This issue can be addressed by ‘big loop recurrence’ at the point of memory retrieval (i.e. recursive interactions between the hippocampus and neocortex, rather than within the hippocampus itself; Kumaran et al., 2012). Specifically, if hippocampal output is iteratively fed back as input to prompt further retrieval, and several retrieved memory patterns can be coactive within the hippocampus at any one time, then it is possible to perform rapid generalisation and inference across the content of those related experiences. Alternatively, the CA1 region of the hippocampus can be re-cast in the role of neocortex, producing overlapping activity patterns for related events from the pattern separated output of CA3, and therefore supporting generalisation and inference despite individual memory traces being stored in an orthogonal manner (Schapiro et al., 2017). This issue can also be addressed at the point of memory encoding, if one assumes that the constituent elements of related experiences are represented invariantly in the hippocampus (e.g. Quiroga et al., 2005; see Section 16.4.4). This preserves the overlap between the representation of related events, supporting generalisation and inference between those events, but places specific constraints on the process of pattern separation.

16.4.6 Hippocampal Contributions to Familiarity-based Recognition

A distinction has been made between episodic memory, semantic memory, and familiarity-based recognition (see Chapter 13). Episodic memory is characterized by the ability to recall detailed information about an event and its context. Semantic memory is characterized by factual knowledge without recall of the individual events and contexts in which it was acquired. Familiarity-based recognition depends on an unattributable feeling of familiarity associated with a stimulus in the absence of detailed information about the event and context in which it was encountered. Of these three processes, the hippocampus has been most strongly associated with episodic memory (e.g. Aggleton and Brown, 1999). As described above, however, it has also been argued that the hippocampus is required to provide associations between representations in disparate cortical areas, while associations within each area can be formed locally. Thus, the familiarity of single stimuli might be supported by the association of elements within each neocortical area, independent of the hippocampus, while recognition of cross-modal associates amongst equally familiar distractors would require the hippocampus.

Consistent with this view, empirical evidence indicates that simple recognition memory does not depend on the hippocampus, but on nearby neocortical areas (Zhu et al., 1996; Murray and Mishkin, 1998; Wan et al., 1999; Aggleton and Brown, 1999; Baxendale et al., 1997; Baddeley et al., 2001), but see also (Manns and Squire, 1999; Zola et al., 2000). In contrast, there is some evidence that recognition of cross-modal associations is impaired by bilateral damage restricted to the hippocampus (Holdstock et al., 2000; Vargha-Khadem et al., 1997), although more extensive unilateral damage may also impair the binding of elements within the same modality (Kroll et al., 1996). The logical extension of this idea is that episodic memory requires the full recollection of an event and its context in all its multi-modal detail and so will require an intact hippocampus.

Much of the analysis of the differential role of the hippocampus in episodic retrieval (or ‘recollection’) and familiarity-based recognition has focused on the idea that these two processes both contribute independently to the performance of recognition memory tests (see

e.g. Yonelinas, 2002). In both forced choice and yes-no recognition paradigms, the recollective component is assumed to be 'all or nothing' and 'high-threshold'. That is, the stimulus is recalled in great detail or not at all, and a novel foil is never falsely recalled. By contrast, the familiarity-based process is more like a signal detection problem: the subject guesses whether the item is familiar or not informed by a noisy measure of familiarity. The hippocampus has been associated with the recollective component (Aggleton and Brown, 1999; Yonelinas et al., 2002) and so should provide a high-threshold, all or nothing mechanism in recognition memory tests (see Rugg and Yonelinas (2003) for a review).

The hippocampal contribution to recognition memory (via 'recollection') can be modelled by using the stimulus-driven medial temporal neocortical representation to retrieve a stored pattern of activation in CA3, and comparing the retrieved activation to the stimulus-driven activation in the EC (Norman and O'Reilly, 2003). By explicitly retrieving an entire stored pattern, the process is all-or-nothing and even foils which resemble previously presented patterns are not falsely recognized. The use of sparse hippocampal representations, orthogonalized via the DG, serves to prevent interference between different stored events. By contrast, familiarity-based recognition is modelled as a 'sharpening' of the neocortical representation. Under this model, while a new item is represented by weak activation of a large number of neocortical neurons, repeated presentation of the item results in strong activation of a smaller number of neocortical neurons via a competitive learning mechanism. The activation of neocortical neurons that fire in response to a given stimulus can then be used as a measure of familiarity-based recognition. A similar model holds that the hippocampus (or adjacent cortical regions) can also signal stimulus or event familiarity by probing the energy function of previously stored memories (Bogacz et al., 2001; Greve et al., 2010). If incoming sensory input is consistent with previously stored activity patterns, then network output will be boosted by strong recurrent connections between active neurons; while novel sensory input will produce lower network activity due to the absence of those connections. As such, simply monitoring overall network output during the initial presentation of a stimulus provides a measure of familiarity.

One advantage of this scheme is that repeated presentations of some items in a list will not impair recognition of the other items in the list, as seen experimentally (Ratcliff, 1990). A characteristic of this model is that hippocampal damage will specifically impair recognition memory when related lures (novel items that resemble previously seen items) are included in the test. The familiarity-based recognition mechanism would produce false alarms to these stimuli, while the hippocampal recollection mechanism would not, consistent with some empirical evidence (Holdstock et al., 2002).

16.4.7 Free Recall and Temporal Context

One of the distinguishing features of episodic memory is the ability to retrieve the ongoing context within which an event occurs (Gardiner and Java, 1993; Tulving, 1993; Knowlton and Squire, 1995), and one suggestion for the role of the human hippocampus is that it provides the spatio-temporal context for episodic memory (O'Keefe and Nadel, 1978). Theoretical analyses of associative memory have also made the distinction between the content of an event and its context (e.g. Raaijmakers and Shiffrin, 1981) or between the record of an event and the 'header' or index term used to reference it (e.g. Morton et al., 1985). One possibility is that the hippocampus serves to associate the content of an event with its context. A related possibility

is that the hippocampus provides a representation of temporal context itself, e.g. dynamically varying patterns of activity generated by its own recurrent connectivity that serve to bind coactive neocortical representations (Buzsaki and Tingley, 2018). This type of model coincides with a considerable psychological literature describing models of memory in which the retrieval of stored items is held to depend, at least in part, on their association to a representation of context that changes slowly with time or experience. In some of these models the context representation changes independently of the stored items (Estes, 1955; Mensink and Raaijmakers, 1988; Davelaar et al., 2005), while in others the context representation is derived from the items themselves (Howard and Kahana, 2001; Howard et al., 2015). A given item will subsequently be retrieved according to the similarity between the context signal at retrieval and that associated with the item. In particular, these models are often used to address free recall paradigms, in which the cue for retrieval is not external (as is typically the case for the canonical model) but internally generated.

One of the first mechanistic models of this type was described by Levy (1996), who demonstrated that a recurrent network model of CA3 with Hebbian plasticity that is repeatedly exposed to sequential activity patterns (i.e. temporally correlated inputs) develops ‘context neurons’. These do not form part of the sequential activity pattern but fire persistently during specific subsections of that pattern, providing a slowly varying context signal which supports sequence retrieval and allows sequences with overlapping sections to be disambiguated (see Fig. 16.11). Wallenstein and Hasselmo (1998) described a similar model which uses context neurons to form associations between temporally discontinuous events (further apart in time than the timescale required for pre- and post-synaptic activity to induce LTP, i.e. more than around one hundred milliseconds, see Chapter 10) and can therefore account for the role of the hippocampus in trace conditioning. Indeed, this type of mechanism also provides a good model for short-term serial recall (Burgess and Hitch, 2005).

The temporal context model (Howard and Kahana, 2001) extended this framework to account for several additional features of experimental data from free recall paradigms. In this model, the context representation is derived from the presented or retrieved items themselves, becoming a recency-weighted sum of the context arising from each item. After a sequence of items has been presented, the context vector will be most similar to that associated with recent items, producing the well-known recency effect; and the context vector for consecutive items in the sequence will be more similar, producing the well-known temporal contiguity effect, both observed during free recall. Finally, because each item that is presented affects the context vector to which subsequent items are associated, the retrieval of immediately subsequent items in the list is more likely. This leads to an asymmetry such that forward associations within the list are stronger than backward associations, creating the characteristic forward-bias of free recall (e.g. Kahana, 1996).

Each of these models is supported by the observation of slow changes in hippocampal activity over time (e.g. Manns et al., 2007; Mankin et al., 2012; Ziv et al., 2013; Cai et al., 2016); and by evidence that representations of spatio-temporal context within the hippocampus are reinstated during retrieval (Manning et al., 2011; Howard et al., 2012; Miller et al., 2013). Intriguingly, it has been demonstrated that a generalised form of the temporal context model is equivalent to a temporal difference algorithm for learning the successor representation, hinting that spatial and episodic context may share a common mechanism (Gershman et al., 2012; see

Section 16.3.2). Similarly, we note that the temporal context model has been extended to account for spatially tuned responses in the hippocampus (Howard et al., 2005; Howard et al., 2014b).

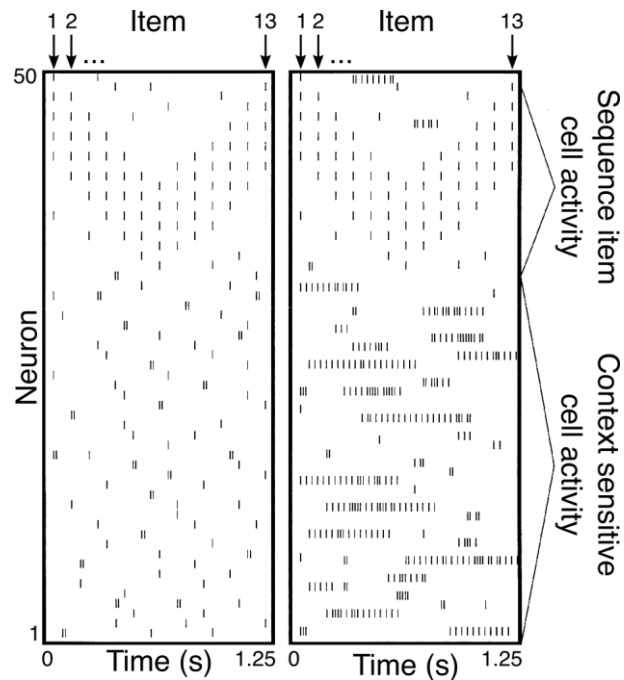


Figure 16.11 Context field development. Each rectangle shows a subset of 50 simulated pyramidal cells firing across time, each action potential represented by a vertical line. A unique group of five cells firing at the same time encodes a single item in a 13-item sequence, see the top portion of each rectangle. This can be thought of as afferent activation of CA3 pyramidal cells due to a specific pattern of sensory events. **A:** Notice the background firing during the first learning trial that does not encode sequence items directly. This stems from activity at recurrent excitatory synapses. Repeated exposure to the same sequence can lead to enhanced synaptic potentiation between cells firing in the background and cells that encode sequence items, due to a simple Hebbian learning rule. **B:** After the fourth learning trial, this repeated potentiation leads to a condition where background cells begin to respond to the appearance of contiguous segments of the entire sequence. The portion of the full sequence to which the cell responds is called the ‘context field’ of the cell. Because the context fields overlap, the entire sequence can be reconstructed by interdigitating them in the proper order. Adapted from (Wallenstein et al., 1998).

In a related model of free recall, Romani et al. (2013) focussed on accounting for a different feature of the experimental data – the relationship between the number of items studied and retrieved (e.g. Murray et al., 1976). They demonstrated that, following the encoding of an arbitrary number of discrete, sparse activity patterns in the asymmetric synaptic connections of a recurrent neural network, a combination of periodic inhibition and firing rate adaptation could be used to produce the sequential reinstatement of those activity patterns. Importantly, transitions tended to occur between activity patterns with the greatest representational overlap and be limited to a subset of activity patterns that followed a power law relationship with the total number of encoded items, as observed in experimental data. Transitions between activity patterns also tended to be cyclical, moving sequentially between the same subset of patterns, and recall was terminated when the system completed one loop. This is consistent with the observation that subjects tend not to recall any further words after they have erroneously recalled a studied item for the second time (Miller et al., 2012). However, this model cannot account for primacy, recency, or temporal contiguity effects.

16.4.8 The Hippocampus as a Generative Model

It has long been known that episodic memory retrieval is reconstructive rather than veridical (e.g. Bartlett, 1932; Shachter and Addis, 2007) and that the hippocampus is implicated in the imagination of new experiences (e.g. ‘episodic future thinking’ or ‘scene construction’; Atance and O’Neill, 2001; Hassabis et al., 2007; see Chapter 13). As noted earlier, hippocampal spatial representations can also be predictive (e.g. Skaggs et al., 1996; Mehta, et al., 1997; Stachenfeld et al., 2017) and – in some circumstances – represent behavioural trajectories that have never been experienced (e.g. Gupta et al., 2010; Dragoi and Tonegawa, 2011; Olafsdottir et al., 2015). These observations have led to the hypothesis that the hippocampus may function as a generative model, making use of existing memories to simulate the sensory and reward contingencies of new experiences. As mentioned above, a generative memory system may offer several advantages for efficiently extracting the descriptive statistics of experience by optimising the ‘training’ of neocortex during long-term consolidation (Shin et al., 2017; van de Ven et al., 2020; Stoianov et al., 2021).

A detailed, neural level model of how the hippocampus might use episodic memories to generate complex visuospatial imagery was described by Becker and Burgess (2001) and subsequently developed by Byrne et al. (2007) and Bicanski and Burgess (2018; see Fig. 16.12). This model explicitly makes use of the constraints associated with spatial information and our detailed knowledge of how it is represented in the brain. Specifically, the representations of spatial layout in long-term memory are assumed to be allocentric (e.g. independent of the orientation of the person) in contrast to the egocentric (i.e. body centred) short-term representations involved in perception, action and working memory (Goodale and Milner, 1992; Milner et al., 1999; Burgess et al., 1999). Transformation between these two reference frames is achieved by a head direction modulated gain field circuit, hypothesised to reside in retrosplenial cortex, and analogous to the role proposed for gain-field neurons found in posterior parietal cortex (e.g. Pouget and Sejnowski, 1997). Again, it is worth noting the incredible correspondence between this model and synaptic connectivity in the fly brain (Lu et al., 2022; see Fig. 16.12B).

During memory encoding, egocentric representations of the distance and direction to spatial boundaries and discrete objects from the current viewpoint are transformed, via the gain field circuit, into allocentric boundary and object vector cell (BVC and OVC) representations in the medial temporal lobe. By extension from spatial models of the hippocampus (Hartley et al., 2000), BVCs are bi-directionally connected to place cells, which form a continuous attractor. When attending to a specific object, OVCs also become bi-directionally connected to the place cells encoding that location, head direction cells encoding the current orientation, and perirhinal neurons encoding object identity. During subsequent memory retrieval - cued by activity in perirhinal neurons encoding object identity, for example, attractor dynamics reinstate activity in OVCs, place cells and head direction cells, as well as BVCs. This allocentric representation of spatial context is transformed into egocentric spatial imagery via the gain field circuit, with a viewing location and orientation dictated by reinstated place and head direction cell activity to be consistent with that during encoding. Importantly, in later versions of the model, mock motor efference from grid cell inputs is also used to support dynamic imagery, planning, and the mental exploration of previously unexplored routes (Bicanski and Burgess, 2018). As such, this model provides a neural level framework for

interpreting the role of hippocampus in reconstructive episodic memory retrieval, episodic future thinking and scene construction.

This model is consistent with the observation of hemi-spatial neglect in imagery following parietal damage (e.g. Bisiach and Luzzatti, 1978) and functional imaging of the retrieval of spatial context (Burgess et al., 2001b). It also provides an explanation for the involvement of Papez's circuit (see Chapter 3) in supporting both episodic memory (see Chapter 13) and the representation of head-direction (see Chapter 11). In addition, several of the allocentric and egocentric spatial responses predicted by this model have subsequently been identified in the rodent brain (Deshmukh and Knierim, 2013; Wang et al, 2018; Hinman et al., 2019; Hoydal et al, 2019; Alexander et al., 2020; see Bicanski and Burgess, 2020 for a review). Finally, we note that the place cells in this model correspond to a latent variable representation of location to which different sensory inputs are mapped (during bottom-up encoding) and can be generated from (during top down imagery). In addition, grid cell inputs allow imagery to be dynamic and generated from new viewpoints, consistent with their proposed role in planning routes through real or conceptual spaces (Bush et al., 2015; Stemmler et al., 2015; Behrens et al., 2018). Intriguingly, a similar transformation circuit and set of allocentric spatial responses are learned by an agent tasked solely with predicting upcoming egocentric visual inputs (Uria et al., 2020). More often, however, generative models go beyond these networks (which predict a single set of inputs) by attempting to learn or approximate the probability distribution from which their inputs (or predicted outputs) are drawn (see e.g. Kingma and Welling, 2014). These themes are explored in more detail in the next section.

