Space and Context in the Rodent Hippocampal Region

Mark C. Fuhs

CMU-CS-06-170

November 2006

School of Computer Science
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

thesis committee
David S. Touretzky (Chair)
James L. McClelland
Tai Sing Lee
Robert U. Muller (SUNY Downstate)

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

This work was supported by an NSF Graduate Research Fellowship, NSF IGERT DGE-9987588, NIH MH59932, and the Pennsylvania Tobacco Settlement Fund. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: entorhinal cortex, hippocampus, place cells, grid cells, path integration, attractor dynamics, model selection.

ABSTRACT

Recently, cells in the dorsal medial entorhinal cortex (dMEC) were found have spatially modulated activity patterns, comprising multiple active regions organized in a hexagonal lattice across the environment. These "grid cells" show the same field spacing in any environment and are sometimes modulated by the rodent's speed and direction of travel, leading to the hypothesis that dMEC subserves a path integration system. Downstream, hippocampal "place cells" are typically active only within a single contiguous region of an environment. Unlike grid cells, changes either in sensory information or to the rodent's task can cause place cells to "remap," or radically change their activity patterns.

The first part of this thesis presents a neural network model of dMEC grid cells that provides both a cogent explanation of the firing properties of the grid cells and a mechanism by which they could satisfy the computational requirements for path integration. The efficiency and properties of a hippocampal spatial code derived from grid cells is also explored.

The second part considers place cell remapping as a method of encoding context and presents a Bayesian statistical model for context learning. Context learning is defined as the problem of decomposing the rodent's history of experiences into temporal windows within which the distribution of sensory and task-related hippocampal input is statistically stationary. Context learning can therefore be understood as a model selection problem: how many contexts make up the rodent's world? The theory provides an understanding of why remapping sometimes develops gradually over many days of experience, why the time course of reversal learning depends on the degree to which the reward contingencies were changed, and why overlapping sequence learning does not consistently result in "context-dependent" sequence representations.

The third part presents an analysis of a hippocampal physiology experiment, in collaboration with Bruce McNaughton (University of Arizona), that pitted sensory information against path integration information. Rats foraged in two identical, connected boxes with either the same or opposite orientations. The observed pattern of place cell responses suggest that a combination of linear and angular path integration eventually overrides sensory cues, even when linear path integration alone does not.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and guidance of a number of people, and I would like to extend my deep gratitude to each one. My advisor, Dave Touretzky, has been very generous with his time and support of this work. Many a late night, impromptu meetings with him have led to detailed dissections of new research ideas that markedly clarified whether there was merit or (often) not in pursuing them. My thanks to Bill Skaggs who, as a co-advisor during my IGERT rotation, gave me a hands-on introduction to experimental neurophysiology. My thanks also to the members of my thesis committee, Jay McClelland, Tai Sing Lee, and Bob Muller, who each contributed a unique perspective to this research, as well as Bruce McNaughton, Amanda Casale and Shea VanRhoads and the NSMA group at the University of Arizona at Tucson for their warm collaboration on the experimental study presented in Chapter 8.

In addition, I have had a number of stimulating interactions with current and former students and faculty, including Aaron Courville, Nathaniel Daw, Bard Ermentrout, Geoff Gordon, Francisco Pereira, Dave Redish and Chuck Rosenberg. All of them and many others I met through the Center for the Neural Basis of Cognition and / or the Computer Science Department, two extraordinary organizations that exemplify the interdisciplinary spirit of research at CMU.

Finally, I would like to thank my parents for their boundless love and their inspiring and unshakable faith in my potential for success.

ABBREVIATIONS

CA1 cornu ammonis (Ammon's horn) subregion 1

CA2 cornu ammonis (Ammon's horn) subregion 2

CA3 cornu ammonis (Ammon's horn) subregion 3

CPFE context pre-exposure facilitation effect

DG dentate gyrus

dMEC dorsolateral medial entorhinal cortex

EC entorhinal cortex

FFX fibria-fornix pathway

HF hippocampal formation

LTP long-term potentiation

LEC / LEA lateral entorhinal cortex (area)

MEC / MEA medial entorhinal cortex (area)

PREE partial reinforcement extinction effect

CONTENTS

1

INTRODUCTION

One of the most widely discussed topics in the philosophy of mind is the question of whether the brain can be understood as a computational process (e.g. Searle, 1980; Pylyshyn, 1989; Dennett, 1994). This question originated in the formal sense out of the Church-Turing thesis which posited that every computable function could be computed using a Turing machine (Kleene, 1943). To this day, no formalism of computation has been offered that defines a class of computations not computable by a suitable Turing machine. That the Turing machine formalism suffices to describe any computation has naturally led to the possibility that, to the extent to which the brain can be described as a computational system, its computations could be understood in algorithmic form.

Discovering the computational processes underlying brain function has proven to be an unparalleled scientific challenge, and, as in other domains of hard science, concrete, quantitative theories are critical to the scientific process. Indeed, major advances in our understanding of brain function have been made by so-called "rational analysis," the exploration of how neural systems may implement an approximation of a computationally ideal system. For example, neuronal responses in the visual and auditory systems have been shown to reflect information theoretic coding principles. The role of the dopamine system in animal learning has been elegantly formalized within a reinforcement learning framework. Examples are numerous.

This thesis advances along that path by contributing to an understanding of the rodent hippocampal region using a variety of computational techniques in machine learning and artificial neural networks. The hippocampal region receives convergent input from a diverse array of brain structures, leading to its implication in a variety of cognitive functions. In recent years, lesion and physiology studies have provided converging evidence that the hippocampus proper is critical for learning associations between sensory stimuli and the rodent's spatial location. Indeed, theories of rodent navigation have centered around the hippocampus and the "place cells" within that show spatially selective firing fields (for review, see Redish, 1999).

While navigation studies have focused on associations with particular locations, "contextual" classical conditioning studies have focused on associations with whole environments, and hippocampal place cells have been shown to

have different activity patterns in different environments. The notion that the hippocampus provides a representation of context has become pervasive, and it has been suggested that hippocampal lesion impairments in other tasks, such as reversal learning and T-maze alternation, derive from the lesioned brain's lack of a representation of context.

This thesis comprises two parts. The first part, Chapters 2-5, provides a review of the experimental literature, while, the second part, Chapters 6-8, covers original research that advances an understanding of hippocampal region function in rodent navigation and in providing a representation of context. In Chapter 2, the neuroanatomy of the hippocampal region is reviewed, including both a description of connectivity within the region and a summary of the afferent and efferent projections of structures within the region. Chapter 3 presents a review of lesion and physiology data involving two structures in the parahippocampal region, the postsubiculum and entorhinal cortex, geared toward an understanding of these structures roles in rodent navigation. Chapter 4 presents a systematic review of the physiology data from the hippocampus, while Chapter 5 presents a review of behavioral data, including the effects of hippocampal lesions, transient neuronal inactivation, and genetic knockouts, focusing where possible on studies that restrict the affected area to particular subregions (DG, CA3, CA1).

Chapter 6 presents a neural network model of a path integration system (Fuhs and Touretzky, 2006), a critical component for rodent navigation. The model connects previous theories that a path integrator must exist within the brain with recent data from the dorsal medial entorhinal cortex (dMEC), proposing a model of dMEC that both explains the intriguing properties of dMEC neurons and demonstrates how cells with such properties could compose a neural path integrator.

Chapter 7 presents a computational theory of context learning in the rodent hippocampus. The theory presents a single, unified notion of context that is equally applicable across the different experimental domains in which context has been posited to play a critical role. The theory formalizes this notion of context within a Bayesian statistical framework, demonstrating that the development of different contextual representations within the hippocampus can be understood as a statistical inference problem.

Chapter 8 presents a physiology experiment that explores the roles of linear and angular path integration in forming spatial representations in the hippocampus (Fuhs et al., 2005). This author was responsible for the complete analysis of the raw data, including the development of a novel inference technique to measure moment-to-moment changes in the rats' representation

of the environment.

Chapter 9 concludes with a discussion of experimental and theoretical questions prompted by this work that may be addressed by future research.

2

NEUROANATOMY

There is some inconsistency regarding the naming conventions involving groups of structures in the hippocampal region. Consistent with Witter and Amaral (2004), I will use the term "hippocampal region" to denote two sub-structures, the archicortex, or "hippocampal formation," and the "parahippocampal region." The hippocampal formation includes the *fascia dentata* (dentate gyrus), *cornu ammonis* (Ammon's Horn, including subfields CA3, CA2 and CA1) and the subiculum. The "parahippocampal region" denotes the entorhinal cortex (EC) and pre- and para-subiculum. The term "hippocampus" is most commonly used to denote Ammon's horn and the dentate gyrus, and it will be used so here. In the remainder of this section, a summary of the intrinsic and extrinsic connectivity is presented based on reviews by Witter et al. (2000; 2004), unless otherwise noted.

The hippocampal formation, situated in the medial temporal lobe in mammals (see Figure 2.1), is part of the limbic system, a phylogenetic predecessor of the neocortex. It is sometimes referred to as "archicortex" because its laminar organization is more primitive than neocortex, containing only three anatomically distinct cell layers and only one layer in which principal cells reside. Similar structural homologues, often referred to as the medial pallium, have been identified across the Chordate phyla, including birds, frogs, and fish.

2.1 Entorhinal Cortex

The dominant afferent input to the entorhinal cortex arises from two neighboring structures, the perirhinal and postrhinal cortices, each of which receives converging input from a diverse panoply of cortical areas. In addition, the entorhinal cortex receives somewhat weaker input from many of these areas as well, including perirhinal and postrhinal cortices; the olfactory bulb, anterior olfactory nucleus, and piriform cortex; agranular insular cortex; medial prefrontal cortex, especially the prelimbic / infralimbic areas; anterior cingulate and posterior cingulate (retrosplenial) cortex; insular cortex; ventral posterior temporal cortex; and posterior parietal cortex.

Afferent connectivity

EC is subdivided into medial and lateral regions. The distribution of afferent projections to these regions is not uniform. Figure 2.2 shows the

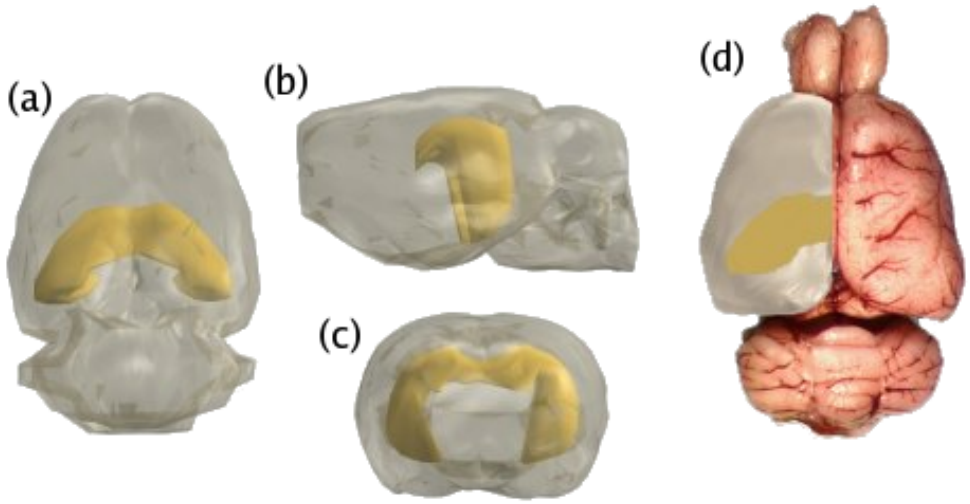

FIGURE 2.1: The position of the hippocampal formation within the rodent brain. In views (a) and (d), the anterior-to-posterior (front-to-back) axis of the brain is top-to-bottom on the page. In (b), the anterior end is toward the left. View (c) is from the anterior end of the brain.

distribution of inputs from other cortical structures. Simplifying this picture somewhat results in the pattern of connections shown in Figure 2.3. Interestingly, the afferent visual and spatial areas (retrosplenial, parietal, and occipital) preferentially target the medial entorhinal cortex (MEC), especially the dorsal portion. Of particular note is the projection from the neighboring presubiculum (including the dorsal part, also referred to as the postsubiculum) whose projection to EC targets layer III of the medial portion almost exclusively.

In addition to the cortical afferents, EC receives projections from a number of subcortical areas as well, including the thalamus, mostly from the nucleus reuniens and nucleus centralis medialis; medial septal complex; amygdala; and hypothalamus, including the supramammillary nucleus and lateral hypothalamic area. Monoaminergic inputs from the central tegmental area, dorsal raphe nuclei, and locus ceoruleus have also been found.

Intrinsic and efferent connectivity

Projections from EC layers II and III form one of the dominant input pathways, the perforant path, to all subregions of the hippocampal formation. Return projections from the hippocampus terminate largely in layer V of EC. The principal cells of layer II, stellate cells, are organized in groups ("cell islands"), whereas cells in other layers are more diffusely arranged. Some cells in layer II also synapse onto other cells within layer II, and, in fewer instances, layer III. Layer III cells give off collaterals within layer III and to

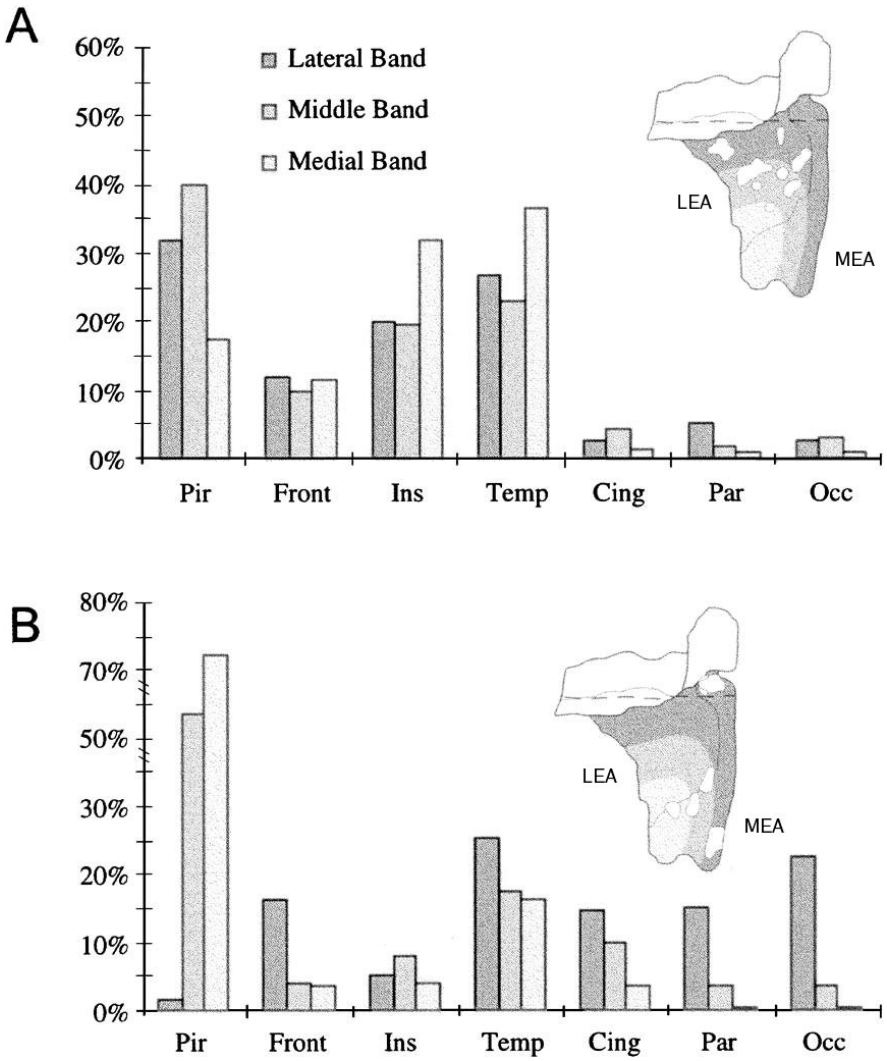

FIGURE 2.2: Distribution of inputs to the LEA (A) and MEA (B) arising from piriform, frontal, insular, temporal, cingulate, parietal and occipital cortex, respectively. The lateral, middle, and medial bands indicate areas of LEA and MEA that project to dorsal, medial, and ventral structures in the hippocampal formation. (Reprinted from Burwell (2000).)

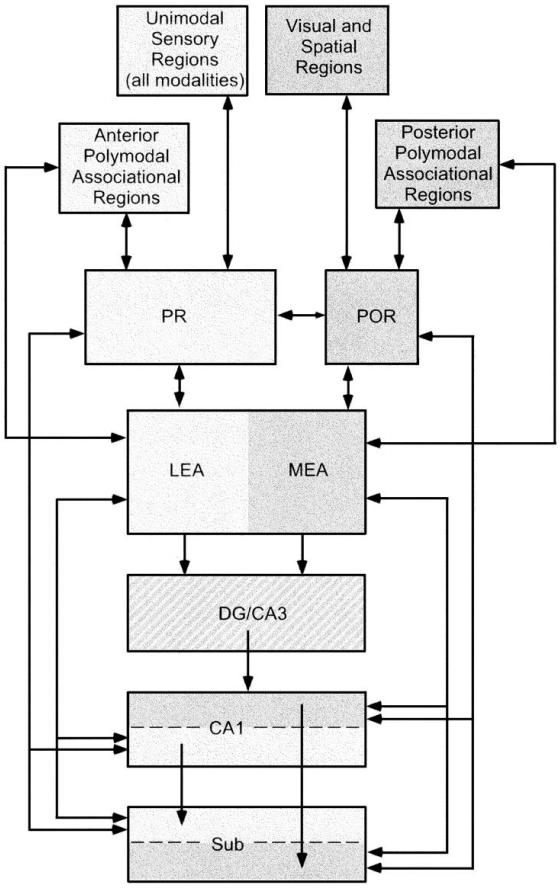

FIGURE 2.3: Projections to LEA and MEA differ. LEA and MEA afferents mix in DG and CA3, but are kept segregated in CA1 and the subiculum. (Reprinted from Burwell (2000).)

layer I, which contains a sparse population of both stellate (excitatory) and GABAergic (inhibitory) horizontal cells that terminate onto the dendrites of layer II stellate cells.

Efferent cortical projections from EC are targeted most strongly at the perirhinal and postrhinal cortices, arising largely from deep EC layers. Perirhinal and postrhinal cortices in turn have projections back to unimodal sensory and polymodal association areas. In addition, direct projections to olfactory areas, prelimbic (medial prefrontal) cortex, orbitofrontal agranular insular cortex, and retrosplenial cortex have been found. Subcortical projections are largely to the septum, mostly lateral, the amygdala, and nucleus accumbens (ventral striatum).

2.2 Hippocampus Formation

The hippocampal formation makes afferent and efferent connections largely, though not exclusively, via two main pathways. The first, the fimbria-fornix (FFX) pathway, connects the HF directly with several subcortical structures, while the second, the entorhinal cortex, provides the gateway to and from a multitude of areas throughout the neocortex. The HF being a bilateral structure, substantial intrinsic interhemispheric projections are also made via the anterior commissure. A summary of the connectivity of the hippocampal formation is shown in Figure 2.4.

Afferent connectivity

The entorhinal cortex provides the dominant cortical input to the HF, a fiber bundle called the perforant path. Perirhinal and postrhinal cortices, which project heavily to EC, also contribute to the perforant path projection. EC layer II projects to the dentate and CA3, while layer III projects to CA1 and the subiculum. (The projection from EC (in medial temporal cortex) to the hippocampus (Ammon's horn) is sometimes referred to as the temporoammonic pathway, often in contrast with perforant path terminals in the dentate.) CA1 and the subiculum both project back to the deep layers of EC, predominantly layer V.

The fimbria-fornix pathway is dominated by GABAergic and cholinergic projections from the medial septum and a neighboring structure, the nucleus of the diagonal band of Broca. The GABAergic projections largely terminate on GABAergic neurons, providing essentially an inhibition of intrahippocampal inhibition. These GABAergic projections provide the substantial theta-frequency (4–12Hz) modulation of hippocampal activity.

Intrinsic and efferent connectivity

Descriptions of the intrinsic circuitry of the hippocampal formation typically involve a reference to the "trisynaptic pathway," denoting a flow of information from EC to DG, CA3, and CA1 (see Figure 2.5). However, as

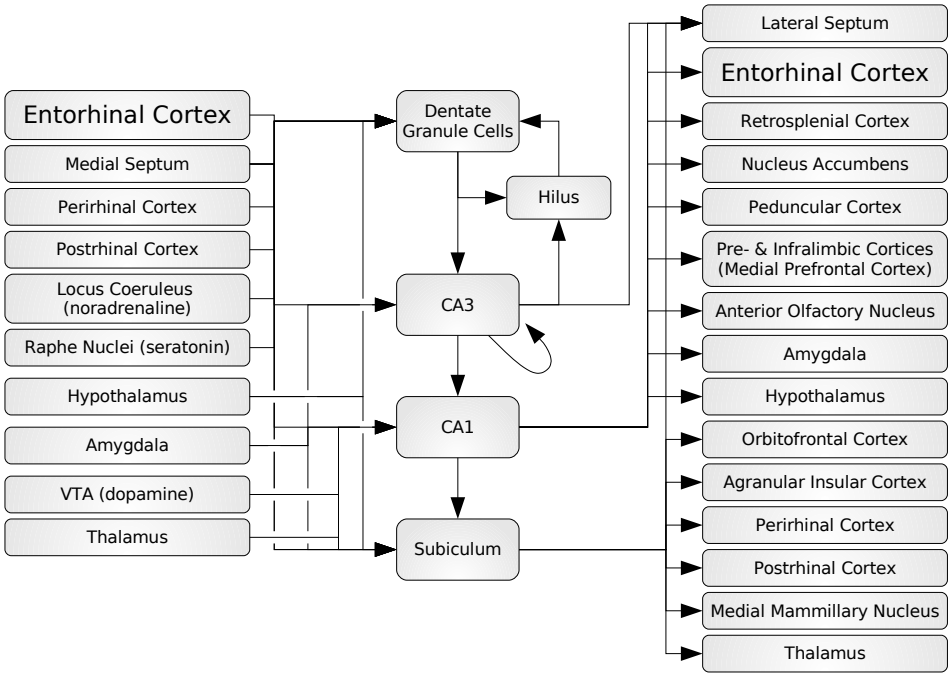

FIGURE 2.4: Intrinsic and extrinsic connections of the hippocampal formation.

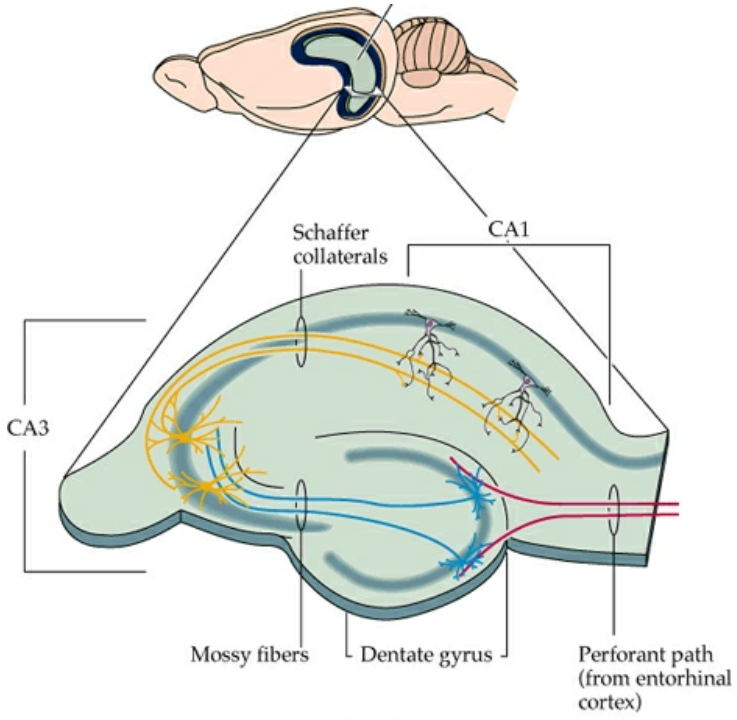

FIGURE 2.5: Anatomically "correct" diagram of the trisynaptic pathway within the hippocampus. Perforant path projections from the entorhinal cortex project to DG; the DG mossy fiber pathway projects to CA3; Schaffer collaterals from CA3 project to CA1.

Figure 2.4 shows, the circuitry is far more complex, and involves multiple sets of recurrent connections both between and within HF regions.

The dentate gyrus contains two major groups of excitatory cells, granule cells in the granule layer and mossy cells in the polymorphic layer. The polymorphic layer is often referred to as the hilus, and has also sometimes (erroneously) been grouped with the hippocampus as region CA4. Granule cells project to both hilar mossy cells and CA3 pyramidal cells via axonal projections referred to as "mossy fibers." Mossy fiber synapses are sizable and quite potent, requiring few mossy fiber action potentials to trigger an action potential in the post-synaptic cell. However, an individual granule cell makes few connections, including only 14–28 CA3 pyramidal cells. Hilar mossy cells send strong projections both back to the granule cells as well as to inhibitory basket cells, forming recurrent circuitry within DG comprised of both excitatory and inhibitory components. The mossy cells also receive projections from CA3, forming a larger recurrent loop encompassing both DG and CA3.

In CA3, in addition to perforant path and mossy fiber input, substantial and highly divergent recurrent collaterals between pyramidal cells are found. These same axons bifurcate to form the Schaffer collateral projection into CA1. Interestingly, CA3 is known to have exactly one extra-hippocampal target, the lateral septum, an important consideration when interpreting subregional lesion studies showing a larger effect of DG or CA3 lesions than CA1.

The Schaffer collateral and perforant path input converge in CA1, and there are no projections back to DG or CA3. CA1 projects both directly and indirectly (via the subiculum) to the entorhinal cortex, as well as directly to several cortical and subcortical structures. The subiculum receives perforant path input as well, and provides the dominant cortical and subcortical output of the hippocampal formation. Because CA1 and the subiculum are the only two areas to have substantial extra-hippocampal projections, interpretation of CA1 lesion studies requires consideration of whether impairment is due to a loss of information processing in these areas or a loss of information propagation to downstream areas.

It is interesting to note the distribution of perforant path projections from medial and lateral EC. As mentioned above, these areas receive somewhat different input from afferent cortical areas, and the LEC and MEC project to non-overlapping portions of CA1 and the subiculum (see Figure 2.3). By contrast, DG and CA3 receive convergent input from both LEC and MEC, suggesting a functional role in unifying or associating information across these two areas.

3 THE PARAHIPPOCAMPAL REGION

3.1 The Head Direction System and Postsubiculum

While this thesis does not directly address the inner workings of the head direction system, a brief summary is provided here as background for future discussions of related systems. For a more comprehensive discussion, the reader is referred to any of several models of the head direction system (Skaggs et al., 1995; Redish et al., 1996; Blair et al., 1997; Goodridge and Touretzky, 2000; Song and Wang, 2005).

Head direction cells are so named because they fire preferentially when the animal's head is facing a particular horizontal bearing (azimuth) in the environment. Cells sensitive to head direction have been found in several regions of the brain, most notably the postsubiculum (PoS), the anterior dorsal nucleus of the thalamus (AD), and the lateral mammillary bodies (Taube et al., 1990a,b; Blair et al., 1997; Blair and Sharp, 1998; Stackman and Taube, 1998). Additionally, medial prestriate and retrosplenial cortex cells were sensitive to head direction and in many cases modulated by one or more of: behavior (left turns, right turns, forward motion), visual cues, and vestibular information (passive rotation) (Chen et al., 1994b,a). Some striatal neurons have also been found to show head direction sensitivity (Wiener, 1993).

The postsubiculum, part of the parahippocampal region, and AD are reciprocally connected, and PoS appears to derive its head direction signal from AD, as AD lesions disrupt the head-direction tuning of PoS cells (Goodridge and Taube, 1994, 1997). AD head direction tuning is in turn dependent on the vestibular system, as vestibular system lesions abolish AD head direction specific firing (Stackman and Taube, 1997). Additionally, it is believed that AD receives motor efference copy, indicative of very-near-future movements, as AD cells are predictive of future head direction by approximately 23 msec (Taube and Muller, 1998).

If the postsubiculum derives its head direction tuning from AD, what purpose does PoS serve? An intact postsubiculum is not necessary for head direction specificity in AD (Goodridge and Taube, 1997), but PoS lesions do disrupt hippocampal place cell firing (Calton et al., 2003) and spatial navigation tasks typically associated with hippocampal function (Taube et al.,

FIGURE 3.1: In a typical awake, behaving, rodent neurophysiology experiment, a microdrive is surgically implanted atop a rat's head, and one or more recording electrodes are lowered into the targeted brain region. In addition, one or two ground wires are positioned nearby but outside the region, away from any cell bodies, to act as a voltage reference. In tasks involving some spatial component, LEDs attached to the top of the microdrive are tracked by a camera that is mounted to the ceiling of the room. This provides for the simultaneous recording of the animal's position, its head direction (in some cases), and the activity of the neurons. (Photo by Eric L. Hargreaves. Reprinted with permission.)