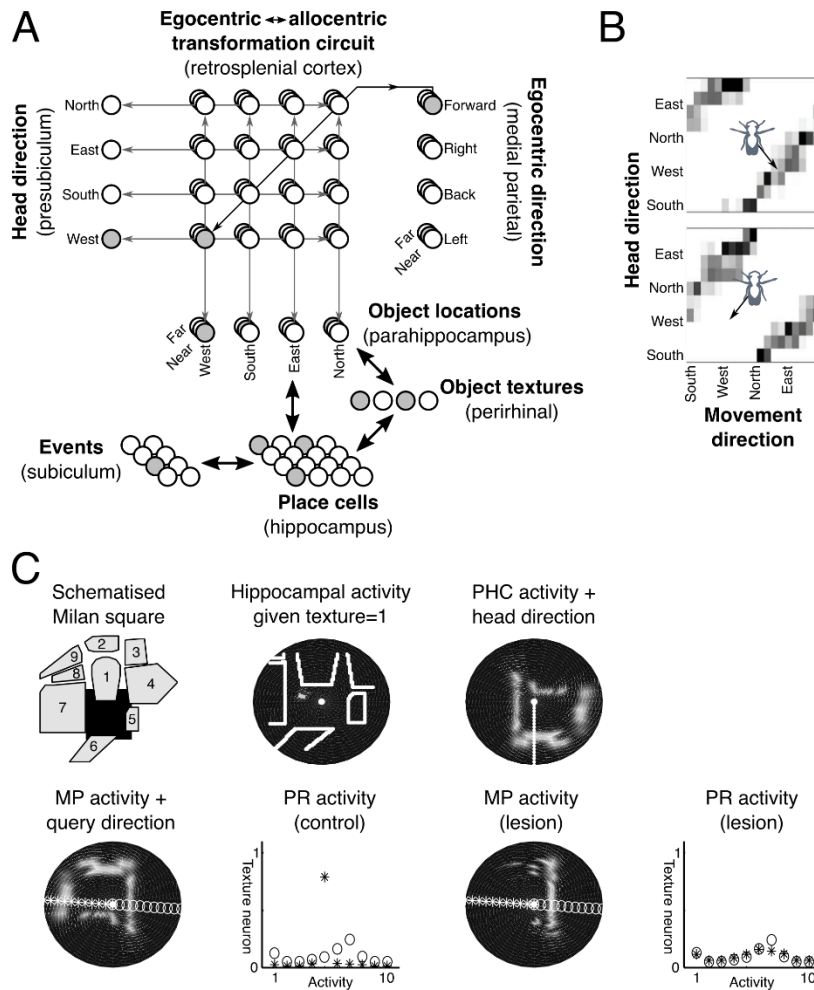


Figure 16.12 A model of episodic memory and visuospatial imagery. **A:** Functional architecture of encoding and retrieving the spatial context of an event. Neurons shown in grey illustrate a possible pattern of activity, corresponding to an imagined location with extended landmarks nearby to the West, far away to the North and at an intermediate distance to the South. The imagined Westerly heading direction means that these landmarks are imagined as far to the right, nearby straight ahead and at an intermediate distance to the left. The transformation between egocentric (left, right, ahead) and allocentric (North, South, East, West) directions to landmarks occurs via a circuit linking these representations and head direction. Adapted from (Burgess et al., 2001a). **B:** Empirical data from *Drosophila*, illustrating the strength of synaptic connections between neurons encoding head direction and allocentric movement direction, and projecting to regions which drive movement in specific egocentric directions (backwards and right, top panel; backwards and left, bottom panel). Note the correspondence with the diagonal synaptic weight matrix in the transformation circuit predicted by the model (shown in panel A). Adapted from Lu et al. (2022). **C:** Simulation of the retrieval of spatial information in the Milan square experiment of Bisiach and Luzzatti (1978). i: Training consists of simulated exploration of the square (shaded area, North is up). The system is cued to imagine being near to the Cathedral (i.e. the perirhinal cell for the texture of building 1 and parahippocampal cell for a building at a short distance North are activated) and the recurrently connected hippocampal-parahippocampal-perirhinal system settles. ii: The hippocampus settles to a location in the Northwest corner of the square (hippocampal cell activity shown as the brightness of the pixel corresponding to the location of each cell's place field). iii: The parahippocampus correctly retrieves the locations of the other buildings (parahippocampal cell activity shown as the brightness of the pixel for the location encoded by each cell, relative to the subject at the centre). A line indicates that the imagined heading direction is South. iv: Medial parietal cell activity: the parahippocampal map has been correctly rotated given head-direction South (straight ahead is up), stars indicate a direction of inspection to the left, circles to the right. v: Perirhinal cell activations correctly showing building 5 to the left and building 7 to the right. vi: Effect of a right parietal lesion on the medial parietal representation (note lack of activation on the left) and perirhinal activations (vii: note decrease in activation of building 5 when inspection is to the left). Adapted from (Becker and Burgess, 2001).

16.5 Hippocampal Function Beyond Space and Memory

In recent years, theoretical models of the hippocampus have been reinvigorated by empirical findings that demonstrate coding for task-relevant non-spatial dimensions (e.g. Constantinescu et al., 2016; Aronov et al., 2017; Bao et al., 2019; Knudsen and Wallis, 2021; Nieh et al., 2021). This has led to a broader conceptualisation of hippocampal function that encompasses and incorporates both spatial cognition and associative memory. As described below, these models propose that the hippocampus encodes low dimensional cognitive maps (Tolman, 1948) which capture relational structure (Eichenbaum and Cohen, 2014) between task-relevant states and can be used to plan behaviour, e.g. via model-based RL (Stachenfeld et al., 2017). This process is facilitated by the formation of efficient representations - both in terms of identifying and encoding the relevant dimensions, and by capturing the underlying structure of experienced transitions or relationships between states to facilitate prediction, generalisation and planning (Behrens et al., 2018; Whittington et al., 2022).

16.5.1 Low Dimensional Latent State Representations

Numerous recent experimental studies have demonstrated that the hippocampal formation can also code for location and movement direction along abstract, continuous dimensions beyond those that represent physical space. For example, Aronov et al. (2017) demonstrated that hippocampal pyramidal cells form receptive fields in auditory space (i.e. firing in response to specific tones) when rodents were trained to manipulate sound frequencies using a joystick in order to obtain reward. Similarly, Nieh et al. (2021) showed that the same neurons jointly encoded location and the accumulation of visual evidence when reward was contingent on turning in the direction of the greatest number of visual cues distributed along the walls of a virtual T-maze. This has led to the theory that the hippocampus attempts to form efficient representations of the structure or ‘state space’ (i.e. a cognitive map) of any given task to support ongoing behaviour. Importantly, this requires identifying the ‘latent variables’ (those that are not directly mapped to sensory inputs, such as location and head direction in the case of spatial navigation) that most efficiently and accurately describe the relevant task dimensions (Gershman and Niv, 2010; Niv, 2019; Radulescu et al., 2021). This formulation provides a unifying explanation for several features of both place cell firing and single unit responses to non-spatial variables. We note, however, that in the rodent hippocampus at least, spatial variables appear to occupy a privileged position, in that they are reliably encoded even in the absence of reward (while responses to non-spatial variables tend to rely on the delivery of reward, potentially indicating task relevance). Indeed, it has been suggested that non-spatial variables are consistently encoded in conjunction with location (O’Keefe and Krupic, 2021; see Chapter 11). Whether this is because rodents consistently expect spatial structure to be task relevant or reflects the evolutionary trajectory of the hippocampus (from coding for space to any task relevant variable) is not yet clear.

In either case, the idea that hippocampal cells code for task relevant latent variables rather than purely location helps to explain several features of spatial firing patterns. For example, ‘splitter cells’ fire in a specific location on the neck of a T maze depending on whether the animal is about to turn right or left on a spatial alternation task (Frank et al., 2000; Wood et al., 2000; Grieves et al., 2016), thereby disambiguating identical sensory inputs depending on

behavioural context. Similar responses have been observed when mice are rewarded on every fourth lap of a circular track, with hippocampal place cells conjunctively coding both location on the track and lap number (Sun et al., 2020). Crucially, this conjunctive code was replaced by a purely spatial code when reward was provided on every lap of the track. Another example is provided by hippocampal remapping (Bostock et al., 1990; see Chapter 11), which can be triggered by changes in the task being executed within a specific environment (e.g. Markus et al., 1995), thereby forming a new representation to support a different behaviour (Sanders et al., 2020). Each of these examples is consistent with the hippocampus identifying higher level task structure that disambiguates perceptually identical visits to the same location according to prospective (or retrospective) behaviour. Conversely, orientation independent firing of place cells in an open field could be understood as resulting from a complementary process by which the hippocampus associates perceptually distinct inputs (i.e. the view in each head direction) with a task state that affords the same contingencies (i.e. a single spatial location). In the non-spatial domain, we note that the receptive fields of time cells – which fire at specific time points during delay periods in which spatial location is approximately constant (Eichenbaum, 2014; Umbach et al., 2020) – can also be understood as encoding latent states. Time is an implicitly latent variable, given that it cannot be directly observed, and these representations encode progress through a delay period prior to the receipt of reward (or in trace conditioning tasks, see Section 16.4.7).

To efficiently encode task structure, hippocampal representations of latent variables should be low dimensional, with activity patterns varying according to changes in task-relevant variables but remaining insensitive to those which are not predictive of reward. Continuous attractor models of head direction, place and grid cells (which encode latent spatial variables) highlight how the firing patterns of very large numbers of neurons can be constrained to exist on a low dimensional manifold (i.e. surface) via specific patterns of recurrent connectivity (see Section 16.2.3, Box 16.3). In those examples, the manifold corresponds directly to the physical variables of heading direction (a 1-D ring) or the (2-D) location of the moving animal on the surface of the ground. More generally, when the state-space of a behavioural task or function can be described in terms of a small number of latent variables, it would be efficient for the firing patterns of neurons supporting that function to form population codes for the values of those latent variables. Such codes, in which populations of neurons with localised tuning curves densely cover the range of values taken by a given latent variable, can encode probability distributions over values of that variable (Zemel et al., 1998) and facilitate appropriate generalisation by being constrained to firing patterns with a correspondence to performance of the task.

How might task-relevant latent variables be identified? As described above, the hippocampal role in episodic memory can be framed as finding and storing compressed representations of sensory experience that can later be used to re-construct the full sensory representation (see Section 16.4.4). In machine learning, the same problem is addressed by ‘autoencoders’ – feed-forward neural networks that are trained to reconstruct input patterns in the output layer via a much smaller intermediate layer of neurons (Hinton, 1989). Gluck and Myers (1993) extended this framework by describing the hippocampus as a ‘predictive autoencoder’: a network that can reproduce the input pattern as well as classifying the ‘outcome’ with which it is associated. This constrains the network to find internal representations that differentiate input patterns which predict different outcomes, serving to dissociate identical perceptual states that are

associated with different actions and thus identify task relevant latent variables (see also Benna and Fusi, 2021).

More recent models have demonstrated similar functionality by incorporating representations of behavioural context. Specifically, latent variables can be identified by accounting for the prospective or retrospective behaviour or neural activity associated with a specific perceptual input. For example, Recatanesi et al. (2021) demonstrated that training a recurrent neural network to predict upcoming sensory inputs produces representations of the low dimensional latent variables defining a range of different tasks, including place cell like responses during spatial tasks. Similarly, George et al. (2021) demonstrated that disambiguating sensory representations according to the recent history of activity successfully disambiguated latent states associated with the same perceptual inputs, accounting for the splitter cell and lap cell responses described above. It is also possible that extracting the most slowly varying features of sensory input may identify underlying statistical regularities which encode useful latent variables. As described earlier, this ‘slow feature analysis’ has been used to account for the emergence of place, grid and head direction coding (i.e. account for the identification of latent spatial variables) from visual input (Franzius et al., 2007; Section 16.2.2).

16.5.2 Predicting Future State Occupancy

In the same way that spatial representations in the rodent hippocampus (see Section 16.2) are believed to support navigation (see Section 16.3), the purpose of the latent variable representations described above is to support efficient planning across a range of task domains by predicting the outcome of different actions. Ideally, the mechanisms that support planning should also generalise easily between domains to make the learning of novel tasks more efficient by leveraging previous experience. As discussed earlier, planning behaviour in arbitrary state spaces has been formalised by RL (e.g. Sutton and Barto, 1988), in which the ‘value’ of states is estimated in terms of the expected future reward to be gained discounted by how far into the future it will be reached. In particular, it has been proposed that hippocampal pyramidal cells might encode a successor representation (SR; i.e. the expected discounted future occupancy of other states from each starting state) to facilitate straightforward planning in arbitrary spatial and non-spatial state spaces (Stachenfeld et al., 2017; see Section 16.3.2). Importantly, the SR is more flexible than model-free RL because it can still be used to compute value if the distribution of rewards changes. However, the SR is biased by the transitions experienced during learning and therefore policy-dependent and sensitive to the distribution of previous rewards. If the agent’s policy or the transition structure of the task changes, the SR will no longer reflect future state occupancy. One solution is to learn the SR under a default policy (e.g. random exploration) and use that ‘default representation’ (DR) for planning under new goal-directed policies, taking account of deviations from the default policy, if those are not too great (Todorov, 2007; Piray and Daw, 2021). We also note the potential role for ‘predecessor’ (Namboodiri and Stuber, 2021) and ‘first occupancy’ (Moskovitz et al., 2021) representations in planning as alternatives to the SR.

However, if the hippocampus encodes the probability distribution of *future* state occupancy under a given policy, how does this relate to the idea that the hippocampus encodes the probability distribution of *current* state occupancy in these arbitrary spaces? Representations

of current and predicted future state occupancy are related by the transition matrix of a given task and policy. Specifically, the predicted occupancy of states at the next time step can be estimated by taking the current state-occupancy distribution and multiplying it by the transition matrix (see Fig. 16.13). Similarly, the SR (or DR) can be estimated by repeatedly multiplying the current state-occupancy distribution by the transition matrix and a temporal discount factor. We note that the transition matrix itself could reflect movements between locations in spatial navigation, or the relations between latent states in any cognitive map (Tolman 1948; Eichenbaum and Cohen, 2014; Behrens et al., 2018). The models in this section assume that the transition matrix is implicitly encoded within the hippocampal representation of current state occupancy. Specifically, as noted in Section 16.3.3, the covariance of state occupancy representations (e.g. place cell firing patterns) effectively encodes the transition matrix between those states – cells encoding adjacent states are more likely to be coactive, and those with distal states less likely to be coactive. As discussed earlier (see Section 16.2.2), there is also an apparent similarity between grid cell firing patterns and the eigenvectors of the covariance matrix between place cell firing patterns (Castro and Aguiar, 2014; Dordek et al., 2016).

Under this formulation, then, the hippocampus provides a probabilistic representation of state occupancy that closely resembles (and may be equivalent to) the SR (or DR), the covariance of this representation encodes the transition matrix between states during exploration, and grid cells represent the eigenvectors of that transition matrix. Appealingly, these eigenvectors have several features that could facilitate planning. Specifically, a weighted sum of their firing rates can represent the current state occupancy distribution (analogous to grid cell to place cell models; see Section 16.2.2). In addition, multiplying the representation of state occupancy by the transition matrix to compute future state occupancy simply corresponds to changing the weights from each grid cell by the eigenvalue corresponding to the eigenvector represented by its firing pattern (see Fig. 16.13). Thus, a simple weighted sum of the firing of grid cells could estimate future state occupancy and compute the SR (or DR) for use in planning (Corneil and Gerstner, 2015; Stachenfeld et al., 2017; Baram et al., 2018). Importantly, this reweighting process does not necessarily involve changes in synaptic weights but could be implemented by simply modulating grid cell firing rates. Indeed, several empirical studies have provided evidence of a role for grid firing patterns in planning movements through abstract state spaces. For example, Constantinescu et al. (2016) studied this kind of ‘conceptual navigation’ using a task in which participants predicted the types of bird that would be reached by a trajectory through a 2-D space of varying neck and leg length. They found patterns of metabolic activity with six-fold (or ‘hexadirectional’) modulation in the entorhinal and medial prefrontal cortices that may be consistent with the presence of grid firing patterns, which also observed during navigation through virtual spatial environments (Doeller et al., 2010).

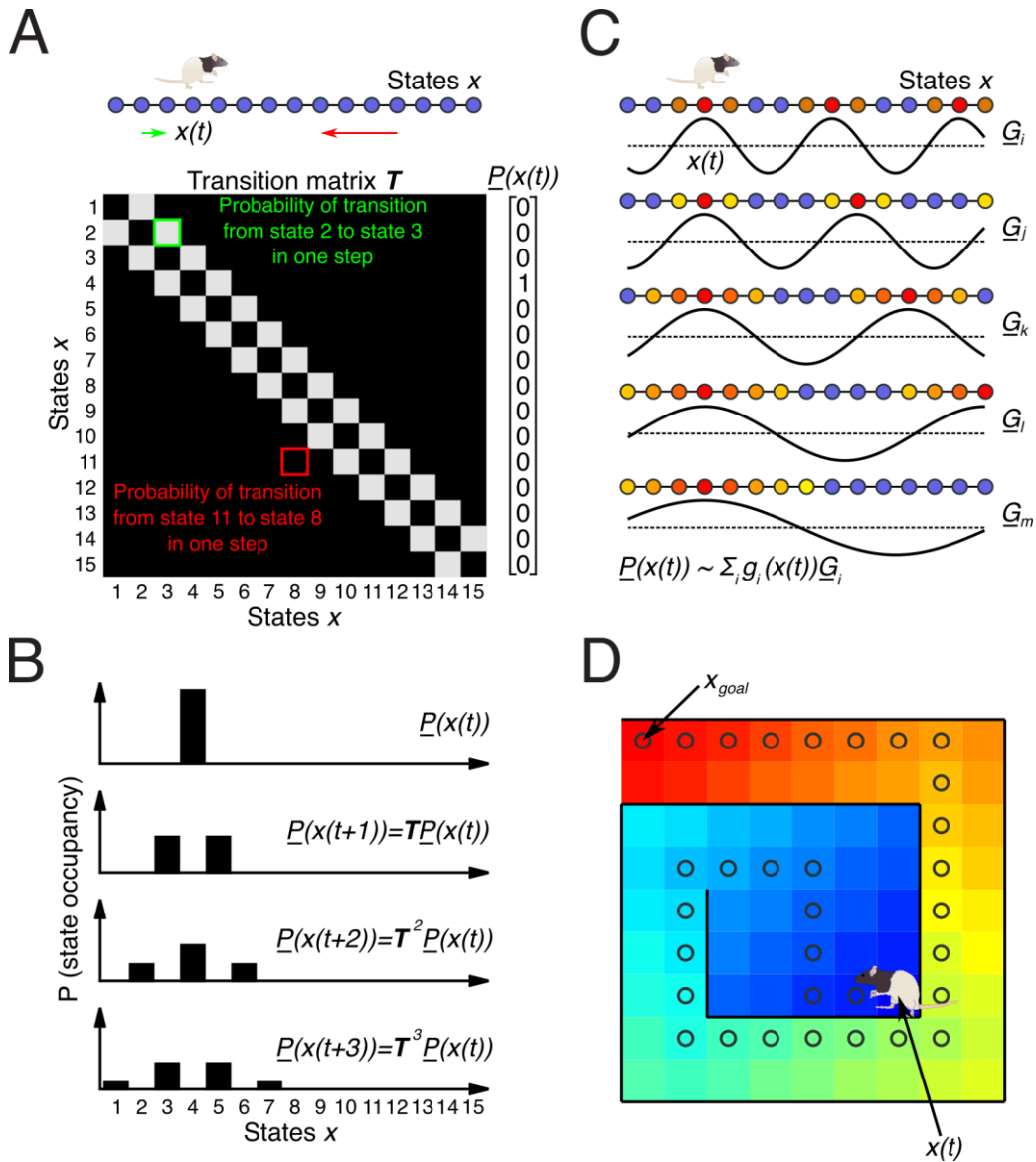


Figure 16.13 Prediction with transition matrices and grid-like eigenvectors. **A:** Many (Markovian) tasks can be conceived as a set of discrete states (for example locations x) where the probability of transitions between states is captured by a transition matrix T . This is illustrated by the transition matrix T for a mouse with a policy of moving one step left or right along a linear track in each time step with equal probability. **B:** Pre-multiplying the current state occupancy distribution $\underline{P}(x(t))$ by T predicts state occupancy at the next time step $\underline{P}(x(t+1))$, or at any subsequent time step. **C:** Grid firing patterns could be eigenvectors of a diffusive transition matrix reflecting random movements (i.e. $T\underline{G}_i = \lambda_i \underline{G}_i$). In this case, using a grid cell population vector (i.e. summing grid cell firing profiles weighted by their firing rates $G_i(x(t))$) to represent current state occupancy (i.e., $\underline{P}(x(t)) \sim \sum_i G_i(x(t)) \underline{G}_i$), means that pre-multiplying the current state occupancy $\underline{P}(x(t))$ by the transition matrix simply re-weights each grid cell's firing rate within the population vector by its eigenvalue (λ_i), so that: $\underline{P}(x(t+1)) \sim \sum_i G_i(x(t)) \lambda_i \underline{G}_i$. This type of weighted sum could be performed by a single layer readout network. **D:** The temporally discounted future state occupancy over multiple steps (i.e. the 'successor representation') $\underline{P}_\gamma(x(\tau \geq t))$ sums the effects of pre-multiplying by the transition matrix multiple times, each time weighted by the discount factor (γ), producing a power series whose value is again a simple reweighting of each grid cell's firing within the population vector. This example shows how the discounted future probability of a goal location x_{goal} is simply a weighted sum of the products of grid cell firing rates at the current and goal locations: $P_\gamma(x(\tau \geq t) = x_{goal}) \sim \sum_i G_i(x(t)) G_i(x_{goal}) / (1 - \gamma \lambda_i)$. It increases as the current location moves towards the goal and could therefore be used to guide navigation if T captures the transition structure of the environment under random exploration. Adapted from (Baram et al., 2018).

A single transition structure - such as that resulting from random exploration - can be used for planning, as described above, but cannot predict the effects of multiple different actions, as in models of path integration in which grid cell firing predicts the effects of successive directed actions on state occupancy. Indeed model-based RL, or dynamic programming, generally depends on exploring the effects of sequences of specific actions. One solution is to create multiple ‘first occupancy’ representations corresponding to different policies, and search through the effects of sequentially combining them (Moskovitz et al., 2021). An alternative solution, using a single set of eigenvectors, is available for translation-invariant transition structures, in which a transition or action has the same effect irrespective of the current state, allowing perfectly factorised state and transition representations (see Section 16.5.3). Fourier basis functions (i.e. plane waves of differing frequency and orientation) are eigenvectors for any translation-invariant transition matrices. Thus, a single Fourier set could act as eigenvectors for multiple translation invariant transition matrices corresponding to different actions, with different eigenvalues for each action; and sums of eigenvectors weighted by the appropriate eigenvalues could represent future state occupancy under different combinations of actions. This provides a potentially unifying framework for grid cell models: both in facilitating prediction and planning, and as the effects of the complex eigenvalues on the phases of the Fourier components corresponds to oscillatory interference models of grid cell firing, while the dynamics of the evolution of state occupancy is equivalent to continuous attractor models of path integration (Yu et al., 2020).

16.5.3 Factorising States and Transition Structure for Prediction

The rapid generalisation of existing knowledge to new situations is a crucial skill. One way to approach this problem is to factorise tasks into the structure of transitions between states and the contents of each state. In this way, common structure can be learned across several tasks in parallel to the contents of each state which can vary from task to task (Behrens et al., 2018). According to the formulation described in the previous section, grid cells in MEC might represent abstract transition structure while environmental sensory inputs are conveyed via lateral entorhinal cortex (LEC), and place cells represent the conjunction between transition-related and environment-related information (Manns and Eichenbaum, 2006; Knierim et al., 2006). This is consistent with results obtained in virtual environments wherein these two sources of (movement and sensory) input can be dissociated, showing that place cell firing is more influenced by environmental visual inputs while grid cell firing is more influenced by physical self-motion (Chen et al., 2019), and generalises those results to non-spatial tasks. The proposed factorisation of transition structure and content applies most completely to situations in which the transition structure is translation invariant and can therefore be learned and applied across both states and tasks.

The pre-eminent example of such a transition structure is Euclidean space, whose rules apply across all spatial environments (e.g. moving one step North, then East, then South, then West always brings you back to the same location). The place, head-direction and grid cell system may therefore have evolved to perform a factorisation of structure and content under pressure to perform spatial navigation in familiar and new environments. This machinery provides a pre-configured representation of structure that can be used to organise external experience (McNaughton et al., 1996; Buzsaki and Tingley, 2018). This mechanism could also confer

rapid generalisation across any tasks with a common relational structure, by attaching contents (or concepts) to locations whose transitions (or relationship) to all other locations are pre-defined, i.e. by constructing a cognitive map (see Tolman, 1948; O’Keefe and Nadel, 1978; Eichenbaum and Cohen, 2014). The formation of an efficient representation of these relationships over different sets of content allows rapid generalisation of the prediction of future states to new situations with similar relational structure (as discussed in Section 16.5.2), and the ability to form efficient structural representations corresponds to the representations of the states themselves lying on a low dimensional manifold (as discussed in Section 16.5.1). These insights apply to continuous inputs and state spaces, but they should also apply more generally to any set of discrete states with graph-like transition structures. However, we note that learning to map a transition-driven structural representation onto a sensory-driven representation of state-content, when both representations may contain noise or uncertainty, can be difficult. This is a well-known problem in robotics (a.k.a. simultaneous localisation and mapping) and may require non-local propagation of mismatch errors, which has been proposed as a potential role for hippocampal replay (Evans and Burgess, 2020).

The ‘Tolman-Eichenbaum Machine’ (‘TEM’, Whittington et al., 2020) aimed to test the idea that optimising neural network parameters for prediction of future state information could explain neural responses in the hippocampal formation (see also Uria et al., 2020) and demonstrate the value of factorising transition structure and state content. This model was comprised of three components: an MEC network with action-dependent recurrent connectivity that encodes and updates an estimate of abstract location (in a manner similar to continuous attractor network models, see Section 16.2.3); an LEC network that encodes sensory inputs; and a recurrently connected hippocampal network that encodes task specific representations of location as conjunctions of the inputs arriving from the MEC and LEC via Hebbian learning (as in models of associative memory, see Section 16.4). The model was trained, in discrete graph-like state spaces, to predict sensory inputs at the next state given previous sensory inputs and actions (i.e. state transitions) via an error-correcting learning rule. Importantly, training was performed across multiple tasks with the same transition structure but different contents and applied to both spatial and non-spatial domains (see Fig. 16.14). After random exploration of these state spaces, the TEM learns representations in the hippocampal network that resemble place cell responses and representations in the MEC network that include grid-like firing patterns. Interestingly, if the agent systematically approaches objects, the firing patterns of some units in the MEC network also resemble object vector cells (Hoydal et al., 2019), while some units in the hippocampal network resemble landmark vector cells (Deshmukh and Knierim, 2013; see Fig. 16.14).

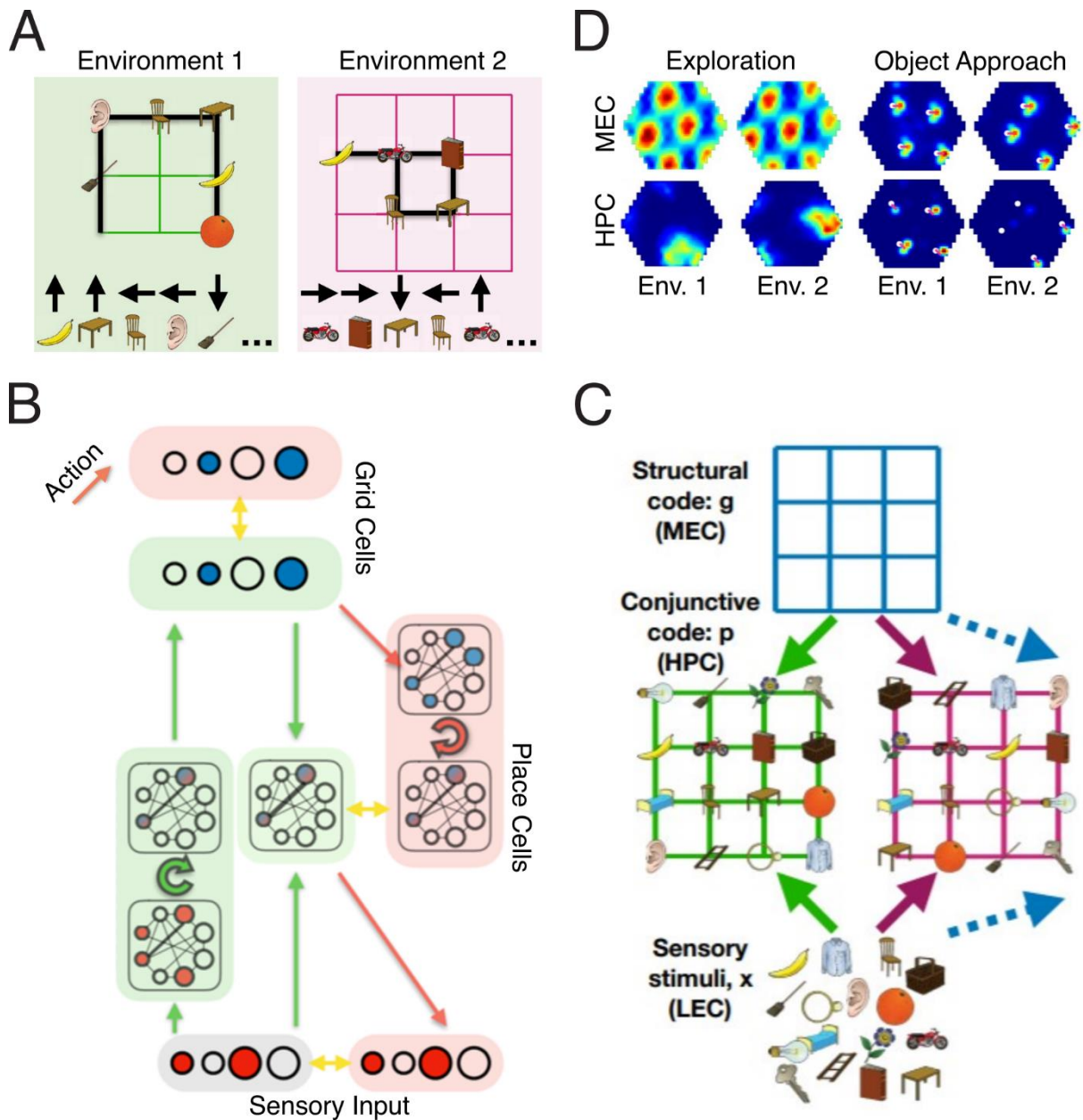


Figure 16.14 The Tolman-Eichenbaum Machine. **A:** The agent learns to predict the next item (e.g. banana, table, chair etc) as a function of the previous item and the transition/action taken (e.g. North, South, East, West) across several environments. The possible actions and underlying structure (i.e. the graph) remain constant, but the allocation of items to states (i.e. graph vertices) varies across environments. An object-location task is shown, but non-spatial structures, e.g. family trees, can also be learnt. **B:** The network architecture can be seen as a variational autoencoder in which ‘encoding’ connections (shown in green) propagate a representation of recent sensory input up through a recurrent place cell layer (‘HPC’) to a recurrent grid cell layer (‘MEC’), and ‘decoding’ connections (shown in red) predict the next grid, place and sensory representation. Recurrent connections between grid cells are action-dependent (allowing prediction, given the action taken in the previous time step, top left corner) and recurrent connections between place cells form a Hopfield-like memory for the item-state associations in the current environment. Yellow arrows show differences between ‘predicted’ and ‘encoded’ representations whose minimization drives learning. **C:** The network develops representations that capture the transition structure in the MEC layer, and representations of states in the HPC layer determined by the conjunction of transitions and sensory information. **D:** During random exploration, MEC neurons develop grid-like responses and HPC neurons develop place-like responses, but if systematically approaching objects, these become more like object vector cell and landmark vector cell responses, respectively. Adapted from (Whittington et al., 2020).

More generally, the ability to predict future state occupancy is important for efficient RL and has motivated models that go beyond the discounted average over all future time steps captured by the SR (or DR). Deep networks that learn from large numbers of training examples to map sensory inputs to actions that maximise return (e.g. Mnih et al., 2015) benefit from an ability to read and write to an external memory system when relevant information is only partially observable (Graves et al., 2016). In RL settings, however, such memory systems do not work well without a predictive internal model which ensures that the format of stored representations allows the prediction of future observations. For example, Wayne et al. (2018) combined an autoencoder with a recurrent network that can learn over long timescales (Hochreiter and Schmidhuber, 1997) in order to predict future sensory inputs from current observations. This necessitated the generation of latent variable representations which could be stored in memory and accessed by a separate policy network that learns to produce optimal actions. Importantly, the autoencoder network was trained to predict both the sensory inputs (requiring it to develop a model of sensory dynamics) and the estimated value of each state (i.e. discounted future reward). This strategy, similar to Gluck and Myers (1993), encourages the emergence of representations which emphasise features that are salient for future reward. This network could subsequently perform latent learning and flexible navigation within virtual environments. Aside from the biologically implausible aspects of training deep networks (error backpropagation requiring a detailed record of non-local neural activity to guide synaptic change), this work highlights the utility of developing mnemonic representations to predict future observations and rewards through unsupervised learning, complementing representations driven by direct reinforcement. Indeed, many of the models described in Sections 16.5.1-3 stress the hippocampal contribution to this role.

16.5.4 What is the Role of Episodic Memory?

Although the models described above can account for hippocampal representations of both spatial and non-spatial state spaces to support planning, prediction, and generalisation across tasks with similar structure, it is less clear how they relate to memory for individual episodes as considered by Marr (1971). Marr (1971) and related models of memory conceive the role of hippocampus in terms of storing and retrieving single episodes (using orthogonal representations to avoid interference), so that statistical structure across episodes could be extracted into semantic memory in neocortex (using invariant representations to support generalisation). This dichotomy (between representing individual events and common structure) raises similar questions as the debate concerning pattern separation, reflecting the theoretical advantages of orthogonal representations for disambiguating similar events versus the apparently invariant representations of event elements observed in place and concept cell recordings (see Section 16.4.4). Hence, it is still unclear how to reconcile the two proposed roles of the hippocampal formation: in generating cognitive maps for planning and in storing the contents of individual episodes. It is possible that the region mediates both roles, as suggested by TEM, by extracting low dimensional latent variable representations (such as those for space and time) across multiple events in entorhinal cortex, while making use of associative memory in the hippocampus proper (i.e. CA3) to store specific experiences as conjunctions of those variables with the sensory contents of each event. The hippocampus could subsequently aid the learning of latent variable representations by replaying specific events until their

statistics are fully captured (see Section 16.4.5). This would result in a separation of ‘predictable’ episodes, which are consistent with prior experience and can be rapidly consolidated (e.g. Tse et al., 2007; Tse et al., 2011); and ‘unpredictable’ episodes, which would degrade the generalisation of entorhinal cognitive maps and should therefore be retained in the hippocampus proper (e.g. Sun et al., 2021).

Importantly, a similar process (i.e. the slow extraction of even longer-range statistical regularities) could operate from entorhinal cortex to more distant cortical structures (such as ventromedial prefrontal cortex) to explain ‘systems consolidation’, by which knowledge becomes fully independent of the hippocampal formation. However, it is possible that use of the latent variables associated with space and time to fully reconstruct images of past events, or to imagine new experiences, remains dependent on the hippocampus, unlike more abstract semantic knowledge (Nadel and Moscovitch, 1997; see Section 16.4.5). We note the similarity with theoretical models that examine the role of episodic memory in supporting planning within an RL framework. In particular, Lengyel and Dayan (2007) described how ‘episodic control’ supports more efficient behaviour than either model-free or model-based RL when the transition and reward structure of a task have been poorly sampled. Specifically, if an agent has discovered any sequence of actions that lead to reward and is under pressure to limit exploration of an environment, then repeating that sequence of actions is more advantageous than trying to construct and exploit a complex model or an incomplete value function (see also Zilli and Hasselmo, 2008; Gershman and Daw, 2017; Pritzel et al., 2017; Botvinick et al., 2019).

16.6 Conclusions

As in other areas of neuroscience, the ability to specify a proposed mechanism of hippocampal function as a computational model has been invaluable in many ways. Firstly, it allows potential mechanisms to be specified precisely, reducing the ambiguity and potential ‘hypothesis drift’ that might afflict verbal descriptions. Secondly, it enables simulation of the resulting behaviour which, at the very least, can serve as a demonstration of the viability of a given proposal. Beyond this it enables quantitative, falsifiable predictions regarding the effects of experimental manipulations to be made at the levels of cells, systems and behaviour. Some of the models reviewed here have been used to make such predictions and have also contributed to the design of experiments to test them. Nonetheless, the value of a computational model lies in the insights it provides, just like other forms of scientific endeavour.

Numerous questions for future research are prompted by these models, not least how to resolve the apparent differences between the proposed hippocampal roles in cognitive mapping and episodic memory. Computational models have been invaluable in attempting to unify these two largely disparate streams of research. Specifically, by suggesting that the hippocampus encodes an index of features that identify the current event amongst perceptually similar events in episodic memory, consistent with representing the current ‘location’ in a low dimensional task-relevant latent state space during cognitive mapping. In addition, cognitive mapping models propose a role for entorhinal cortex in embedding these latent state representations within the structure of transitions or relationships between them. Beyond the efficient storage and retrieval of experience via latent states (i.e. a memory system), this supports planning: the ability to infer

the effects of applying transition or relational operators to the latent state representations (equivalent to path integration in the spatial domain). In situations in which the transition or relational structure is conserved across states or tasks spatial navigation being the preeminent example), the factorisation of states and transitions allows rapid generalisation. This framework is consistent with the original idea of the hippocampus providing a cognitive map or efficient representation of states or concepts and the relationships between them for planning and inference. Nonetheless, many questions remain - such as how this framework relates to the consolidation of mnemonic information. Future modelling will undoubtedly refine our understanding of the specific mechanisms at work, by informing empirical studies and being modified by their results.