1992). The postsubiculum has direct projections to the hippocampal region, specifically the medial entorhinal cortex, and is likely critical for updating the representation of space in MEC during movement (see Section 3.2 and Chapter 6).

The information flow from PoS to AD is critical for updating AD based on associations of head direction with external landmarks. Normally, familiar external cues are able to reorient the head direction system to reestablish correspondence between an observed cue and its previously learned compass direction (Goodridge and Taube, 1995). However, cue control of cells in AD is largely abolished when PoS is lesioned (Goodridge et al., 1998).

3.2 PATH INTEGRATION AND THE ENTORHINAL CORTEX

3.2.1 *Behavioral path integration*

While animal navigation based on external landmarks is well documented (e.g. Collett et al., 1986), it has also been shown that a wide variety of animals can navigate in the absence of spatially informative cues (for review, see Etienne and Jeffery, 2004). Path integration experiments typically rely on the nesting behavior of the species being studied. After searching in an environment in a long circuitous path for a target object, perhaps some food or a lost pup, the animal returns directly to its nest. Several experiments have been carried out to dissociate navigation by external spatial cues from navigation using an internal path integration mechanism (Mittelstaedt, 1962;

Mittelstaedt and Mittelstaedt, 1980; Etienne et al., 1986, 1988, 1996).

As a classic example, Mittelstaedt and Mittelstaedt (1980) tested path integration in a pup retrieval task. A gerbil's pup was displaced from its mother's nest at the edge of an arena into the center. The mother gerbil then departed from the nest, searched the environment, and located the pup. The pup was placed on an area of the floor that was slowly rotated while the mother was upon it. Since the speed of rotation was below detection by the mother's vestibular system, the mother set off in a direction different from the actual location of the nest. The deviation from the correct nest location was predicted by the amount of rotation, indicating that the return path was not guided by external cues but rather by some internal navigation system.

This behavioral evidence led to a variety of theories that postulated some form of neural path integration system (for discussion, see Chapter 6). Recent evidence suggests that the entorhinal cortex provides a representation of space based on idiothetic (self-motion) information that could serve as the basis for path integration.

3.2.2 *Entorhinal cortex*

The spatial properties of cells in the entorhinal cortex were first explored by Quirk et al. (1992), who recorded from medial entorhinal cortex. The cells from which they recorded were spatially selective, showing a "place field," a spatially localized area of activity within the environment surrounded by quiescence. When a prominent cue card affixed to the wall of the arena was rotated, MEC place fields rotated with it. However, removing the cue had little impact on the cells' activity. Moreover, in contrast to place cells in the hippocampus (see Section 4.1), MEC place fields were similar in arenas with two different shapes (square and cylindrical), suggesting that they represented a "universal" (environment independent) map of space.

Grid cells

More recently, a series of papers from Moser and colleageus have reported in detail about the physiological properties of MEC neurons (Fyhn et al., 2004; Hafting et al., 2005; Sargolini et al., 2006). Fyhn et al. (2004) systematically recorded from layer II neurons along the dorsal-to-ventral axis of MEC. Fascinatingly, the firing fields in the dorsal subregion (dMEC), nearest the border with postrhinal cortex, had multiple peaks, or nodes, and the structure of the nodes obeyed the planar symmetry group p6m, the most symmetric of the 17 groups (Grünbaum and Shephard, 1987). Of p6m's many rotations, reflections and glide reflections, the most notable is the order six rotational symmetry: firing fields rotated about any node by any multiple of 60 ^{o} were similar to the original field. Each node was surrounded by six equally spaced nodes around it. Treating nodes as points in a point lattice, one may therefore

describe the point lattice as hexagonal, as the Voronoi polygons of the lattice are hexagons, or equilateral triangular, as the Delaunay triangulation comprises equilateral triangles. These cells have been called "grid cells."

As recording distance is increased from the postrhinal border along the dorsal-ventral axis, both the size of fields and spacing between them increased proportionately. In the ventral portion of dMEC, little or no spatial specificity was observed. However, the increase of field sizes and spacing suggests that the cells may not be spatially non-selective; rather, their spatial selectivity may be observable only over arena sizes in excess of what was used.

Postrhinal cortex provides the dominant visuospatial input to dMEC and therefore its strongest source of spatial information. However, Fyhn et al. (2004) report that fields in postrhinal cortex showed little spatial information content or spatial correlation across sessions. Additionally, lesions of DG, Ammon's horn and the subiculum did not destroy the spatial specificity of dMEC fields either. The lesions did not significantly impact spatial information rate, the median number of peaks or the mean field size, though trends toward larger, less spatially informative fields were present. These lesions did impact the spatial coherence and the uniformity of dispersion of the firing fields, though fields were still more dispersed than chance. Thus, the two most obvious sources of afferent spatial information for dMEC cells do not in fact appear necessary for spatially localized dMEC fields.

Grid cells as a universal map

Hafting et al. (2005) reported several additional properties of layer II dMEC cells. Of particular interest was the behavior of simultaneously recorded cells in proximity to one another (recorded on the same tetrode). First, the spacing of nodes varied by 30 cm across the dorso-ventral span from which they recorded. However, the spacing of neighboring neurons varied very little, on average only 2.1 cm. Similarly, the orientations of grid cell fields observed across rats were uniformly distributed, but, among neighboring cells, the orientations were tightly coupled. In contrast, neighboring grid cell field phases showed no statistical correlation. Taken together, these results suggest that local connectivity couples the behavior of proximal units, whereas more distal neurons showed no evidence joint participation within the same neural network. Whether these local neural networks are related to the stellate "cell islands" observed in layer II has yet to be determined.

Hafting et al. (2005) also explored the effects of several environmental manipulations on layer II dMEC cells. First, in a novel environment, grid cell fields were observed immediately. The first two minutes show significant correlations with later recording, though they are not as sharp as subsequent two minute blocks. Second, when the size of the arena was varied, the size and spacing of the nodes remained constant. (Subsequent preliminary

data suggest that grid cells do not distinguish between arena shape or color either (Hafting et al., 2004; Fyhn et al., 2005).) Third, when all lights were extinguished, grid spacing, mean firing rate and spatial information remained unchanged, though there was a mildly weaker spatial correlation across sessions. If one takes these cells to indicate the rat's internal representation of his location, then this weaker spatial correlation may simply reflect the darkness-induced error in updating that spatial representation during travel. Taken together, these data suggest that the first-order hexagonal structure of the grid cell fields is sensory independent. These properties are similar to the head direction system and suggest that internal motion information, whether from the vestibular system or motor efference copy, likely plays a significant role in updating the dMEC representation. For this reason, the grid cells have been conjectured to underlie the behavioral ability to path integrate.

Grid cells are not immune to all cue changes, however. When a prominent orienting cue card was rotated 90°, grid cell fields rotated equivalently. Such field rotation presumably occurs in an effort to realign the internal spatial map with the external world, as is observed in the head direction system.

Beyond the first-order hexagonal structure, it was observed that the individual peak firing rates of different nodes varied, and these variations were systematic. Repeated visits to the same environment showed that peak firing rates were positively correlated across sessions (r = 0.35). Whether this modulation reflects the impact of afferent input or some internal dynamic of dMEC is not known. However, the wealth of afferent input to MEC suggests that these peak variations may in fact represent a superposition of afferent input and the dMEC representation of space.

Grid cells and path integration

Sargolini et al. (2006) recorded from layers III, V and VI to assess the degree to which cells in these other layers show similar properties to the layer II cells. They found that, while most cells in layer II show p6m periodicity, only a minority of cells in III, V and VI do. Of those that did, similar properties were found. Node spacing increased with distance from the postrhinal border. Grid orientations of nearby, simultaneously recorded cells were consistent, while grid phases of nearby, simultaneously recorded cells were uniformly distributed.

While cells in layer II showed almost no directional tuning, approximately 70% of cells in each of the other recorded layers did. The distribution of preferred directions appeared uniform, and, among simultaneously recorded neurons, the preferred directions were heterogeneous. Moreover, the population of grid and directional neurons were not disjoint. While neurons in layer II showed no directional preference, 66% of layer III and 90% of layer

V neurons showed both directional and grid activity patterns. In layer VI, that percentage drops to 28%. When a prominent orienting cue card was rotated 90°, the grid fields and direction tuning rotated together, consistent with the notion that a change in the internal compass of the rat (e.g. in the head direction system) underpins both the direction preference of the cells and the dynamics of the grid cell fields.

Finally, Sargolini et al. (2006) note that there is a positive correlation between firing rate and speed. Speed and directional preference together compose a motion vector, the fundamental building block of a path integration system. This existence of this information in dMEC was predicted by the theoretical work of Fuhs and Touretzky (2006), which is presented in Chapter 6.

While all of these studies have focused on MEC, Hargreaves et al. (2005) recorded from several areas in the hippocampal region, including CA1, MEC, LEC, the parasubiculum and perirhinal cortex (the major input to LEC). They found that cells in CA1, MEC and parasubiculum show spatial selectivity, whereas cells in LEC and perirhinal cortex largely do not. Both spatial information (bits per spike) and spatial correlations between sessions are lower for LEC and perirhinal. Taken together with the Fyhn et al. (2004) data showing little spatial information in postrhinal cortex, it appears that the spatial representation in dMEC is created therein and propagated into the hippocampal formation.

4

THE HIPPOCAMPAL FORMATION: NEUROPHYSIOLOGY

4.1 Place Cells

Well before dMEC grid cells were discovered, DG, CA3 and CA1 principal cells had been known to have spatially localized firing fields, which has led to their being described as "place cells" (O'Keefe and Dostrovsky, 1971; Jung and McNaughton, 1993). In contrast to dMEC, a cell usually has at most one field in an environment, though granule cells more often have multiple fields (see Figure 4.1; Jung and McNaughton 1993; Wilson and McNaughton 1993). Individual place cells therefore code more clearly for a particular place than do grid cells. Hippocampal place fields are slightly larger and fire at lower rates than those in dMEC, and they are generally quite stable across sessions, somewhat more so than dMEC grid cells (Fyhn et al., 2004). Also, in addition to the animal's location, hippocampal principal cells are modulated by a variety of environmental stimuli and internal variables.

Most neuronal recordings are performed toward the dorsal end of hippocampus; in the ventral portion, fewer cells show spatial selectivity and those that do show more diffuse fields. However, recent evidence suggests that ventral cells may simply have very large place fields (Moser, 2006).

4.1.1 *Place cell maps and remapping*

In dMEC, grid cells appear to be active in all environments, and neighboring grid cell fields appear to have a fixed phase relationship: when the environment changes, the phase relationship among nearby grid cells does not (Fyhn et al., 2005). By contrast, most place fields are inactive in most environments, and, when active, the place fields of hippocampal neurons randomly reorganize between environments (see Figure 4.2; O'Keefe and Conway (1978); Muller and Kubie (1987); Thompson and Best (1989); Wilson and McNaughton (1993)). Thompson and Best (1989) identified hippocampal complex-spike burst (pyramidal) cells under barbiturate anesthesia, which causes all cells to become active. They then recorded from these cells while rats explored three different environments, a radial maze, a cylindrical environment, and a rectangular environment. Only 37% of cells showed a place field in any of the environments, and these cells were often active in only

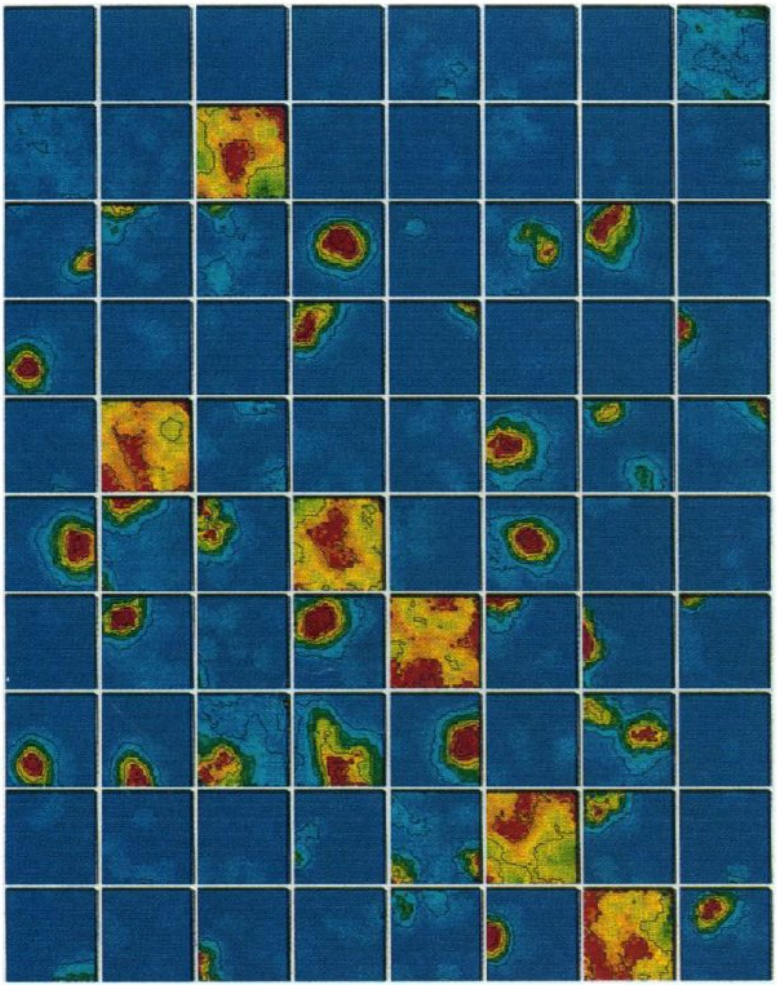

FIGURE 4.1: Simultaneously recorded cells in the dorsal CA1 region of the hippocampus while a rat foraged for food in a square arena. While many pyramidal or "place" cells are inactive or have a single spatially localized field, interneurons are active throughout the environment. (Reprinted from Wilson and McNaughton, 1993.)

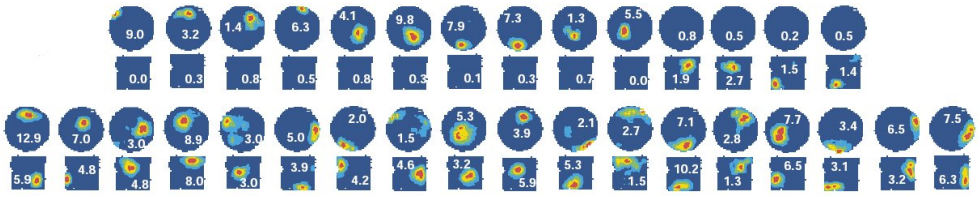

FIGURE 4.2: Place cells "remap" between familiar square and cylindrical arenas. (Reprinted from Lever et al., 2002.)

one of the three environments. The remaining two-thirds of the isolated cells were not active in any environment. Statistical tests showed that cells were selected as if at random to be active in each environment. Thus, the subset of cells active in each environment is sparse (few in number) and random (different in each environment).

Muller and Kubie (1987) studied how changing the shape of the arena affected hippocampal place cell activity. They recorded from two familiar arenas, one cylindrical and one square, and again found that different but overlapping populations of place cells were active in each arena. Moreover, of those cells that were active in both arenas, the place fields were different in an apparently random way. They coined the term "remapping" to describe this place cell selection and field reorganization phenomenon. A recent meta-analysis by Redish et al. (2001) showed that there were no anatomical correlates to how cells remapped: anatomically nearby cells (< 1400 microns apart) were no more likely to have spatially correlated firing fields than more distant cells recorded on different electrodes.

These results yield two critical insights. First, while the dMEC creates a universal map of space, the hippocampal representation is context-dependent. Second, while dMEC cells are active in all environments, the hippocampal encoding is based on a representation that is sparse. This encoding sparsity is achieved in two ways. First, the subset of active cells changes between contexts. Second, the subset of active cells *in any one place* changes between contexts. If two place cells are active in two different contexts and they happen to be active in the same location in one context, they are nonetheless unlikely to be active in the same location in the second context. Thus, even among cells active in multiple contexts, the population activity at any location within either context is quite different. This sparse encoding minimizes interference between contexts, which is critical for storing and recalling multiple activity patterns in an associative neural network (Marr, 1971; Hopfield, 1982, 1984). Since an animal would typically experience many contexts over the course of a day or week, the sparsity therefore likely reflects

an encoding strategy to maximize the number of contexts simultaneously representable within the hippocampus. Sparse encoding will be discussed in more detail in Chapter 7.

Long-term (in)stability of place fields

Place fields are generally stable across recording sessions, and, in one study, a place field was shown to be stable for 153 days (Thompson and Best, 1990). However, place field stability has been shown to be dependent on NMDA-dependent long-term potentiation (LTP) of synaptic efficacies. Normally, LTP is triggered subsequent to an influx of Ca^{2+} through NMDA receptors, which are sensitive both to the potential of the postsynaptic cell and to presynaptic release of glutamate (or exposure to the synthetic amino acid N-methyl-D-aspartate; NMDA). The Ca^{2+} activates the calcium-calmodulin-dependent kinase II (CaMKII) enzyme, resulting in an immediate potentiation via the CaMKII phosphorylation of nearby glutamate-sensitive AMPA receptors to make them more permeable (Silva et al., 1992). Additionally, longer-term structural changes in the synapse requiring protein synthesis can more permanently cement the change in efficacy (Toni et al., 1999). Rotenberg et al. (1996) tested a mutant strain of mice that expressed a Ca^{2+}-independent form of CaMKII, and found that place fields were less spatially localized and less stable between sessions.

Subsequent studies have confirmed this instability. Barnes et al. (1997) compared the stability of place fields in young and old rats, LTP deficiency having been implicated in the latter group. While the place fields of young rats were reliable across multiple sessions, between which the rats were removed from the maze, old rats sometimes showed the same fields but sometimes did not. A similar result was observed by Kentros et al. (1998), who created an LTP-deficient population by systemic pharmacological blockade of NMDA receptors. In addition, when protein synthesis is inhibited, blocking the expression of late-phase LTP, place fields show short-term stability but long-term instability (Agnihotri et al., 2004). Interestingly, when LTP-deficiency is localized to CA3 or CA1, CA1 place fields remain stable across sessions (McHugh et al., 1996; Nakazawa et al., 2002).

4.1.2 *Sensory-vestibular conflict I: Cue rotations and the head direction system*

In contrast to the substantial environmental changes used to promote complete remapping, a variety of more subtle experimental manipulations have been performed. One class of these manipulations involves environments in which the directional cues (i.e. those that indicate in which direction "north" is) are rotated, repositioned, or removed.

Rotated cue card

One of the first observations about cue control of place fields was that, when the external landmarks are rotated, so do the animal's place fields

(O'Keefe and Conway, 1978). Muller and Kubie (1987) trained rats to forage for food in a gray, cylindrical arena with a white cue card attached to one side. When the cue card was rotated 90°, place fields rotated by the same amount. A similar result has been found for head direction cells in the postsubiculum and thalamus Goodridge and Taube (1995). In addition, when the orienting cue was removed, place fields were observed to rotate by a random angle, though they remained otherwise intact in most cases.

Disorientation

A more direct link between place cells and the head direction system was observed by Knierim et al. (1995), in a curious experiment involving animal disorientation. Knierim et al. (1995) studied rats who were either disoriented or not disoriented before entering a cylindrical environment with an orienting cue card affixed to the arena wall. They recorded simultaneously from place cells and at least one thalamic head direction cell. For rats who received disorientation before each session, eventually (by the seventh recording session), place cells and head direction cells no longer aligned their fields with the cue card. While most place cells appeared to have rotated fields, other cells (especially in one rat) remapped. Nonetheless, of simultaneously recorded cells that showed rotated fields, the amount of rotation was approximately equal, as was the rotation of any head direction cells also recorded. Thus, even when a cue loses control over place cells and head direction cells, the two spatial representations maintain coherent firing fields with respect to each other.

Cues for estimating compass bearing

In order to explore the types of cues most strongly used to orient the rat in space, Cressant et al. (1997) trained rats to randomly forage in a cylinder that contained two or three objects. For the group of rats for which the objects were near the center of the arena, rotation of the objects rarely exerted control over the angular location of the place fields. By contrast, for the group of rats for which the objects were near the edge of the arena, either apart or together, almost perfect cue control was observed. They also observed that a cue card, added to the objects-in-center group, acquired control of the place fields after a week of exposure without rotations; however, subsequent rotation of the objects in the absence of the cue card still did not result in place field changes. Together, these results suggest that, to shift the angular location of place fields, the cue must be as distal as possible.

What is particularly interesting about the Cressant et al. (1997) study is that distal cues are the best possible cues for determining one's bearing, but the worst possible cues for determining one's position. Thus, the impact of a cue rotation on the place code is likely due to its impact on the head direction system, and not vice versa, since it is the head direction system which would be expected to orient based on distal cues.

Explicit visual-vestibular conflict

To further explore the relationship between the head direction system and the hippocampal place code, Knierim et al. (1998) created an explicit conflict between vestibular and visual information in order to examine how the conflict was resolved in each system. Rats were placed in a cylindrical arena with a white cue card, and the entire arena, including the rat, was rotated either slowly or quickly by either a small amount (45 degrees) or a large amount (180 degrees). In some rats, recording electrodes were in the hippocampus; in others, they were in the anterior thalamus. For slow or small rotations, head direction tuning curves and place fields rotated with the cue card. For large, fast rotations, head direction cells sometimes maintained their orientation in the room frame or otherwise altered their tuning curves relative to the position of the cue card, suggesting that the system accounted for the vestibular sense of rotation in preference to the location of the cue card. In a similar percentage of cases, place fields remapped after the large, fast rotation.

Interpreting the exact cause of the remapping is difficult, as place cells and head direction cells were not simultaneously recorded. One explanation is that visual information is encoded in an allocentric representation, i.e. the cue's bearing is encoded as "north," where the direction of north is defined by the head direction system. After the rotation, the rotated cue is now represented as a different card located "south." This difference in afferent input to the hippocampus leads to the observed remapping. An alternate explanation is that the fast rotation is (incorrectly) perceived by the rat to include a translational component. Such a possibility is reasonable considering that even the rotational component was not integrated particularly accurately, as judged by the tuning curves of head direction cells after the rotation. This perceived translation would then alter the path integration representation in dMEC, resulting in remapping in CA1. By either explanation, internal integration of perceived movement plays a critical role in determining the hippocampal place code. Consistent with this, disabling the vestibular system disrupts both postsubicular head direction cells and CA1 place cells (Stackman et al., 2002).

While the Knierim et al. (1998) study created a perceived vestibular sense of motion in an environment that was visually unchanged, Jeffery and O'Keefe (1999) explored the effect of witnessing the cue being moved in the absence of perceived self-motion (see Figure 4.3). They ran five sessions, each containing two parts. During the first part (card-only rotations), Covered rats were either covered while the cue card was rotated (n=2) or this part was omitted (n=2); in either case, rats never explicitly witnessed the card being moved. During the second part (card+rat rotations), *all rats were covered*, and the rats and the

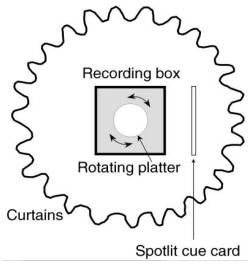

FIGURE 4.3: The recording apparatus used by Jeffery and O'Keefe (1999). The rat could be rotated on the center platter in addition to the cue card being moved. (Reprinted from Jeffery and O'Keefe, 1999.)

card were rotated (slowly) by different amounts (either 90º or 180º). A third group, Briefly Covered rats, received the same training as the Covered Rats, except all were covered for only 30s (instead of 2+min) during the card-only rotations. They found two main effects.

First, during the card-only rotations, the place fields of Covered rats always rotated with the card, whereas the place fields of the Uncovered rats rarely did, sometimes remaining stable in the room frame, sometimes remapping. This effect was mirrored in the card+rat rotations, in which the place fields of Covered rats always realigned to the card, whereas those of the Uncovered rats progressively decreased the rate of re-alignment (81% in session 1, 33% in session 5). Interestingly, though the influence of the card in the covered card+rat rotations progressively declined for Uncovered rats, the cue showed control of the place code more often in the card+rat rotations than the card-only rotations, suggesting that, though the card was being ignored when visibly moved, it still retained some control over the place code when invisibly moved.

Second, they found that the longer the duration of covering during the card-only trials, the more impact the card had on the place code. This suggests that rats maintain both a head direction estimate and a sense of precision (or confidence) in that estimate. When covered, the estimate will tend to drift over time, and the perception of a dearth of viable sensory information to maintain an accurate estimate may contribute to a lower perceived precision in that estimate. Consequently, covering the rat for longer periods may shift the balance of reliance on idiothetic and sensory information, biasing the system to re-orient when sensory information again becomes available.

Cue rotations and task performance

Does the response of place fields to a rotated cue affect performance of a spatial task? Lenck-Santini et al. (2002) trained rats on three versions of a place preference task in which they had to navigate to and remain briefly

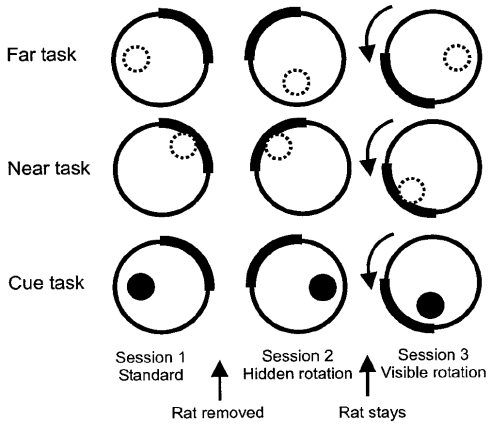

FIGURE 4.4: The three tasks used by Lenck-Santini et al. (2002). The rotation disrupted performance in the Far task, but not the other two. (Reprinted from Lenck-Santini et al., 2002.)

within a small area in order to receive a food reward, which was randomly scattered into the cylindrical arena (see Figure 4.4). In all three versions, three sessions were run in which a cue card was rotated 90 degrees between sessions. Rats were removed from the arena between sessions 1 and 2, but not between 2 and 3, resulting in one hidden rotation and one visible rotation. In the Cue task, a visible floor cue demarcated the reward area, which was unrelated to the location of the cue card. In the Near and Far tasks, the reward area was unmarked and was located either very close to or on the opposite side of the arena from the cue card, respectively.

In general, fields almost always rotated subsequent to the hidden rotation, but rotated less frequently with the visible rotation. As observed by Jeffery and O'Keefe (1999), fields were progressively less likely to rotate with the card the more often it was observed to move. Interestingly, fields were more likely to rotate with the cue during the hidden rotation if the cue was predictive of reward area. In the Far task, performance was substantially disrupted when fields did not rotate with the card. In the Near task, performance was only mildly disrupted, though rats did make more entries into the field-relative reward area than chance, suggesting a minimal hippocampal component to behavior. Thus, place fields are predictive of behavior when the rat must use a locale navigation strategy, which involves determining one's location in space based on a "cognitive map" incorporating multiple landmarks (Tolman, 1948; O'Keefe and Nadel, 1978), instead of a taxon (cue-approach) strategy.

Local cues, distal cues and place cells

While visual and/or vestibular rotations can cause conflicting estimates of head direction and disrupt both head direction and hippocampal place

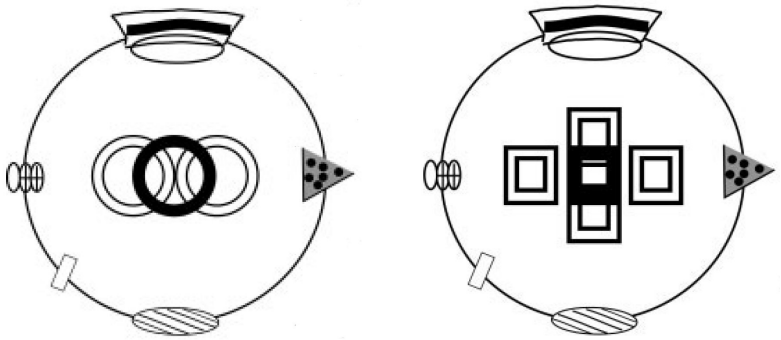

FIGURE 4.5: Translational shifts of the track performed by Knierim and Rao (2003). The track was surrounded by curtains with distal objects affixed to or near them. (Reprinted from Knierim and Rao (2003).)

cells, Cressant et al. (1997) noted that cues in the center of an area, those that would serve as poor reference points for head direction, rarely influenced place cells when rotated. To more explicitly explore the roles of proximal and distal landmarks on place cells, Knierim and Rao (2003) trained rats to forage for food on circular or rectangular tracks (with no surface cues). They translated the tracks north, south, east, and west in the room (see Figure 4.5), but found that, with the exception of one rat on the rectangular maze, rats showed little to no evidence of remapping during the translations. The exception rat showed substantial remapping on the rectangular maze, but none on a circular maze a few days later. However, when the distal landmarks were rotated by 45^0 while rats were foraging on the circular track, place fields clearly rotated in accord, suggesting that rats did process the distal landmarks, but predominantly for head direction instead of position.

4.1.3 *Sensory-vestibular conflict II: Path integration*

While studies discussed in the previous section explored the relationship between the hippocampal place code and the head direction system, other studies have explored the relationship between the hippocampal place code and the internal path integration system.

Place field realignment

Gothard et al. (1996a) trained rats to depart from a holding box, travel down to the end of a linear track to receive a food reward, and then return to the holding box. Rats were initially given 6 to 7 training sessions with the holding box at a fixed distance from the end of the track. After initial training, test sessions were conducted in which, after the rat departed from the box, it was moved to a different location. There were five possible locations: the original location (P1), farthest from the end of the track, and four locations

closer to the end (P2–P5). They observed that, when the track length was shortened slightly (P2), the locations of place fields were shifted to compress them all within the shorter length. However, for even shorter track lengths, they observed a discontinuity. Place cells would initially fire as if the rat were leaving from box position P1. However, after covering part of the track length, place cell firing would abruptly realign with the rat's position relative to the track and room. During the return journey, place cells fired unaltered until just before entering the box, at which point they abruptly switched back to representing the animal's track position relative to the box.

The authors make two key observations. First, at any one time, all simultaneously recorded cells appeared to reflect the same position on the track, irrespective of whether that position was determined the rat's position relative to the box, the room, or both. Second, when departing from the box, the box was behind the animal and therefore unlikely to be providing a strong sensory influence, whereas, when returning to the box, the box was in front of the animal. By contrast, in shortened trials (P3-P5), rats showed longer maintenance of the box-relative place cell activity during the outgoing than the returning portion of the trip. This suggests that the primary influence on prolonging the box-relative place cell activity was an internal spatial representation updated by the animal's movements: a path integration system.

Gothard et al. (2001) further explored this paradigm in two ways. They recorded from both DG and CA1 cells to determine to what extent the areas differed, possibly suggesting a role for the hippocampus proper in path integration. However, firing patterns in the two areas were similar. They also recorded during trials in which the lights were extinguished, to explore how distal sensory cues influenced the place cell representation. As expected, room-aligned place fields were observed later in dark trials than in light trials, suggesting that visual cues hasten the transition to room-alignment.

To test the behavioral significance of box-room transitions, Rosenzweig et al. (2003) trained rats to find a room-defined location on a linear track, starting from, as before, a moving start box. They found an age-related deficit in locating the room-relative place and in transitioning from box-aligned to room-aligned place fields. Young adult rats generally transitioned to a room-relative reference frame faster than old rats, and they slowed down more in the goal position relative to a control position, indicating knowledge of the reward position (at which they had to sit and wait for medial forebrain bundle stimulation reward). These results confirm the notion that this experimental paradigm contrasts learning-dependent sensory and learning-independent path integration mechanisms for localization: sensory information has a

weaker impact on the hippocampal place code in old rats, presumably due to the age-related deficiency in LTP.

Two identical boxes

In another experiment designed to study hippocampal activity when path integration and sensory information are in conflict, Skaggs and McNaughton (1998) trained rats to forage for food in two identical boxes, each with an opening on the east side and connected by a corridor. While the Gothard et al. studies suggested that sensory information should eventually overcome (and presumably reset) the path integration system, Skaggs and McNaughton observed something entirely unexpected: a partial remapping between the boxes. While some place cells showed identical fields in both boxes, others cells clearly differentiated the boxes. Different rats showed different degrees of partial remapping, and, for a particular rat, the degree of partial remapping varied between days. Nonetheless, there was no systematic trend of more or less remapping over the course of the experiment.

Subsequent work by Fuhs et al. (2005), which represents a portion of this thesis and is discussed in detail in Chapter 8, did not confirm these findings. In the part of our experiment that replicated the Skaggs and McNaughton experiment, rats did not remap at all between the boxes. The significance of this difference and some preliminary data suggesting its underlying cause is discussed in Chapter 8.

4.1.4 *Local cues and local place field changes*

While cue rotations appear to have their most direct influence on the head direction system, other studies have addressed environment manipulations that do not involve rotations, thereby factoring out any effect of the head direction system. Gothard et al. (1996b) trained rats to depart from a holding box, retrieve a food reward at a goal location defined by its proximity to two local cues, and then return to the holding box. The goal location (and the cues that demarcate it) were moved before each trial began, and the holding box was moved between the rat's departure and its return. They found that, while some place cells represented position in the room reference frame, other place cells showed coherent fields relative to the holding box location or the goal location. Interestingly, while room-relative place fields were distributed over the entire arena, box or goal fields were observed only in proximity to the box or goal locations. They note that the data are consistent with (though not demonstrably indicative of) the rats switching between three contexts, one each for the room, the box and the goal.

Barriers

An earlier experiment by Muller and Kubie (1987) considered a simpler task in which a barrier was added at varying locations within a cylindrical arena. They found that place fields observed where there was previously no

barrier were largely suppressed when the added barrier encroached upon the field, and new place fields were observed where the barrier was added. Fields more distant from the added barrier were not significantly different, a locality effect that has also been observed in a cue deletion experiment (Hetherington and Shapiro, 1997). The opacity of the barrier did not impact Muller and Kubie's results: a clear plastic barrier yielded similar results. However, when just the weighted base of the barrier was instead used, the impact on place cell activity was minimal, suggesting that the place field changes were due to the obstruction of the barrier rather than its visual impact.

In a subsequent study, Rivard et al. (2004) trained rats in an environment with a barrier in a fixed position. During subsequent sessions in which the barrier was rotated or translated, cells showed heterogeneous responses. Some fields were observed to follow the barrier, some fields were suppressed when the barrier was moved onto them, some fields showed only a rate change, and some, even those juxtaposed with the barrier, did not react to the barrier movement at all. Thus, it was clear that the arena and the barrier were represented concurrently instead of as separate contexts ("near the barrier" "elsewhere in the arena") active at different times, as observed by Gothard et al. (1996b). Interestingly, when introduced to a second environment that also contained the barrier, some cells that were active in proximity to the barrier in the first environment showed similar fields in the second environment. Thus, hippocampal place cells represented both the arena, the barrier, and, in some cases, the conjunction of the two.

Object replacement

If the presence or absence of an object induces place field changes, what effect is caused by the substitution of one object with another. Lenck-Santini et al. (2005) recorded from rats while they explored a cylindrical environment containing two objects. When the objects were rotated 90° about the center of the arena, place fields near the objects showed a complex pattern of change, while fields far from the objects showed only a mild decrement in firing rate. Fields near the objects did not rotate with the objects, per se, frequently shutting off or possibly "following" one of the objects. By contrast, when the familiar object was switched with a new object (without rotation), no hippocampal unit activity differences were observed. This is consistent with the earliest barrier experiments in which the sensory qualities of the object seemed of little importance compared to their impact on navigation (Muller and Kubie, 1987), as well as a lesion study implicating the hippocampus in detecting changes to an object's position but not to its identity (Lee et al., 2005).

Annular water maze

Much as the insertion of an object evokes local changes to the hippocampal

place code, the introduction of some behavioral significance to a particular location can also impact place cells. A pair of studies involved training rats to find a hidden platform in a ring-shaped pool (or annular water maze) (Hollup et al., 2001b; Fyhn et al., 2002). Hollup et al. (2001b) found that, when systematically trained to find the platform in one location, more than twice the place fields were found at the platform location as in any other segment of the maze. When rats were trained with randomly distributed platform locations, no such accumulation of place fields was observed. For the fixed-location group, there was also a slightly elevated number of place fields in the segment just before the platform, suggesting an encoding of expectation of the platform.

In a subsequent study, Fyhn et al. (2002) first trained rats to find the platform at a fixed location and then moved the platform to a different location. They made a number of observations:

- Roughly a third of the cells active at the platform location showed fields as soon as the rat climbed onto the platform at the new location, but with repeated training at the new location, this effect went away.
- Only a few fields elsewhere were changed due to the change in platform location.
- Of cells active at the new platform location whose recording was maintained through training, some fields (4/10) became inactive with repeated visits to the new location, others (2/10) remained active only on the platform, and others (4/10) became active during the swim phase before the platform was reached, perhaps indicating a predictive hippocampal representation.
- The effect was not observed when rats were trained with random platform locations, suggesting that the effect was not based purely on novel platform / location associations.
- The effect was not observed when a task-irrelevant stimulus was used (a water jet at the bottom of the tank).
- Interneurons were less active immediately after the platform change, but return to baseline during subsequent training. This is consistent with Wilson and McNaughton (1993), and may suggest an increase in LTP.
- They suggest that the change in activity may reflect the exploration of the platform, not simply the mismatch itself, because the decay of

platform-specific fields at the new platform position was consistent with the time course of exploration of novel objects.

Taken together, these results clearly demonstrate that significant locations are preferentially represented, even in the absence of a local cue at the site, and that, similar to the barrier experiments, place cells will, at least for some period of time, flexibly relocate when the goal location is moved.

4.1.5 *Delayed and gradual remapping*

Delayed abrupt remapping

Bostock et al. (1991) recorded from CA1 and CA3 in 12 rats while they randomly foraged for food in a cylindrical arena painted grey. The northern quadrant of the cylinder was initially covered by a white cue card. After one to four weeks of training in the white card environment, the white card was replaced in some sessions by a black card. Ten of thirteen rats did not remap during the first black card session. However, during subsequent sessions recorded from seven rats, six of whom initially did not remap, four rats were observed to remap after repeated exposure to both the white and black card environments. In cases in which two neurons were simultaneously recorded, they always both either remapped or remained unchanged, suggesting that, when it occurred, the remapping was population-wide.

Double rotation experiments

In an experimental manipulation described as a "double rotation", rats run on a track (plus maze or circular maze) with multiple local texture and scent cues; these track cues are referred to as the "local cues." Large visual distal cues are mounted to curtains surrounding and at some distance from the track. The original experiments were performed on a plus maze with 180° double rotations: one set of cues was rotated 90° clockwise, while the other set of cues was rotated 90° counterclockwise (Tanila et al., 1997; Shapiro et al., 1997). These double rotations produced discordant responses among place cells. Some fields rotated consistently with the distal cues, others rotated with the local cues, others did not rotate at all, and others remapped. Subsequent parallel recordings from tens of cells by Knierim (2002) confirm this finding, as well as extended it to double rotations of lesser magnitude.

There was also an experience-dependent effect on remapping. Shapiro et al. (1997) reported that, while the majority of cells rotated with the distal cues initially, repeated exposure to the double rotation condition (as well as less regular cue scrambling and cue deletion conditions) resulted in progressively fewer cells rotating with the distal cues and progressively more cells remapping. Brown and Skaggs (2002) attempted to reproduce these findings, but observed remapping in only two of four rats. Interestingly, neither of those two rats remapped during the first session, but both did by

day 8. In one rat, the remapping was abrupt, starting on day 8. In the other rat, lost data prevents knowing the exact time course of remapping. In both cases, remapping was only partial, even after at least 11 days of training.

Lee et al. (2004b) trained rats in two double rotation and three standard sessions per day for eight days. One of four double rotation angles was used (45°, 90°, 135°, and 180°). They observed no increase in remapping over the course of their experiment (Lee & Knierim, personal communication). However, it should be noted that, with only four exposures to each double rotation angle, it is possible that the experiences were insufficient to trigger increased remapping.

Gradual remapping

In a vivid demonstration of gradual remapping, Lever et al. (2002) studied rats alternately exposed to cylindrical and square environments, for 4-6 trials per day. In contrast to previous experiments such as Muller and Kubie (1987), rats were given no prior experience in either environment. They found that there was little remapping initially; however, each successive day of exposure to both environments yielded a higher percentage remapping. There were two particularly interesting details in their data. First, the remapping was gradual; for several days in each rat, a varying but increasing degree of partial remapping was observed. Second, they show the slow steady formation or destruction of individual fields during the course of a session and, in some cases, over several days, suggesting that a learning process is intimately tied to this remapping.

Taken together, the experiments in this section clearly demonstrate the powerful effect previous experience can have on the hippocampal representation.

4.1.6 *Hysteresis, pattern separation and pattern completion*

Hysteresis

One of the first demonstrations of hysteresis in the hippocampal place code was the study by Quirk et al. (1990), who recorded from place cells in light and dark conditions. When the lights were extinguished subsequent to the animals' entry into the arena, 24 of 28 cells maintained their firing fields. However, when the lights were extinguished prior to the animals' entry, 14 of 22 cells remapped in comparison to previous lights-on sessions that day, and, in the majority of cases, these remapped fields persisted when the lights were subsequently turned on. Interestingly, they found cases in which, among two simultaneously recorded cells, one remapped in the entry-in-darkness condition and one did not. Thus, the firing of some place cells in the lights-on condition depended upon whether the animal entered in the light or darkness, while for others it did not.

Pattern separation and completion in CA3

The vast majority of hippocampal neurophysiology studies involve record-

ing from CA1. While lowering recording electrodes from the top of the skull, CA1 is the first layer of hippocampal cells reached, and they are generally reported to be the most likely to be active in any given environment (e.g. Vazdarjanova and Guzowski, 2004; Leutgeb et al., 2005b). However, in three recent studies, the activity of cells in CA1 and CA3 were simultaneously studied to see how they reacted to changes in the environment.

In the first study, Leutgeb et al. (2004) recorded in arenas in two rooms. They found that, when the arena shapes in the two rooms were the same, CA1 remapping was only partial, while CA3 remapping was complete. When the arena shapes in the two rooms were made progressively more different, CA1 cells responded with increased remapping. They suggest that the room difference alone is sufficient for CA3 to form a completely distinct representation of the two rooms, whereas CA1 is more reflective of the current sensory experience. This claim is bolstered by their showing that CA1 fields are relatively stable in novel environments, whereas CA3 fields change and develop over the course of the first 30 min of exposure.

In the second study, Lee et al. (2004b) found essentially the opposite results. They performed a double rotation experiment (see Section 4.1.5) using four double rotation angles (45º, 90º, 135º, and 180º). In confirmation of previous double rotation studies, some CA1 fields rotated with distal cues, other rotated with local cues, and others remapped. By contrast, CA3 showed a more homogeneous response. Three times as many CA3 fields rotated with local cues, fewer fields rotated with distal cues, and there was less remapping and fewer "ambiguous fields."

In a related analysis, Lee et al. (2004a) examined how place fields evolve over the course of an experimental session. On a linear track, place fields in CA1 are known to slide backward and develop a negative skew relative to the direction of travel (Mehta et al., 1997, 2000). Lee et al. (2004a) found that place fields also moved backward in CA3, but only during the first few exposures to each double rotation angle. On days 3 and 4, CA3 fields did not move, and the fields' skewness was already judged to be high at the beginning of each session. By contrast, place fields in CA1 did not shift on day 1, but did on subsequent days. These results are in agreement with Leutgeb et al. (2004).

In the third study, Vazdarjanova and Guzowski (2004) used an *in situ* hybridization technique that binds immediate-early gene (IEG) mRNA to fluorescent markers within the cell. Riboprobes for two IEGs with different temporal expression profiles were used. The first, *Arc*, is transcribed about 2-15 minutes after cell activity, while the second, *Homer 1a*, is transcribed about 25-40 minutes after activity. When exposed to two different environments,

one in each time window, the technique can therefore distinguish between cell activity in each environment. They exposed rats to a standard environment (A) containing distal room cues, arena walls, and objects within the arena and one of four variations: different within-arena objects (A_{obj}), different within-arena object locations (A_{conf}), a different room with different distal room cues (A_B), and a different room with a different arena and objects (B_B). They found that, for the smaller variations (A_{obj}, A_{conf}, and A_B), the representation was more similar to the environment A representation in CA3 than CA1. For the largest variation (B_B), the representation was more distinct in CA3 than CA1.

The Vazdarjanova and Guzowski (2004) study essentially unified the results of Leutgeb et al. (2004) and Lee et al. (2004b). While CA1 shows an immediate representation of an environment that is more directly representative of the sensorium, CA3 constructs representations which attempt to more clearly distinguish different environments, while representing similar environments with greater similarity (Guzowski et al., 2004). This is known in the associative memory literature as pattern separation and pattern completion, respectively, and the significance of these findings is discussed in Chapter 7.

To explore the pattern separation effect in more detail, Leutgeb et al. (2005c) trained rats to forage in arenas that varied either in shape (square and cylinder) or color (black and white) in the same room. They found that, while cells change rate, often to the degree that activity in one arena is largely absent, spikes are nonetheless localized to the same area within both arenas, i.e. the place fields don't move. They coined the term "rate remapping" to describe this. By contrast, when recording in two different rooms, place field correlation coefficients (which are firing rate independent), are much lower, a phenomenon they call "global remapping." They confirm the results of Leutgeb et al. (2004) that global remapping in CA3 is much more pronounced than in CA1. Nonetheless, in both regions of the hippocampus, the observed global remapping is caused by place field shift in addition to rate change.

Morph experiments

As many place field experiments have involved training in square and cylindrical arenas, Wills et al. (2005) explored how rats would represent a series of arenas whose shapes were morphs between a square and cylinder. During the first three days of training, rats were trained in two environments that differed in shape (square vs. cylinder), color (brown vs. white), and texture (masking tape vs. white-painted wood). The square arena was actually a "morph box," whose walls were composed of vertical slats that allowed the box to be morphed into various shapes. A white painted-wood cylinder was used for the other arena. Rats ran three sessions per

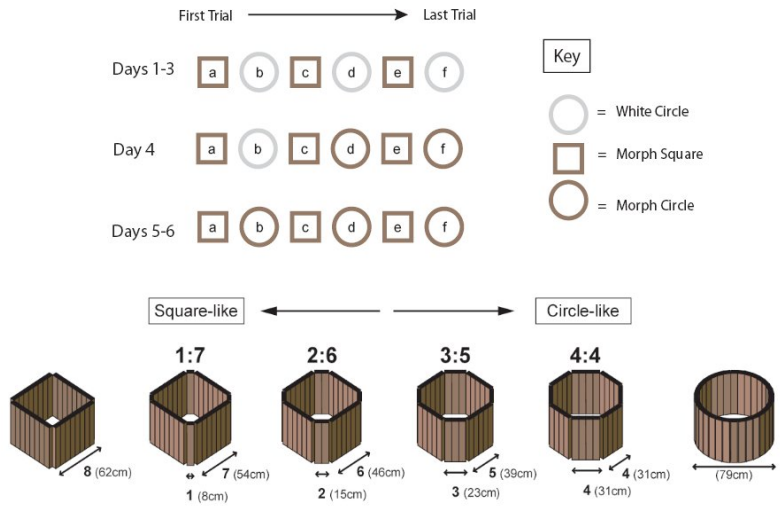

FIGURE 4.6: (Top) The six days of pretraining used to cause rapid remapping between the square and cylindrical arenas. (Bottom) The six arena shapes used in the morph training. (Adapted and reprinted from Wills et al., 2005).

day alternating between the two environments and completely remapped between them. In the fourth and sixth sessions of the fourth day and on days 5-6, the white cylinder was replaced with the morph box, now reshaped into a cylinder; complete remapping between arena shapes persisted (see Figure 4.6, top). They then trained rats in a randomized sequence of six arena shapes (see Figure 4.6, bottom): square, morph 1:7, morph 2:6, morph 3:5, morph 4:4 (octagon), and cylinder. During the first pass, CA1 representations in the morph environments in all rats were highly similar to the square map for 1:7 and 2:6 and to the circle map for 3:5 and 4:4. During the second pass, two of the four rats switched maps earlier (rat 2 at morph 1:7; rat 3 at 2:6). This pattern completion is initially present, and the hippocampal representations of the morph environments become even more similar to the square or cylindrical representations over the course of the first two minutes of recording.

Leutgeb et al. (2005b) performed a very similar experiment that yielded the opposite results: a gradual transition in representation between the two morphs, both in CA1 and in CA3. Leutgeb et al. (2005b) first pretrained rats for 16-19 days to forage for food in square and cylindrical arenas in the same room (see Figure 4.7). Unlike Wills et al. (2005), the two arenas were not initially distinguished by color and material, so fields presumably gradually remapped over the course of pretraining. The rats were pretrained for eight sessions per day, 10 min each, in a random sequence of the two arenas. Then,

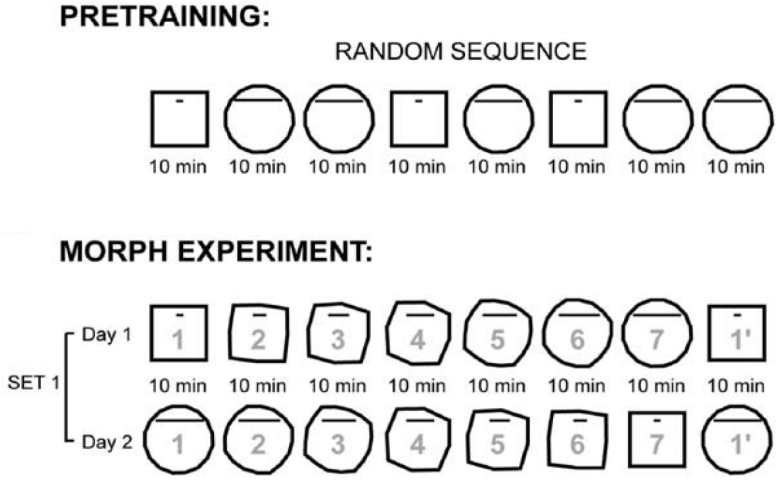

FIGURE 4.7: The pretraining and morph experiment sequences used by Leutgeb et al. (2005b). Unlike Wills et al. (2005), their pretraining sequence involved only the two arenas later used during the morph experiment proper. The horizontal lines indicate orienting cue cards whose widths vary in different arena shapes. (Reprinted from Leutgeb et al., 2005b).

rats foraged for food in sequences of "morph" arenas. Again, eight sessions were run. In sessions 1 and 8, either the square or cylindrical arena was used, and, in session 7, the other familiar arena was used. Sessions 2-6 were arenas whose shape was morphed from the pretrained shape in session 1 to the pretrained shape in session 7.

They found two interesting results, both involving hysteresis. First, the CA3 place code during morph session 7 is more similar to the place code from session 1 than what was observed during the non-morphed pretraining sessions, even though the arenas in sessions 1 and 7 are the pretrained arenas. This effect was more pronounced during the first morph set, though still statistically distinguishable during later morph sets. CA1 showed no such effect. They argue that the session-by-session similarity of the morph sets facilitates a CA3 representation of each arena that is somewhat more similar to the previous arena. When the arena is switched between the pretrained arenas for the last session, CA3 shows no hysteresis.

Why do the results of Wills et al. (2005) and Leutgeb et al. (2005b) differ? Preliminary evidence from Leutgeb et al. (2006b) suggests that rats in the Wills et al. (2005) study adopted a different sets of path integrator coordinates for each environment, whereas rats in the Leutgeb et al. (2005b) used the same set of path integrator coordinates in both environments. Leutgeb et al. (2006b) trained two groups of rats: group 1 was trained in cylindrical and

square arenas in the same location, while group 2 was trained in cylindrical and square arenas in two separate locations connected by a corridor. Both groups were then tested on the same sequence of morph arenas. Group 1 reproduced the earlier findings of Leutgeb et al. (2005b). Group 2, who had been introduced to the two arenas in two different room locations, showed the sharp changes in hippocampal representation typical of the Wills et al. (2005) findings.

In addition to the morph results, Leutgeb et al. (2005b) report a second major finding. Unlike CA3, the CA1 representations during the first session of each sequence were different from those of the last session, even when the training sequence consisted only of a series of cylindrical and square environments. Unfortunately, they do not explore the nature of the change, i.e., whether the two representations become more similar over time or more different, so it is difficult to speculate on what computational process might underpin this change.

4.1.7 *Replay*

When two CA1 cells have overlapping place fields in an environment, there will naturally be a correlation between their activity patterns. Wilson and McNaughton (1994) discovered that the same correlations are observed subsequent to this awake exploration during sleep. By contrast, in a sleep session that preceded the exploration, no such correlations were observed, suggesting that the correlated activity resulted in learning which led to repetition of these activity patterns. Shen et al. (1998) extended this finding to DG.

Skaggs and McNaughton (1996) examined the temporal relationship of pairs of cells on a 1D track, triangular or square in shape. As expected, cells with partially overlapping place fields had temporally asymmetric cross-correlations: when one cell's place field preceded another on the track, the temporal cross-correlation reflected this bias. During sleep sessions after running on the track but not before, the temporal cross-correlation was also observed to be both non-zero and asymmetric. Thus, they concluded that replay during sleep preserves some temporal structure of the awake experience.

A subsequent pair of studies demonstrated that this reactivation shows a precise temporal structure, suggesting that entire sequences are replayed during sleep. Louie and Wilson (2001) examined rapid eye movement (REM) sleep, which shows strong theta frequency (4-12Hz) modulation, much like when the rat is active. They found that windows of activity during REM sleep matched well with sequences of activity recorded previously during awake

behavior, and these sequences were replayed during sleep at a similar time scale to their occurrence during behavior. Lee and Wilson (2002) examined slow wave sleep (SWS) and found that such sequences were replayed during SWS on a much shorter time scale (~200 ms vs. ~6 s).

When the rat is not active, EEG power in the theta band is low, and the EEG signal (and corresponding brain state) is described as low-amplitude irregular activity (LIA). Foster and Wilson (2006) trained rats to shuttle back and forth on a linear track, receiving a food reward at each end. Rats frequently entered LIA during consumption of the reward, and Foster and Wilson discovered that the sequence of place fields on the track was replayed *backwards* during EEG ripple events, which occurred several times during each period of LIA. This result has several profound implications. First, since replay was observed after the very first traversal of a novel track, sequences of activity can be encoded after a single "episode" of experiences. Second, the reverse direction of replay, especially in the context of receiving a food reward, suggests a biological solution to the credit assignment problem: replaying recent events permits the discovery of how the reward was achieved.

4.2 Place Cells and Multiple Behavioral Goals

While Section 4.1 discussed the behavior of hippocampal principal cells while the animal performed a single task, often random foraging, this section addresses neurophysiology studies involving multiple tasks or tasks involving multiple stages. In general, cells respond conjunctively to the animal's place and to a variety of variables associated with the details of the task and / or which task is being performed.

4.2.1 *Multiple tasks or task stages*

Delayed (non-)match to sample

Hampson et al. (1993) trained rats to perform a spatial delayed-match-to-sample (DMS) task. The task consists of three phases. In the sample phase, one of two levers in an operant chamber was extended, and the rat had to press the lever. In the delay phase, the rat had to return to a nose-poke port at the opposite side of the operant chamber and remain there repeatedly nose-poking the port throughout the delay. In the choice phase, both levers were extended, and the rat is rewarded for choosing the same lever as in the sample phase. (In a delayed-non-match-to-sample task (DNMS), the rat would be rewarded for choosing the other lever.)

Hampson et al. (1993) report that, as expected, CA3 and CA1 cells are sensitive to place; however, they found that cells conjunctively encoded stages of the task as well. They report several classes of cell behaviors.

Sample-match cells respond to both the sample and match lever presses. Match-only cells respond to the match lever press but not the sample lever press. Sample-match-response cells respond to both lever presses as well as the reward.

Similarly, Otto and Eichenbaum (1992) trained rats in an odor-guided continuous DNMS task, and report that cells often responded to the match or non-match between the current and previous trial. They did not observe persistent activity during the delay which would implicate the hippocampus directly in working memory. In a follow-up study, Wood et al. (1999) report that cells show correlations with many aspects of the task: the rat's position, the rat's approach to the odor cup, the odor being sniffed, and the trial type (match or non-match observed).

Multiple spatial tasks on a track

Markus et al. (1995) trained rats to either randomly forage or shuttle between four locations for food on an open field. They report that, when rats are performed the shuttle task, more place fields were directional than in the random foraging task. This directionality was reminiscent of what is observed on a linear track or multiple-arm maze, though fewer cells had directional fields in the open-field shuttle task than on an 8-arm radial maze. This suggests that the animal's navigational route (likely represented elsewhere in the brain) causes a directional representation similar to the representation of the environmental constraints of the radial maze.