Acknowledgements

The authors wish to thank Andrej Bicanski, Jesse Geerts, Zilong Ji, Eleanor Spens, and Oliver Vikbladh for useful discussions and feedback during the preparation of this chapter.

Text Boxes

Box 16.1 Learning via synaptic modification: 'Hebbian' learning rules

In the simplest type of neural network model, the firing rate or 'activity' of a neuron (a) is simply a function (the 'transfer function' f) of the net current coming into the neuron, which in turn is simply a weighted sum of the firing rates (u_i) of the neurons connecting to it. That is: $a=f(\sum_i w_i u_i)$, often written as: $a=f(\underline{w} \cdot \underline{u})$, where \underline{w} is the vector of connection 'weights' modelling the strengths (e.g. net synaptic efficacy) of connections from the input neurons, and \cdot is the vector dot product. With the simplest, linear, transfer function, the activation is given by:

$$a = \underline{w} \cdot \underline{u}$$

Equation 1

In such networks, 'learning' corresponds to modification of the connection weights \underline{w} . Below we discuss some of the 'Hebbian' learning rules mentioned in the rest of the chapter, and their effects, following the discussion in (Dayan and Abbott, 2002), where further details can be found.

A learning rule directly implementing Hebb's (1949) postulate of coincident firing leading to increased coupling between neurons describes the change in connection weights in terms of the product of pre- and post-synaptic firing rates:

$$\tau \frac{dw_i}{dt} = au_i, \text{ or } \tau \frac{d\underline{w}}{dt} = \underline{a}\underline{u},$$

Equation 2

where τ gives the rate of change of connection weights with time. When this rule is applied to a 'training set' of n example input patterns of activity \underline{u}^μ , each presented for an equal duration over a total time T , we can integrate Equation 2 to see the total change in \underline{w} :

$$\underline{w} \rightarrow \underline{w} + \frac{T}{\tau} \sum_{\mu} a^{\mu} \underline{u}^{\mu},$$

Equation 3

where $a^{\mu} = \underline{w} \cdot \underline{u}^{\mu}$ from Equation 1. If the connection weights are only updated after presentation of all of the input patterns then we can say:

$$\underline{w} \rightarrow \underline{w} + \frac{T}{\tau} \sum_{\mu} (\underline{w} \cdot \underline{u}^{\mu}) \underline{u}^{\mu} = \underline{w} + \frac{nT}{\tau} \underline{Q} \underline{w},$$

Equation 4

where \underline{Q} is the correlation matrix of the input patterns ($\underline{Q} = \langle \underline{u} \underline{u} \rangle$ where $\langle \rangle$ denotes the average over input patterns, and $\underline{u} \underline{u}$ is the outer product of \underline{u} with itself). Thus simple Hebbian learning rules are also known as correlation based learning rules. Inspection of Equation 4 indicates that the weight vector \underline{w} , if plotted in the same space as the input vectors \underline{u}^{μ} , will eventually follow the principal eigenvector of the correlation matrix, i.e. it will lie along the direction from the origin to the mean input pattern ($\langle \underline{u} \rangle$) or, if $\langle \underline{u} \rangle$ is at the origin, along the first principal component of the set of input patterns. However, this learning rule is not stable: large weights produce large output activations which produce large increases in weights and so on. More formally, it can be seen from the dot product of \underline{w} with Equation 2 that the length of the weight vector increases whenever the output neuron is active:

$$\frac{d |\underline{w}|^2}{dt} = 2 \underline{w} \cdot \frac{d \underline{w}}{dt} = \frac{2 a \underline{w} \cdot \underline{u}}{\tau} = \frac{2 a^2}{\tau}.$$

Equation 5

One way to introduce balance into the learning rule is to allow for a connection weight to increase or to decrease according to the levels of pre- and post- connection activity, by analogy with long-term potentiation and depression (LTP and LTD, see Chapter 10). In this way Equation 2 could become:

$$\tau \frac{d \underline{w}}{dt} = (a - \theta) \underline{u}, \text{ or } \tau \frac{d \underline{w}}{dt} = a(\underline{u} - \underline{\phi}),$$

Equation 6

where either a postsynaptic threshold or a set of pre-synaptic thresholds are applied to determine the sense and size of weight changes (respectively: θ - the level postsynaptic activity must surpass for the connection to increase rather than decrease; or $\underline{\phi}$ - the vector of activity levels each input neuron must surpass). The most obvious choice of threshold for the pre- or post-synaptic neuron is its average activity over the training set. In this case, following a similar derivation to Equation 4, both versions produce the same learning rule:

$$\underline{w} \rightarrow \underline{w} + \frac{nT}{\tau} \underline{C} \underline{w},$$

Equation 7

where C is the covariance matrix of the input patterns: $C = \langle (\underline{u} - \langle \underline{u} \rangle)^2 \rangle$. These learning rules are also known as covariance rules. Inspection of Equation 7 indicates that the weight vector will eventually follow the principal eigenvector of the covariance matrix, i.e. it will lie along the direction of the first principal component of the set of input patterns. It should be noted that these rules are also not stable, in this case $d|w|^2/dt$ is proportional to the variance ($\langle a^2 \rangle - \langle a \rangle^2$) of the output activity over the training set.

The BCM learning rule, derived from experimental investigation of visual cortical plasticity (Bienenstock et al., 1982), proposes that:

$$\tau \frac{dw}{dt} = au(a - \theta(a)).$$

Equation 8

This requires both pre- and post- synaptic activity for modification of a connection weight (unlike the rules in Equation 6), and also involves a sliding post-synaptic threshold ($\theta(a)$) which varies with post synaptic activity. So long as the post-synaptic threshold increases as a power of a greater than 1 (typically following a time-averaged estimate of a^2), it can ensure stability of the learning rule: effectively increasing the threshold to an overactive output neuron so that connection weights to it tend to be reduced.

The other common way in which Hebbian learning rules are stabilized (e.g. in Rumelhart and Zipser's (1986) 'competitive learning' algorithm) is to use divisive normalization: explicitly constraining the length of the weight vector to remain constant during learning by dividing all weights by $|w|$. Although this is a non-local operation, since synaptic strengths must be altered according to the state of other, distant, synapses, a similar effect can be achieved by the local learning rule of Oja (1982):

$$\tau \frac{dw}{dt} = au - \beta a^2 w.$$

Equation 9

A similar analysis to Equation 5 shows that $|w|^2$ will tend to a value $1/\beta$ under repeated application of this rule.

As well as being stable, the BCM and normalized Hebbian learning rules involve competition between connections: increasing some of the connection weights onto a neuron will lead to a decrease in the others. Under the BCM rule this occurs due to increased activity leading to a higher post-synaptic threshold and thus an increased incidence of LTD versus LTP. With normalization this occurs directly due to the increase in the length of the weight vector. These learning rules tend to allow a neuron to become tuned to respond to specific patterns of input activation, see text. There is at least some evidence for competitive interaction between synapses such that increasing the strengths of one set of synapses leads to a decrease in the strengths of others so as to normalize the total synaptic strength onto the neuron (Royer and Pare, 2003). We note, however, that evidence for the dependence of synaptic plasticity on the precise timing of pre- and post-synaptic activity (see Chapter 10) changes the likely nature of

learning rules based on LTP and LTD, as in the effects of temporal asymmetry noted in Section 16.2.2.

Box 16.2 Population Vectors

Let us assume a population of neurons in which each fires according to the distance between some variable and a preferred value of that variable. This could be the case for place cells if, for example, we assume that the firing rates r_i are simply a Gaussian function of the distance between the animal's current location and preferred locations x_i . We can subsequently estimate or 'decode' the animal's current location from the firing rate weighted average of this vector, also known as the 'population vector' (following Georgopolous et al., 1986):

$$\frac{\sum r_i x_i}{\sum r_i}$$

Equation 10

This estimate is simple to compute, requiring only the firing rates and preferred values, but will be biased unless the preferred values evenly cover the range of possible values. Such even coverage might be true for head direction cells, but a population vector of place cell firing recorded within an open arena will be biased towards the centre (Muller et al., 1987). More generally, population vector estimates will be less accurate than methods that take into account the function relating firing rates to the encoded variable (i.e. the shape of the firing rate profile; see Dayan and Abbott, 2002).

Box 16.3 Attractors in memory, neural coding and path integration

Point attractors and memory

A network of recurrently connected neurons can be arranged so that a finite number of discrete patterns of activation across the neurons are stable states or 'attractors'. This means that any pattern of activation similar enough to one of these attractors will evolve into that attractor pattern under the dynamics of the network. These patterns of activation are 'stored' in the network in the sense that they will be 'retrieved' from any initial pattern that is similar enough. Such networks are also referred to as 'auto-associative' and are an example of a 'content-addressable' memory, in that a pattern of activity is retrieved by a pattern of similar content rather than, say, an unrelated index term or the address of a storage location.

In one of the simplest models (Hopfield, 1982), activity of neuron I is modelled as $a_i = \pm 1$. Connections between neurons I and j are symmetric, with synaptic 'weight' $w_{ij} = w_{ji}$. The dynamics of the network are given by: $a_i(t+1) = \text{sign}(\sum_j w_{ij} a_j(t))$, such that the 'energy' or Lyapunov function of the network:

$$E \propto -\sum_{ij} w_{ij} a_i a_j ,$$

Equation 11

can only reduce. If connection weights undergo a form of Hebbian learning when the to-be-stored patterns of activation (\underline{a}^μ , say) are present, such that $w_{ij} \propto \sum_\mu a_i^\mu a_j^\mu$, then these representations will become attractors, so long as the number of stored patterns is not too large (less than around $0.14N$, where N is the number of neurons, in this case). That is, a similar enough pattern of activation will converge onto the stored pattern under the dynamics of the network (see also Cohen and Grossberg, 1983). This situation is often visualized by imagining how the 'energy' of the network varies as a function of the network 'state' $\underline{a} = (a_1, a_2, \dots, a_N)$ – the attractor states being local minima of the energy surface to which nearby states will evolve under the network's dynamics (see Fig. 16.3A). Similar behaviour is also shown by more biologically realistic models (Amit, 1992; Treves and Rolls, 1992; McClelland et al., 1995; see also Fig. 16.9). We note that 'modern' Hopfield networks or 'dense associative memories' (e.g. Krotov and Hopfield, 2016; Demircigil et al., 2017) exhibit much greater storage capacity, in addition to reduced interference between stored patterns, but biologically realistic implementations of those networks have yet to be proposed (but see Krotov and Hopfield, 2020).

Line attractors and neural coding

The value of a continuous variable (or 'stimulus' s) often seems to be represented in the firing rates of a population of neurons, each of which is tuned to respond preferentially to a single 'preferred' value. For example, 'head-direction cells' can be thought of in this way, with s representing the rat's heading. The pattern of activation of the population is often visualized by imagining the neurons arranged so that their location reflects their preferred values: showing a smooth bump of activity across the neurons peaked at the actual value of the stimulus. However, if the firing rates are noisy it will be difficult to estimate the precise value of the stimulus. The presence of recurrent connections between neurons, arranged so that the weight of the connection between each pair is simply a decreasing function of the difference in their preferred values (or physical separation when arranged as above), can help by ensuring that the firing pattern takes the shape of a smooth bump¹ (see Fig. 16.3B). With the appropriate choice of recurrent connections, such a network can perform optimal decoding (Latham et al., 2003), including the situation where the representation is formed from different unreliable sources of information (Deneve et al., 2001).

The patterns of activation comprising a smooth bump can be thought of as a line in the N dimensional state space $\underline{a} = (a_1, a_2, \dots, a_N)$ of the network. Each point on the line corresponds to a different estimate of s (referred to as \hat{s}) Conversely, all of the possible noisy patterns of activation that end up producing the same \hat{s} lie on an $N-1$ dimensional subspace within which the action of the recurrent connections corresponds to convergence onto the line (see Fig. 16.3C). An important aspect of these networks is that, while the recurrent connections ensure that patterns of activation move onto the line attractor, movement along it, corresponding to changing \hat{s} , is not affected by the recurrent connections (since the connections between a pair

¹ these patterns have low 'energy' as activation is concentrated in nearby neurons, which have the strongest interconnections.

of neurons depends only on the *difference* in their preferred values, not what those preferred values are).

Line attractors and path integration

Since the (symmetric) recurrent connections provide no resistance to motion of the bump of activity along the line attractor, its position is easily moved by asymmetric connections from each neuron to neighbours further along the line (Zhang, 1996; Skaggs et al., 1995). The size of the asymmetric connections (which should correspond to the spatial derivative of the symmetric connections for the bump to move without changing shape; Zhang, 1996) compared to the symmetric ones dictates the movement of the bump (see Figs. 16.3-5). Thus, if the strength of the asymmetric connections is proportional to angular velocity, the location of the bump of activity in a ring of head-direction cells will track the head direction of the animal – performing angular 'path integration' (see Redish et al., 1996 for an alternative model using only neurons with conjunctive tuning to head direction and angular velocity).

As noted by Zhang (1996) and McNaughton et al. (1996), the angular path integration models of head-direction cell firing can be extended to path integration models of place cell firing. In this case, the place cells are imagined as a 2-D array, so that the location of each neuron corresponds to the location of its place field in the environment (see Fig. 16.4). Again, symmetrical connections decreasing in strength with the physical separation of the pre- and post-synaptic neurons can ensure that neural activity forms a single-peaked bump over the array, while asymmetric connections from each neuron to its neighbours along a given direction will cause the bump to shift in that direction (see Figs. 16.3-4). In this case, to perform path integration of position, the strength of the asymmetric connections between a pair of neurons displaced in a given direction needs to be proportional to the velocity of the rat in that direction (see Samsonovich and McNaughton, 1997 for a more detailed model; Droulez and Berthoz, 1991, and Dominey and Arbib, 1992, for related earlier models; and Conklin and Eliasmith, 2005; Burak and Fiete, 2009 for accurate path integration using only neurons with conjunctive tuning to location and velocity).

References

- Aggleton JP, Brown MW (1999) Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioural Brain Sci* 22: 425-490.
- Agmon H, Burak Y (2020) A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. *Elife* 9: e56894
- Alexander AS, Carstensen LC, Hinman JR, Raudies F, Chapman GW, Hasselmo ME (2020) Egocentric boundary vector tuning of the retrosplenial cortex. *Science Advances* 6 (8), eaaz2322
- Alme CB, Miao C, Jezek K, Treves A, Moser EI, Moser M-B (2014) Place cells in the hippocampus: eleven maps for eleven rooms. *PNAS* 111(52): 18428-35
- Almog N, Tocker G, Bonnevie T, Moser EI, Moser MB, Derdikman D. 2019. During hippocampal inactivation, grid cells maintain synchrony, even when the grid pattern is lost. *eLife* 8:e47147
- Alvarez P, Squire LR (1994) Memory consolidation and the medial temporal lobe: a simple network model. *Proc Natl Acad Sci U S A* 91: 7041-7045.
- Alvernhe A, Save E, Poucet B (2011) Local remapping of place cell firing in the Tolman detour task. *Eur J Neurosci* 33(9):1696-705
- Alyan S, McNaughton BL (1999) Hippocampectomized rats are capable of homing by path integration. *Behav Neurosci* 113: 19-31.
- Amaral DG, Ishizuka N, Claiborne B (1990) Neurons, numbers and the hippocampal network. *Prog Brain Res* 83: 1-11.
- Ambrose RE, Pfeiffer BE, Foster DJ (2016) Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. *Neuron* 91(5):1124-1136
- Amit DH (1989) *Modelling Brain Function*. Cambridge University Press.
- Amit DJ (1992) *Modeling Brain Function : The World of Attractor Neural Networks*. Cambridge University Press.
- Anderson JA (1995) *An Introduction to Neural Networks*. MIT Press.
- Arleo A, Gerstner W (2000) Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biol Cybern* 83(3):287-99
- Aronov D, Nevers R, Tank DW (2017) Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature* 543(7647):719-722
- Atance CM, O'Neill DK (2001) Episodic future thinking. *Trends in Cognitive Sciences* 5(12): 533-539
- Baddeley AD, Vargha-Khadem F, Mishkin M (2001) Preserved recognition in a case of developmental amnesia: implications for the acquisition of semantic memory? *Journal of Cognitive Neuroscience* 13: 357-369.