Markus et al. (1995) also report that some place fields remap between the foraging and shuttle tasks, suggesting that the task itself results in two similar but distinct hippocampal representations. Oler and Markus (2000) explore this task-dependent remapping by training rats on a plus maze that could, with two additional pieces of track, be extended into a figure-8 maze (see Figure 4.8). Rats were trained on both a win-shift task on the plus maze and a figure-8 trajectory on the figure-8 maze. Critically, since the figure-8 maze was composed in part by the plus maze, place fields on the plus maze could be directly compared while rats performed each task. They again found a partial remapping between the tasks. Oler and Markus studied both middle-aged and old rats, and found no behavioral difference between them on the plus-maze, a working memory task; however, old rats showed less remapping between tasks, further reflecting an inability of hippocampal cells in old rats to encode salient behavioral aspects of an environment (Redish et al., 1998). Both studies recorded from both CA1 and DG/CA3 (they could not distinguish between these two areas), and found no differences between the two cell groups.

Remapping and task performance

The aforementioned studies recorded from rats already familiar with the task or tasks. To address how the hippocampal representation changes

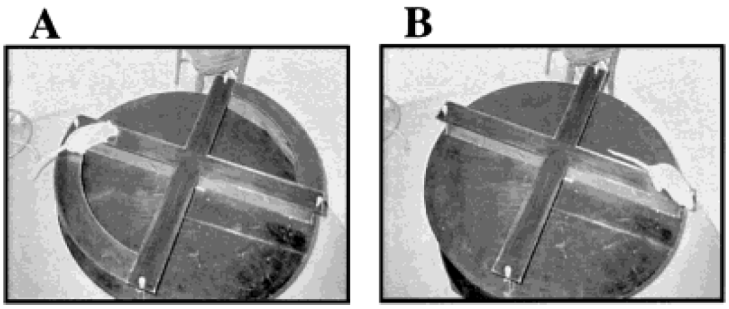

FIGURE 4.8: The two configurations of the apparatus used by Oler and Markus (2000): (A) the figure-8 configuration and (B) the plus maze configuration. (Reprinted from Oler and Markus, 2000.)

during training, Masters and Skaggs (2001) trained rats on a place preference task in a cylinder with white or black cue cards. Briefly pausing in an unmarked area triggered reward; the location of the reward area depended on the color of the cue card. Two of the four rats completely remapped upon first exposure to the black-card environment; one rat never remapped; one rat initially did not remap at all, but after many days exposure to both environments, substantially remapped (at least 2/3 of the cells) between environments. The rats' remapping between the environments seemed to be both necessary and sufficient for the animal to learn the task; rats that didn't remap were unable to acquire the task and rats that did remap were able to acquire the task. Incredibly, the one rat that did not remap until many days into training was unable to learn the task prior to remapping, but improved considerably during the same session as the first remapping. It is difficult to draw causal inferences from this study, since it is unclear whether other brain areas ostensibly involved in learning the task (e.g. striatum and pre- and infralimbic cortices) prompted the remapping or were simply able to learn the task subsequent to it. Nonetheless, the experiment demonstrates the intimate association between multiple contextual representations (maps) in the hippocampus and multiple behavioral goals.

Attentional effects

While these studies have explored the conjunctive encoding of the task or task variables with the first-order place representation, other studies have explored the impact of the task on the quality of the place representation itself. Zinyuk et al. (2000) trained rats either on a place preference task (defined in relation to the room) or to randomly forage. In their version, the platform on which each rat was trained was either stationary or continuously rotating. In both groups, fields were found that were stable in the room reference frame, the rotating reference frame, or the combination (conjunctive encoding); other

fields were disrupted, showing no clear field in either frame. Compared to foragers, navigators showed more room frame-stable or conjunctive cells and fewer disrupted cells; cells tied to the rotating frame only were rare in either group. Thus, conflicting reference frames tended to disrupt the spatial locality of place fields, but making the animal's location behaviorally significant improved hippocampal spatial coding for both reference frames.

Kentros et al. (2004) trained mice in three tasks of increasing spatial demand: a null task (do nothing), a random foraging task, and a place preference task. (Unlike rats, mice are spontaneously active, even without an explicit motivation, so place fields can still be mapped out; rats would just doze off.) They report several observations:

- Place cell stability across days increased with task demand and novelty. (The instability in mice during random foraging is not typical of rats.)
- Good performers of the place preference task had much more stable fields across days than bad performers.
- Place cells in place preference performers and non-performers were of equal size and information content, but performers' fields were more coherent (smoothly changing) and fired more strongly.
- Prior to place preference task training, mice did not preferentially explore a novel object vs. a familiar one. After training, performing mice showed substantially more exploration of the novel object than the familiar, whereas non-performers showed equal exploration of both.
- A dopamine D1/D5 receptor agonist increases stability in "no task" mice (with little baseline field stability) and a dopamine D1/D5 receptor antagonist decreases stability in "foraging" mice (with moderate baseline field stability).

Thus, the stability of place fields was reflective of a mouse's ability to perform a place preference task. They attribute the stability to an attentional mechanism, showing that the increased attention to the environment facilitates novel object recognition, and they argue that this attentional mechanism is mediated in the hippocampus by dopamine, showing that place field stability can be artificially manipulated using a dopamine agonist or antagonist.

4.2.2 *Spatial alternation*

If place cells conjunctively encode the task being performed or different stages of the DNMS task, do place cells also differentially encode travel along

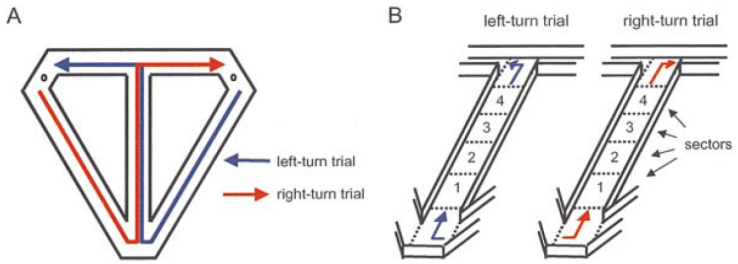

FIGURE 4.9: A) The figure-8 maze used by Wood et al. (2000). B) To ensure that they measured only place cell activity indicating future or past behavior, they restricted analysis to activity in sectors 2 and 3 on the center arm to exclude any differences due to the rat's current position. (Reprinted from Wood et al., 2000)

different but overlapping routes? Wood et al. (2000) explored this question by training rats on a continuous T-maze alternation task on a figure-8 maze, shown in Figure 4.9. They found that firing rate modulation of some place fields along the center arm reliably differentiated between left-turn and right-turn trials. Frank et al. (2000) performed a similar analysis on data recorded while animals traveled about a W-shaped maze. They observed both prospective coding (place field modulation dependent on the animal's future location) and retrospective coding (place field modulation dependent on the animal's past location).

Other studies have failed to find route-based modulation of place cell activity. Lenck-Santini et al. (2001) trained rats on a continuous alternation task on a Y-maze. One arm was designated the goal arm (G, where rewards were provided), while the other two arms (A and B) were to be visited in alternation, resulting in a G-A-G-B-G-A-G-B pattern of arm visits. No differential encoding of either the prior or next arm was found on the goal arm. Similarly, Hölscher et al. (2004) trained rats on a continuous T-maze alternation task on a figure-8 maze and found that only 4 of 45 cells showed any route dependence. Importantly, despite the lack of route-dependent encoding, rats in these studies were capable of performing the alternation.

In an elegant exploration of the discrepancies between these studies, Bower et al. (2005) trained rats to traverse routes through an open field composed of an overlapping sequence of straight-line route segments; at least one segment was used twice within the overall route (see Figure 4.10). When food was available at the end of each route segment and the rat's path remained unfettered during the course of training, no differential encoding of the sequence stage was observed, though rats were able to learn the sequence. However, when the training paradigm was adapted to include the use of

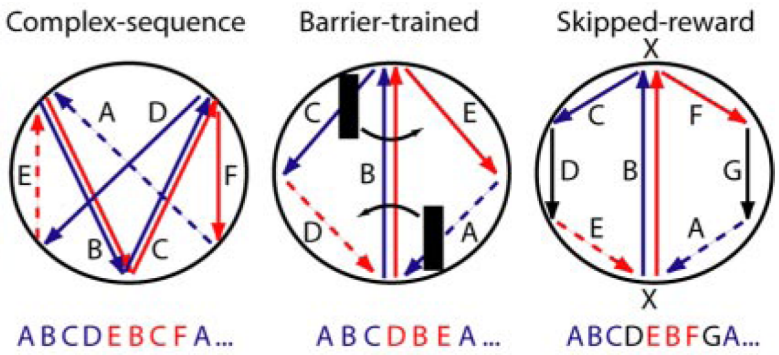

FIGURE 4.10: Three different training paradigms were used by Bower et al. (2005). In each case, they contained at least one overlapping segment. (Reprinted from Bower et al., 2005.)

barriers to guide the rat through the sequence during early training, or when reward at the beginning and end of the common segment was removed, differential encoding was observed. As the authors point out, these two alternative training paradigms were inspired by their original use in Wood et al. (2000) and suggest that differential encoding, where observed, is likely a conjunctively encoded state-dependent task representation projected from some afferent brain area.

None of these studies contained a delay component. Recently, a pair of studies by two different groups demonstrated the effect of adding a delay just before the center arm in a T-maze alternation task (Ainge et al., 2005; Robitsek et al., 2005). Both groups trained the rats in a way that evoked route-dependent encoding along the center arm, and they both observed that, when the delay was added, the route-dependent encoding was almost completely abolished. Nonetheless, rats continued to perform the alternation. Thus, the cause of the route-dependent encoding is dissociable from the performance of the task, even in rats who have previously developed such a conjunctive encoding.

4.2.3 *Serial reversal learning*

Instead of alternating between two behaviors on every trial, others have examined the hippocampal representation during alternation between blocks of one behavior and blocks of another. In the conditioning literature, this is referred to as serial reversal learning, reflecting the reversal of behavior at the start of each block of trials.

Ferbinteanu and Shapiro (2003) recorded from CA1 place cells while performing a periodically reversed spatial win-stay task on a plus maze. At

the beginning of each block, one of two possible goal arms (East or West) was chosen as the rewarded arm. Rats were started randomly on either the north or south arm, and learned to go to the end of the rewarded arm. When rats performed well, a new block was started and the task was reversed: the other goal arm was now rewarded instead. Thus, on either start arm, the rat could be headed to one of two goal locations (prospective), and, on each goal arm, the rat could have been coming from either start arm (retrospective).

While some cells were sensitive to place alone, others were conjointly sensitive to the start or goal arm of that trial. Over half of the cells coded either retrospectively or prospectively, with a somewhat higher percentage of journey-dependent cells coding for goal location. In contrast to the alternating T-maze studies, decreased prospective coding during some trials was significantly correlated with the rat choosing the incorrect goal arm. During such error trials, prospective coding decreased prominently while the decrease in retrospective coding was more modest.

Smith and Mizumori (2006) performed a very similar experiment, differing largely in two ways. First, the inactive goal arm could serve as a start arm, resulting in three possible start arms for each of the two goal arms. Second, they began trained rats on a randomly rewarded version of the task, recording from the hippocampus from the beginning of training. As with Ferbinteanu and Shapiro (2003), they report a complex pattern of activity. They show that these activity patterns are often block-specific and show several examples: a cell coding for one place and one start arm in east-goal-arm blocks, and a different place and start arm in west-goal-arm blocks, a cell with an event-related response in east-goal-arm blocks and a place field in west-goal-arm blocks, etc. During prior training in random reward sessions, these differential encodings did not develop, suggesting that it was the statistical structure of the reversal task, not the behaviors associated with traversing the arms, that engendered the development of block-specific representations.

4.3 Task-related Activity Outside the Hippocampus

The observation that hippocampal cells respond in relation to the animal's behavior, not just its sensory experience, suggests that other areas of the brain provide a task representation to the hippocampus. Here is a brief review of recent neurophysiology studies in brain areas implicated in task performance with afferent and or efferent connectivity to the hippocampal region.

Hippocampal studies reporting an accumulation of place fields at a goal

location are complemented by a study of neurons in the prelimbic / infralimbic (PL / IL) area of medial prefrontal cortex (mPFC). Hok et al. (2005) trained rats on a place preference task in a circular arena; the reward location was unmarked. Spending two seconds within the reward location resulted in the release of a food pellet, which scattered around the arena. Cells in PL / IL areas showed spatially localized firing fields, mostly either at the reward location, or the pellet landing location, which they presumably deduced from the sound of the pellet hitting the floor. Interestingly, the cells with landing location fields did not change when an orienting cue card was rotated. However, when the ceiling-mounted feeder was repositioned, the location of the landing field changed.

Jung et al. (1998) recorded from deep-layer mPFC neurons during three tasks: 8-arm radial maze win-shift task (don't visit the same arm twice), a figure-eight maze spatial alternation task, and random foraging. Within the sample of cells recorded, correlations to nearly any behavior they could imagine were found. Correlates on the radial maze included: direction of movement on arm, entering an arm, running, turning, between turning and reaching a new goal, firing during the intertrial interval, etc. Classical "place fields" were not typically observed. Correlates on the figure-eight maze included: arriving at goal, leaving goal, moving toward goal, returning to center arm, and, occasionally, center arm memory of the prior path. In a phenomenon similar to hippocampal remapping, coding of behaviors in mPFC showed no systematic similarities between the two tasks. Jung et al. (2000) found that neighboring neurons in mPFC that code for different task stages showed no evidence of coupling within a local neural network.

In nucleus accumbens and the ventromedial caudate nucleus, which both receive afferent input *from* the hippocampus, Mulder et al. (2004) report that some neurons code entire segments of travel in a plus maze, i.e. from one goal location (at the end of an arm) to the next, or from a goal location to the center of the maze. They argue that the spatial and temporal profiles of neural activity are consistent with the hypothesis that these areas serve as a bridge between the limbic and motor systems.

Chang et al. (2002) recorded from the medial prefrontal cortex (mPFC), nucleus accumbens and dorsal striatum while rats performed a delayed match-to-position task. They found that neurons in all three areas showed the greatest differences between the sample and match phases of the task, though a robust lever phase was also observed. Differential activity was also observed in all regions during various portions of the delay period. However, most of the cells in mPFC that showed differential activity during the delay period fired consistently throughout the duration of the delay period, unlike

the other areas; many other cells in mPFC showed a gradual ramping of activity through the duration of the delay period.

Baeg et al. (2003) recorded from multiple mPFC units while rats learned a serial alternation task in a figure-eight maze. Bayesian decoding of multiple units active on the center arm was performed to analyze cell ensemble responses during individual passes down the center arm. Since individual passes were analyzed, both correct (left-to-right and right-to-left) trials and error trials (left-to-left and right-to-right) could be analyzed, which made it possible to decouple retrospective and prospective coding on the figure-eight maze. The analysis revealed that the cell ensembles showed both retrospective and prospective coding. The decoding accuracy increased from near chance levels on day 1 to peak accuracy by day 8, as did behavioral task performance. Interestingly, during the first few days, cells showed retrospective coding at the earlier part of the center arm, but not the latter part, suggesting that further training allowed the mPFC network to extend the retrospective coding down the length of the center arm. Prospective coding along the center arm increased concomitantly with the retrospective coding in the latter portion of the center arm.

During error trials on days 5-8, cells showed accurate coding of the prior arm, suggesting that errors were not due to forgetting the prior arm. Activity during the early center-arm portion of error trials was similar to activity during correct trials with the same prior arm, suggesting good retrospective coding. However, all late center-arm activity during error trials was slightly different from correct trials, and behavior prediction was at chance. Some cells were continuously active along the entire center arm, while others were not. Behavioral prediction was still possible (though with some reduced accuracy) by examining cells with more spatially (temporally) localized activity, suggesting that mPFC may chain together units along the center arm to construct a temporally continuous representation of the task. It would be interesting to know whether adding a delay would disrupt such a representation.

5

THE HIPPOCAMPAL FORMATION: LESIONS, KNOCKOUTS AND INACTIVATION

While chapter 4 reviews the complex neurophysiological properties of cells in the hippocampus, this chapter considers the complementary question: what is the impact of these neurons on the animal's behavior? Data from three types of studies will be reviewed here:

Tissue lesions. These studies involve localized destruction of tissue, either by physical removal of the tissue (aspiration lesions), heat (radiofrequency lesions), or the use of a toxic agent to destroy just the neurons. All of these techniques have their drawbacks. Physical destruction of the neural tissue (aspiration, radiofrequency) can also destroy fibers of passage through the lesioned area that do not directly connect to the neurons therein. Toxic agents usually promote cell death by causing an influx of calcium into the neuron. This influx strongly stimulates the neurons which in turn powerfully stimulate downstream structures. This downstream stimulation may produce behavioral effects apart from the cell death in the targeted area.

Inactivation. These studies involve injecting receptor antagonists into brain tissue that either completely block neural activity or disable particular receptors. Particularly common are inactivation studies involving AP5. AP5 blocks NMDA receptors, which have been implicated in synaptic plasticity.

Genome knockouts. These studies involve strains of mice that are bred to have changes to specific genes. Often, these changes involve disabling genes that contribute to the formation of NMDA receptors.

Where possible, I will focus on studies limiting the manipulation to a single subregion (DG, CA3 or CA1). Throughout this chapter, the term "lesion" may be used more generally to refer to all three types of studies.

This chapter divides these studies into three parts. Section 5.1 considers the effects of hippocampal lesions on contextual conditioning studies, where conditioned associations are shown to be differentially expressed in different environments. Section 5.2 discusses reversal learning and the impairments in reversal learning observed subsequent to hippocampal lesions. Section 5.3 focuses on the role of the hippocampus in forming associations within a context, especially associations involving spatial location or associations across time.

5.1 Conditioning and Environmental Contexts

While the hippocampus is not required for forming basic associations (e.g., a tone with a shock), the environmental context in which associations are formed play a role in where the associations are later expressed. This section reviews studies on the context specificity of fear conditioning, the context specificity of the extinction of conditioned fear, and the context specificity of latent inhibition.

5.1.1 *Contextual fear conditioning*

In a classical fear conditioning experiment, a neutral conditioned stimulus (CS) such as a tone is paired with an aversive unconditioned stimulus (US) such as a shock. Animals learn to associate the CS with the US, resulting in a conditioned fear response (CR), which is typically measured by the amount of time during which the animal is motionless (freezing). One also finds that the background cues of the environment, often referred to as the context, can influence the perception of fear.

Conditioning and context

As discussed in Section 4.1.1, the hippocampus represents different environments with highly orthogonal activity patterns, an encoding that first appears within the hippocampus itself. It would therefore seem reasonable to believe that the hippocampus is involved in the role of context in fear conditioning. Indeed, Phillips and Le Doux (1992) found that, while both cued and contextual associations required an intact amygdala, fear of the context in the absence of the cue required the hippocampus. In a related study, Honey and Good (1993) found that the CS–US pairings show some context specificity: rats who learned the CS–US pairing in one context showed a weakened association (less freezing) when presented with the CS in a different context. Finally, Selden et al. (1991) found that hippocampal lesions impaired the rats' ability to choose a safe environment instead of the environment in which it was shocked; however, cued fear conditioning was spared.

Unfortunately, hippocampal lesions also cause hyperactivity, which interferes with the measurement of conditioned freezing. A study by Maren et al. (1997) is illuminating. Pre-training excitotoxic lesions of the dorsal hippocampus did not significantly impair contextual fear recall. However, post-training excitotoxic lesions of the dorsal hippocampus resulted in substantial deficits. This suggests that, when available, the hippocampal representation is the basis of the contextual association, but, when unavailable, some other (perhaps elemental) representation of the environment is used. When electrolytic lesions were instead made, conditioned freezing was decreased, though they show that pre-shock activity is elevated in lesioned rats, confounding

measurement of the CR.

Richmond et al. (1999) studied the effects of excitotoxic lesions of the complete, dorsal or ventral hippocampus on contextual fear conditioning. In agreement with Maren et al. (1997), they found that pre-training dorsal lesions did not cause an impairment. However, ventral or complete hippocampal lesions did, though they argue that the impairments in their study were also due to hyperactivity.

Context pre-exposure facilitation effect

A subsequent set of studies have explored the role of the hippocampus in a related behavioral phenomenon: the context pre-exposure facilitation effect (CPFE; Fanselow, 1990). When rats are very briefly exposed to a context and shocked, they later show almost no fear of the context. However, when they are first pre-exposed to the context, the shock elicits a conditioned fear response when subsequently returned to the shock context. This experimental design allows a disambiguation between the time in which the representation of the context is formed (pre-exposure) and the time in which the context-shock association is made.

In a series of experiments, Matus-Amat et al. (2004) explored the effect of inactivation of the dorsal hippocampus using the $GABA_A$ agonist muscimol. They found that muscimol injected into the dorsal hippocampus prior to context pre-exposure nullified the effect of the pre-exposure. Additionally, muscimol injected just prior to the shock inhibited the association of the context with the shock. They argue from these findings that context pre-exposure allows the hippocampus time to construct a contextual representation of the environment. When the representation is inhibited from forming, no contextual representation can form. When the representation is inhibited from being recalled, the shock cannot be associated with it.

One criticism of their study might be that muscimol injections during training might lead to other side effects. Barrientos et al. (2002) administered the protein synthesis inhibitor anisomycin into the dorsal hippocampus immediately after pre-exposure to the context. Late stage LTP (> 4 hrs) requires protein synthesis and has been shown to be inhibited in hippocampal slices by application of anisomycin (Mochida et al., 2001). Barrientos et al. found that pre-exposed rats did not show facilitation when anisomycin was administered, supporting the notion that the relevant contextual representation is found in dorsal hippocampus, and that late-stage LTP is required in order for the newly-formed contextual representation to be preserved for recall during the shock trial.

Rudy and Matus-Amat (2005) carried out similar studies in the ventral hippocampus, finding congruent results. Inactivation of the ventral hippocampus prior to pre-exposure reduced the CPFE, as did injections of

anisomycin into the ventral hippocampus following context pre-exposure.

Consolidation

While these studies have considered the role of the hippocampus in contextual fear conditioning over relatively short time scales, studies of human amnesia have led to theories of hippocampal function that predict that the role of the hippocampus in any declarative learning should be time-limited (McClelland et al., 1995; McClelland and Goddard, 1996). Young et al. (1994) explored this possibility within the contextual conditioning paradigm. When rats trained 7, 14 or 28 days after hippocampectomy were given fear conditioning training, the contextual association was not observed at any post-surgical latency. However, rats who were exposed to the context 28 days before lesioning were able to associate the context with the CS.

In a related study, Shimizu et al. (2000) tested mice with reversible CA1 NMDA receptor knockouts. Adding doxycycline to the mice' water caused NMDAR1 protein expression to be suppressed within 3-5 days, and NMDA currents and LTP were both mostly abolished in doxycycline treated mice. The protein began to reappear within 5 days of the end of treatment. Contextual fear conditioning impairments were caused by treatment during the first two weeks but not the fourth week *after* training; mice were tested one month after training. Since AMPA receptors were not affected, they suggest that the CA1 context encoding is repeatedly potentiated during the consolidation process, and interference with that re-potentiation disrupts consolidation.

This time-limited role of the hippocampus in contextual fear conditioning was also underscored by studies of neurogenesis and environmental enrichment. Abraham et al. (2002) studied perforant-path elicited LTP in DG, finding that shorter high-frequency stimuli produced LTP that lasted for a few days, while longer stimuli produced LTP that lasted for at least two months (over a year in the one recorded case). However, this long-lasting LTP diminished with exposure to an enriched environment containing a variety of objects that were rearranged each day. They posit that the cause of the diminishing LTP they observe is due to DG neurogenesis and associated pruning, triggered by experience in the enriched environments.

Feng et al. (2001) tested the role of neurogenesis more directly, by creating knockout mice lacking the *Presenilin-1* (PS1) gene, implicated in adult neurogenesis in the forebrain region (cerebral cortex, hippocampus, striatum, amygdala, etc., but not the olfactory bulb, thalamus, brainstem or cerebellum). The dentate gyrus is the cite of the highest levels of adult neurogenesis, so neurogenesis downregulation would be expected to have a particularly strong effect on hippocampal-dependent tasks.

While some mice were exposed to enriched environments, including "various toys, spin wheels, small tunnels, and houses," (Feng et al., 2001, p. 913)

others were not. Without enriched environment exposure, they found very little neurogenesis in either group (no difference). Enrichment triggered substantial neurogenesis in both knockouts and controls, but knockout mice showed a 37% reduction in new cells relative to controls. Electrophysiological properties of DG cells were not distinguishable between groups, nor were there performance differences across a variety of tasks (object novelty, Morris water maze, contextual and cued fear conditioning).

With two weeks of *prior* experience in enriched environments, both contextual and cued conditioning were enhanced. With two weeks of experience in enriched environments *subsequent* to conditioning, the contextual but not cued conditioning was retained more strongly in knockout mice than control mice. They propose that this difference is due to impaired enrichment-triggered DG neurogenesis in the knockouts; the knockouts failed to "clear out" old memories by replacing older DG cells with newer "naive" cells. Unlike DG, the amygdala does not show adult neurogenesis, and cued conditioning, which is amygdala but not hippocampal dependent, did not differ between groups.

In summary, the hippocampus is implicated in the formation of a contextual representation with which associations, such as a shock, may be made; however, once consolidated, the recognition of context no longer depends on intact hippocampal function. Suppression of NMDA function subsequent to acquisition impairs the maintenance of the contextual representation, and stimulation of DG neurogenesis hastens its clearance from the hippocampus.

5.1.2 *Extinction*

The extinction paradigm consists of first training the animal on a CS–US pairing, and then repeatedly presenting the CS in the absence of the US. After initial training, the animal will show a conditioned response to the CS. However, as the extinction phase proceeds, the response will lessen, until no CR is observed.

Extinction and hippocampal lesions

The functional underpinnings of conditioning in general and extinction in particular are most generally associated with structures in the basal ganglia, including the striatum and amygdala, and prefrontal cortex (Maren, 2001; Daw et al., 2005). The impact of hippocampal lesions on extinction varies substantially. While some studies have found little or no impact (Webster and Voneida, 1964; Berger and Orr, 1983; Corcoran et al., 2005; Ji and Maren, 2005), others have found more severe deficits (Kimble, 1968; Chan et al., 2003).

Partial reinforcement extinction effect

Even more perplexing are the lesion studies of the partial reinforcement extinction effect (PREE; Wike, 1953; Pennes and Ison, 1967). When rats are

given only partial reinforcement – some trials are rewarded, others are not – the CS–US association is more difficult to extinguish. While combined hippocampus / subiculum and EC / subiculum lesions both abolish the PREE, hippocampal lesions alone do not (Jarrard et al., 1986). Subiculum lesions alone also do not abolish the effect (Sinden et al., 1988), but transection of the subicular fibers to accumbens does (Rawlins et al., 1989), as do lesions of the nucleus accumbens (Tai et al., 1991). In short, the role of the hippocampus in extinguishing an association is murky at best.

Renewal of extinction

The role of the hippocampus in providing context specificity for the extinction has been more conclusively demonstrated. This context specificity is demonstrated most clearly during renewal experiments that vary the context during different parts of the experiment.

While one would expect extinction to simply decrement the associative strength between the CS and US until the association was "unlearned," converging evidence suggests that some component of extinction involves new learning (for reviews, see Bouton, 1993; Bouton et al., 2006). This is particularly well illustrated by the renewal phenomenon, of which there are several variants. Each variant takes the form x-y-z, where x denotes the context in which the original association is learned, y denotes the context in which the association is extinguished, and z denotes the context in which probe trials are given to test for a conditioned response. There are three variants: ABA, ABC and AAB. In ABA renewal, rats are conditioned to the CS–US pairing in context A, then extinguished in context B. When returned to context A after extinction, rats nonetheless continue to show a CR. The extinction is largely learned as a second context-specific association that supplements the original association. (However, the degree of CR observed in the test context is nonetheless not as strong as is found when there is no extinction, so some context-independent learning must occur.) The AAB variant is particularly intriguing. Though rats acquire and extinguish the association in the same context, the original association is more general, while the extinction is highly context-specific. Thus, a conditioned response is observed in the novel context B.

This immediately raises questions about the relationship between the CS, the US and the context. In particular, does the context have a direct additive effect on prediction of the US, or does the context modulate the CS–US association? Evidence suggests the latter. A direct positive association between the conditioned context and the US is belied by the observation of renewal in a novel context (i.e. the ABC variant). Neither is there a direct negative association between the extinguished context and the US: extinguishing one CS in an environment does not cause suppression of

another CS–US association in the same environment (Bouton and King, 1983).

When fimbria-fornix or excitotoxic hippocampal lesions were administered before the animals were shocked, no deficits were observed (Wilson et al., 1995; Frohardt et al., 2000). Similar to the aforementioned studies involving contextual fear conditioning, alternative (possibly elemental) environmental associations are constructed in the absence of an available hippocampal representation. However, dorsal hippocampal lesions or inactivation lead to a varied set of results. Lesions or muscimol inactivation (a $GABA_A$ agonist) before renewal testing result in generalization of the extinction outside of the extinction context (Corcoran and Maren, 2001, 2004; Ji and Maren, 2005). The same effect was observed when the dorsal hippocampus was lesioned prior to conditioning (Ji and Maren, 2005), as well as when the ventral hippocampus was inactivated (Hobin et al., 2006). However, inactivation before the extinction phase produced the opposite results: fear was expressed in all contexts. In essence, the extinction learning was lost.

One explanation for these findings may lie in the impact of hippocampal manipulations on learning the association and its extinction. Rats with pre-training lesions showed just as much freezing during the conditioning session as sham rats; however, on the next day, lesioned rats showed less freezing during the first extinction session (Ji and Maren, 2005). This suggests that, while hippocampal damage may not have interfered with the acquisition of the association, it may have interfered with maintenance of the fear memory after conditioning (Wittenberg et al., 2002). Lesions or inactivation before conditioning or after extinction would equate the impact of the hippocampus on the relative strengths of the original and extinction associations. In these circumstances, the extinction association, the more recently formed of the two, prevailed, irrespective of the context. By contrast, inactivation before extinction would result in maintenance of a strong fear association but a relatively weaker extinction association. (These experiments employed training paradigms based on a fixed number of trials.) In this circumstance, the fear association prevailed.

5.1.3 *Latent inhibition*

Latent inhibition and hippocampal lesions

Latent inhibition is essentially the same paradigm as extinction, only in reverse order. First, the animal is exposed to the CS in the absence of reinforcement; then, the CS is paired with a US. The unreinforced exposure to the CS decreases the CR that develops due to the subsequent CS–US training.

Early reports suggested that the hippocampus was responsible for the

latent inhibition effect (Ackil et al., 1969; Solomon and Moore, 1975). However, subsequent studies using more selective lesioning techniques have led to the hypothesis that it is in fact the entorhinal cortex that serves as the basis of latent inhibition (Honey and Good, 1993; Purves et al., 1995; Shohamy et al., 2000). Nonetheless, like extinction, latent inhibition shows a strong contextual specificity.

Renewal of latent inhibition

Honey and Good (1993) showed that latent inhibition is context specific. Rats were exposed to an unreinforced sound X in context A and an unreinforced sound Y in context B. Then, two types of appetitive reinforced trials in context A: AX+ and AY-. In control rats, prior unreinforced exposure to the AX trials resulted in a weaker conditioned response to the AX+ trials than the AY+ trials. In rats with excitotoxic hippocampal lesions, the latent inhibition generalized across contexts: lesioned rats showed the same weak response to both trials that control rats showed to the AX+ trials.

Holt and Maren (1999) performed a similar experiment using muscimol inactivation instead of lesions. Rats were given unreinforced CS trials either in context A or B, followed by fear conditioning trials (CS→shock) in context C. Freezing was then measured in context A subsequent to CS presentations. Control rats who received unreinforced trials in context B or no unreinforced trials at all showed similar high levels of freezing, while rats who received unreinforced trials in context A showed less freezing. Rats who received dorsal hippocampal inactivation prior to final testing also showed latent inhibition, comparable to that observed in control rats. However, the inhibition was not context specific: even when unreinforced trials were given in context B, latent inhibition was still observed in context A.

The role of the hippocampus in latent inhibition therefore closely parallels that of extinction.

5.2 Reversal Learning

This section first reviews the behavioral literature on various forms of reversal learning, including serial reversal learning and partial reinforcement reversal learning. The impact of hippocampal lesions on reversal learning performance is then considered.

5.2.1 *Behavior*

In the simplest case, reversal learning involves training an animal on an initial discrimination between two choices, A and B, in which one choice, say A, is rewarded and the other is not. The reward contingencies are then reversed: B is rewarded and A is not. There are several experimental

variants. In serial reversal learning, multiple reversals are performed, usually after a fixed number of trials have elapsed or after the animal reaches some predetermined threshold of performance on the current discrimination. In partial reinforcement reversal learning, the probability of reward for at least one of the choices is between 0 and 1. The standard notation includes the probability of reward, as a percentage, for each choice, separated by a colon: 100:0 therefore denotes full reinforcement, while 67:33 denotes partial reinforcement of both choices.

Serial reversal learning

In a classic work, Brunswick (1939) explored both serial and partial reinforcement reversal learning. Rats were trained on an elevated T-maze in which they had to choose the left or right arm for reward. Two general patterns emerged. First, when multiple reversals were performed between 100:0 and 0:100, rats reversed their choice in fewer trials with each subsequent reversal. This pattern is clearly observable even after the second reversal – when the rat first returns to the original discrimination – suggesting that the behavioral strategy to prefer choice A is not lost when, after the first reversal, the subsequent strategy to prefer choice B is adopted. Rather, the animal learns that there are in fact two patterns of reinforcement, switching between them progressively more quickly with each reversal. By the fifth reversal, most of the improvement occurs between the first and second trials after the reversal. While Brunswick (1939) trained rats roughly to a performance criterion before reversal, Gatling (1952) trained rats using an equal number of trials per reversal and found similar results. Buytendijk (1930) showed single-trial reversals after five or fewer reversals, and Dufort et al. (1954) observed single-trial reversals after at most six.

Partial reinforcement reversal learning

The second major finding by Brunswick (1939) is that more ambiguous choices lead to a slower reversal of choice. Full reinforcement reversals (100:0 $\rightarrow$ 0:100) lead to a relatively fast reversal of the rat's choice. However, more ambiguous reinforcement paradigms (50:0 $\rightarrow$ 0:50, 75:25 $\rightarrow$ 25:75, 100:50 $\rightarrow$ 50:100) lead both to a slower acquisition of the original discrimination and a slower reversal of choice when the reward contingencies are reversed. For sufficiently subtle differences in reinforcement (67:33 $\rightarrow$ 33:67), rats showed no choice preference during either the original discrimination or the reversal.

The partial reinforcement reversal results of Brunswick (1939) were criticized because of the ambiguity about whether the slower reversal was due to the partial reinforcement of the original discrimination, of the reversal discrimination, or both. Subsequently, Wike (1953) trained rats in a T-maze discrimination using either full (100:0) or partial (50:0) reinforcement. Both discriminations were reversed to 0:100. Again, rats in the partial reinforcement group were slower to learn both the initial discrimination and the

reversal. Thus, even though both groups of rats were trained on the same reversal condition, prior experience differences with the initial reinforcement condition influenced their subsequent acquisition of the reversal. Subsequent studies have confirmed these results Grosslight et al. (1954); Wise (1962); Pennes and Ison (1967).

Elam and Tyler (1958) studied partial reinforcement reversal learning in rhesus monkeys, varying the probability of reward of the less profitable choice instead of the more profitable choice. Monkeys were presented with two objects, under which food was sometimes placed. They compared two initial reward schedules 60:0 and 60:40. The reversed reward schedule was always 0:100. In the initial discrimination, the monkeys were able to show some choice preference for the more rewarded object; however, preference in the 60:40 condition was mild. When reversed, monkeys in the 60:40 condition showed slower reversal of behavior.

Pubols (1962) studied serial reversal learning, varying the number of trials per reversal (10, 20 or 40). When looking at performance as a function of the number of reversals, the 10 trials per reversal group is initially better, but the performance of the three groups eventually switches order, reflecting that the increased training per reversal improves the rats' performance. However, when looking at performance as a function of the total number of trials, the 10 trials per reversal group showed better performance than the 20 or 40 trials per reversal groups; the difference between the groups decreased with increased training. This may reflect a perseveration effect: after many trials with the same outcome, the contextual modulation of the current behavior may be diminished, resulting in more errors at the beginning of a new (reversed) block of trials.

Overtraining reversal effect

At the other extreme, Reid (1953) found that, when rats were significantly overtrained on the initial discrimination, they learned the reversal more quickly. Rats were first trained to criterion, and then given 0, 50 or 150 further trials. Then, the discrimination was reversed. While overtrained rats initially showed more perseveration for the initially rewarded choice, they eventually reversed more quickly. These effects were stronger for the 150 overtraining-trials group than the 50 overtraining-trials group.

In a further exploration of the overtraining reversal effect, Mackintosh (1962) first trained rats to distinguish cue card color (solid black vs. solid white); one was rewarded, the other was not. Mackintosh was able to reproduce the reversal results of Reid (1953), but Mackintosh also found that, if the nature of the task was changed to a task involving discriminating horizontal vs. vertical lines, the overtraining that sped reversal now interfered with the task switch: more overtraining on the black-white discrimination

caused slower acquisition of the horizontal-vertical discrimination. Thus, the overtraining reversal effect appears to be an attentional effect. Under a reversal, overtraining causes the reward to be more strongly correlated with the attended features (color) than with other features. However, when the relevant features change, overtraining impairs the animals' ability to switch consideration to new features of the environment.

5.2.2 *Lesions*

Not only does reversal learning lead to separate hippocampal representations of each condition (see Section 4.2.3), reversal learning has consistently been shown to be impaired by hippocampal lesions (Kimble and Kimble, 1965; Silveira and Kimble, 1968; Olton, 1972; Winocur and Olds, 1978; Berger and Orr, 1983; McDonald et al., 2002; Ferbinteanu and Shapiro, 2003) or fimbria-fornix transection (Neave et al., 1994; Aggleton et al., 1995; Fagan and Olton, 1986). For example, Kimble and Kimble (1965) trained rats with hippocampal lesions and controls to serially reverse a position preference on a Y-maze. Rats were given ten trials per day for 10 days during which time they learned to choose the arm based on location (or route to goal). Each reversal commenced upon reaching a predefined performance criterion. The median number of successful reversals for controls and hippocampals was 9 and 1, respectively. This substantial impairment is common across studies, as is the lack of significant impairment in hippocampal rats when learning the initial discrimination.

Berger and Orr (1983) studied rabbits in a non-spatial Pavlovian conditioning paradigm and reversal: one of two tones predicted an air puff to the eye. They measured the nictitating membrane closure as a conditioned response. Rabbits with hippocampal lesions showed dramatically slower learning on the reversal, though not on the original association. Interestingly, hippocampal rabbits did not show impairments on extinction, again disconfirming the hypothesis that the hippocampus' role lies in the inhibition of previously learned responses.

Winocur and Olds (1978) trained rats to pass through the correct door (out of two choices) in order to receive a reward. As expected, hippocampal lesions impaired performance when the reward was switched to the other door. However, if the reversal was carried out in a second room, the impairment was greatly reduced, suggesting that the room cues provided an external contextual cue with which to disambiguate current and past learning. McDonald et al. (2001, 2002), exploring the same paradigm, found that hippocampal rats, whether in the same room or a different room, initially perseverated on the reversal; however, both lesion and sham groups reached asymptotic

performance in the different room much more quickly than either group in the same room. Thus, an explicit change in context that accompanies the reversal supersedes any inference process about a possible change in behavioral context (reward contingency).

The notion that each new environment provides a context with which to associate a particular reward contingency is bolstered by Fagan and Olton (1986). Rats were trained to choose between two arms, one rewarded and one not, in several different rooms. The rewarded arm depended on the room. The sequence of training days was: D1, D2, D3, R3, D4, D1 (again), D5, D2 (again), where the number indicates the room number, D indicates the original discrimination in that room, and R indicates its reversal. Fimbria-fornix lesions caused some impairment in initial learning, an effect expected since rats were required to select the arm based on external room cues. Moreover, substantial impairment was observed when the reward contingencies were reversed in R3. Subsequent new discriminations D4 and D5 were learned much more quickly. Though not significantly different, there was a trend toward better performance in the D1 (again) and D2 (again) sessions than the D4 and D5 sessions, suggesting that, even in lesioned rats, some context-specific savings occurs when each context is explicitly differentiated (e.g. by different room cues).

In addition to the hippocampus, lesions of the medial prefrontal cortex have been found to impair reversal learning (Aggleton et al., 1995). However, reversal learning impairments were not found due to inactivation of the more specific prelimbic-infralimbic (PL-IL) area (Ragozzino et al., 1999). Interestingly, PL-IL inactivation was found to have an effect on an intermodal task shift: switching from a place strategy (go to a position) to a response preference (always turn right) on a four-arm radial maze was impaired (Ragozzino et al., 1999). This sets the stage for the possibility that the overtraining reversal effect is PL-IL mediated.

5.3 Associations within a Context

While the previous sections have discussed conditioned associations to an environmental or behavioral context, this section will focus on associations formed between elements (a place, an odor, a reward, etc.) within a context. Recent methodological advances have made it possible to lesion or inactivate single subregions within the hippocampal formation, as well as inactivate NMDA-dependent plasticity within one or more of these subregions. Much of this review will focus on those subregional studies in order to best illustrate what is known about the contribution of each subregion to overall

hippocampal function.

5.3.1 *Reference memory*

Reference memory and whole hippocampus studies in the water maze. Because of the strongly spatial nature of hippocampal representations, a variety of studies have examined the role of the hippocampus in forming associations involving spatial location. One of the most popular assessments of spatial learning is the Morris water maze task (Morris, 1981). A pool filled with an opaque liquid (often water with dried milk added) contains a single platform. The rat is released from a random starting location in the pool and must swim to the platform in order to end its swim. In the cued version of the task, the platform is visible just above the water, while in the hidden version the platform is slightly submerged, and the rat must learn to navigate to the location of the platform using distal visual cues outside the pool. The experiment is intended to highlight the rat's ability to navigate based on distal cues, and it is essentially a better-controlled version of the place preference task, discussed in the previous chapter. The liquid prevents navigation based on local odor cues, an important control since odor cues have been shown to stabilize hippocampal place fields (Save et al., 2000).

In a landmark result, lesions of the entire hippocampus have been shown to disrupt the hidden but not the cued platform version of the task (Morris et al., 1982). Moser et al. (1995) lesioned progressively larger segments from the dorsal (septal) and ventral (temporal) poles of the hippocampus. They found that rats were impaired in learning the platform location only when at least the dorsal 40-60% of the hippocampus was destroyed. As long as at least ~26% of the dorsal hippocampus remained, no impairment was observed, a result consistent with the observation that place cells are more spatially selective in the dorsal portion of the hippocampus.

Water maze variants

In a variant of the water maze task, Hollup et al. (2001a) trained rats in an annular water maze, essentially a circular water track. Rats typically swim in laps around the track, guaranteeing that they come into contact with the location of the platform in a reasonable amount of time. Rats with hippocampal lesions, even if pretrained before surgery in either a standard Morris water maze or the annular water maze, do not slow down and search for the platform at the correct location. Thus, the hippocampal place code is necessary for the recognition of a place-goal association.

In another variant of the water maze task, McGregor et al. (2004) trained rats in a rectangular water maze with no other cues. The NW and SE corners were therefore indistinguishable, as were the NE and SW corners, though the two groups of corners were distinguishable from each other (short wall

on the left vs. right). For each rat, one pair of corners was chosen, and the hidden platform was randomly alternated between them. After surgery, rats were trained on a cued version of the task, and all rats performed very well. However, on subsequent days when the cue was removed, rats with hippocampal lesions could not learn to distinguish the (possibly) correct corners from the (always) incorrect corners. They found equivalent results when four objects denoting the corners of the rectangle were used in a cylindrical pool. Rats could not learn to preferentially approach the two possibly correct goal locations near each corner. When trained in a square in which opposing walls were the same color (white or black) and neighboring walls were of opposite color, control and lesioned rats both performed very well, suggesting that they are able to solve the task when it involves a simple visual discrimination (white on left; black on right). They therefore attribute the deficit to an inability to process spatial relationships among cues.

Alternative training strategies

The water maze task is fairly complicated, and subsequent studies have explored how alternative training strategies affect the hippocampal dependency of the task. Morris et al. (1990) found that, if trained long enough, even hippocampal rats eventually perform as well as controls. Whishaw et al. (1995; 1996) first trained fimbria-fornix or hippocampus lesioned rats on the visible-cue variant of the water maze task. With sufficient training, rats learned to correctly navigate to the platform, even without the cue. However, the same rats showed no ability to transfer learning to a new location, while control rats were able to flexibly learn new platform locations. Day et al. (1999) trained rats using progressively smaller platforms, another effective training paradigm. However, rats tended to overshoot the location of the platform, suggesting that they had learned to navigate by learning the direction to the platform, but they were unable to recognize its location (using the distal landmarks) once there.

These studies suggest that multiple learning systems are present within the rodent brain, and, in the absense of the hippocampus, other systems attempt to compensate. An elegant demonstration of this hypothesis was provided by McDonald and White (1993), who tested rats with lesions to hippocampus, amygdala, or dorsal striatum on three different 8-arm radial maze tasks. In a spatial *win-shift task*, rats were trained to visit each arm once for reward. This is a spatial working memory task, and only hippocampal lesions caused impaired performance. In the *conditioned cue preference* task, rats were confined to "light" or "dark" arms, only one of which was rewarded. Only rats with amygdala lesions showed no preference for the previously rewarded arm when allowed to choose between them. In a non-spatial *win-stay task*, rats were trained to approach a light for food reward. Only rats with

dorsal striatal lesions were impaired. Taken together, the authors suggest that different learning systems handle spatial working memory (win-shift task), cue-reward associations (conditioned cue preference), and cue-response associations (win-stay task).

Declarative vs. procedural learning

Is the role of the hippocampus simply to rapidly learn goal-place associations, or does it also play a role in procedural learning? For example, one might dissociate the strategy for navigating to the platform from the specific knowledge of the platform's location. When first learning the water maze task, rats execute an innate thigmotaxic strategy, in which they travel around the edge of the pool, whiskers in physical contact with the pool walls, searching for an exit. (This is often described as an arbitrary behavioral strategy, though the notion that a rat should instead swim out into the center of a body of water in order to find an exit from it is unintuitive.) Over the course of multiple trials, repeated failure of this strategy and discovery of the hidden platform lead to acquisition of the task. Day et al. (1999) also examined, after the shrinking platform training, how quickly rats with and without hippocampal lesions would switch back to a thigmotaxic strategy when the platform was moved to random positions at the edge of the pool. Hippocampal rats were significantly slower switching back to the thigmotaxic strategy.

To address the procedural learning question further, Micheau et al. (2004) trained rats on a variant of the water maze task in which rats had to dwell in the location of the hidden platform for a prescribed amount of time before the platform would rise. While rats with continuous (chronic micropump) hippocampal inactivation showed some location preference, they displayed a complete failure to dwell in the correct location. In a second experiment, half of the control rats from the first experiment received continuous inactivation subsequent to training (during consolidation). The inactivated rats still displayed a dwell response, indicating intact procedural memory, but forgot where the platform was located, indicating impaired declarative memory.

Despite the distinction between declarative and procedural memory, it's important to recognize their essential dependency: learning a procedural strategy through trial and error requires a declarative (perhaps episodic) memory of recent successes and failures to guide improvement of the behavior. This dependency may underly the results of Micheau et al. (2004). The memory of recent trials likely guides the switch from the failing thigmotaxic strategy to a dwelling-at-a-place strategy and back. However, while the incremental improvement of the strategy likely requires the recall of recent experiences (see Section 4.1.7), the resultant strategy is not dependent on hippocampal consolidation.

A pair of studies has explored the role of NMDA-dependent LTP in water maze learning (Bannerman et al., 1995; Saucier and Cain, 1995). In each case, they found that naive rats were impaired in water maze learning during systemic NMDA receptor inactivation; however, rats given pretraining to find a platform before inactivation were unimpaired, even if, during pretraining, the platform moves each trial and the water maze is in a different room. Thus, they argue that the NMDA receptor contribution to water maze strategy learning can be dissociated pharmacologically from its contribution to platform location memory.

A subsequent study has clouded this interpretation. Rampon et al. (2000) showed that CA1 NMDA receptor knockout mice show deficits in contextual fear conditioning and social transmission of food preference; however, with prior experience in enriched environments, knockouts showed no performance deficit. Critically, they show that, even in the knockouts, synaptic density is higher in CA1 following exposure to enriched environments, indicating that NMDA-independent plasticity is possible in CA1. (An *in vitro* study showed that LTP is easier to induce in hippocampal slices from enriched mice (Duffy et al., 2001), though this larger potentiation has been associated with NMDA receptors (Tang et al., 2001).) To what extent the benefits of water maze pretraining (Bannerman et al., 1995; Saucier and Cain, 1995) may be due to a similar enrichment effect is not known.

DG-specific impairments

Reference memory and subregional studies in the water maze. Subregional lesions paint a more nuanced picture of hippocampal function. McNaughton et al. (1989) showed that colchicine lesions of DG impaired water maze performance, as well as two other spatial tasks (hole preference on a circular platform, and new-daily spatial memory in a radial maze). Interestingly, spatially selective firing in CA3 and CA1 was not disrupted, suggesting the deficit did not arise from the lack of a spatial representation. Xavier et al. (1999) found that rats with DG lesions could eventually show preference for the general area (quadrant) of the platform, but not its specific location, and tended to take more circuitous routes to get there. Lassalle et al. (2000) blocked the output of DG by temporarily inactivating the mossy fiber synapses, which caused an impairment only when the mossy fibers were blocked during learning. When blocked during recall (after a week of training) or during consolidation, no effect was observed. Thus, it seems DG is critical for acquisition (but not performance of) the water maze task, at least when the procedural aspect of the task is unfamiliar.

Otnaess et al. (1999) blocked LTP of perforant path synapses to DG by electrically saturating them. Comparing low-frequency and high-frequency

stimulation, they found that only high-frequency stimulation saturates LTP, and only high-frequency stimulation disrupts learning the water maze task. (What other efferent areas are affected by this stimulation was not discussed.) However, if rats are pretrained on another water maze in a different room with different distal cues and a different platform location, high-frequency stimulation has no effect on task performance.

The Otnaess et al. (1999) results are similar to the aforementioned systemic NMDA-LTP inactivation results (Bannerman et al., 1995; Saucier and Cain, 1995). It should be noted that mossy fiber LTP is associative but not NMDA-dependent (Schmitz et al., 2003), suggesting that, when the procedure is familiar, mossy fiber LTP (and any non-NMDA dependent plasticity in CA3 or CA1) is sufficient for reference-memory learning of a platform location. Hoh et al. (1999) attempted to address the procedural learning question by inactivating NMDA receptors during non-spatial (random platform location) pretraining as well as reference-memory water maze training. They report no impact of systemic NMDA inactivation; however, their data suggest that the inactivation window was not sufficient to block learning during later trials.

CA3-specific impairments

Brun et al. (2002) found that rats with excitotoxic lesions of CA3 were also impaired in learning the water maze platform location. However, in the annular water maze, CA3 lesions had no effect. This result is in contrast to whole hippocampus lesions which do impair performance in the annular water maze (Hollup et al., 2001a). CA1 place fields were slightly less sharp but otherwise normal without CA3, even after manual transection of all fibers between CA3 and CA1. Thus, in sufficiently constrained circumstances, the DG-CA3 network is not necessary to demonstrate a place response.

Nakazawa et al. (2002) studied the performance of mutant mice lacking functional NMDA receptors (and therefore NMDA-dependent LTP) only in CA3. In the standard version of the task, four extra-maze cues were provided to aid the rats' navigation, and knockout mice performed as well as controls. When three of the four cues were removed, CA3-NMDAR knockout mice performed as badly as when no cues were present, but control mice performed as well as if all four cues were present. With only one cue present, CA1 place fields of mutants were smaller and less robust than controls; however, CA1 fields returned to their previous strength under full cue conditions. Thus, whereas the mossy fiber pathway appears necessary for platform location encoding, the perforant path and recurrent input to CA3 appears necessary for maintaining a coherent spatial representation in the face of environmental changes.

CA1-specific impairments

By contrast, Tsien et al. (1996) showed that NMDA-dependent LTP in CA1 is necessary for learning the water maze under normal conditions. They

created CA1-specific NMDA receptor knockout mice, and found that, during probe trials, CA1-KO mice showed no preference for the quadrant of the pool in which the platform was previously located. McHugh et al. (1996) report that CA1 place fields in such mice are stable but larger and more diffuse, and cells with overlapping place fields show reduced temporal correlation compared to controls.

To examine the role of consolidation in water maze learning, Shimizu et al. (2000) created reversible CA1 NMDA receptor knockout mice. Adding doxycycline (doxy) to rats' water caused NMDAR1 protein expression to be suppressed within 3-5 days, and NMDA currents and LTP were both mostly abolished in doxy treated rats. The protein began to reappear within 5 days of the end of doxy treatment. Doxy treatment during training caused impairments in water maze performance (as it did with contextual fear conditioning). Doxy treatment during the first but not second week *after* training caused impaired water maze performance; mice were tested two weeks after training. Since AMPA receptors were not affected, they suggest that CA1 memory encodings are repeatedly potentiated during the consolidation process, and interference with that re-potentiation disrupts consolidation.

Object replacement and displacement

Reference memory and other tasks. DeCoteau and Kesner (1998) trained rats on one of three go / no-go discrimination tasks in which rats were presented with a standard "scene" comprising a series of objects at particular places in the arena or a perturbed scene, which could involve replacement of an object, displacement of an object, or both. Rats learned to approach the standard scene but not the perturbed scene for a reward. Rats were trained prior to lesioning. Both hippocampal and parietal lesions caused rats to incorrectly "go" when the object was displaced but not replaced. The results for the replacement / displacement task were more complex. Controls learned to solve this task spatially; they performed well when probed on displacement only scenes, but not on replacement only scenes. Interestingly, hippocampal rats re-learned to solve the task based on object identity, showing strong performance in replacement-only probe trials.

Configural learning

Gilbert and Kesner (2003) trained rats to associate odors or small toys (A and B) with positions (1 and 2) in an apparatus. Objects were placed atop a small food cup filled with sand, requiring displacement of the object to retrieve the reward. Odors were mixed with the sand. Food rewards were burried in the sand, requiring digging to retrieve the reward. On each trial, a single food cup was presented. This is a biconditional discrimination task: A1 and B2 were rewarded; A2 and B1 were not. Rats with excitotoxic lesions of the dorsal CA3, but not dorsal DG or dorsal CA1, were much slower to

learn the object-place associations and were unable to learn the odor-place associations.

What is surprising about this result is that CA1 lesions do not cause impairment. CA1 is typically considered to be the output pathway of the DG-CA3 circuit, suggesting that lesions of CA1 should prevent the CA3 representation from reaching efferent areas. The authors point out that pilot studies involving complete CA1 lesions also do not cause impairment, suggesting that the extrahippocampal pathway from CA3 to the lateral septum mediates (or is sufficient for) performance of this task.

The biconditional discrimination task used by Gilbert and Kesner (2003) is an example of a so-called "configural association" task (Rudy and Sutherland, 1995). The hallmark of configural tasks is that they can be learned only by constructing reward associations with conjunctions of cues, e.g. odor and position, as opposed to single (elemental) cues. Unfortunately, the inconsistencies among hippocampal lesion studies addressing configural association learning are substantial, and others have suggested that these inconsistencies are indicative of some of the tasks having an underlying spatial or temporal component (Jarrard, 1993; Redish, 2001). As such, the configural learning lesion literature has tended as much to obscure as to clarify the function of the hippocampus and it will not be further addressed here.

5.3.2 *Short and intermediate-term memory*

Intermediate-term memory. One of the difficulties in interpreting results involving the reference-memory water maze task is that only one platform location is ever learned, and performance is generally measured over several days of training. An alternative testing paradigm is to change the location of the platform each day. While "working memory" is typically used to describe an association that is acquired and applied within a single trial, such as in a DNMS task, intermediate-term memory testing involves multiple (typically 4 or more) trials using the same platform location.

During the first trial after the platform is moved, perseveration on the previous location is typically observed, leading to poor performance. The key feature of this testing paradigm is the observation of improvements on subsequent trials that demonstrate rapid encoding of the new platform location. This paradigm has the additional benefit that it decouples learning of the path to the goal from learning the location of the goal itself, since improvement during the second trial requires both that the animal has learned the new platform location (to some degree) and that the animal does not take the first-trial path to get there.

Lesion studies

Hippocampal lesions severely impair intermediate-term place learning in the water maze (Whishaw et al., 1995; Whishaw and Jarrard, 1996). While DG lesions moderately impair reference-memory place responses, intermediate-term memory for the platform is completely abolished (Xavier et al., 1999). Rats with DG lesions show no preference for the platform position on the day before or improvement in finding the platform over the 4 trials each day.

Similar to the water maze, the 8-arm radial maze is sometimes used to test spatial memory. McNaughton et al. (1989) exposed rats to a single sample arm during the first trial each day. On subsequent trials, spaced 15 min apart, the animal could explore all eight arms, but only the sample arm was rewarded. A new sample arm was chosen each day. DG lesions caused more errors in returning to the sample arm after the first trial.

NMDA inactivation studies

Steele and Morris (1999) pretrained rats on the intermediate-term water maze task and then studied how hippocampal lesions, systemic NMDA receptor inactivation, or intrahippocampal NMDA receptor inactivation effected performance on subsequent testing. Between the first and second trial, they inserted a delay of 15 sec, 20 min or 2 hours. Hippocampal lesions completely disrupted performance at all delays. NMDA receptor inactivation, whether systemic or intrahippocampal disrupted performance at the 20 min and 2 hour delays, but not the 15 sec delay. NMDA-inactivated rats did show some improvement with each trial, and perseverated on the previous day's platform location, suggesting that some degree of gradual learning was still possible.