- Banino A et al. (2018) Vector-based navigation using grid-like representations in artificial agents. *Nature* 557(7705):429-433
- Bao X, Gjorgieva E, Shanahan LK, Howard JD, Kahnt T, Gottfried JA (2019) Grid-like neural representations support olfactory navigation of a two dimensional odor space. *Neuron* 102: 1066–1075
- Baram AB, Muller TH, Whittington JCR, Behrens TEJ (2018) Intuitive planning: global navigation through cognitive maps based on grid-like codes. *bioRxiv*
- Barry C, Burgess N (2007) Learning in a geometric model of place cell firing. *Hippocampus* 17: 786-800
- Barry C, Ginsberg LL, O’Keefe J, Burgess N (2012) Grid cell firing patterns signal environmental novelty by expansion. *PNAS* 109: 17687 – 17692
- Barry C, Hayman R, Burgess N, Jeffery KJ (2007) Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience* 10: 682-684
- Bartlett FC (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press
- Bassett JP, Wills TJ, Cacucci F (2018) Self-organized attractor dynamics in the developing head direction circuit. *Current Biology* 28: 609-615
- Battaglia, F.P. et al. (2004) Hippocampal sharp wave bursts coincide with neocortical ‘up-state’ transitions. *Learn. Mem.* 11, 697–704
- Baxendale SA, Van Paesschen W, Thompson PJ, Duncan JS, Shorvon SD, Connelly A (1997) The relation between quantitative MRI measures of hippocampal structure and the intracarotid amobarbital test. *Epilepsia* 38: 998-1007.
- Baxter MG, Murray EA (2001) Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. *Hippocampus* 11: 61-71.
- Becker S, Burgess N (2001) A model of spatial recall, mental imagery and neglect. *Advances in neural information processing systems* 13: 96-102.
- Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018) What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100 (2): 490-509
- Bellmund JLS, de Cothi W, Ruitter TA, Nau M, Barry C, Doeller CF (2020) Deforming the metric of cognitive maps distorts memory. *Nature Human Behaviour* 4: 177–188
- Benna MK, Fusi S (2021) Place cells may simply be memory cells: Memory compression leads to spatial tuning and history dependence. *PNAS* 118(51): e2018422118
- Bi G-Q, Poo M-M (1998) Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *Journal of Neuroscience* 18(24): 10464-10472

- Bicanski A, Burgess N (2016) Environmental anchoring of head direction in a computational model of retrosplenial cortex. *The Journal of Neuroscience* 36(46): 11601-11618
- Bicanski A, Burgess N (2018) A neural-level model of spatial memory and imagery. *eLife* 7: e33752
- Bicanski A, Burgess N (2019) A computational model of recognition memory via grid cells. *Current Biology* 29: 1–12
- Bicanski A, Burgess N (2020) Neuronal vector coding in spatial cognition. *Nature Reviews Neuroscience* 21: 453–470
- Bienenstock EL, Cooper LN, Munro PW (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci* 2: 32-48.
- Bisiach E, Luzzatti C (1978) Unilateral neglect of representational space. *Cortex* 14: 129-133.
- Blair HT, Cho J, Sharp PE (1998) Role of the lateral mammillary nucleus in the rat head direction circuit: a combined single unit recording and lesion study. *Neuron* 21: 1387-1397.
- Blair HT, Gupta K, Zhang K (2008) Conversion of a phase- to a rate-coded position signal by a three stage model of theta cells, place cells, and grid cells. *Hippocampus* 18: 1239-55
- Blair HT, Lipscomb BW, Sharp PE (1997) Anticipatory time intervals of head-direction cells in the anterior thalamus of the rat: implications for path integration in the head-direction circuit. *J Neurophysiol* 78: 145-159.
- Blair HT, Sharp PE (1995) Anticipatory head direction signals in anterior thalamus: evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. *J Neurosci* 15: 6260-6270.
- Blair HT, Wu A, Cong J (2014) Oscillatory neurocomputing with ring attractors: a network architecture for mapping locations in space onto patterns of neural synchrony. *Philosophical Transactions of the Royal Society B* 369(1635): 20120526
- Blum KI, Abbott LF (1996) A model of spatial map formation in the hippocampus of the rat. *Neural Comput* 8: 85-93.
- Boccaro CN, Nardin M, Stella F, O'Neill J, Csicsvari J (2019) The entorhinal cognitive map is attracted to goals. *Science* 363 (6434), 1443-1447
- Boccaro CN, Sargolini F, Thoresen VH, Solstad T, Witter MP, Moser EI, Moser MB (2010) Grid cells in pre- and parasubiculum. *Nature Neuroscience* 13: 987–994
- Bogacz R, Brown MW, Giraud-Carrier C. 2001. Model of familiarity discrimination in the perirhinal cortex. *J Comput Neurosci* 10:5–23
- Bonnevie T, Dunn B, Fyhn M, Hafting T, Derdikmann D, Kubie JL, Roudi Y, Moser EI, Moser M-B (2013) Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience* 16: 309-317

- Bostock E, Muller RU, Kubie JL (1991) Experience-dependent modifications of hippocampal place cell firing. *Hippocampus* 1: 193-205.
- Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, Hassabis D (2019) Reinforcement learning, fast and slow. *Trends in cognitive sciences* 23(5), 408-422
- Brandon MP, Bogaard AR, Libby CP, Connerney MA, Gupta K, Hasselmo ME (2011) Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science* 332: 595–599
- Brandon MP, Koenig J, Leutgeb JK, Leutgeb S (2014) New and distinct hippocampal place codes are generated in a new environment during septal inactivation. *Neuron* 82(4): 789-96
- Brown MA, Sharp PE (1995) Simulation of spatial learning in the Morris water maze by a neural network model of the hippocampal formation and nucleus accumbens. *Hippocampus* 5: 171-188.
- Brunel N, Trullier O (1998) Plasticity of directional place fields in a model of rodent CA3. *Hippocampus* 8: 651-665.
- Buetfering C, Allen K, Monyer H (2014) Parvalbumin interneurons provide grid cell-driven recurrent inhibition in the medial entorhinal cortex. *Nat Neurosci* 17(5):710-8
- Burak Y, Fiete IR (2009) Accurate path integration in continuous attractor network models of grid cells. *PLoS Computational Biology* 5: e1000291
- Burgess CP, Burgess N (2014) Controlling phase noise in oscillatory interference models of grid cell firing. *Journal of Neuroscience* 34: 6224-6232
- Burgess N (2008) Grid cells and theta as oscillatory interference: Theory and predictions. *Hippocampus* 18: 1157 - 1174
- Burgess N, Barry C, O'Keefe J (2007) An oscillatory interference model of grid cell firing. *Hippocampus* 17: 801 – 812
- Burgess N, Becker S, King JA, O'Keefe J (2001a) Memory for events and their spatial context: models and experiments. *Philos Trans R Soc Lond B Biol Sci* 356: 1493-1503.
- Burgess N, Hartley T (2002) Orientational and geometric determinants of place and head-direction. In: *Neural information processing systems* 14 pp 165-172. MIT Press.
- Burgess N, Hitch GJ (2005) Computational models of working memory: putting long term memory into context. *Trends Cogn Sci* 9: 535-541.
- Burgess N, Jeffery KJ, O'Keefe J (1999) Intergrating hippocampal and parietal functions: a spatial point of view. In: *The hippocampal and parietal foundations of spatial cognition* (Burgess N, Jeffery KJ, O'Keefe J, eds), pp 3-29. Oxford University Press.
- Burgess N, Maguire E, O'Keefe J (2002) The human hippocampus and spatial and episodic memory. *Neuron* 35: 625-641.

- Burgess N, Maguire EA, Spiers HJ, O'Keefe J (2001b) A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage* 14: 439-453.
- Burgess N, O'Keefe J (1996) Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus* 6: 749-762
- Burgess N, O'Keefe J (2011) Models of place and grid cell firing and theta rhythmicity. *Current Opinion in Neurobiology* 21: 734-744
- Burgess N, Recce M, O'Keefe J (1994) A model of hippocampal function. *Neural Networks* 7: 1065-1081.
- Bush D, Barry C, Burgess N (2014) What do Grid Cells Contribute to Place Cell Firing? *Trends in Neuroscience* 37: 136-145
- Bush D, Barry C, Manson D, Burgess N (2015) Using Grid Cells for Navigation. *Neuron* 87: 507-520
- Bush D, Burgess N (2014). A hybrid oscillatory interference/continuous attractor network model of grid cell firing. *Journal of Neuroscience* 34(14), 5065-5079
- Bush D, Burgess N (2020) Advantages and Detection of Phase Coding in the Absence of Rhythmicity. *Hippocampus* 30: 745-762
- Bush D, Philippides A, Husbands P, O'Shea M (2010) Dual Coding with STDP in an Auto-associative Network Model of the Hippocampus. *PLoS Computational Biology* 6(7): e1000839
- Butler WN, Hardcastle K, Giocomo LM (2019) Remembered reward locations restructure entorhinal spatial maps. *Science* 363: 1447-1452
- Butler WN, Smith KS, van der Meer MA, Taube JS (2017) The head-direction signal plays a functional role as a neural compass during navigation. *Current Biology* 27: 1259-1267
- Buzsáki G, Tingley D (2018) Space and Time: The Hippocampus as a Sequence Generator. *Trends Cogn Sci* 22(10):853-869
- Byrne P, Becker S, Burgess N (2007) Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological Review* 114: 340-375
- Cacucci F, Lever C, Wills TJ, Burgess N, O'Keefe J (2004) Theta-modulated place-by-direction cells in the hippocampal formation in the rat. *J Neurosci* 24: 8265-8277.
- Cai DJ et al. (2016) A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* 534(7605):115-8
- Carr, M.F. et al. (2011) Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nat. Neurosci.* 14, 147–153
- Cartwright BA, Collett TS (1983) Landmark learning in bees. *Journal of Comparative Physiology* 151: 521-543

- Castro L, Aguiar P (2014) A feedforward model for the formation of a grid field where spatial information is provided solely from place cells. *Biol Cybern* 108:133-43
- Chance FS (2012) Hippocampal Phase Precession from Dual Input Components. *Journal of Neuroscience* 32(47): 16693-16703
- Chaudhuri R, Gercek B, Pandey B, Peyrache A, Fiete I (2019) The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience* 22: 1512-1520
- Chavarriaga R, Strösslin T, Sheynikhovich D, Gerstner W (2005) A computational model of parallel navigation systems in rodents. *Neuroinformatics* 3: 223–241
- Chen G, Lu Y, King JA, Cacucci F, Burgess N (2019) Differential influences of environment and self-motion on place and grid cell firing. *Nature Communications* 10(1), 1-11
- Chen G, Manson D, Cacucci F, Wills TJ (2016) Absence of visual input results in the disruption of grid cell firing in the mouse. *Current Biology* 26(17), 2335-2342
- Chen X, He Q, Kelly JW, Fiete IR, McNamara TP (2015) Bias in Human Path Integration Is Predicted by Properties of Grid Cells. *Current Biology* 25(13): 1771-1776
- Cheng S, Frank LM (2011) The structure of networks that produce the transformation from grid cells to place cells. *Neuroscience* 197: 293-306
- Climer JR, Newman EL, Hasselmo ME (2013) Phase coding by grid cells in unconstrained environments: two-dimensional phase precession. *European Journal of Neuroscience* 38(4): 2526–2541
- Cohen MA, Grossberg S (1983) Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans Sys Man and Cybernetics*, 13: 815-821.
- Conklin J, Eliasmith C (2005) A controlled attractor network model of path integration in the rat. *Journal of Computational Neuroscience* 18: 183-203
- Connelly CI, Burns JB, Weiss R (1990) Path planning using Laplace's equation. *Proc 1990 IEEE Int Conference on robotics and automation* 2102-2106.
- Constantinescu AO, O'Reilly JX, Behrens TEJ (2016) Organizing conceptual knowledge in humans with a gridlike code. *Science* 352 (6292), 1464-1468
- Corneil DS, Gerstner W (2015) Attractor Network Dynamics Enable Preplay and Rapid Path Planning in Maze-like Environments. *NIPS* 28
- Couey JJ, Witoelar A, Zhang S-J, Zheng K, Ye J, Dunn B, Czajkowski R, Moser M-B, Moser EI, Roudi Y, Witter MP (2013) Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature Neuroscience* 16: 318-324
- Courellis HS, Nummela SU, Metke M, Diehl GW, Bussell R, Cauwenberghs G, et al. (2019) Spatial encoding in primate hippocampus during free navigation. *PLoS Biol* 17(12): e300054

- Cressant A, Muller RU, Poucet B (1997) Failure of centrally placed objects to control the firing fields of hippocampal place cells. *J Neurosci* 17: 2531-2542.
- Cueva CJ, Wei XX (2018) Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. *ICLR*
- Czurko A, Hirase H, Csicsvari J, Buzsaki G (1999) Sustained activation of hippocampal pyramidal cells by 'space clamping' in a running wheel. *Eur J Neurosci* 11: 344-352.
- D'Albis T, Kempter R (2017) A single-cell spiking model for the origin of grid-cell patterns. *PLoS Comput Biol* 13(10): e1005782
- Damasio AR (1989) The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation* 1: 123-132.
- Davelaar EJ, Goshen-Gottstein Y, Ashkenazi A, Usher M (2005) A context activation model of list memory: Dissociating short-term from long-term recency effects. *Psychol Rev* 112: 34-53.
- Dayan P (1991) Navigating through temporal difference. In: *Neural information processing systems 3* (Lippman RP, Moody JE, Touretzky DS, eds), pp 464-470.
- Dayan P (1993) Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation* 5(4): 613 - 624
- Dayan P, Abbott LF (2002) *Computational neuroscience*.
- de Cothi W, Nyberg N, Griesbauer E-M, Ghanamé C, Zisch F, Lefort JM, Fletcher L, Newton C, Renaudineau S, Bendor D, Grieves G, Duvelle E, Barry C, Spiers HJ (2021) Predictive Maps in Rats and Humans for Spatial Navigation: The Successor Representation Explains Flexible Behaviour. *bioRxiv*
- de Lavilléon G, Lacroix MM, Rondi-Reig L, Benchenane K (2015) Explicit memory creation during sleep demonstrates a causal role of place cells in navigation. *Nature Neuroscience* 18: 493–495
- Dehaene S (1997) *The number sense: How the mind creates mathematics*. Oxford University Press, Oxford, UK
- Demircigil M, Heusel J, Löwe M, Uppang S, Vermet F (2017) On a Model of Associative Memory with Huge Storage Capacity. *Journal of Statistical Physics* 168, 288–299
- Deneve S, Latham PE, Pouget A (2001) Efficient computation and cue integration with noisy population codes. *Nat Neurosci* 4: 826-831.
- Deng W, Aimone J, Gage F (2010) New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nat Rev Neurosci* 11, 339–350
- Deshmukh SS, Knierim JJ (2013) Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus* 23(4):253-67
- Deuker L et al. (2013) Memory Consolidation by Replay of Stimulus-Specific Neural Activity. *Journal of Neuroscience* 33(49): 19373-19383

- Dhillon A, Jones R (2000) Laminar differences in recurrent excitatory transmission in the rat entorhinal cortex in vitro. *Neuroscience* 99: 413–422.
- Doeller CF, Barry C, Burgess N (2010) Evidence for grid cells in a human memory network. *Nature*, 4:463 (7281), 657-661
- Doeller CF, Burgess N (2008) Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *PNAS* 105: 5909-5914.
- Dollé L, Sheynikhovich D, Girard B, Chavarriaga R, Guillot A (2010) Path planning versus cue responding: a bio-inspired model of switching between navigation strategies. *Biological Cybernetics* 103 (4), 299-317
- Dominey PF, Arbib MA (1992) A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cereb Cortex* 2: 153-175.
- Domnisoru C, Kinkhabwala AA, Tank DW (2013) Membrane potential dynamics of grid cells. *Nature* 495: 199-204
- Dordek Y, Soudry D, Meir R, Derdikman D (2016) Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *eLife* 5: e1009
- Douchamps V, Jeewajee A, Blundell P, Burgess N, Lever C (2013) Evidence for Encoding versus Retrieval Scheduling in the Hippocampus by Theta Phase and Acetylcholine. *Journal of Neuroscience* 33: 8689-8704
- Dragoi G, Tonegawa S (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469: 397–401
- Droulez J, Berthoz A (1991) A neural network model of sensoritopic maps with predictive short-term memory properties. *Proc Natl Acad Sci U S A* 88: 9653-9657.
- Dupret, D., O’Neill, J., Pleydell-Bouverie, B. & Csicsvari, J. The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat. Neurosci.* 13, 995–1002 (2010)
- Duvelle, E. et al. Insensitivity of place cells to the value of spatial goals in a two-choice flexible navigation task. *J. Neurosci.* 39, 2522–2541 (2019)
- Edvardson V, Bicanski A, Burgess N (2019) Navigating with grid and place cells in cluttered environments. *Hippocampus* 30(3): 220-232
- Ego-Stengel, V. and Wilson, M.A. (2010) Disruption of ripple associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* 20, 1–10
- Eichenbaum H (2014) Time cells in the hippocampus: A new dimension for mapping memories. *Nature Reviews Neuroscience* 15:732-744
- Eichenbaum H, Cohen NJ (2014) Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron* 83(4):764-70
- Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003) Cellular networks underlying human spatial navigation. *Nature* 425: 184-188.

- Ekstrom AD, Meltzer J, McNaughton BL, Barnes CA (2001) NMDA receptor antagonism blocks experience-dependent expansion of hippocampal "place fields". *Neuron* 31: 631-638.
- Eliav T, Geva-Sagiv M, Yartsev MM, Finkelstein A, Rubin A, Las L, Ulanovsky N (2018) Nonoscillatory Phase Coding and Synchronization in the Bat Hippocampal Formation. *Cell* 175(4): 1119-1130
- Erdem UM, Hasselmo M (2012) A goal-directed spatial navigation model using forward trajectory planning based on grid cells. *European Journal of Neuroscience* 35: 916-931
- Estes WK (1955) Statistical theory of spontaneous recovery and regression. *Psychological Review* 62, 145–154
- Evans T, Burgess N (2020) Replay as structural inference in the hippocampal-entorhinal system. *bioRxiv*
- Fahlman S, Lebiere C (1990) The cascade-correlation learning architecture. In: *Neural Information Processing Systems 2* (Touretzky DS, ed), pp 524-532. Morgan Kaufmann.
- Feng T, Silva D, Foster DJ (2015) Dissociation between the Experience-Dependent Development of Hippocampal Theta Sequences and Single-Trial Phase Precession. *Journal of Neuroscience* 35 (12) 4890-4902
- Fenton AA, Csizmadia G, Muller RU (2000) Conjoint control of hippocampal place cell firing by two visual stimuli I. The effects of moving the stimuli on firing field positions. *J Gen Physiol* 116: 191-209.
- Fenton AA, Kao HY, Neymotin SA, Olypher A, Vayntrub Y, Lytton WW, Ludvig N (2008) Unmasking the CA1 ensemble place code by exposures to small and large environments: more place cells and multiple, irregularly arranged, and expanded place fields in the larger space. *Journal of Neuroscience* 28: 11250-11262
- Fiete IR, Burak Y, Brookings (2008) What grid cells encode about rat location. *J. Neuroscience* 28, 6856-6871
- Fortin NJ, Agster KL, Eichenbaum HB (2002) Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience* 5(5):458-62
- Foster DJ, Morris RG, Dayan P (2000) A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* 10: 1-16.
- Foster, D.J., and Wilson, M.A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440, 680-683
- Franjic D et al. (2021) Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* (in press)
- Frank LM, Brown EN, Wilson M (2000) Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* 27: 169-178.
- Franzius M, Sprekeler H, Wiskott L (2007) Slowness and Sparseness Lead to Place, Head-Direction, and Spatial-View Cells. *PLoS Comput Biol* 3(8): e166

- Frean MR (1990) The Upstart algorithm: A method for constructing and training feedforward neural networks. *Neural Comput* 2: 198-209.
- Fuchs EC, Neitz A, Pinna R, Melzer S, Caputi A, Monyer H (2016) Local and Distant Input Controlling Excitation in Layer II of the Medial Entorhinal Cortex. *Neuron* 89(1):194-208
- Fuhs MC, Touretzky DS (2000) Synaptic learning models of map separation in the hippocampus. *Neurocomputing* 32: 379-384.
- Fuhs MC, Touretzky DS (2006) A spin glass model of path integration in rat medial entorhinal cortex. *Journal of Neuroscience* 26: 4266 – 4276
- Fyhn M, Hafting T, Treves A, Moser MB, Moser EI (2007) Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*. 446: 190-4
- Fyhn M, Hafting T, Witter MP, Moser EI, Moser MB (2008) Grid cells in mice. *Hippocampus* 18: 1230-8
- Gaffan D (1994) Dissociated effects of perirhinal cortex ablation, fornix transection and amygdectomy: evidence for multiple memory systems in the primate temporal lobe. *Exp Brain Res* 99: 411-422.
- Gaffan D, Gaffan EA (1991) Amnesia in man following transection of the fornix. *Brain* 114:2611–18
- Gallant, S. I. Three constructive algorithms for network learning. 652-660. 1986. Proc. 8th Annual Conf. of Cognitive Science Society.
- Gardiner JM, Java RI (1993) In: *Theories of memory* (Collins A, Gathercole S, Morris P, eds), pp 168-188. Hillsdale NJ: Erlbaum.
- Gardner RJ, Hermansen E, Pachitariu M, Burak Y, Baas NA, Dunn BA, Moser MB, Moser EI (2022) Toroidal topology of population activity in grid cells. *Nature* 602: 123–128
- Gardner RJ, Lu L, Wernle T, Moser M-B, Moser EI (2019) Correlation structure of grid cells is preserved during sleep. *Nature Neuroscience* 22: 598–608
- Gardner-Medwin AR (1976) The recall of events through the learning of associations between their parts. *Proc R Soc Lond B Biol Sci* 194: 375-402.
- Geerts JP, Chersi F, Stachenfeld KL, Burgess N (2020). A general model of hippocampal and dorsal striatal learning and decision making. *PNAS* 202007981
- Geisler C, Diba K, Pastalkova E, Mizuseki K, Royer S, Buzsáki G (2010) Temporal delays among place cells determine the frequency of population theta oscillations in the hippocampus. *PNAS* 107(17): 7957-62
- George D, Rikhye RV, Gothoskar N, Guntupalli JS, Dedieu A, Lázaro-Gredilla M (2021) Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nature communications* 12(1), 1-17
- Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. *Science* 233: 1416-1419.

Gershman SJ, Daw ND (2017) Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual Review of Psychology* 68, 101-128

Gershman SJ, Moore CD, Todd MT, Norman KA, & Sederberg PB (2012) The successor representation and temporal context. *Neural Computation* 24, 1553-1568

Gershman SJ, Niv Y (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology* 20, 1-6

Gerstner W, Abbott LF (1997) Learning navigational maps through potentiation and modulation of hippocampal place cells. *Journal of Computational Neuroscience* 4(1) 79–94

Gillespie AK, Astudillo Maya DA, Denovellis EL, Liu DF, Kastner DB, Coulter ME, Roumis DK, Eden UT, Frank LM. Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron*. 2021 Aug 23;. doi: 10.1016/j.neuron.2021.07.029

Giocomo LM, Hussaini SA, Zheng F, Kandel ER, Moser MB, Moser EI (2011) Grid cells use HCN1 channels for spatial scaling. *Cell* 147(5): 1159-70

Giocomo LM, Zilli EA, Fransen E, Hasselmo ME: Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science* 2007, 23:1719-1722

Girardeau, G. et al. (2009) Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci.* 12, 1222–1223

Gluck MA, Myers CE (1993) Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus* 3(4):491-516

Gluck MA, Myers CE (1996) Integrating behavioral and physiological models of hippocampal function. *Hippocampus* 6: 643-653.

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends Neurosci* 15: 20-25.

Goodridge JP, Touretzky DS (2000) Modeling attractor deformation in the rodent head-direction system. *J Neurophysiol* 83: 3402-3410.

Gorchetchnikov A, Grossberg S (2007) Space, time and learning in the hippocampus: how fine spatial and temporal scales are expanded into population codes for behavioural control. *Neural Networks* 20: 182-193

Gorchetchnikov A, Hasselmo ME (2002) A model of hippocampal circuitry mediating goal-driven navigation in a familiar environment. *Neurocomputing* 44-46: 423-427.

Gothard KM, Skaggs WE, McNaughton BL (1996) Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *J Neurosci* 16: 8027-8040.

Graham KS, Hodges JR (1997) Differentiating the roles of the hippocampus complex and the neocortex in long-term memory storage: Evidence from the study of semantic dementia and Alzheimer's disease. *Neuropsychology* 11: 77-89.

- Graves A et al. (2016) Hybrid computing using a neural network with dynamic external memory. *Nature* 538: 471–476
- Greve A, Donaldson DI, van Rossum MCW (2010) A Single-Trace Dual-Process Model of Episodic Memory: A Novel Computational Account of Familiarity and Recollection. *HIPPOCAMPUS* 20: 235–251
- Grieves RM, Dudchenko PA (2013) Cognitive maps and spatial inference in animals: Rats fail to take a novel shortcut, but can take a previously experienced one. *Learning and Motivation* 44(2), 81-92
- Grieves RM, Wood ER, Dudchenko PA (2016) Place cells on a maze encode routes rather than destinations. *Elife* 5, e15986
- Guanella A, Kiper D, Verschure P (2007) A model of grid cells based on a twisted torus topology. *International Journal of Neural Systems* 17: 231-240
- Guazzelli A, Bota M, Arbib MA (2001) Competitive Hebbian learning and the hippocampal place cell system: modeling the interaction of visual and path integration cues. *Hippocampus* 11: 216-239.
- Guazzelli A, Bota M, Corbacho FJ, Arbib MA (1998) Affordances, Motivations, and the World Graph Theory. *Adaptive Behavior* 6(3-4): 435-471
- Gupta AS, van der Meer MAA, Touretzky DS, Redish AD (2010) Hippocampal replay is not a simple function of experience. *Neuron* 65(5):695-705
- Gurney K (1997) *An Introduction to Neural Networks*. Taylor and Francis, PA, USA
- Gustafson NJ, Daw ND (2011) Grid Cells, Place Cells, and Geodesic Generalization for Spatial Reinforcement Learning. *PLoS Comput Biol* 7(10): e1002235
- Hafting T, Fyhn M, Bonnevie T, Moser MB, Moser EI (2008) Hippocampus-independent phase precession in entorhinal grid cells. *Nature* 453: 1248-52
- Hafting T, Fyhn M, Molden S, Moser MB, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436: 801-806.
- Hannula DE, Tranel D, Cohen NJ (2006) The Long and the Short of It: Relational Memory Impairments in Amnesia, Even at Short Lags. *Journal of Neuroscience* 26(32): 8352-8359
- Hardcastle K, Ganguli S, Giocomo LM (2015) Environmental boundaries as an error correction mechanism for grid cells. *Neuron* 86(3), 827-839
- Hardcastle K, Maheswaranathan N, Ganguli S, Giocomo LM (2017) A Multiplexed, Heterogeneous, and Adaptive Code for Navigation in Medial Entorhinal Cortex. *Neuron* 94(2):375-387
- Harris KD, Henze DA, Hirase H, Leinekugel X, Dragoi G, Czurko A, Buzsaki G (2002) Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* 417: 738-741.

Hartley T, Bird CM, Chan D, Cipelotti L, Husain M, Vargha-Khadem F, Burgess N (2007). The hippocampus is required for short-term topographical memory in humans. *Hippocampus* 17(1), 34-48

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

Harvey CD, Collman F, Dombeck DA, Tank DW (2010) Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* 461: 941-946

Hassabis D, Kumaran D, Vann SD, Maguire EA (2007) Patients with hippocampal amnesia cannot imagine new experiences. *PNAS* 104 (5): 1726-1731

Hasselmo ME (1999) Neuromodulation: Acetylcholine and memory consolidation. *Trends in Cognitive Sciences* 3: 351-359

Hasselmo ME (2005) A model of prefrontal cortical mechanisms for goal directed behavior. *Journal of Cognitive Neuroscience* 17(7):1115-29

Hasselmo ME (2008) Grid cell mechanisms and function: Contributions of entorhinal persistent spiking and phase resetting. *Hippocampus* 18: 1116 – 1126

Hasselmo ME, Bodelon C, Wyble BP (2002) A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation* 14: 793:817

Hasselmo ME, Brandon MP (2012) A model combining oscillations and attractor dynamics for generation of grid cell firing. *Frontiers in Neural Circuits* 6: 30

Hasselmo ME, Fehlau BP (2001) Differences in time course of ACh and GABA modulation of excitatory synaptic potentials in slices of rat hippocampus. *J Neurophysiol* 86: 1792-1802.

Hasselmo ME, McClelland JL (1999) Neural models of memory. *Curr Opin Neurobiol* 9: 184-188.

Hasselmo ME, Schnell E, Barkai E (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *J Neurosci* 15: 5249-5262.

Hasselmo ME, Wyble BP (1997) Free recall and recognition in a network model of the hippocampus: simulating effects of scopolamine on human memory function. *Behav Brain Res* 89: 1-34.

Hasselmo ME, Wyble BP, Wallenstein GV (1996) Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus* 6: 693-708.

Hawkins J, Lewis M, Klukas M, Purdy S, Ahmad S (2019) A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex. *Front. Neural Circuits*

Hayman R, Jeffery K (2008) How heterogeneous place cell responding arises from homogeneous grids - a contextual gating hypothesis. *Hippocampus* 18:1301-13

- Hebb DO (1949) *The organisation of behavior*. New York: Wiley.
- Hertz J, Krogh A, Palmer R (1990) *Introduction to the Theory of Neural Computation*. Perseus Books.
- Hinman JR, Chapman GW, Hasselmo ME (2019) Neuronal representation of environmental boundaries in egocentric coordinates. *Nature Communications* 10(1):2772
- Hinton GE (1989) Connectionist learning procedures. *Artificial Intelligence* 40(1–3): 185-234
- Hinton GE, Sejnowski TJ (1999) *Unsupervised learning*. MIT Press.
- Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Computation* 9(8): 1735–1780
- Hok, V. et al. Goal-related activity in hippocampal place cells. *J. Neurosci.* 27, 472–482 (2007)
- Holdstock JS, Mayes AR, Cezayirli E, Isaac CL, Aggleton JP, Roberts N (2000) A comparison of egocentric and allocentric spatial memory in a patient with selective hippocampal damage. *Neuropsychologia* 38: 410-425.
- Holdstock JS, Mayes AR, Roberts N, Cezayirli E, Isaac CL, O'Reilly RC, Norman KA (2002) Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus* 12: 341-351.
- Hollup, S. A., Molden, S., Donnett, J. G., Moser, M. B. & Moser, E. I. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *J. Neurosci.* 21, 1635–1644 (2001)
- Hölscher C, Anwyl R, Rowan MJ (1997) Stimulation on the Positive Phase of Hippocampal Theta Rhythm Induces Long-Term Potentiation That Can Be Depotentiated by Stimulation on the Negative Phase in Area CA1 In Vivo. *Journal of Neuroscience* 17(16): 6470-6477
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* 79: 2554-2558.
- Hoppensteadt FC (1986) *An introduction to the mathematical properties of neurons*. Cambridge: Cambridge University Press.
- Hori E, Tabuchi E, Matsumura N, Tamura R, Eifuku S, Endo S, Nishijo H, Ono T (2003) Representation of place by monkey hippocampal neurons in real and virtual translocation. *Hippocampus* 13: 190-196.
- Horiuchi TK, Moss CF (2015) Grid cells in 3-D: Reconciling data and models. *Hippocampus* 25(12):1489-500
- Horner AJ, Bisby JA, Bush D, Lin W-J, Burgess N (2015) Evidence for Holistic Episodic Recollection via Hippocampal Pattern Completion. *Nature Communications* 6: 7462

Howard LR, Javadi AH, Yu Y, Mill RD, Morrison LC, Knight R, Loftus MM, Staskute L, Spiers HJ (2014a) The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. *Current Biology* 24: 1331-1340

Howard MW, Fotedar MS, Datey AV, Hasselmo ME (2005) The Temporal Context Model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. *Psychol Rev* 112: 75-116.

Howard MW, Kahana MJ (2001) A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology* 46: 269-299.

Howard MW, MacDonald CJ, Tiganj Z, Shankar KH, Du Q, Hasselmo ME, Eichenbaum H (2014b) A Unified Mathematical Framework for Coding Time, Space, and Sequences in the Hippocampal Region. *J Neurosci* 34(13): 4692–4707

Howard, M.W., Shankar, K.H., Aue, W.R., and Criss, A.H. (2015). A distributed representation of internal time, *Psychological Review*, 122, 24-53

Howard, M.W., Viskontas, I.V., Shankar, K.H., and Fried, I. (2012) Ensembles of human MTL neurons “jump back in time” in response to a repeated stimulus. *Hippocampus* 22: 1833-1847

Høydal ØA, Skytøen ER, Andersson SO, Moser M-B, Moser EI (2019) Object-vector coding in the medial entorhinal cortex. *Nature* 568: 400–404

Huerta PT, Lisman JE (1995) Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro. *Neuron* 15(5):1053-63

Hulse BK, Jayaraman V (2020) Mechanisms Underlying the Neural Computation of Head Direction. *Annual Reviews of Neuroscience* 43: 31-54

Huxter J, Burgess N, O'Keefe J (2003) Independent rate and temporal coding in hippocampal pyramidal cells. *Nature* 425: 828-832.

Ismakov R, Barak O, Jeffery J, Derdikman D (2017) Grid Cells Encode Local Positional Information. *Current Biology* 27(15): 2337-2343

Jacobs J, Weidemann CT, Miller JF, Solway A, Burke JF, Wei X, Suthana N, Sperling MR, Sharan AD, Fried I, Kahana MJ (2013) Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience* 16: 1188-1190

Jaramillo J, Schmidt R, Kempter R (2014) Modeling Inheritance of Phase Precession in the Hippocampal Formation. *Journal of Neuroscience* 34(22): 7715-7731

Jeewajee A, Barry C, Douchamps V, Manson D, Lever C, Burgess N (2014) Theta phase precession of grid and place cell firing in open environments. *Philosophical Transactions of the Royal Society B* 369: 20120532

Jensen O, Lisman JE (1996) Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn Mem* 3: 279-287.

Jensen O, Lisman JE (2000) Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding. *J Neurophysiol* 83: 2602-2609.

Jercog PE, Ahmadian Y, Woodruff C, Deb-Sen R, Abbott LF, Kandel ER (2019) Heading direction with respect to a reference point modulates place-cell activity. *Nature Communications* 10: 2333

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

Johnson A, Redish AD (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience* 27(45): 12176-89

Jones VJ (1976) A fragmentation hypothesis of memory: cued recall of pictures and of sequential position. *Journal of Experimental Psychology: General* 105: 277-293.

Julian JB, Keinath AT, Frazzetta G, Epstein RA. Human entorhinal cortex represents visual space using a boundary-anchored grid. *Nat Neurosci.* 2018 21(2):191-194

Jung MW, Wiener SI, McNaughton BL (1994) Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J Neurosci* 14: 7347-7356.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24, 103–109

Kali S, Dayan P (2000) The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model. *J Neurosci* 20: 7463-7477.

Kali S, Dayan P (2004) Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nat Neurosci* 7: 286-294.

Kang L, De Weese MR (2019) Replay as wavefronts and theta sequences as bump oscillations in a grid cell attractor network. *eLife* 8: e46351

Kartsounis LD, Rudge P, Stevens JM (1995) Bilateral lesions of CA1 and CA2 fields of the hippocampus are sufficient to cause a severe amnesic syndrome in humans. *J Neurol Neurosurg Psychiatry* 59(1): 95-8

Kaufman AM, Geiller T, Losonczy A (2020) A Role for the Locus Coeruleus in Hippocampal CA1 Place Cell Reorganization during Spatial Reward Learning. *Neuron* 105(6):1018-1026

Kay K, Chung JE, Sosa M, Schor JS, Karlsson MP, Larkin MC, Liu DF, Frank LM. Constant Sub-second Cycling between Representations of Possible Futures in the Hippocampus. *Cell*. 2020 Feb 6;180(3):552-567

Keith JR, McVety KM (1988) Latent place learning in a novel environment and the influence of prior training in rats. *Psychobiology* 16: 146-151.

Kentros C, Hargreaves E, Hawkins RD, Kandel ER, Shapiro M, Muller RV (1998) Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science* 280: 2121-2126.