Interestingly, since rats were pretrained on the task, this result stands in contrast to reference-memory tests that did not observe a significant impairment (Bannerman et al., 1995; Saucier and Cain, 1995). The different results underscore the role of the hippocampus in the rapid acquisition of new associations. In the reference memory version, the goal position may be acquired gradually over several trials. In the present experiment, performance was measured on the second trial, and, indeed, most improvement was shown between trials 1 and 2. Steele and Morris (1999) argue that such one-trial improvement requires NMDA-dependent one-shot associations with the hippocampal map, whereas more gradual learning requires the map, but not NMDA-mediated platform associations with points on the map.

In a related experiment, Nakazawa et al. (2003) trained wild-type and CA3 NMDA receptor knockout mice on an intermediate-term water maze task. After three days of pretraining on a cued version of the task, days 1-12 consisted of "training" days in which mice were placed on the platform if they did not find it. During these days, both groups of mice steadily improved, presumably improving their search strategy and learning the

procedural aspects of the task. Days 13-16 consisted of "test" days. Wild-type and knockout mice showed the same latencies to find the platform during the first trial, but wild-type mice improved significantly more during subsequent trials than knockout mice. Knockout mice showed no more perseveration for the previous day's platform location than wild-type mice, suggesting they learned the rule-based volatility of the platform, but weren't able apply the "one-shot" learning from trial 1 during trial 2.

The obvious interpretation of these results is that CA3 NMDA receptor function is necessary for proper recall of the current platform location; however, this is not the case. When mice were instead presented during days 13-16 with the locations of 9-12, respectively, knockouts performed as well as controls. Thus, CA3 NMDA receptor function was necessary only when the platform location was novel. Place cell recordings in CA1 of the knockout mice during the exploration of familiar and novel portions of a track show that place fields are larger, more diffuse, and often multi-peaked in just the novel portion. On the next day, place fields look the same in both portions of the track. This suggests that CA3 NMDA-dependent synaptic plasticity is critical for rapidly creating novel representations, including place-goal associations, a hypothesis further supported by DNMS studies discussed in the following section.

Short-term memory and delayed-(non-)match-to-sample. In the intermediate-term memory studies above, several trials were given on the same goal, but performance improved (and was sometimes judged) mostly from the degree of improvement from the first to the second trials. Naturally, the hippocampal dependence on such "one-shot" learning would be suspected to carry over to tasks in which the goal changes each trial.

Complete hippocampal lesions

Long and Kesner (1996) trained rats to perform either a go / no-go working memory or reference memory task requiring the discrimination of different (allocentric) distances between objects. In the reference memory task, rats were rewarded when the objects were 7 cm but not 2 cm apart (or vice versa), and hippocampal lesions did not impair rats at this task. In the working memory task, a study phase presented objects at one of these distances, and a test phase presented either the same distance or the other distance. Rats were trained to approach the objects only when at the same distance as in the study phase for a food reward. Complete hippocampal lesions impaired rats' performance in this task. While spatial relationships underlie this task, the task is not strictly a spatial task in the navigational sense, and, consistent with this, lesions of just the dorsal hippocampus, which impairs water maze performance, did not impair this task. (Lesions limited to ventral

hippocampus did not cause impairment either.)

In a related study, Long and Kesner (1998) trained rats to perform either go / no-go working memory or reference memory tasks testing egocentric distance or spatial discrimination. The test for egocentric distance placed an object at one of two distances from the rat and required it to identify the correct distance. In the reference memory version, one of the distances was always correct. In the working memory version, a DNMS task, a study phase was presented in which an object was placed at some distance from the rat and, after a delay, the object at either the same or a different distance was presented. The test for spatial discrimination involved placing an object at one of two locations in the arena. In the reference memory version, rats had to discriminate between the fixed rewarded and unrewarded locations. In the working memory version, rats had to "go" to the object during the test phase if and only if it was in the same location as when presented in the study phase. Rats were impaired at working-memory versions of both the egocentric distance and spatial discrimination tasks, but were unimpaired at either reference memory task.

Clark et al. (2001) trained rats on an object-based DNMS task, and found that, even with extended training, hippocampal rats performed poorly at 1 min and 2 min delays, but were unimpaired at 4 sec or 30 sec delays. Their results stand in contrast to Duva et al. (1997), who found no impairment, but Clark et al. note that the lesions performed by Duva et al. (1997) were not as complete, perhaps permitting unimpaired performance.

Intra-hippocampal NMDA inactivation

Day et al. (2003) trained rats on a one-trial food-place paired associate task. Rats were presented with two samples phases and a choice phase. During each sample phase, rats were presented with food of a randomly selected flavor (F1, F2) buried at a randomly located food well (L1, L2). During the choice phase, rats were given a taste of one of the foods (e.g. F1) and had to navigate to the correct food well (e.g. L1) for reward. Many flavors and positions were used over the course of the experiment.

Rats with intra-hippocampal AP5 (NMDA antagonist) injections before the first sample phase but not before the choice phase performed at chance. Rats with intra-hippocampal CNQX (AMPA antagonist) injections at either time performed at chance. One might suspect that CNQX caused general spatial impairments by blocking all hippocampal activity, including general place-specific activity (apart from any place-goal association). However, rats repeatedly trained on the same two associations (instead of associations that varied with each trial) performed above chance despite either antagonist. Thus, interference with NMDA-dependent synaptic plasticity appears to effect the encoding of food-place associations, and CNQX inactivation of the

Subregion	Delay	Environment	NMDA inactivation	Dorsal lesion
	10 sec	familiar	none	transient
DG	10 sec	novel	none	partial
	5 min	familiar	(unknown)	*sustained*
	10 sec	familiar	none	transient
CA3	10 sec	novel	partial	partial
	5 min	familiar	transient	*sustained*
	10 sec	familiar	none	none
CA1	10 sec	novel	none	none
	5 min	familiar	*sustained*	*sustained*

TABLE 5.1: A summary of impairments by Lee and Kesner (2002; 2003), categorized as one of four types. A transient impairment is an initial performance accuracy decrease subsequent to the lesion or inactivation which eventually dissipated. A partial impairment is a performance decrease with some subsequent improvement, but not a return to baseline accuracy. A sustained impairment is an initial deficit that does not improve.

hippocampus impairs their recall.

Subregional lesions and inactivation

In a systematic study of the role of each hippocampal subregion, Lee and Kesner (2002; 2003) lesioned or inactivated each subregion after training rats on a spatial DNMS task in an 8-arm radial maze. For each trial, rats first explored a randomly-selected study arm, and, after a delay, were then allowed to enter either the study arm (unrewarded) or a neighboring choice arm (rewarded). There were three variants of the task. In the familiar variant, the delay was 10 sec and the room used was the same as that during training. In the novel variant, the delay was still 10 sec but the task was performed in a different, novel room. In the delay variant, the delay was increased to 5 min and the task was performed in the familiar room.

Their results are summarized in Table 5.1, and there are two particularly interesting findings. First, in a novel but not familiar environment, a sustained partial impairment is observed if DG is lesioned or CA3 is lesioned or inactivated. Together with the results of Nakazawa et al. (2003), this suggests that DG is necessary for creating a task-relevant representation of the new environment, and CA3 must learn this representation using NMDA-dependent synaptic plasticity. Second, the 5 minute delay requires an intact DG, CA3 and CA1, as well as intact learning in CA1 but not CA3, since the CA3 learning deficit was transient. (AP5 leakage from DG to CA1 prevents interpretation of the DG inactivation data.)

The difference in results with the longer delay time highlights the dis-

tinction between two independent memory processes (Hebb, 1949): active maintenance and associative learning. Active maintenance refers to the sustained activation of neurons in order to maintain a representation of some memory; disruption of this activation results in the memory being lost. Associative learning involves adapting synaptic strengths in a neural network so that the memory can later be recalled (Hopfield, 1982, 1984). Since the memory is encoded in the weights for later recall, the memory can be recalled so long as the synaptic strengths are maintained; sustained activity is not necessary. (The role of the hippocampus in associative memory is discussed in detail in Chapter 7.) The hippocampus has long been implicated in spatial associative memory (e.g. McNaughton and Morris, 1987), and there is converging evidence for active maintenance of spatial and other memories in frontal and parietal areas across the mammalian phylogenetic tree (Cohen et al., 1997; Awh et al., 1998; Courtney et al., 1998; Postle et al., 2004; see also Section 4.3).

Interpreted in the context of these two memory processes, the Lee and Kesner results suggest that an extrahippocampal active maintenance buffer is sufficient to bridge the 10 sec delay, whereas associative memory storage in the hippocampus is necessary to bridge the 5 min delay. Disruption of performance by NMDA inactivation in CA1 but not CA3 suggests that, whereas plasticity in CA3 is necessary for the creation of an appropriate representation, plasticity in CA1 is necessary for preserving each trial-specific memory during the delay period.

A series of related studies by Gilbert et al. (1998; 2001; 2006) further elucidates the function of each subregion. They trained rats to perform a spatial DNMS task in an open field. During the sample phase, rats retrieved a reward covered in a food well that was placed at one of 15 locations in the apparatus. During the choice phase, the sample phase well was rebaited and an unbaited distractor well was also covered at a variable distance (15 cm, 37.5 cm, 60 cm, 82.5 cm, or 105 cm) from the rewarded well. Rats with CA3 lesions were impaired at this task irrespective of the distance between the two wells. By contrast, rats with DG lesions were impaired in a distance-dependent manner: while no significant impairment was observed at the farthest distance, progressively stronger impairments were observed at closer distances, and these impairments did not ameliorate over time. Rats with CA1 lesions showed no impairment.

The lack of effect of a CA1 lesion stands in contrast to the 8-arm radial maze DNMS study (Lee and Kesner, 2002, 2003). The difference may lie in the way in which rats solve each problem. The open field task may be solved as a novelty problem: which of the two target objects during the test phase

is in a novel position? This hypothesis is consistent with impairments in novelty detection following DG and CA3 but not CA1 lesions (see Section 5.3.4). Novelty detection is not applicable to the 8-arm task and the water maze task, in which there is no sensory comparison to be made; rather, the rat must recall its previous behavior. In these latter tasks, an intact CA1 is crucial to good performance.

Just as Nakazawa et al. (2002) found that NMDA-dependent LTP in CA3 was necessary for reference memory water maze performance in cue-degraded environments, Gold and Kesner (2005) found that CA3 lesions disrupt a spatial DNMS task when environmental cues are removed during the testing phase. Since task training occurred before the lesion and cues were removed only during the test phase, these data further support CA3 being integral to recall of an intact spatial representation under landmark perturbation.

Spatial alternation

Short-term memory and overlapping sequences. Section 4.2.2 discussed the inconsistently observed physiological representations of spatial alternation (e.g. on a T-maze). Perhaps the most interesting aspect of the physiology data is that, when a delay is imposed at the start of the "common" segment of the T-maze, differentiation of the left-to-right and right-to-left routes is largely abolished (Ainge et al., 2005; Robitsek et al., 2005). It is therefore somewhat surprising that hippocampal lesions impair T-maze alternation only when the delay is present (Ainge and Wood, 2003; see also Neave et al., 1994; Aggleton et al., 1995); with no delay, normal rats remap but hippocampal rats are not impaired. Thus, the conditions under which the hippocampus physiologically differentiates the common segment based on the rat's route is dissociated from the conditions under which hippocampal function is necessary to follow that route.

Odor sequences

In a non-spatial odor-based version of the DNMS task, Agster et al. (2002) trained rats on two overlapping six-odor sequences, as shown in Figure 5.1. Learning of the sequences was measured by correctly choosing the odor during the fifth stage, which depended upon which odors had been rewarded during the first and second stages. Rats were pretrained on the task prior to lesions, so the purpose of the study was not to demonstrate acquisition of the sequences themselves (reference memories), but to determine whether the hippocampus is involved in recalling the particular sequence in each trial.

In all cases, the inter-stage interval was less than one minute. When rats were trained on alternating sequences presented in relatively rapid succession (1 minute inter-trial interval), hippocampal rats initially performed at chance levels. Rats with radiofrequency hippocampal lesions improved to near-

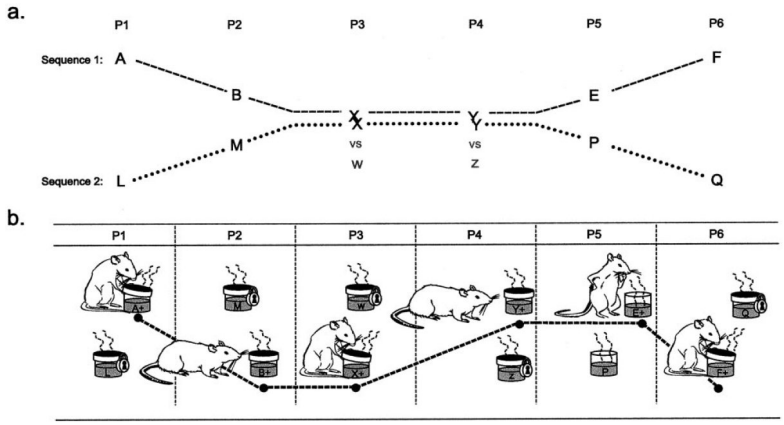

FIGURE 5.1: Overlapping sequences used by Agster et al. (2002). A) The initial difference between the sequences (P1 and P2) indicates how rats should respond on P5. B) All but choice P5 is forced. The position of the odor cups is randomized to ensure no spatial relationships can be used to solve the task. (Reprinted from Agster et al., 2002.)

control levels, whereas rats with ibotenic acid lesions did not. (The exact cause of the difference is not clear.) When the inter-trial interval was increased to 15 minutes and sequence order was randomized, rats with radiofrequency lesions performed as well as controls. However, when a 30 minute delay was inserted before the fifth stage, hippocampal rats returned to near chance levels, despite the inter-trial interval. Thus, when sequences are presented in quick succession or when a pre-choice delay is imposed, the hippocampus plays a critical role.

5.3.3 *Temporal associations*

Trace conditioning

Reference memory. In delay conditioning, a CS (e.g. a tone) is continuously presented over some short window of time, near the end of which, a US (e.g. a shock) is given. Animals learn the temporal offset between the beginning of the CS and the time of the US, displaying a CR at the expected time of the US. A variant of this, called trace conditioning, uses a short CS, followed by a "trace" period in which nothing happens, followed by the US. As with delay conditioning, normal animals learn to predict the US.

Hippocampal lesions impair the acquisition of trace but not delay conditioning (Solomon et al., 1986). Huerta et al. (2000) trained wild-type and CA1 NMDA receptor knockout mice in delay and trace conditioning paradigms, and found that the knockout mice were slower than controls to learn only the trace conditioning paradigm; however, unlike rats with complete hippocam-

pal lesions, asymptotic performance of the knockout mice was equivalent to controls.

Beylin et al. (2001) explored whether the hippocampus would be implicated if the delay task were simply more difficult. They double the length of the delay period and found that rats with hippocampal lesions took four times the number of trials to reach the control animals' asymptotic performance; however, the deficit was mild. Hippocampal rats trained on the trace conditioning paradigm showed almost no freezing after the same number of trials (1200). Interestingly, they also found that, once hippocampal rats were trained on the delay conditioning paradigm, performance transferred to subsequent trace conditioning, suggesting that, once constructed, the CS-US association can be maintained even with the trace.

Bangasser et al. (2006) further explored the role of the hippocampus in trace conditioning by training rats on trace, delay and a contiguous trace conditioning (CTC) paradigm. In the CTC paradigm, the CS and trace period are as in standard trace conditioning, but the CS is presented a second time simultaneously with the US. Thus, while the first presentation of the CS is not directly linked to the US as it is in delay conditioning, it is indirectly linked through the second presentation of the CS. Hippocampal lesions impaired the trace but not the CTC paradigm, suggesting that the critical role of the hippocampus is to bridge the CS-US association across the trace period.

Transitive inference

Dusek and Eichenbaum (1997) tested rats on a transitive inference task. Rats were trained on pairs of odors (A+ vs. B-, B+ vs. C-, C+ vs. D-, D+ vs. E-). Two types of lesions were performed: aspiration lesions of the perirhinal and entorhinal cortex and radiofrequency lesions of the fornix. Performance of the two lesion groups did not appear to differ from each other. After surgery, lesioned rats continued to perform indistinguishably from controls on the training pairs (B+ vs. C-, C+ vs. D-), but fell to chance levels on the "relational probe pair" (B+ vs. D-). When trained on new pairs, both control and hippocampal rats failed to perform well during initial training, demonstrating that the transitive inference deficit was not due to the control animals quickly learning the correct response for the new BD pair.

Van Elzakker et al. (2003) revisited the odor-based transitive inference task, varying the number of odor pairs. The first task contained four pairings (A+B-, B+C-, C+D- and D+E-), and they reproduced the findings of Dusek and Eichenbaum (1997), namely that rats correctly solved B+D-. However, when trained on a second task using five pairings (A+B-, B+C-, C+D-, D+E-, E+F-), rats could correctly solve B+E- but not B+D-. Interestingly, in both tasks, rats learned the last pairing (task 1: D+E-; task 2: E+F-) in the sequence better than any of the others. The authors suggest that rats are learning to

solve the last pairing purely as an avoidance problem: avoid E- (task2: F-) and ignore the positive odor with which it is paired. Thus, while the D+ (task 2: E+) should get positive valence from this pairing to offset its negative valence in the other pairings, it in fact does not. The B+D- pairing in the first task and B+E- pairing in the second task elicit transitive responses, but the B+D- in the second task does not, because it is E that is undervalued in the second task, not D. Thus, Van Elzakker et al. (2003) conclude that rats cannot really solve transitive inference problems, even with a hippocampus.

Spatial sequences

Single trial memory for sequence order. Gilbert et al. (2001) tested for learning the temporal order of a sequence of arm visits in an 8-arm radial maze. Rats were forced to visit each arm in a sequence that varied each day (one trial per day). Two of the arms, with temporal separation of 0, 2, 4 or 6, were then opened and rats were rewarded for entering the arm occurring earlier in the sequence. Rats with dorsal CA1 and dorsal CA3 (Kesner et al., 2004) but not dorsal DG lesions performed near chance, even for the longest separation (6 arms). Previous studies have demonstrated the same result for complete hippocampal lesions (Chiba et al., 1994). Also, DeCoteau and Kesner (2000) showed that, for fixed sequences, rats optimize their path through the arms for speed (procedural memory), and this optimization is suppressed by striatal lesions, specifically the medial caudoputamen.

Odor sequence learning

In an impressive display of odor memory, two groups have shown that the hippocampus is necessary for learning the temporal order of a sequence of odors (Fortin et al., 2002; Kesner et al., 2002). Fortin et al. (2002) trained rats on one-trial sequences of five odors, and then tested them on either the presentation order of two odors in the sequence or on whether an odor was part of the sequence (see Figure 5.2).

When tested on sequence order – a Fruit Loop was hidden in the less recently presented odor cup – rats with hippocampal damage (radiofrequency lesions) performed at near chance levels on almost all pairings. (Only A vs E was significantly different from chance, though performing at less than 70% correct). However, when tested on recognition – a Fruit Loop was hidden in the cup whose odor was not presented in the previous sequence – performance of hippocampal rats was indistinguishable from controls. Both groups performed at ~80% on the earliest odor comparisons (A vs. X) and at ~90% on the most recent odor comparisons (E vs. X), so hippocampal rats appear to have the same temporal bias as control rats. The authors suggest that they therefore have "normal access to differences in trace strengths for the odors." However, the observation that traces degrade does not imply that animals can "measure" this degree of degradation in order to infer

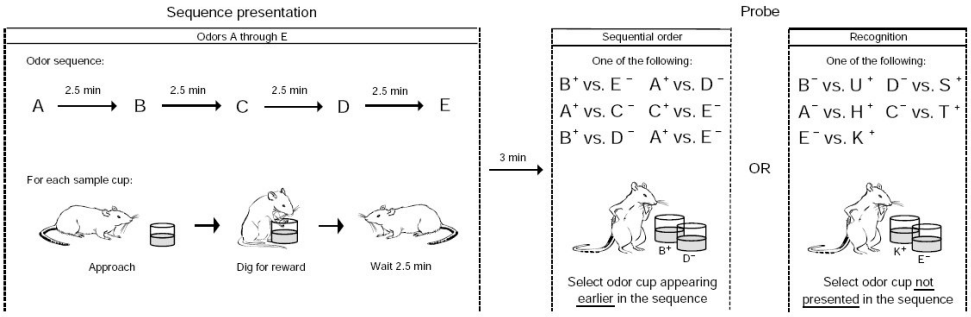

FIGURE 5.2: Details of the sequence presentation and probe trials of Fortin et al. (2002). For each trial, 5 odors were randomly selected from a set of 20. A variable (average 2.5 minute) inter-stimulus interval was interposed between stimulus presentations. The probe test consisted of two sand-filled odor cups presented in one of three locations in the home cage, in order to randomize the position of the cups. The inter-trial interval was 30-60 minutes. (Reprinted from Fortin et al., 2002.)

stimulus order, nor that the degradation of odor traces within any particular trial is gradual; a bi-stable recognition memory system, one whose traces remained strong until they vanished suddenly after some degree of time, could nonetheless produce behavior consistent with the present data.

5.3.4 *Novelty*

Clark et al. (2000) tested rats on an unrewarded version of the DNMS task. Rats were exposed to two copies of an object during the sample phase. After a delay, the test phase began, in which a copy of the familiar object and a novel object were presented. Rats (and many other animals) preferentially explore a novel object in an environment without being trained to do so, so they used exploration time as their dependent variable. Rats with radiofrequency and ibotenic acid lesions were impaired at longer delays (10 min, 1 hr) relative to controls. At shorter delays, little (1 min) to no (10 sec) impairment was observed. Rats with fornix lesions were not impaired. The delay-dependent effect is similar to that observed by Lee and Kesner (2002; 2003) and likely underpins the extent to which a sustained-activity working memory buffer can provide recognition memory over short delays in the absence of hippocampal function.

In another test of novelty detection, Lee et al. (2005) explored whether DG, CA3 or CA1 lesions affected object replacement or displacement. Figure 5.3 shows the seven sessions of the study. During sessions 5-6, object E was generally not re-explored, even by controls. However, control rats showed substantial re-exploration of object D. All three lesion groups showed de-

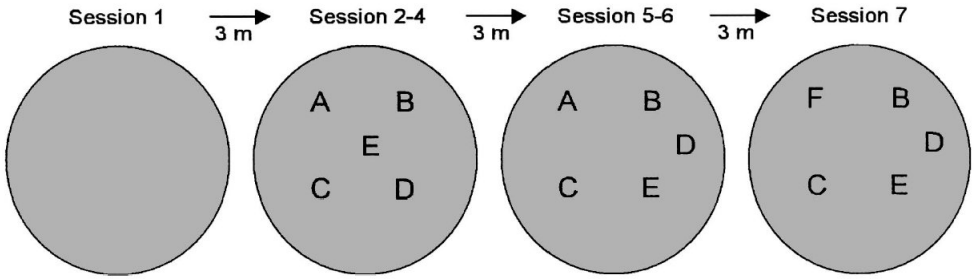

FIGURE 5.3: The seven sessions of the novelty study by Lee et al. (2005). In sessions 5-6, object E was moved to object D's location, and object D was moved to a previously open location. In session 7, object A was replaced with a novel object F. Each session lasted 6 minutes, with a 3-minute inter-session interval. (Reprinted from Lee et al., 2005.)

creased exploration relative to controls, but, whereas the DG and CA3 group differences were quite significantly different, the CA1 group difference was not. Exploration of the novel object F was elevated in all groups, though the 3-minute inter-session interval was perhaps insufficient to expect a significant effect, as seen in Clark et al. (2000).

In contrast to Lee et al. (2005), Rampon et al. (2000) report an impairment in object novelty detection in CA1 NMDA receptor knockout mice. However, pre-exposure to enriched environments improved performance, though not entirely to the level of control mice.

5.4 SUMMARY

Hippocampal lesions result in a complex pattern of impairments. However, converging evidence suggests that the hippocampal lesions cause three classes of deficits. First, as reviewed in Section 5.1, studies of contextual conditioning, extinction and latent inhibition suggest that the hippocampus contributes an encoding of the context in which the CS-US association is made. This is consistent with a wealth of hippocampal physiology studies showing different place cell activity patterns in different environments. Second, Section 5.2 discusses the role of the hippocampus in reversal learning tasks, where separating the task into different reward contingencies, each represented by a different behavioral context, is critical to solving the task. Hippocampal lesions severely disrupt learning reversal learning tasks. Third, Section 5.3 discusses the role of the hippocampus in forming associations involving spatial location or associations across time, especially when these associations must be learned quickly and / or re-learned repeatedly.

While it is difficult to clearly dissociate the roles of DG, CA3 and CA1, a series of subregional studies have demonstrated that some distinctions may be drawn in their functional roles. DG appears critical for the formation of new associations, especially those involving precise spatial locations. In CA3, NMDA-mediated synaptic plasticity is crucial for the rapid formation of novel associations and performance in novel environments, as well as for detecting novel spatial arrangements of landmarks. In addition, CA3 lesions or NMDA inactivation impair pattern completion: familiar environmental contexts are not accurately recalled when environmental cues are removed. While plasticity in CA3 is critical in novel situations, plasticity in CA1 appears necessary for maintaining frequently changing associations across a delay, a function consistent with physiological studies showing that potentiation in CA1 but not CA3 appears to fade by the next day. CA3 and CA1 both appear critical for learning temporal sequences.

This concludes the review of prior experimental studies. The following three chapters present the original research contributions of the thesis.

6

SPACE AND THE DORSAL MEDIAL ENTORHINAL CORTEX

> "There are ever appearing in the world men who, almost as soon as they are born, take a bee-line to the rack of the inquisitor, the axe of the tyrant...."
> – Ralph Waldo Emerson (Society and Solitude, 1870)

An earlier version of this chapter appeared in the Journal of Neuroscience as "A spin-glass model of path integration in rat medial entorhinal cortex" by Mark C. Fuhs and David S. Touretzky (2006, vol. 26, pp. 4266–4276; Copyright © 2006, Society for Neuroscience; reprinted with permission).

6.1 Path Integration and Animal Navigation

Even before the first modern scientific debates about animal path integration (Wallace, 1873a; Darwin, 1873b; Wallace, 1873b; Darwin, 1873a), the image of a bee returning directly to its nest after foraging for pollen engendered its own English language verb: beeline (Oxford English Dictionary, 1989). However, only in roughly the last 30 years has there been strong scientific evidence to support the theory that a wide variety of animals are capable of navigation based on an internal sense of motion (for review, see Section 3.2.)

Early models of path integration, largely in insects, were based on some representation of the distance traveled from a start location (Mittelstaedt, 1962; Müller and Wehner, 1988). Mittelstaedt (1962) proposed a simple model in which the animal's current location was represented in Cartesian coordinates as a 2-vector that represented the current position of the animal relative to its home base. This 2-vector was to be updated based on the distance and bearing in which each step is taken. A more neurally plausible implementation of this theory was developed by Touretzky et al. (1993). Instead of Cartesian coordinates, Müller and Wehner (1988) suggested a model based in polar coordinates using a particular formulation of update rule that reproduced the systematic path integration errors observed in rats.

In all of these early theories, external sensory information played no role in path integration. Given some level of noise within the path integration system, error would therefore be expected to accumulate over time. To make

path integration workable, theories of animal navigation (McNaughton, 1989; O'Keefe, 1989; McNaughton et al., 1991, 1994, 1996; Touretzky and Redish, 1996; Redish and Touretzky, 1997; Samsonovich and McNaughton, 1997) began to explore neurobiological solutions to a problem known within the robotics community as the "simultaneous localization and mapping" (SLM) problem (Smith et al., 1990; Montemerlo et al., 2002). The goal of a navigation system that integrates sensory and idiothetic (self-motion) information is to be able to use sensory information regarding the locations of landmarks to infer one's current location; this is localization. However, in a novel environment, the locations of landmarks are not known, requiring the creation of a correspondence between sensory landmark information and position within the environment; this is mapping.

Two groups have proposed detailed computational theories of rodent navigation that deserve particular mention (Touretzky and Redish, 1996; Redish and Touretzky, 1997, 1998; Samsonovich and McNaughton, 1997). Samsonovich and McNaughton (1997) proposed that place cell maps were represented as 2-dimensional plane attractors within the recurrent CA3 network. In their theory, an activity bump on the surface of the plane represents the position of the animal, and this position can be updated either via learned associations between the bump's position and the current sensorium or based on prewired idiothetic motion information. Their theory required that each plane attractor be created during development, and, during the developmental process, idiothetic input would be organized to produce movement of the bump.