- Killian NJ, Jutras MJ, Buffalo EA (2012) A map of visual space in the primate entorhinal cortex. *Nature* 491(7426):761-4.
- Kingma DP, Welling M (2014) Auto-Encoding Variational Bayes. arXiv
- Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser M-B (2008) Finite scale of spatial representation in the hippocampus. *Science* 321(5885):140-3
- Knierim JJ, Kudrimoti HS, McNaughton BL (1995) Place cells, head direction cells, and the learning of landmark stability. *J Neurosci* 15: 1648-1659.
- Knierim JJ, Lee I, Hargreaves EL (2006) Hippocampal place cells: Parallel input streams, subregional processing, and implications for episodic memory. *Hippocampus* 16: 755-764
- Knowlton BJ, Squire LR (1995) Remembering and knowing: two different expressions of declarative memory. *J Exp Psychol Learn Mem Cogn* 21: 699-710.
- Knudsen EB, Wallis JD (2021) Hippocampal neurons construct a map of an abstract value space. *Cell* 184(18):4640-4650
- Koenig J, Linder AN, Leutgeb JK, Leutgeb S (2011) The spatial periodicity of grid cells is not sustained during reduced theta oscillations. *Science* 332: 592–595
- Kohonen T (1972) Correlation matrix memories. *IEEE Trans Comp C-21*: 353-359.
- Kornienko et al. (2018) Non-rhythmic head-direction cells in the parahippocampal region are not constrained by attractor network dynamics. *eLife*
- Kosslyn SM, Ball TM, Reiser BJ (1978) Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance* 4: 47-60
- Kroll NE, Knight RT, Metcalfe J, Wolf ES, Tulving E (1996) Cohesion failure as a source of memory illusions. *Journal of Memory and Language* Vol 35: 176-196.
- Kropff E, Carmichael JE, Moser M-B, Moser EI (2015) Speed cells in the medial entorhinal cortex. *Nature* 523: 419–424
- Kropff E, Treves A (2008) The emergence of grid cells: intelligent design or just adaptation? *Hippocampus* 18: 1256-1269
- Krotov D, Hopfield JJ (2016) Dense Associative Memory for Pattern Recognition. arXiv
- Krotov D, Hopfield JJ (2020) Large Associative Memory Problem in Neurobiology and Machine Learning. arXiv
- Krupic J, Bauza M, Burton S, Barry C, O’Keefe J (2015) Grid cell symmetry is shaped by environmental geometry. *Nature* 518: 232–235
- Kubie JL, Fenton AA (2012) Linear look-ahead in conjunctive cells: an entorhinal mechanism for vector-based navigation. *Frontiers in Neural Circuits* 6: 20

- Kumaran D, Hassabis D, McClelland DL (2016) What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences* 20(7): 512-334
- Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, 119(3), 573-616.
- Kunz L et al. (2021) A neural code for egocentric spatial maps in the human medial temporal lobe. *Neuron* 109 (17):2781-2796
- Langston RF, Ainge JA, Couey JJ, Canto CB, Bjerknes TL, Witter MP, Moser EI, Moser MB (2010) Development of the spatial representation system in the rat. *Science* 328: 1576-1580
- Latham PE, Deneve S, Pouget A (2003) Optimal computation with attractor networks. *J Physiol Paris* 97: 683-694.
- Laurens J, Angelaki DE (2018) The brain compass: a perspective on how self-motion updates the head direction cell attractor. *Neuron* 97 (2), 275-289
- LeCun, Y. et al. (2015) Deep learning. *Nature* 521, 436–444
- Lee I, Griffin AL, Zilli EA, Eichenbaum H, Hasselmo ME (2006) Gradual translocation of spatial correlates of neuronal firing in the hippocampus toward prospective reward locations. *Neuron* 51 (5), 639-650
- Leibold C, Monsalve-Mercado MM (2017) Traveling Theta Waves and the Hippocampal Phase Code. *Sci Rep* 7(1): 7678
- Lengyel M, Dayan P (2007) Hippocampal Contributions to Control: The Third Way. *NIPS*
- Lengyel M, Szatmary Z, Erdi P (2003) Dynamically detuned oscillations account for the coupled rate and temporal code of place cell firing. *Hippocampus* 13: 700-714.
- Leutgeb JK, Leutgeb S, Treves A, Meyer R, Barnes CA, McNaughton BL, Moser MB, Moser EI (2005) Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron* 48: 345-358.
- Lever C, Burton S, Jeewajee A, O'Keefe J, Burgess N (2009) Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience* 29: 9771-9777
- Lever C, Wills T, Cacucci F, Burgess N, O'Keefe J (2002) Long-term plasticity in the hippocampal place cell representation of environmental geometry. *Nature* 416: 90-94.
- Levy WB (1996) A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus* 6: 579-590.
- Lieblich I, Arbib MA (1982) Multiple representations of space underlying behavior. *Behav Brain Sci* 5: 627-659
- Lu J, Behbahani AH, Hamburg L et al. (2022) Transforming representations of movement from body- to world-centric space. *Nature* 601: 98–104

- Lubenov EV, Siapas AG (2009) Hippocampal theta oscillations are travelling waves. *Nature* 459(7246): 534-9
- Ludvig N, Tang HM, Gohil BC, Botero JM (2004) Detecting location-specific neuronal firing rate increases in the hippocampus of freely-moving monkeys. *Brain Res* 1014: 97-109.
- Mallot HA, Gillner S (2000) Route navigating without place recognition: what is recognised in recognition-triggered responses? *Perception* 29: 43-55.
- Mamad O, Stump L, McNamara HM, Ramakrishnan C, Deisseroth K, Reilly RB, et al. (2017) Place field assembly distribution encodes preferred locations. *PLoS Biol* 15(9): e2002365.
- Mankin EA, Sparks FT, Slayyeh B, Sutherland RJ, Leutgeb S, Leutgeb JK (2012) Neuronal code for extended time in the hippocampus. *PNAS* 109: 19462-19467
- Mankin EA, Thurley K, Chenani A, Haas OV, Debs L, Henke J, Galinato M, Leutgeb JK, Leutgeb S, Leibold C (2019) The hippocampal code for space in Mongolian gerbils. *Hippocampus* 29(9):787-801
- Manning JR, Polyn SM, Litt B, Baltuch G, Kahana MJ. 2011. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proc Natl Acad Sci USA* 108:12893–12897
- Manns JR, Eichenbaum H (2006) Evolution of declarative memory. *Hippocampus* 16 (9), 795-808
- Manns JR, Hopkins RO, Squire LR (2003) Semantic memory and the human hippocampus. *Neuron* 37:127–33
- Manns JR, Howard MW, Eichenbaum H (2007) Gradual changes in hippocampal activity support remembering the order of events. *Neuron* 56 (3), 530-540
- Manns JR, Squire LR (1999) Impaired recognition memory on the Doors and People Test after damage limited to the hippocampal region. *Hippocampus* 9: 495-499.
- Mao D, Avila E, Caziot B, Laurens J, Dickman JD, Angelaki DE (2021) Spatial modulation of hippocampal activity in freely moving macaques. *Neuron* 109(21):3521-3534
- Markus EJ, Qin YL, Leonard B, Skaggs WE, McNaughton BL, Barnes CA (1995) Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *Journal of Neuroscience* 15 (11): 7079-7094
- Marr D (1970) A theory for cerebral cortex. *Proc R Soc Lond B Biol Sci* 176: 161-234.
- Marr D (1971) Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci* 262: 23-81.
- Martinet L-E, Sheynikhovich D, Benchenane K, Arleo A (2011) Spatial Learning and Action Planning in a Prefrontal Cortical Network Model. *PLoS Comput Biol* 7(5): e1002045
- Mathis A, Herz AV, Stemmler M (2012) Optimal population codes for space: grid cells outperform place cells. *Neural Computation* 24: 2280-2317

- Mattar MG, Daw ND (2018) Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience* 21: 1609–1617
- Maurer AP, Van Rhoads SR, Sutherland GR, Lipa P, McNaughton BL (2005) Self-motion and the Origin of Differential Spatial Scaling Along the Septo-Temporal Axis of the Hippocampus. *Hippocampus* 15: 841-852.
- McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* 102: 419-457.
- McClelland JL, Rumelhart DE (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol 2: Psychological and Biological Models*. MIT Press.
- McClelland, J.L. (2013) Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *J. Exp. Psychol. Gen.* 142, 1190–1210
- McCloskey M, Cohen NJ (1989) Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. *Psychology of Learning and Motivation* 24: 109-165
- McHugh TJ, Blum KI, Tsien JZ, Tonegawa S, Wilson MA (1996) Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. *Cell* 87(7): 1339-49
- McNaughton BL, Barnes CA, Gerrard JL, Gothard K, Jung MW, Knierim JJ, Kudrimoti H, Qin Y, Skaggs WE, Suster M, Weaver KL (1996) Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *J Exp Biol* 199: 173-185.
- McNaughton BL, Barnes CA, O'Keefe J (1983) The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp Brain Res* 52, 41–49
- McNaughton BL, Barnes CA, O'Keefe J (1983) The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp Brain Res* 52: 41-49.
- McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser MB (2006) Path integration and the neural basis of the cognitive map. *Nature Reviews Neuroscience* 7: 663 – 678
- McNaughton BL, Morris RG (1987) Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences* 10: 408-415.
- McNaughton BL, Nadel L (1990) Hebb-Marr networks and the neurobiological representation of action in space. In: *Neuroscience and connectionist theory* (Gluck MA, Rumelhart DE, eds), pp 1-63. Hillsdale N.J.: Lawrence Erlbaum Assoc.
- Mehta MR, Barnes CA, McNaughton BL (1997) Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc Natl Acad Sci U S A* 94: 8918-8921.
- Mehta MR, Lee AK, Wilson MA (2002) Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* 417: 741-746.

- Mehta MR, Quirk MC, Wilson MA (2000) Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* 25: 707-715.
- Mensink GJ, Raaijmakers JG (1988) A model for interference and forgetting. *Psychological Review* Vol 95: 434-455.
- Mezard M, Nadal J-P (1989) Learning in feedforward layered networks: the Tiling algorithm. *J Physics A* 22: 2191-2203.
- Mhatre H, Gorchetchnikov A, Grossberg S (2012). Grid cell hexagonal patterns formed by fast self-organized learning within entorhinal cortex. *Hippocampus* 22: 320-334
- Milford MJ, Wyeth GF (2008) Mapping a Suburb With a Single Camera Using a Biologically Inspired SLAM System. *IEEE TRANSACTIONS ON ROBOTICS* 24(5): 1038-1053
- Miller JF, Neufang M, Solway A, Brandt A, Trippel M, Mader I, Hefft S, Merkow M, Polyn SM, Jacobs J, Kahana MJ, Schulze-Bonhage A (2013) Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science* 342: 1111-1114
- Miller JF, Weidemann CT, Kahana MJ (2012) Recall termination in free recall. *Memory & Cognition*, 1-11
- Milner AD, Dijkerman HC, Carey DP (1999) Visuospatial processing in a case of visual form agnosia. In: *The Hippocampal and Parietal Foundations of Spatial Cognition* (Burgess N, Jeffery KJ, O'Keefe J, eds), pp 443-466. Oxford: Oxford University Press.
- Mnih V et al. (2015) Human-level control through deep reinforcement learning. *Nature* 518: 529-533
- Moll M, Miikkulainen R (1997) Convergence-Zone episodic memory: Analysis and simulations. *Neural Networks* 10: 1017-1036.
- Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16: 1936-1947.
- Morris RGM, Garrud P, Rawlins JN, O'Keefe J (1982) Place navigation impaired in rats with hippocampal lesions. *Nature* 297: 681-683.
- Morton J, Hammersley RH, Bekerian DA (1985) Headed records: a model for memory and its failure. *Cognition* 20: 1-23.
- Moskovitz T, Wilson SR, Sahani M (2021) A First-Occupancy Representation for Reinforcement Learning. *arXiv*
- Muller RU, Kubie J, Ranck JB (1987) Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *Journal of Neuroscience* 7(7): 1935-1950
- Muller RU, Kubie JL (1987) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J Neurosci* 7: 1951-1968.
- Muller RU, Kubie JL, Saypoff R (1991) The hippocampus as a cognitive graph (abridged version). *Hippocampus* 1: 243-246.

- Muller RU, Stead M, Pach J (1996) The hippocampus as a cognitive graph. *J Gen Physiol* 107: 663-694.
- Murray DJ, Pye C, Hockley WE 1976. Standing's power function in long-term memory. *Psychol Res* 38: 319–331
- Murray EA, Mishkin M (1998) Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *J Neurosci* 18: 6568-6582.
- Murre JM (1996) TraceLink: a model of amnesia and consolidation of memory. *Hippocampus* 6: 675-684.
- Nadel L, Moscovitch M (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* 7: 217-227.
- Nakashiba, T. et al. (2009) Hippocampal CA3 output is crucial for ripple-associated reactivation and consolidation of memory. *Neuron* 62, 781–787
- Nakazawa K, Quirk MC, Chitwood RA, Watanabe M, Yeckel MF, Sun LD, Kato A, Carr CA, Johnston D, Wilson MA, Tonegawa S (2002) Requirement for Hippocampal CA3 NMDA Receptors in Associative Memory Recall. *Science*.
- Namboodiri VMK, Stuber GD (2021) The learning of prospective and retrospective cognitive maps within neural circuits. *Neuron* 109(22):3552-3575
- Nau M, Navarro Schröder T, Bellmund JLS, Doeller CF (2018) Hexadirectional coding of visual space in human entorhinal cortex. *Nature Neuroscience* 21: 188–190
- Navratilova Z, Giocomo LM, Fellous JM, Hasselmo ME, McNaughton BL (2012b) Phase precession and variable spatial scaling in a periodic attractor map model of medial entorhinal grid cells with realistic after-spike dynamics. *Hippocampus* 22: 772-789
- Navratilova Z, Hoang LT, Schwindel CD, Tatsuno M, McNaughton BL (2012a) Experience-dependent firing rate remapping generates directional selectivity in hippocampal place cells. *Front Neural Circuits* 6: 6
- Nieh EH, Schottdorf M, Freeman NW, Low RJ, Lewallen S, Koay SA, Pinto L, Gauthier JL, Brody CD, Tank DW (2021) Geometry of abstract learned knowledge in the hippocampus. *Nature* 595(7865):80-84
- Niv Y (2019) Learning task-state representations. *Nature neuroscience* 22 (10), 1544-1553
- Norman KA, Detre G, Polyn SM (2008) Computational models of episodic memory. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 189–225). Cambridge University Press
- Norman KA, O'Reilly RC (2003) Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol Rev* 110: 611-646.
- O'Keefe J, Krupic J (2021) Do hippocampal pyramidal cells respond to nonspatial stimuli? *Journal of Physiology* 101(3): 1427-1456

- Ocko SA, Hardcastle K, Giocomo LM, Ganguli S (2018) Emergent elasticity in the neural code for space. *PNAS* 115(50): E11798-E11806
- Oja E (1982) A simplified neuron model as a principal component analyzer. *J Math Biol* 15: 267-273.
- O'Keefe J, Burgess N (1996) Geometric determinants of the place fields of hippocampal neurons. *Nature* 381: 425-428.
- O'Keefe J, Burgess N (2005) Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* 15: 853 - 866
- O'Keefe J, Burgess N (2005) Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* 15: 853-866.
- O'Keefe J, Conway DH (1978) Hippocampal place units in the freely moving rat: why they fire where they fire. *Exp Brain Res* 31: 573-590.
- O'Keefe J, Nadel L (1978) *The hippocampus as a cognitive map*. Oxford: Oxford University Press.
- O'Keefe J, Recce ML (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3: 317-330.
- O'Keefe J, Speakman A (1987) Single unit activity in the rat hippocampus during a spatial memory task. *Exp Brain Res* 68: 1-27.
- Ólafsdóttir HF, Barry C, Saleem AB, Hassabis D, Spiers HJ (2015) Hippocampal place cells construct reward related sequences through unexplored space. *Elife* 4: e06063
- Olafsdottir HF, Bush D, Barry C (2018) The Role of Hippocampal Replay in Memory and Planning. *Current Biology* 28: pR37-R50
- Ólafsdóttir HF, Carpenter F, Barry C (2016) Coordinated grid and place cell replay during rest. *Nature Neuroscience* 19(6): 792-794
- Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., & Verfaellie, M. (2006) Working memory for conjunctions relies on the medial temporal lobe. *Journal of Neuroscience* (17), 4596–4601
- O'Neill J, Boccara CN, Stella F, Schoenenberger P, Csicsvari J (2017) Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science* 355(6321):184-188
- Orchard J (2015) Oscillator-interference models of path integration do not require theta oscillations. *Neural Computation* 27(3): 548-60
- Ormond J, O'Keefe J (2021) Hippocampal place cells use vector computations to navigate. *bioRxiv*
- Packard MG, McGaugh JL (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol Learn Mem* 65: 65-72.

- Page H, Wilson J, Jeffery KJ (2018) A dual-axis rule for updating the head direction cell reference frame during movement in three dimensions. *J Neurophysiol* 119(1): 192-208
- Pastoll H, Solanka L, van Rossum MCW, Nolan MF (2013) Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron* 77: 141-154
- Pavlidis C, Greenstein YJ, Grudman M, Winson J (1988) Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. *Brain Res* 439(1-2):383-7.
- Payne HL, Lynch GF, Aronov D (2021) Neural representations of space in the hippocampus of a food-caching bird. *Science* 373(6552): 343-348
- Pérez-Escobar JA, Kornienko O, Latuske P, Kohler L, Allen K (2016) Visual landmarks sharpen grid cell metric and confer context specificity to neurons of the medial entorhinal cortex. *eLife* 5: e16937
- Pertsov, Y., Miller, T. D., Gorgoraptis, N., Caine, D., Schott, J. M., Butler, C., & Husain, M. (2013). Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. *Brain* 136(8), 2474–2485
- Peyrache A, Lacroix MM, Petersen PC, Buzsáki G (2015) Internally organized mechanisms of the head direction sense. *Nature Neuroscience* 18: 569
- Piray P, Daw ND (2021) Linear reinforcement learning in planning, grid fields, and cognitive control. *Nat Commun* 12(1):4942
- Pouget A, Sejnowski TJ (1997) A new view of hemineglect based on the response properties of parietal neurones. *Philos Trans R Soc Lond B Biol Sci* 352: 1449-1459.
- Pritzel A, Uria B, Srinivasan S, Puigdomènech A, Vinyals O, Hassabis D, Wierstra D, Blundell C (2017) Neural Episodic Control. *arXiv*
- Qasim SE, Fried I, Jacobs J (2021) Phase precession in the human hippocampus and entorhinal cortex. *Cell* 184(12): 3242-3255
- Quiroga RQ (2012) Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience* 13(8), 587-597
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435 (7045), 1102-1107
- Raaijmakers JG, Shiffrin RM (1981) Search of associative memory. *Psychological Review* 88: 93-134.
- Radulescu et al., 2021
- Ratcliff R (1990) Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychol Rev* 97: 285-308.

Raudies F, Brandon MP, Chapman GW, Hasselmo ME (2015) Head direction is coded more strongly than movement direction in a population of entorhinal neurons. *Brain Research* 1621: 355-67

Recatanesi et al. (2021)

Redish AD (1999) *Beyond the Cognitive Map: From Place Cells to Episodic Memory*. Cambridge MA: MIT Press.

Redish AD (2016) Vicarious trial and error. *Nat Rev Neurosci* 17(3):147-59

Redish AD, Elga AN, Touretzky DS (1996) A coupled attractor model of the rodent head direction system. *Network* 7: 671-685.

Redish AD, Elga AN, Touretzky DS (1996) A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems* 7: 671–685

Redish AD, Rosenzweig ES, Bohanick JD, McNaughton BL, Barnes CA (2000) Dynamics of hippocampal ensemble activity realignment: time versus space. *J Neurosci* 20: 9298-9309.

Redish AD, Touretzky DS (1998) The role of the hippocampus in solving the Morris water maze. *Neural Comput* 10: 73-111.

Redondo R, Morris RGM (2011) Making memories last: the synaptic tagging and capture hypothesis. *Nat Rev Neurosci* 12, 17–30

Remme, M. W., Lengyel, M., & Gutkin, B. S. (2010). Democracy-independence trade-off in oscillating dendrites and its implications for grid cells. *Neuron*, 66(3), 429-437

Rennó-Costa C, Tort ABL (2017) Place and Grid Cells in a Loop: Implications for Memory Function and Spatial Coding. *J Neurosci* 37(34):8062-8076

Rich PD, Liaw HP, Lee AK (2014) Place cells. Large environments reveal the statistical structure governing hippocampal representations. *Science* 345(6198):814-7

Robinson NTM et al. (2020) Targeted Activation of Hippocampal Place Cells Drives Memory-Guided Spatial Behavior. *Cell* 183 (6): 1586-1599

Rolls ET, Kesner RP (2006) A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology* 79(1): 1-48

Rolls ET, Robertson RG, Georges-Francois P (1997) Spatial view cells in the primate hippocampus. *Eur J Neurosci* 9: 1789-1794.

Rolls ET, Stringer SM, Elliot T (2006) Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning. *Network* 17: 447-65

Rolls ET, Treves A (1997) *Neural Networks and Brain Function*. Oxford University Press.

Romani S, Pinkoviezky I, Rubin A, Tsodyks M (2013) Scaling laws of associative memory retrieval. *Neural computation* 25(10), 2523-2544

- Rosenbaum RS, Priselac S, Kohler S, Black SE, Gao F, Nadel L, Moscovitch M (2000) Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. *Nat Neurosci* 3: 1044-1048.
- Royer S, Pare D (2003) Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* 422: 518-522.
- Rugg MD, Yonelinas AP (2003) Human recognition memory: a cognitive neuroscience perspective. *Trends Cogn Sci* 7: 313-319.
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: *Parallel distributed programming Vol 1: Foundations* (Rumelhart DE, McClelland JL, eds), pp 318-364. MIT Press.
- Rumelhart DE, McClelland JL (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol I: Foundations*. MIT Press.
- Rumelhart DE, Zipser D (1986) Feature discovery by competitive learning. In: *Parallel distributed programming Vol 1: Foundations* (Rumelhart DE, McClelland JL, eds), pp 151-193. MIT Press.
- Samsonovich A, McNaughton BL (1997) Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* 17: 5900-5920.
- Sanders H, Wilson MA, Gershman SJ (2020) *eLife* 9: e51140
- Sarel A, Finkelstein A, Las L, Ulanovsky N (2017) Vectorial representation of spatial goals in the hippocampus of bats. *Science* 355: 176-180
- Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser MB, Moser EI (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* 312: 758 - 62
- Save E, Nerad L, Poucet B (2000) Contribution of multiple sensory information to place field stability in hippocampal place cells. *Hippocampus* 10: 64-76.
- Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK (2012) The Future of Memory: Remembering, Imagining, and the Brain. *Neuron* 76: 677-694
- Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1481), 773–786
- Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB (2014) The necessity of the medial temporal lobe for statistical learning. *J Cogn Neurosci* 26(8):1736-47
- Schapiro, A.C., McDevitt, E.A., Rogers, T.T., Mednick, S.C., & Norman, K.A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications* 9(1): 3920
- Schapiro, A.C., Turk-Browne, N.B., Botvinick, M.M., & Norman, K.A. (2017). Complementary learning systems within the hippocampus: A neural network modelling

approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B* 372(1711): 20160049

Schlesiger MI, Cannova CC, Boublil BL, Hales JB, Mankin EA, Brandon MP, Leutgeb JK, Leibold C, Leutgeb S (2015) The medial entorhinal cortex is necessary for temporal organization of hippocampal neuronal activity. *Nat Neurosci* 18(8):1123-32

Schmidt-Hieber C, Häusser M (2013) Cellular mechanisms of spatial navigation in the medial entorhinal cortex. *Nature Neuroscience* 16: 325-31

Scholkopf B, Mallot HA (1995) View-based cognitive mapping and path planning. *Adaptive behavior* 3: 311-348.

Schreiner T, Staudigl T (2020) Electrophysiological signatures of memory reactivation in humans. *Philos Trans R Soc Lond B Biol Sci* 375(1799):20190293

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275: 1593-1599.

Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry* 20: 11-21.

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

Sharp PE (1991) Computer simulation of hippocampal place cells. *Psychobiology* 19: 103-115.

Sharp PE (1996) Multiple spatial/behavioral correlates for cells in the rat postsubiculum: multiple regression analysis and comparison to other hippocampal areas. *Cereb Cortex* 6: 238-259.

Sharp PE, Blair HT, Brown M (1996) Neural network modeling of the hippocampal formation spatial signals and their possible role in navigation: a modular approach. *Hippocampus* 6: 720-734

Sherrill KR, Erdem UM, Ross RS, Brown TI, Hasselmo ME, Stern CE (2013) Hippocampus and retrosplenial cortex combine path integration signals for successful navigation. *Journal of Neuroscience* 33: 19304-19313

Shin H, Lee JK, Kim J, Kim J (2017) Continual Learning with Deep Generative Replay. *NIPS*

Shipston-Sharman, O., Solanka, L., & Nolan, M. F. (2016). Continuous attractor network models of grid cell firing based on excitatory-inhibitory interactions. *The Journal of physiology*, 594(22), 6547–6557

Si B, Treves A (2013) A model for the differentiation between grid and conjunctive units in medial entorhinal cortex. *Hippocampus* 23(12):1410-24

Siapas AG, Wilson MA (1998) Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* 21(5):1123-8

- Sirota, A. et al. (2003) Communication between neocortex and hippocampus during sleep in rodents. *PNAS* 100, 2065–2069
- Skaggs WE, Knierim JJ, Kudrimoti H, McNaughton BL (1995) A model of the neural basis of the rat's sense of direction. In: *Neural Information Processing Systems 7* (Hanson SJ, Cowan JD, Giles CL, eds), pp 173-180. MIT Press.
- Skaggs WE, McNaughton BL (1998) Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. *J Neurosci* 18: 8455-8466.
- Skaggs WE, McNaughton BL, Wilson MA, Barnes CA (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6: 149-172.
- Skaggs, W.E., and McNaughton, B.L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271, 1870-1873
- Solanka, L., van Rossum, M. C., & Nolan, M. F. (2015). Noise promotes independent control of gamma oscillations and grid firing within recurrent attractor networks. *eLife*, 4, e06444
- Solomon, P. R, Vander Schaaf, E. R., Thompson, R. F., & Weisz, D. J. (1986). Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behavioral Neuroscience*, 100(5), 729–744.
- Solstad T, Boccara CN, Kropff E, Moser MB, Moser EI (2008) Representation of geometric borders in the entorhinal cortex. *Science* 322: 1865-1868
- Solstad T, Moser EI, Einevoll GT (2006) From grid cells to place cells: a mathematical model. *Hippocampus* 16: 1026-31
- Speakman A, O'Keefe J (1990) Hippocampal complex spike cells do not change their place fields if the goal is moved within a cue controlled environment. *Eur J Neurosci* 7: 544-555.
- Spiers HJ, Maguire EA, Burgess N (2001) Hippocampal amnesia. *Neurocase* 7: 357-382.
- Sprekeler H, Michaelis C, Wiskott L (2007) Slowness: An Objective for Spike-Timing-Dependent Plasticity? *PLoS Computational Biology* 3(6):e112
- Squire LR, Genzel L, Wixted JT, Morris RG (2015) Memory consolidation. *Cold Spring Harb Perspect Biol.* 7(8):a021766
- Squire LR, Stark CEL, Clark RE (2004) The Medial Temporal Lobe. *Annu. Rev. Neurosci.* 2004. 27:279–306
- Squire LR, Wixted JT (2011) The Cognitive Neuroscience of Human Memory Since HM. *Annu. Rev. Neurosci.* 2011. 34:259–88
- Sreenivasan S, Fiete I (2011) Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nature Neuroscience* 14: 1330-1337

- Stachenfeld KL, Botvinick MM, Gershman SJ (2017) The hippocampus as a predictive map. *Nature neuroscience* 20 (11), 1643-1653
- Staresina BP, Alink A, Kriegeskorte N, Henson RN (2013) Awake reactivation predicts memory in humans. *PNAS* 110 (52), 21159-21164
- Stemers B, Vicente-Grabovetsky A, Barry C, Smulders P, Schröder TN, Burgess N, Doeller CF (2016) Hippocampal Attractor Dynamics Predict Memory-Based Decision Making. *Curr Biol* 26(13):1750-1757
- Stemmler M, Mathis A, Herz AV (2015) Connecting multiple spatial scales to decode the population activity of grid cells. *Science Advances* 1(11): e1500816
- Stensola H, Stensola T, Solstad T, Frøland K, Moser MB, Moser EI (2012) The entorhinal grid map is discretized. *Nature* 492: 72-78
- Stoianov I, Maisto D, Pezzulo G (2021) The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning. *bioRxiv*
- Sun C, Yang W, Martin J, Tonegawa S (2020) Hippocampal neurons represent events as transferable units of experience. *Nat Neurosci* 23(5):651-663
- Sun W, Advani M, Spruston N, Saxe A, Fitzgerald JE (2021) Organizing memories for generalization in complementary learning systems. *bioRxiv*
- Sun Y et al. CA1-projecting subiculum neurons facilitate object–place learning. *Nat. Neurosci.* 22: 1857–1870 (2019)
- Sutton MA, Schuman EM (2006) Dendritic Protein Synthesis, Synaptic Plasticity, and Memory. *Cell* 127 (1): 49-58
- Sutton RS, Barto AG (1988) Reinforcement learning: an introduction. MIT Press.
- Taube JS (1998) Head direction cells and the neuropsychological basis for a sense of direction. *Prog Neurobiol* 55: 225-256.
- Taube JS, Muller RU (1998) Comparisons of head direction cell activity in the postsubiculum and anterior thalamus of freely moving rats. *Hippocampus* 8: 87-108.
- Taube JS, Muller RU, Ranck JB, Jr. (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J Neurosci* 10: 436-447.
- Teng E, Squire LR (1999) Memory for places learned long ago is intact after hippocampal damage. *Nature* 400: 675-677.
- Teyler TJ, DiScenna P (1986) The hippocampal memory indexing theory. *Behav Neurosci* 100: 147-154.
- Todorov E (2007) Linearly-solvable Markov decision problems. *NIPS*
- Tolman EC (1948) Cognitive maps in rats and men. *Psychol Rev* 55: 189-208.

- Touretzky DS, Redish AD (1996) Theory of rodent navigation based on interacting representations of space. *Hippocampus* 6: 247-270.
- Trappenberg T (2002) *Fundamentals of Computational Neuroscience*. Oxford University Press
- Trettel SG, Trimper JB, Hwaun E, Fiete IR, Colgin LL (2019) Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nature Neuroscience* 22: 609–617
- Treves A, Rolls ET (1992) Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 2: 189-199.
- Trullier O, Wiener SI, Berthoz A, Meyer JA (1997) Biologically based artificial navigation systems: review and prospects. *Prog Neurobiol* 51: 483-544.
- Tse, D. et al. (2007) Schemas and memory consolidation. *Science* 316, 76–82
- Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H and Morris RGM (2011) Schema-Dependent gene activation and Memory Encoding in Neocortex. *Science* 333: 891-895
- Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL (1996) Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* 6: 271-280.
- Tulving E (1993) What is episodic memory? *Curr Perspect Psychol Sci* 2: 67-70.
- Ulanovsky, N., & Moss, C. F. (2007). Hippocampal cellular and network activity in freely moving echolocating bats. *Nature neuroscience* 10(2), 224-233
- Umbach G, Kantak P, Jacobs J, Kahana M, Pfeiffer BE, Sperling M, Lega B (2020) Time cells in the human hippocampus and entorhinal cortex support episodic memory. *PNAS* 117 (45) 28463-28474
- Uria B, Ibarz B, Banino A, Zambaldi V, Kumaran D, Hassabis D, Barry C, Blundell C (2020) The Spatial Memory Pipeline: a model of egocentric to allocentric understanding in mammalian brains. *bioRxiv*
- van de Ven, G.M., Siegelmann, H.T. & Tolia, A.S. (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nat Commun* 11, 4069
- Vanderwolf CH (1969) Hippocampal electrical activity and voluntary movement in the rat. *EEG Clinical Neurophysiology* 26: 407 – 418
- Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Van Paesschen W, Mishkin M (1997) Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277: 376-380.
- Vikbladh OM, Meager MR, King J, Blackmon K, Devinsky O, Shohamy D, Burgess N, Daw ND (2019) Hippocampal Contributions to Model-Based Planning and Spatial Memory. *Neuron* 102(3):683-693

- Wallenstein GV, Eichenbaum H, Hasselmo ME (1998) The hippocampus as an associator of discontiguous events. *Trends Neurosci* 21: 317-323.
- Wallenstein GV, Hasselmo ME (1997) GABAergic modulation of hippocampal population activity: sequence learning, place field development, and the phase precession effect. *J Neurophysiol* 78: 393-408.
- Wan H, Aggleton JP, Brown MW (1999) Different contributions of the hippocampus and perirhinal cortex to recognition memory. *J Neurosci* 19: 1142-1148.
- Wan, H. S., Touretzky, D. S., & Redish, A. D. (1994). Towards a computational theory of rat navigation. In *Proceedings of the 1993 connectionist models summer school* (pp. 11-19).
- Wang, C. et al. (2018) Egocentric coding of external items in the lateral entorhinal cortex. *Science* 362, 945–949
- Watrous, A. J., Lee, D. J., Izadi, A., Gurkoff, G. G., Shahlaie, K., & Ekstrom, A. D. (2013). A comparative study of human and rat hippocampal low-frequency oscillations during spatial navigation. *Hippocampus*, 23(8), 656-661.
- Wayne G et al. (2018) Unsupervised Predictive Memory in a Goal-Directed Agent. arXiv
- Weber SN, Sprekeler H (2018) Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity. *Elife* 7:e34560
- Welday AC, Shlifer IG, Bloom ML, Zhang K, Blair HT (2011) Cosine directional tuning of theta cell burst frequencies: evidence for spatial coding by oscillatory interference. *Journal of Neuroscience* 31: 16157–16176
- Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, Behrens TEJ (2020) The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* 183(5):1249-1263
- Whittington JCR, McCaffary D, Bakermans JJW, Behrens TEJ (2022) How to build a cognitive map: insights from models of the hippocampal formation. arXiv
- Widlowski J, Foster DJ (2022) Flexible rerouting of hippocampal replay sequences around changing barriers in the absence of global place field remapping. *Neuron* (in press)
- Wikenheiser AM, Redish AD (2015) Hippocampal theta sequences reflect current goals. *Nature Neuroscience* 18: 289–294
- Wilkie DM, Palfrey R (1987) A computer simulation model of rat's place navigation in the Morris water maze. *Behav Res Meth Instrum Comput* 19: 400-403.
- Wills T, Lever C, Cacucci F, Burgess N, O'Keefe J (2005) Attractor Dynamics in the Hippocampal Representation of the Local Environment. *Science* 308: 873-876.
- Wills TJ, Cacucci F, Burgess N, O'Keefe J (2010) Development of the hippocampal cognitive map in pre-weanling rats. *Science* 328: 1573–1576
- Willshaw DJ, Buckingham JT (1990) An assessment of Marr's theory of the hippocampus as a temporary memory store. *Philos Trans R Soc Lond B Biol Sci* 329: 205-215.

- Willshaw DJ, Buneman OP, Longuet-Higgins HC (1969) Non-holographic associative memory. *Nature* 222: 960-962.
- Wilson, M.A., and McNaughton, B.L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science* 265, 676-679
- Winocur G, Moscovitch M, Bontempi JB (2010) Memory formation and long-term retention in humans and animals: convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia* 48:2339–56
- Winter, S. S., Mehlman, M. L., Clark, B. J., & Taube, J. S. (2015). Passive transport disrupts grid signals in the parahippocampal cortex. *Current Biology*, 25(19), 2493-2502.
- Wiskott L, Sejnowski TJ (2002) Slow feature analysis: unsupervised learning of invariances. *Neural Comput* 14(4):715-70
- Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H (2000) Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* 27(3):623-33
- Yartsev MM, Ulanovsky N (2013) Representation of three-dimensional space in the hippocampus of flying bats. *Science* 340: 367 – 372
- Yartsev MM, Witter MP, Ulanovsky N (2011) Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature* 479: 103 – 107
- Yonelinas AP (2002) The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language* 46: 441-517.
- Yonelinas AP, Kroll NE, Quamme JR, Lazzara MM, Sauve MJ, Widaman KF, Knight RT (2002) Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nat Neurosci* 5: 1236-1241.
- Yoon K, Buice MA, Barry C, Hayman R, Burgess N, Fiete IR (2013) Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience* 16: 1077-1084
- Yu C, Behrens TEJ, Burgess N (2020) Prediction with directed transitions: complex eigenstructure, grid cells and phase coding. *arXiv*
- Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron* 75(1): 168-179
- Zemel RS, Dayan P, Pouget A (1998) Probabilistic interpretation of population codes. *Neural Comput.* 10(2):403-30
- Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci* 16: 2112-2126.
- Zhu XO, McCabe BJ, Aggleton JP, Brown MW (1996) Mapping visual recognition memory through expression of the immediate early gene c-fos. *Neuroreport* 7: 1871-1875.

Zilli EA, Hasselmo ME (2008) Analyses of Markov decision process structure regarding the possible strategic use of interacting memory systems. *Front Comput Neurosci* 2:6

Zipser D (1985) A computational model of hippocampal place fields. *Behav Neurosci* 99: 1006-1018.

Zipser D (1986) Place recognition. In: *Parallel distributed programming Vol 2: Psychological and biological models* (McClelland JL, Rumelhart DE, eds), pp 432-470. MIT Press.

Ziv Y, Burns LD, Cocker ED, Hamel EO, Ghosh KK, Kitch LJ, El Gamal A, Schnitzer MJ (2013) Long-term dynamics of CA1 hippocampal place codes. *Nat Neurosci* 16(3):264-6

Zola SM, Squire LR, Teng E, Stefanacci L, Buffalo EA, Clark RE (2000) Impaired recognition memory in monkeys after damage limited to the hippocampal region. *J Neurosci* 20: 451-463.

Zola-Morgan S, Squire LR, Ramus SJ (1994) Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system [see comments]. *Hippocampus* 4: 483-495.

Zugaro MB, Monconduit L, Buzsáki G (2005) Spike phase precession persists after transient intrahippocampal perturbation. *Nature Neuroscience* 8: 67-71