In order to avoid this complicated construction, Redish and Touretzky proposed that the path integrator was to be found outside the hippocampus (Touretzky and Redish, 1996; Redish and Touretzky, 1997, 1998). (Presciently, their theory posited that superficial EC was one of the regions implicated in path integration.) In their construction, the path integrator circuit involved a *single* 2-dimensional plane attractor, prewired to move a bump across the plane based on idiothetic information. This neural path integrator was bidirectionally connected to the hippocampus, and the hippocampus was responsible for associating input from the path integrator with input from the sensorium. Upon entering a familiar environment, pattern completion in the hippocampus resulted in the recall of that environment's map, which in turn reset the activity bump of the path integrator. When entering a new environment, the essentially random location of the activity bump in the path integrator resulted in a highly novel input pattern, leading to a new place cell map.

Both theories propose essentially the same solution to the SLM problem:

gradually expand a spatial map using the noisy path integration information to explore new areas and form new sensory associations; then, return to familiar territory (e.g. a nest) to correct the path integrator for drift. As new areas become familiar, the map can be expanded further without fear of corruption from path integrator drift. A handful of behavioral studies have looked at this so-called "home base" behavior in animals, and they find that rodents return to the same home base location repeatedly during exploration (Eilam and Golani, 1989; Golani et al., 1993), providing the opportunity to reset the path integration system with each return. Note that solving the SLM problem does not alone solve the original behavioral path integration problem. Navigation using a spatial map instead of a simple distance vector requires a separate system to decode the map and create an appropriate homing vector; this system is sometimes referred to as a praxic navigation system.

Recent electrophysiological recording experiments in the dorsocaudal region of rat medial entorhinal cortex (dMEC) strongly suggest that the population of dMEC grid cells could serve as a distributed encoding of a spatial map (for review, see Section 3.2). This chapter presents the work of Fuhs and Touretzky (2006), in which a neural network model is developed that is consistent with dMEC serving as the path integrator within a locale navigation system. The model provides both a cogent explanation of the firing properties of grid cells and a mechanism by which such cells could satisfy the computational requirements for path integration. Redish and Touretzky (1997) posited these to include the following: (1) spatially localized firing fields that are universal across environments; (2) activity patterns updated based on self-motion information; (3) activity patterns reset during reentry into a familiar environment; and (4) population activity patterns coding for position over a large area. What role the dMEC might play in praxic navigation is not addressed here.

It is first shown that hexagonally spaced activity bumps can arise spontaneously on a sheet of neurons in a spin glass-type neural network model (Hopfield, 1982, 1984). Introducing an asymmetric form of the connection matrix and assigning a preferred direction of motion to each cell, with a corresponding velocity-dependent input such as might be provided by the head direction system (Sharp et al., 2001; Wiener and Taube, 2005), allows the bumps to shift in any direction and gives individual units hexagonally repeating firing fields. A collection of grids with different scales and orientations is then shown to allow an efferent structure such as the hippocampus to construct place specific representations covering a substantially larger area than the period of the largest grid. Finally, it is shown that "sensory"

patterns can be superimposed onto the network, modulating the strengths of the firing fields of the cells without disrupting their hexagonal structure.

6.2 Spin Glass Model

To model the formation of a grid of bumps, a network of nonlinear neurons is used, arranged on a two-dimensional sheet. The weight matrix governing their mutual interactions is symmetric, and neuron interactions are constrained to be local: a neuron interacts only with those within a neighborhood around it on the sheet. This type of network is known as a spin glass model, by analogy with statistical mechanics models of frustration in magnetic systems (Hopfield, 1982, 1984). Spin glass models exhibit stable states (attractors) that correspond to local minima of an energy function. They will settle into one of these minima from any nearby starting state. Adding noise helps the system escape from shallow energy minima (frustrated states) and settle into a global minimum energy state. Attractor networks have been widely used to model the hippocampus (Samsonovich and McNaughton, 1997; Redish, 1999; Káli and Dayan, 2000; Touretzky et al., 2005; see also Zhang, 1996).

The construction of the weight matrix was inspired by the observation, originally proven by (Thue, 1892, 1910), that the optimal packing of equally sized circles in a plane is a hexagonal lattice. Similarly, hexagonal activity patterns on a sheet can be explained as the result of competition between each neuron and those neighbors that are within a certain radius around it. Such competition was created by composing weights based on two properties. First, weights are proportional to a periodic function of the distance between units on the sheet; this creates cooperation between units of similar "phases" and competition between units out-of-phase. Second, unit interactions are constrained to be local. The result yields a symmetric weight matrix that produces multiple activity bumps that arrange themselves in a hexagonally periodic lattice attributable to the locality and radial symmetry of the recurrent connections.

6.2.1 *Structure of the network*

Let ξ_i be the membrane voltage of neuron i, and let f_i be its firing rate. We use square root as the nonlinear transfer function that converts membrane voltage to firing rate, with a threshold of 0:

$$f_i = \begin{cases} \sqrt{\xi_i} & \xi_i > 0 \\ 0 & \xi_i \leqslant 0 \end{cases} \tag{6.1}$$

Ermentrout (1994) showed that this transfer function was consistent with that of a conductance-based model with class I membranes, i.e., neurons whose firing rate can vary continuously from 0.

The evolution of the membrane voltage ξ_i over time is governed by a differential equation that includes an integration time constant τ, recurrent connections with weights W_{ij}, a velocity input ν_i, and a Gaussian noise term ϵ:

$$\tau \frac{d\xi_i}{dt} = -\xi_i + \sum_j W_{ij} f_j + \nu_i + \epsilon \tag{6.2}$$

To allow for translation of the hexagonal activity patterns across the sheet, the full weight matrix W_{ij} is composed of symmetric and asymmetric components. The symmetric matrix establishes hexagonal periodicity, whereas the asymmetric matrix provides directional biases to translate the pattern. To calculate W_{ij}, let d_{ij} be the distance between pairs of units on this sheet. Each unit is assigned an integer coordinate pair (x_i, y_i) whose x_i and y_i values range from 1 to N_{diam}; the circular sheet of neurons used in these simulations is 61 units in diameter, 2861 units total. The Euclidean distance between units i and i is then $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. The symmetric matrix for connections between neurons i and j has connection strengths:

$$W_{ij}^{sym} = \alpha_{sym} \gamma_j \Psi(\omega d_{ij}) \tag{6.3}$$

The structure of the connection strengths is principally determined by $\Psi(\omega d_{ij})$, a local, periodic function of the distance between units. This function is shown in Figure 6.1E and derived in the Section 6.3.2. The spatial frequency rescales the function, thus determining the number of bumps that form along one axis of the grid. The term γ_j decreases the projection strengths of neurons near the edge of the sheet, limiting the boundary effects. Figure 6.1A shows the symmetric weight matrix for the central unit in the sheet.

The asymmetric weight matrix W^{asym}, which is purely inhibitory, has a similar form but is offset from the center, as shown in Figure 1B:

$$W_{ij}^{asym} = \begin{cases} \alpha_{asym} \gamma_j \Psi(\omega \delta_{ij}) & \omega \delta_{ij} < \psi_1 \\ 0 & \omega \delta_{ij} \geqslant \psi_1 \end{cases} \tag{6.4}$$

Because the asymmetric weights serve only to translate the activity pattern across the sheet, a small region of inhibition is sufficient. We therefore restricted the weights to include only the portion of up to its first 0 crossing, ψ_1.

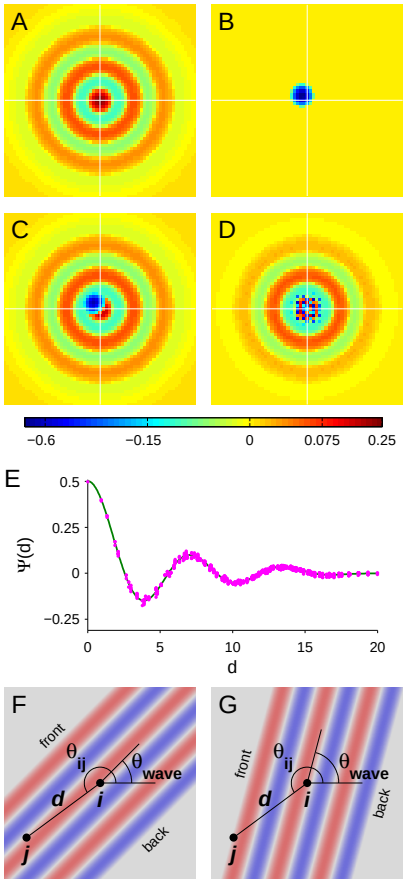

FIGURE 6.1: Recurrent weight matrix W_{ij} contains symmetric and asymmetric components. Shown here are the weights for the central unit in the sheet. A) Symmetric component contains angular rings of excitation. B) Asymmetric component contains a ring of inhibition, offset slightly from the center, opposite the preferred direction ϕ_i of the unit. C) The output weights of a unit (a column of W_{ij}) are the sum of the weights in (A) and (B). D) The input weights for the unit (a row of W_{ij}) are approximately symmetrical; the "noise" reflects the variation in preferred directions of the afferent cells. All weights have been raised to the 0.5 power in these plots to better reveal the structure in both components, which differ in magnitude by a factor of 3. E) The structure of the symmetric component is determined by the wave function Ψ, which depends solely on the distance between units on the sheet. The magenta points indicated values that were learned in a neural network simulation, and the green line indicates the function based on numerical integration that was used to construct the symmetric weight matrix W^{sym}_{ij}. F, G) These diagrams illustrate how the temporal phase of unit j changes as a function of the direction of propagation of the wave packet, θ_{wave}, for a given phase of unit i (see Section 6.3.2). (Reprinted from Fuhs and Touretzky, 2006.)

The offset distance function δ_{ij} is biased in a direction opposite the preferred movement direction of the cell i, offset by approximately one-eighth wavelength:

$$\delta_{ij} = \sqrt{\left(x_i - x_j - \delta_j^x\right)^2 + \left(y_i - y_j - \delta_j^y\right)^2} \tag{6.5}$$

The offsets are $\delta_j^x = \beta/\omega \cos\phi_j$ and $\delta_j^y = \beta/\omega \sin\phi_j$. In theory, the ϕ_i values can be random and uniformly distributed around the circle. However, because of the relatively small number of units in our simulation, we found it advantageous to use just four preferred directions, 90° apart, assigned in an alternating manner across the grid so that, at every point at which four pixels meet, all four preferred directions are represented. This approach ensures a smooth distribution of preferred directions so that the local bump representation is not biased toward motion in any particular direction.

Figure 6.1C shows the sum of the symmetric and asymmetric projection weights of the central unit, $W_{ij} = W_{ij}^{sym} + W_{ij}^{asym}$. Although the efferent weight matrix for each unit is asymmetric, the afferent connections of the unit are approximately symmetric, because it receives projections from units with all possible preferred directions. Figure 6.1D shows the input weights for the central unit of the sheet; the apparent graininess in this plot is attributable to the variation in preferred directions across units. The projection strengths of units near the edge of the sheet were faded to 0 by the γ_j term in Equations 6.3 and 6.4 to ameliorate edge effects. These edge effects were caused by units on the edge receiving unbalanced input: unlike units closer to the center whose inputs were from units in all directions, units near the edge received no input from beyond the edge of the sheet. This imbalance resulted in bumps of activity preferentially forming at the edges of the sheet, constraining the formation thereafter of bumps in the interior. For sheets with non-circular boundaries, this caused the lattice of bumps to form only at orientations that would maximize the number of bumps aligned along the edges of the sheet. For example, a square sheet elicits grid bump lattices with orientations of 0 or 90°. Also, the inter-bump distance along one axis is distorted by a factor of 3 relative to the other, reflecting the stretching of the lattice necessary for bumps to be located along all four edges. The existence of these minimum energy states prevents the network from exhibiting the variety of grid orientations observed by Hafting et al. (2005). Moreover, when the bumps are moved across the sheet, they tend to bunch up along the edge rather than smoothly sliding off, which destroys the hexagonal symmetry. Switching to a toroidal topology would resolve this latter problem but would not eliminate the tendency toward axis alignment. The γ_i term

Spin glass model	
N_{diam}	61
Wave function frequency ω	0.67
Weight parameters:	
Symmetric amplitude α_{sym}	0.5
Asymmetric amplitude α_{asym}	-1.5
First cycle cutoff ψ_1	2.55
Weight fadeout annulus σ_γ	$30/\sqrt[4]{5}$
Asymmetric offset β	1.5
Head direction sharpness σ_{hd}	0.245
Gaussian noise ϵ	0 ± 0.2, mean $\pm$ SD
Integration time constant τ	10
Weight function Ψ	
Training wave frequency κ	$9\pi/31$
Number of waves per packet N_w	3
Tonic firing rate f_{tonic}	1.0
Multiple grids	
Global inhibition coefficients c_2, c_1, c_0	-0.062, 2.103, 1.946

TABLE 6.1: Parameter values

solves both problems. It eliminates distortion caused by edge effects because the progressively weaker weights fail to reinforce the bumps as they move toward the edge, so that they smoothly "fade away" rather than abruptly "fall off." Also, the annular shape eliminates the bias in favor of axis-aligned grids. The γ_j term is defined as follows:

$$\gamma_j = \exp\left[-\left(\frac{1}{\sigma_\gamma}\sqrt{(x_j - r)^2 + (y_j - r)^2}\right)^4\right] \tag{6.6}$$

where $r = \left\lceil \frac{1}{2} N_{diam} \right\rceil$.

The top three panels of Figure 6.2A show the formation of a hexagonal lattice of bumps across the neural sheet. The noise term ϵ in Equation 6.2 serves to break the initial symmetry, and the phase and orientation of the bumps is established within the first few dozen time steps. After 200 time steps, the result is a robust lattice of bumps. Parameter values for this simulation are given in Table 6.1.

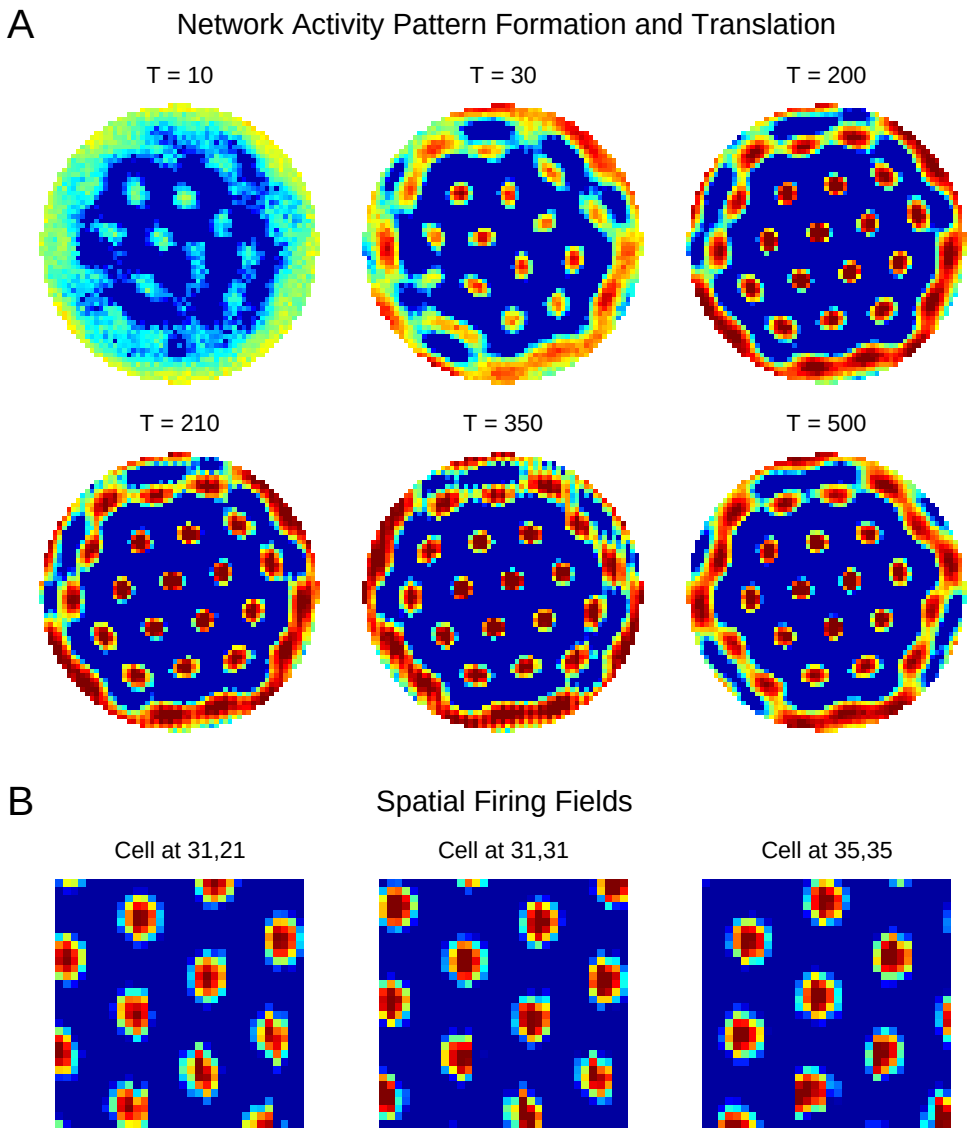

FIGURE 6.2: Formation and translation of the bump grid. A) Starting from an all-zero state, at T = 30, the bump array is somewhat disordered; there is a heptagon in the top left quadrant. By T = 200, a regular hexagonal pattern has been established. The next three panels show translation of the grid to the left. B) Spatial firing fields of three grid cells in a simulated square arena. (Reprinted from Fuhs and Touretzky, 2006.)

6.2.2 *Path integration*

The central hypothesis explored by this model is that path integration can be achieved by moving the bump array around on the sheet. A variety of schemes have been proposed for shifting an attractor bump around a ring (Skaggs et al., 1995; Redish et al., 1996; Zhang, 1996; Goodridge and Touretzky, 2000; Sharp et al., 2001; Hahnloser, 2003) or over a sheet (Samsonovich and McNaughton, 1997; Conklin and Eliasmith, 2005). The model presented here differs from these previous models in that, rather than moving a single bump, it is designed to move multiple bumps simultaneously while enforcing a particular spatial relationship between them.

To move the bumps in direction Φ, the velocity input ν_i to units with preferred direction ϕ_i close to Φ is increased, whereas the input to units with preferred direction nearly opposite Φ is decreased. The asymmetric inhibitory projections, because they are offset from the center of each unit, inhibit one flank of the bumps more than the opposite flank, causing the bumps to shift.

Let s be the animal's current speed, normalized to lie between 0 and 1, and let Φ be its current direction of motion. Then the velocity input to each unit is calculated as follows:

$$\nu_i = 0.5 + 2s\left(\exp\left(-\sin^2\left(\frac{\Phi - \phi_i}{2}\right)/\sigma_{hd}^2\right) - 0.25\right) \tag{6.7}$$

The value of ν_i ranges between 0 and 2; it is 0.5 when the animal's speed is 0. A plot of ν_i as a function of ϕ_i resembles the head direction cell tuning curves seen in postsubiculum (Taube et al., 1990a,b) or anterior dorsal thalamus (Blair and Sharp, 1995; Taube, 1995), which have a width of 90-100°. However, the model also functions properly with broader tuning functions; previous tests used a σ_{hd} value of 0.633, which produced a much wider curve.

The bottom three panels of Figure 6.2A show the bump array being shifted to the right. The hexagonal periodicity of the bumps is preserved during translation. Interestingly, except at the edges of the sheet, the velocity modulation causes little change in the firing rates of the cells, suggesting that the dMEC could perform path integration even if firing rates were only weakly velocity dependent. Figure 6.2B shows the place fields of three units in a square arena. To generate these fields in an efficient and systematic manner, the lattice of bumps was shifted to the right to correspond to leftward travel along the top edge of the arena. Then, at each of 75 positions along the top edge, the network was reset to its previous state at that position, and the activity pattern was shifted upward to correspond to downward travel. As

the activity pattern shifted across the neural sheet, individual cells showed a corresponding pattern of activity across space. The firing fields differ only as to phase; the grid spacing and orientation is constant.

6.2.3 *Resetting the path integrator*

The coupling of path integration and place systems facilitates both the correction of inaccurate path integration information by sensory cues and the representation of place in cue-deprived conditions. The ability of stable cues to reset the place code (Knierim et al., 1995; Jeffery and O'Keefe, 1999) and to influence praxic navigation (Etienne and Jeffery, 2004) is well established. The integration of these two systems suggests that recall of a hippocampal map when, for example, the rat returns to a familiar environment, should lead to reset of the path integrator to agree with the previously constructed place code (Touretzky and Redish, 1996; Redish and Touretzky, 1998).We assume that the projection from CA1 and subiculum to the deep layers of entorhinal cortex can influence the attractor bumps, and that these connections are established through Hebbian learning when the environment is novel.

To simulate recall of a specific activity pattern, we supplied additional external input to each of the cells equal to the value of the firing rate f_i of that cell in that pattern. The input was supplied continuously for 200 time steps via an additional term to Equation 6.2, and it was always successful in resetting the state of the network. However, weaker inputs, typically those below 0.25 f_i, did not always succeed, and, because of the nonlinear properties of the attractor network, the effect of this input was state dependent. If the input pattern was close to the current state, the network settled into a new minimum energy state that mirrored the input. However, if the input was nearly orthogonal to the current state, it had little effect, and the bumps did not move. This suggests that, after brief exposure to an environment or under conditions in which long-term potentiation (LTP) is impaired, a grid is more likely to be reset during the rat's reentry into the environment if the phase to which it is being reset is similar to its phase just before entry.

6.3 *In Vivo* Development of the Weight Matrix

6.3.1 *General constraints*

Although a developmental model of dMEC is beyond the scope of this work, albeit an intriguing avenue for future research, some insights are provided here into the developmental processes potentially underlying the construction of such a matrix.

Patches of cells within dMEC must first be organized into two-dimensional sheets. One way to achieve this is by innervation by an afferent cortical sheet in a localized manner: each dMEC cell receives projections from a small region within the sheet. Such projections would need to be only approximately topographic, i.e., the "logical" position of the dMEC cell in the sheet, as dictated by the source of its afferent projections, does not bear a strict correspondence with its physical location within the brain. Early in development within the visual system, a similar imprecision appears to exist: the chemical markers underlying axonal wiring yield an approximately retinotopic projection that is then improved based on the spatiotemporal coactivity of the neurons (Miller et al., 1989; Wong, 1999). This assumption is important, because anatomically proximate dMEC cells (recorded on the same tetrode) show firing fields at different phases. However, only a small amount of jitter in the relationship between logical and physical position is sufficient to account for the observed heterogeneity of field phase in neighboring neurons. To illustrate this point, consider the second and third cells in Figure 6.2B. Although close to each other on the sheet, they have completely different firing field phases.

To construct the symmetric portion of the weight matrix via Hebbian learning, dMEC cells would require coincident activity between units that varied as a radially symmetric, locally weighted periodic function of their distance. It is important to note that, in the context of Hebbian learning, coincident activity is a statistical construct. On average, the correlation between units at distance d should be proportional to $\Psi\left(\omega d_{ij}\right)$. However, many possible sets of activity patterns could give rise to such a correlation structure, and individual activity patterns need not look anything like the weight matrix itself. For example, the weight function used in these simulations was derived from multi-wave packets of activity propagating across the sheet of dMEC neurons (see Section 6.3.2).

Furthermore, the specific quantitative formulation of used in these simulations is not critical to capturing the activity patterns of the grid cells. Although the used shows two "rings" of excitatory connections, we also tried functions composed of one and three rings (two- and four-wave packets; see Section 6.3.2), both of which produced hexagonal activity patterns. The choice of two rings was based only on our subjective impression that it produced the most accurate path integration. We also tried a cropped version of the 0th-order Bessel function of the first kind in which values beyond the fourth 0 crossing were set to 0; this also produced hexagonal activity patterns. In fact, in early versions of our model, a completely different function was successfully used, one based on the product of a sinusoid with a Gaussian.

Thus, any function approximately of the form that we have described should suffice. The asymmetric portion of the weight matrix is simpler in form and serves only to translate the bumps across the sheet. Models of the head direction system have used various weight matrices to achieve translating of an activity bump around a ring, and our asymmetric weight function is unlikely to be unique. Hahnloser (2003) showed that integration of an activity bump on a ring can be learned by a form of anti-Hebbian learning. A similar form of learning could likely be generalized to a two-dimensional space. Because the asymmetric connectivity used here only creates interactions between neighboring neurons within the same activity bump, the asymmetric connections translate bumps independently of each other, whereas the symmetric connections enforce the hexagonal lattice structure. Thus, teaching a network to translate a single bump across a sheet would be sufficient for it to later translate a lattice of bumps.

6.3.2 *Derivation of the weight function*

During the development of the visual system, it has been known for some time that spontaneous waves of activity propagate across the surface of the retina (for review, see Shatz, 1996; Wong, 1999). These waves have been observed in several mammals, including rodents. They occur one at a time (one wave fully propagates across the retina before another is formed), and the direction of propagation of each wave varies randomly from one to the next. The waves are believed to serve several roles, including the refinement of topographical specificity of axonal projections to the LGN and beyond, as well as ocular specificity by layer in the LGN and by column in visual cortex.

For the purpose of learning the symmetric weight matrix, we consider a variation on this theme: a "wave packet," composed of multiple contiguous waves, propagating across a dMEC sheet, each wave packet traveling in a different random direction. By passing a wave packet across the sheet, connections are learned both between units coactivated by the same wave and between units activated by other waves in the packet. Because the direction of propagation varies from packet to packet, units develop radially symmetric weights: one annulus of excitation and inhibition for each wave in the packet.

We simulated a square sheet of 31×31 dMEC units. Each unit i was assigned integral (x, y) positions on the grid from (15, 15) to (15, 15), which were converted to polar coordinates (r_i, θ_i). The temporal phase of each unit was defined as follows:

$$\zeta_i(t) = t - \kappa r_i \sin(\theta_i - \theta_{\text{wave}}) \tag{6.8}$$

where κ determines the width of each wave in the packet, and θ_{wave} specifies the direction of travel of the wave packet; a different random value of wave was picked for each packet. As time t increased, the temporal phase of each unit advanced. Each wave packet was propagated across the sheet by clamping the firing rate f_i of each unit to a sinusoidal function of its temporal phase:

$$f_i(t) = \begin{cases} f_{tonic} + \sin(\zeta_i(t)) & 0 \leqslant \zeta_i(t) \leqslant 2\pi N_w \\ f_{tonic} & \text{otherwise} \end{cases} \tag{6.9}$$

where N_w is the number of waves in the packet. For values of $\zeta_i(t) < 0$, the wave packet had yet to reach the unit; for values between 0 and $2\pi N_w$, the unit was somewhere within the packet, and, for values greater than $2\pi N_w$, the entire packet had already passed over the unit. Outside the packet, the firing rates of the units were set to a baseline tonic activity, f_{tonic}, of 1. Within the packet, the firing rate of each unit varied from 0 to 2 as a sinusoidal function of its temporal phase. After each packet finished propagating across the network, t was reset, a new random value for θ_{wave} was chosen, and propagation of the next packet began.

A form of Hebbian learning was used to shape connections between dMEC units in the network:

$$\tau_W \frac{dW_{ij}^{learn}}{dt} = \left(f_i - \bar{f}_i\right) f_j - W_{ij}^{learn} \tag{6.10}$$

This learning rule is similar to the Bienenstock-Cooper-Munro (BCM) learning rule (Bienenstock et al., 1982) in that the average postsynaptic activity $\bar{f}_i$ determines the threshold between weight increase and decrease. However, the nonlinearity of the BCM rule was removed, making analytic analysis of the asymptotic weight values tractable. In these simulations, $\bar{f}_i$ was set to f_{tonic}, the true mean activity of each unit. The learning rate τ_W was slowed exponentially from 2,000 to 100,000 during the course of training to increase accuracy by averaging over many packets once the weights were close to their asymptotic values.

One thousand wave packets, each containing three waves, were successively driven across the network. A one-dimensional projection of the learned weights of the center unit is shown in Figure 6.1E (magenta points); weights of other units were similar. The magnitude of the weight is plotted as a function of the distance between the center unit and the others in the network. The three waves in each packet gave rise to a weight profile with a strong center peak and two progressively weaker peaks at larger distances.

As learning progresses, W_{ij}^{learn} converges to the expected value of $\left(f_i - \bar{f}_i\right) f_j$

over all wave packets. This expected value can be determined analytically as a function of the distance d between units i and j. Let ζ_i be the temporal phase of some unit i. Outside the packet, $f_i = \bar{f}_i$, so $dW_{ij}^{learn}/dt = 0$ We therefore only consider values of ζ_i between 0 and $2\pi N_w$. Let ζ_i be the temporal phase of some unit j. Figure 6.1F shows an example of units i and j, both within the wave packet. Let us consider unit i to be at some fixed temporal phase ζ_i within the packet; thus, varying θ_{wave} results in the packet rotating about unit i. The temporal phase of unit j can therefore be expressed as a function of ζ_i, θ_{wave}, and the relative positions of the two units on the sheet:

$$\zeta_j = \zeta_i - \kappa d_{ij} \sin\left(\theta_{ij} - \theta_{wave}\right) \tag{6.11}$$

where θ_{ij} is the angle of unit j relative to unit i. By averaging over the ranges of ζ_i and θ_{wave}, one can calculate the asymptotic values of W_{ij}^{learn} as a function of the distance between units:

$$\Psi(\kappa d) = \frac{1}{2\pi N_w} \int_0^{2\pi N_w} \frac{1}{2\pi} \int_0^{2\pi} \left(f_i - \bar{f}_i\right) f_j \, d\theta_{wave} d\zeta_i \tag{6.12}$$

The outer integral averages over the range of temporal phases of unit i as the packet passes over it. The inner integral averages over the various phases of unit j at a given distance d by averaging over all possible directions of wave propagation.

To evaluate Equation 6.12, we must consider the piecewise nature of the firing rate function (see Equation 6.9). Although we assume by construction that unit i is always within the packet, unit j may or may not be, depending on the value of wave. Figure 6.1G shows the result of increasing θ_{wave} to the point at which unit j is at the front edge of the wave packet. This value of θ_{wave} will be referred to as $\theta_{front,1}$. Rotating the wave packet further results in unit j being outside the wave packet. Rotating it further, unit j rejoins the wave packet at $\theta_{front,2}$. Rotating further, unit j remains within the packet until $\theta_{back,1}$ and is behind the packet until $\theta_{back,2}$. Thus, wave values may be divided into four intervals: $\left[\theta_{back,2}, \theta_{front,1}\right]$ and $\left[\theta_{front,2}, \theta_{back,1}\right]$ when unit jis within the packet; $\left[\theta_{front,1}, \theta_{front,2}\right]$ when the packet has yet to pass over unit j; and $\left[\theta_{back,1}, \theta_{back,2}\right]$ when the packet has finished passing over unit j. This

leads to the following substitution of Equation 6.9 into Equation 6.12:

$$\begin{aligned}\Psi(\kappa d) = \frac{1}{4\pi^2 N_w}\int_0^{2\pi N_w}\Bigg[&\int_{back,2}^{front,1}\left(f_{tonic}+\sin\zeta_j\right)\sin\zeta_i\,d\theta_{wave}\\ &+\int_{back,2}^{front,1}\left(f_{tonic}+\sin\zeta_j\right)\sin\zeta_i\,d\theta_{wave}\\ &+\int_{front,1}^{front,2} f_{tonic}\sin\zeta_i\,d\theta_{wave}\\ &+\int_{back,1}^{back,2} f_{tonic}\sin\zeta_i\,d\theta_{wave}\Bigg]\,d\zeta_i \end{aligned} \tag{6.13}$$

The inner integral in Equation 6.12 has been divided into a sum of integrals in Equation 6.13, each spanning one of the four aforementioned intervals of wave. The first and second inner integrals are equal and so they are collapsed together in Equation 6.14. The third and fourth inner integrals do not contribute to the result because the range of the outer integral is a multiple of two. Simplifying Equation 6.12 and substituting in Equation 6.11 yields the following:

$$\Psi(\kappa d) = \frac{1}{l\pi}\int_0^{l}\int_{back,2}^{front,1}\left(f_{tonic}+\sin\left(\zeta_i+\kappa d\sin\left(\theta_{ij}-\theta_{wave}\right)\right)\right)\sin\zeta_i\,d\theta_{wave}\,d\zeta_i \tag{6.14}$$

where $l = 2\pi N_w$. Although the intervals of θ_{wave} and the value θ_{ij} depend on the position of unit j relative to unit i, the value of the inner integral depends only on the distance between the units because of the rotational symmetry underlying its construction. Hence, Ψ is simply a function of d.

Equation 6.14 was integrated numerically for many values of d, resulting in the green line shown in Figure 6.1E. The results of these integrations were used in the construction of the weight matrices shown in Figure 6.1A-D and in the simulations shown in Figure 6.2.

6.4 Multiple Grids

As Hafting et al. (2005) point out, a single hexagonal grid of the scale observed in dMEC is insufficient for path integration because the bump pattern soon repeats itself, leading to ambiguities in the rat's location. However, multiple grids, with different scales and/or orientations, can encode a much larger space without repetition. More ventrally located cells in dMEC have larger firing fields and proportionately greater spacing between them; the distance

between field peaks varies by at least a factor of two, and cells in different regions of the dMEC show different grid orientations.

To determine to what extent a conjunctive encoding of multiple grids could produce unique positional encodings over a large space, we generated a set of grids of varying spatial frequencies and orientations and computed their activity patterns at all possible positions in environments of varying sizes. We used these patterns to drive a population of 2000 simulated "place cells." The place cell population activity vectors were then correlated across positions to quantify how well the population encoding could distinguish positions over a large space when driven by a set of grid patterns that repeated over smaller spaces.

6.4.1 *Simulating multiple grids*

For computational efficiency, simulated grid cell activities were calculated in closed form. We first defined three basis vectors $\vec{b}_k$, 60° apart:

$$\vec{b}_k = \begin{bmatrix} \cos(k\pi/3) \\ \sin(k\pi/3) \end{bmatrix} \text{ for } 0 \leqslant k \leqslant 2 \tag{6.15}$$

Each grid g was assigned a random orientation θ_g, defining a rotation matrix R_g that was used to rotate the basis vectors:

$$R_g = \begin{bmatrix} \cos\theta_g & -\sin\theta_g \\ \sin\theta_g & \cos\theta_g \end{bmatrix} \tag{6.16}$$

Each grid was also assigned a unique random spatial frequency ω_g between 1 and 2. The grid was represented by an 8×8 array of units. Each unit i had a phase offset vector p_i that represented its position within the grid; values ranged from $(0, 0)$ to $(4\pi\sqrt{3}, 4\pi\sqrt{3})$. Three amplitude functions $z_{g,i,k}$ were calculated for each grid unit as a function of the environment location, $\vec{x}$:

$$z_{g,i,k}(\vec{x}) = R_g \vec{b}_k (\omega_g \vec{x} + p_i) \tag{6.17}$$

The variable $\vec{x}$ ranged over a square space with side length L times the period of the lowest frequency grid, i.e., its values ranged from $(0, 0)$ to $(4\pi\sqrt{3}L, 4\pi\sqrt{3}L)$; values of L up to 10 were simulated. The firing rates of the grid cells, which were hexagonally periodic functions, were calculated based on the sum of the three amplitude functions:

$$f_{g,i}(\vec{x}) = \left[\cos z_{g,i,0}(\vec{x}) + \cos z_{g,i,1}(\vec{x}) + \cos z_{g,i,2}(\vec{x})\right]_+ \tag{6.18}$$

where $[\ldots]_+$ is the semi-linear threshold function that maps negative values to 0.

To construct a place cell representation based on the conjunctive activities of the grids, each of the $n = 2000$ place cells received connections from these grid cells, at most one grid cell per frequency. The weights W were all of unit strength, and the phases of grid cells selected to project to each place cell were chosen at random. The activity of each place cell was calculated as $f_j = \left[\xi_j - \xi_{98\%}\right]_+$, where

$$\xi_j = \sum_{g,i} W_{j,g,i} f_{g,i} \tag{6.19}$$

and $\xi_{98\%}$ is a global inhibition term.

For the place representation to be sparsely coded, as seen in the hippocampus, a simple feed-forward inhibition mechanism was used to reduce the number of active place cells at any one position to approximately 2% of the population. The global inhibition term, $\xi_{98\%}$, was calculated as a function of the average weighted input to the place cell population:

$$\bar{\xi} = \frac{1}{n} \sum_j \xi_j \tag{6.20}$$

The 98th percentile of ξ_j values was then well approximated by a quadratic function of $\bar{\xi}$: $\xi_{98\%} = c_2 \bar{\xi}^2 + c_1 \bar{\xi} + c_0$, with coefficients as shown in Table 6.1. $\xi_{98\%}$ is monotonically increasing and slightly sublinear over the range of values in these simulations.

6.4.2 *Similarity of population codes within an arena*

Grid and place cell activities were calculated over square spaces whose side length, L, was up to 10 times the period of the lowest spatial frequency grid. At each position in the space, the place cell population vector was compared with the vectors at every other position, calculating a correlation coefficient between each pair of vectors. To measure how well the place cells uniquely represented each of the positions in the space, each position was associated with a second position (at least one highest-frequency period away) whose population vector was the most similar to the first. Correlation coefficients between pairs of "most similar positions," r_{max}, indicated how well the population represented the space. Values of r_{max} above 0.5 were considered to indicate inadequate distinction between positions. Although this threshold is somewhat arbitrary, other thresholds yielded the same pattern of results.

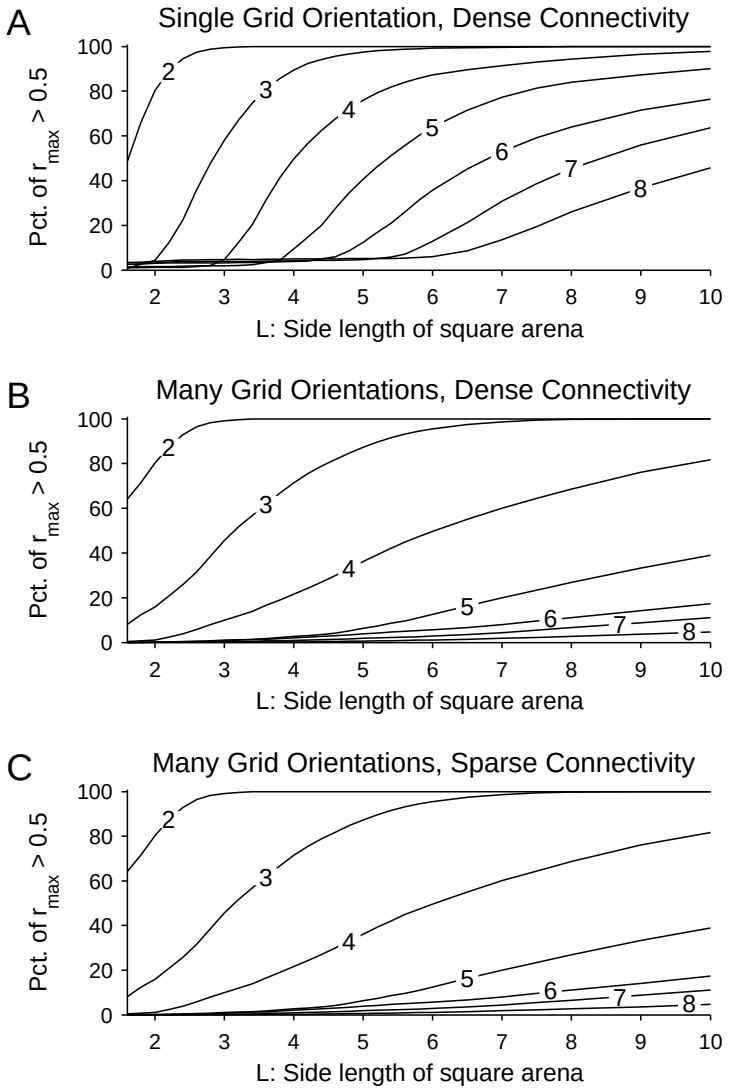

FIGURE 6.3: Multiple grids yield unique place codes. A) Correlations between place cell population vectors decrease as the number of grids increases. Each line shows how the percentage of positions for which r_{max} exceeds 0.5 varies as the size of the space is increased. The number superimposed on each line indicates the number of grids used as input to the place cells. B) Correlations are further reduced when the grid orientations vary. C) When grid cells project sparsely to the place fields (two grids per cell), place coding does not deteriorate. (Reprinted from Fuhs and Touretzky, 2006.)

Figure 6.3A shows how the percentage of positions for which $r_{max} > 0.5$ varies with the size of the space and the numbers of grids. For this experiment, the orientations of the grids were the same. With only two grids, repetition within even smaller spaces was inevitable; population vectors at each position quickly began to reoccur as the size of the space increased. However, as more grids were used, progressively larger spaces could be represented without similar population vectors occurring at multiple positions.

When grid orientations were allowed to vary, additional decorrelation between population vectors was observed. Figure 6.3B shows the distribution of r_{max} when grid orientations were randomized. Clearly, using grids at varying orientations provides a substantial benefit in representing a large number of locations.

In the simulations above, each place cell received a projection from a grid cell in every grid. To explore the effects of a sparser connection structure, the projection density was reduced to two randomly chosen grids per place cell. Figure 6.3C shows the effect of this sparsity on the efficacy of the place code. The results are nearly indistinguishable from the dense connectivity results in Figure 6.3B.

Place cells tended to have multiple firing fields in the larger environments, and, when many grids were sampled, the firing fields were randomly distributed. This would be expected in any sparse code in which individual units are reused as part of a larger set of patterns, and this has been observed experimentally (Gerrard et al., 2001). When only a few grids were sampled, the place cells displayed a somewhat hexagonal arrangement of firing patterns, but this regularity was still limited to localized portions of the space. In no case did the place fields show the hexagonal regularity of the grid cells across the entire space.

6.4.3 *Partial remapping*

With multiple independent grids, resetting the path integrator would require resetting all of the grids. If only some grids were reset, the result should be what (Muller et al., 1991) called "partial remapping." To quantify this effect, we calculated place cell population vectors for each position in a space four times the period of the lowest-frequency grid. The population vectors were calculated twice: during the second run, the phases of N_{reset} of the grids were reset to their previous value, whereas the remaining grids were set to random phase values unrelated to their values during the first run. The simulations used eight grids of varying spatial frequencies and orientations, and place cells received either sparse (three grids) or dense (all eight grids) projections from the grid cell population. The population vector correlations

between the first and second runs are shown in Figure 6.4A. As the number of reset grids increased, the similarity of the place code representation also increased. The number of grids sampled (three vs all eight) had a negligible effect.

6.4.4 *Discordant remapping*

Several studies have reported that place cells show discordant remapping during "double rotations," when two sets of cues are rotated in equal and opposite directions (for review, see Section 4.1.5). Specifically, some place fields rotated with the local track cues, other fields rotated with distal room cues, and others remapped. Moreover, the activity patterns in CA3 and CA1 differed. Lee et al. (2004b) found that, whereas CA3 fields rotated predominantly with one set of cues, CA1 fields rotated in approximately equal numbers with each set of cues, suggesting that the CA3 and CA1 representations may be formed independently.

The discordant remapping must initially be driven by sensory information. However, early experiments in darkness suggest that, once formed, some hippocampal place fields in CA1 persist without the support of visual sensory information (Quirk et al., 1990). Is it possible that discordant remapping could persist in darkness as well?

With multiple independent grids, some grid orientations could align with respect to the local cues, whereas others could align with respect to the distal cues. To understand the effects of such grid discordance on a place cell population, we simulated the activity of place cells under two conditions. In the "standard condition," grid cell networks were initialized to random orientations. In the "rotated condition," the grid networks were divided into two sets, and the orientations in one set were rotated by 90° with respect to those of the other set. Place field correlations were calculated to determine whether the fields in the rotated condition were similar to those in the standard condition (modulo rotation) or whether they bore little resemblance to the standard configuration field at either rotation angle, i.e., they remapped.

Figure 6.4B shows the results. When the number of grids in each set is equal (four grids each) and place cells are sparsely connected, the distribution of place cell responses to the double rotation is consistent with those found by Lee et al. (2004b) in CA1. As one set grows larger than the other, place cells tend to show fields consistent with the orientation of that set. In this figure, place field correlations less than 0.5 were taken to indicate remapping. Other values would change the percentages somewhat; however, values around 0.5 are typical in physiology studies.

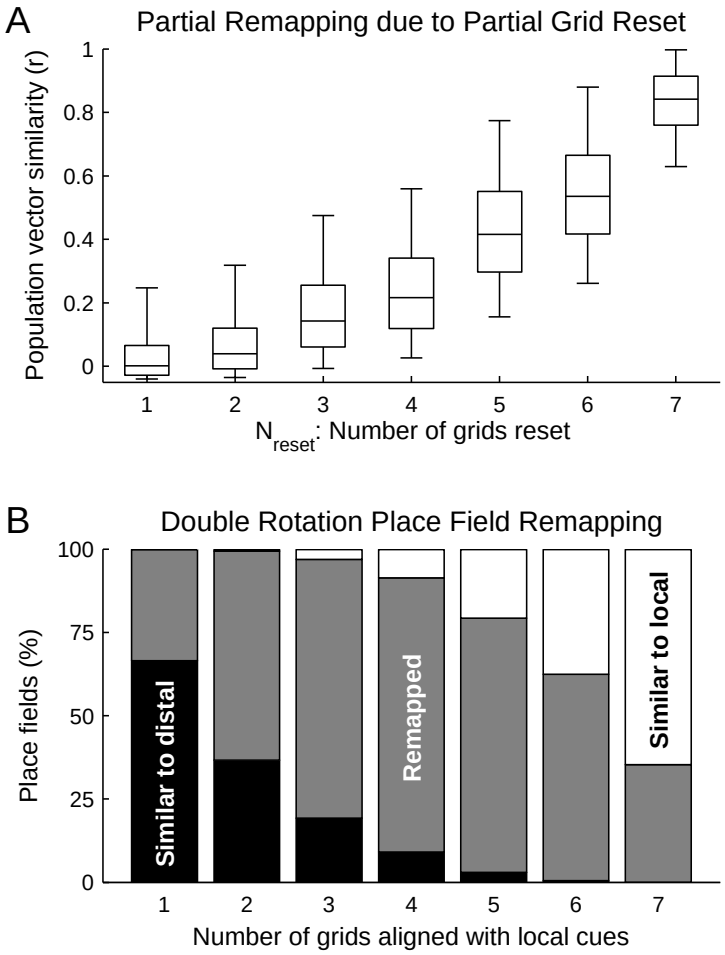

FIGURE 6.4: Resetting only a subset of the grids during recall of an environment results in partial remapping. A) Similarity to the original place code increases with the number of grids that are reset. The joint distribution over 20 runs, each with different grid spacings and orientations, is shown for each value of N_{reset}. Box plot tails indicate the central 95% of the joint distribution. B) Double-rotation experiments can produce a range of remapping effects depending on the number of grids aligned with each set of cues. Distributions were based on the average over 20 runs, each with different grid spacings and orientations. (Reprinted from Fuhs and Touretzky, 2006.)

6.5 Sensory Modulation of Activity Patterns

Hafting et al. (2005) found that dMEC cells exhibited different peak firing rates in their various spatial firing fields, and these variations were reproducible during a second session in the same environment. We interpret this in the following way: dMEC receives sensory or hippocampal input that varies depending on the animal's location, and this secondary input exerts a modulatory influence on the activity levels of dMEC cells without disrupting the hexagonal activity pattern.

To test this interpretation, we explored whether such a secondary input could modulate the rates of different neurons in our model in a reproducible way. A set of 100 random external input patterns was generated, abstractly representing the input from sensory cortex or hippocampus. A pattern of hexagonal firing fields was allowed to form in the network. After the firing fields stabilized, input patterns were applied to the network in the following way: (1) apply random noise for 10 time steps, (2) apply one of the input patterns for 10 time steps, (3) take a snapshot of the vector of firing rates, and repeat. Thus, every 20 time steps, a new pattern was presented for 10 time steps. The purpose of the intervening noise was to eliminate temporal correlations between successive patterns attributable to hysteresis in the network. The entire sequence of 100 patterns was presented twice.

The mean of the firing rates across patterns was subtracted from each snapshot, so that the snapshots reflect input-modulated changes from the average hexagonal grid pattern. The snapshots from the first and second presentations were then cross-correlated (Pearson's r).

Figure 6.5A shows the correlation coefficients across patterns, comparing the first presentation with the second. The dark diagonal suggests that the second presentation evokes very similar network activity to the first presentation and is unrelated to that of the other patterns. Figure 6.5B (top) is a histogram of the correlation coefficients between the first and second presentations of the same pattern (i.e., the diagonal elements in Fig. 6.5A); all correlations are tightly clustered above 0.8. The middle histogram shows correlation coefficients between the first presentation of pattern i and the second presentation of another pattern, that which evoked the closest network activity to that of pattern i. In all cases, the next closest matching pattern has a dramatically lower correlation coefficient. The bottom histogram shows all off-diagonal correlation values from the correlation coefficient matrix shown in Figure 6.5A.

To assess the ability of individual hippocampal cells with limited fan-in to discriminate between these sensory patterns imposed on the same array of

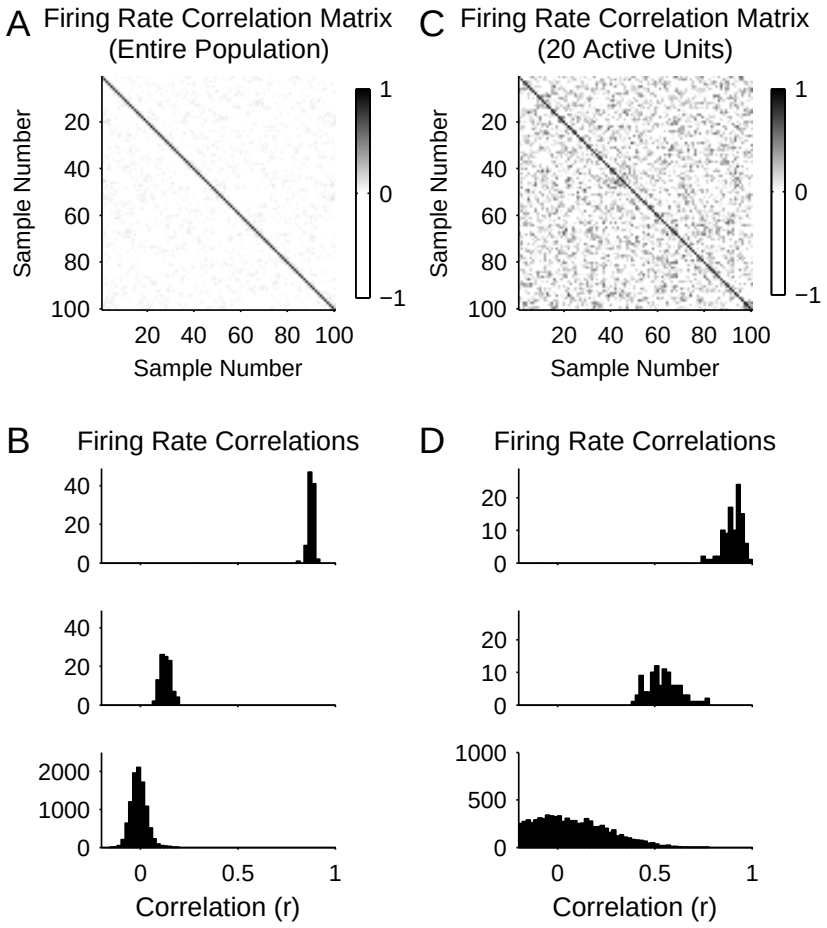

FIGURE 6.5: Sensory modulation of grid cell activity. A) One hundred random input patterns produce population activity vectors that are clearly distinguishable from each other. B) Histogram of correlation coefficients between pattern i and the second presentation of the same pattern (top), the second presentation of the next closest matching pattern (middle), and all patterns other than i (bottom). C, D) Same as A and B except the results were recalculated using a 20 unit subset of the place cell population. (Reprinted from Fuhs and Touretzky, 2006.)

activity bumps, we recalculated the correlations using a randomly selected subset of 20 units with nonzero activity instead of the full population. Pattern discrimination is still feasible, and the results are shown in Figure 6.5C-D.

6.6 Discussion

6.6.1 *Summary of properties of dMEC cells*

The following points summarize the account of our model of the extant data on grid cells in dMEC and present predictions from the model for future studies.

1. Dorsal MEC cells exhibit multiple spatial firing fields arranged in a hexagonal lattice. These are expressed immediately in a novel environment, show a constant phase relationship between cells across environments, and do not change scale with the size of the arena.

Explanation: Local, radially symmetric on-center / off-surround connections produce an attractor network whose stable states are a two-dimensional toroidal manifold of hexagonally periodic activity patterns. (Note that, whereas the state space manifold is toroidal, the network connectivity is a simple sheet.) Different neurons in the same network are most active at different phases of the periodic pattern, but the firing fields of all units will show the same orientation and spacing.

The model occasionally settles into stable states with local irregularities in the network activity pattern, e.g., a pentagonal or heptagonal bump cluster. Such irregularities are caused by frustration among the arrangement of bumps as they form, not properties of specific neurons in the system. From this, an important prediction follows: when such irregularities are observed, they should be environment specific. Although not explicitly addressed by Hafting et al. (2005), the cell shown in their Figure 6 appears to bear this out: the field of the cell looks hexagonal in the familiar environment and heptagonal in the novel environment.

2. The size and spacing of fields is similar among neighboring neurons but varies systematically along the dorsoventral axis.

Explanation: As Hafting et al. (2005) point out, neighboring neurons in dMEC are likely coupled together as part of one of many local neural networks. Simulations show that multiple independent networks provide a basis for constructing a representation of space over a substantially larger domain than the periodicity of any one network.

The size and spacing of fields is determined by two factors in our model. The first is the parameter, which determines the spatial frequency of bumps on the neural sheet. The second is the strength of velocity, which determines how far the bumps are translated across the sheet per unit of travel in the environment. Increasing the rate of bump shift relative to physical movement shrinks the size and spacing of nodes in the field of a cell. Decreasing this ratio expands the node size and spacing. One should be able to determine whether velocity modulation differences account for field size variations: cells with tightly packed fields should show stronger changes in firing rate as a function of the animal's velocity than cells with larger, more dispersed fields. Evidence of this has been observed downstream in the hippocampus (Maurer et al., 2005).

Amaral et al. (1990) estimated that there were 200,000 layer II entorhinal cortical cells that projected to the dentate gyrus (DG). If the number of cells in dMEC is approximately one quarter of that, then 17 networks of the size simulated here could be embedded within layer II. However, grid cells have also been found in layers III, V, and VI of dMEC Sargolini et al. (2006), so the number of grid networks, or the number of cells per network, could be substantially higher than our model assumes.

3. The orientation of fields is similar among neighboring neurons but varies over the dorsoventral axis and is not preserved across environments. (But see Section 6.6.2.)

Explanation: Radially symmetric connections allow each network to settle into a bump array with any orientation. Simulations show that this heterogeneity of orientations provides a superior representation for constructing a place representation over a large space.

4. Across multiple sessions in the same environment, phase and orientation of a firing field of a cell, and firing rate differences between its peaks, are all reproducible, whereas in different environments, they are different.

Explanation: The ability to restore the state of a grid cell network to an earlier state is similar to the previously studied recall problem in single-bump attractor networks. Without such coherent input, the network settles into a random stable state. Failing to reset some of the networks or resetting them discordantly may reinforce partial or discordant remapping in the hippocampus. Variations in peak firing rates among individual nodes of activity can be caused by afferent input without disrupting the hexagonal firing patterns in the network, and these inputs can be distinguished from one another based on the firing rates of the network.

5. In darkness, spacing, mean firing rate, and spatial information did not change, but the fields shifted somewhat relative to the lighted condition.

Explanation: Relative phases and orientations of different grid fields should persist in the dark because they are a product of attractor dynamics, not sensory input. However, field drift attributable to path integration error will accumulate in the absence of sensory information that can maintain alignment of the fields with external landmarks.

6. Grid cells show weak velocity modulation.

Explanation: In order to perform path integration, the model critically relies on a velocity signal, so the existence of a similar signal *in vivo* was a strong prediction of the model. However, the attractor dynamics of a network tend to dampen the impact of the velocity input on the firing rate of a cell, predicting that grid cells will show some velocity modulation, but it will be small relative to the range of firing rates of the cell. It has since been confirmed that cells in layers III, V and VI are modulated both by the animal's speed (weakly) and direction (to varying degrees) (Sargolini et al., 2006; see Section 3.2.2).

6.6.2 *Neural architecture of grid cell networks*

In the idealized architecture presented here, all units use the same weight pattern, rotated to reflect their preferred direction, and all cells show velocity tuning. There are asymmetric connections between cells of similar directional bias, with progressively more symmetric connections between cells of differing directional biases. The Sargolini et al. (2006) results suggest an alternative architecture that divides the roles of the two connection types. In this scenario, two populations of dMEC cells would interact. The symmetrically weighted cells would enforce the hexagonal periodicity of firing patterns, whereas the asymmetric cells would shift these patterns during movement. Only the latter population should show velocity modulation.

A recent computational model of dMEC by McNaughton et al. (2006) explores this architecture. However, there are several other more fundamental differences as well. The McNaughton et al. (2006) model is constructed using a set of hexagonal activity patterns to train a hexagonally toroidal network. This results in a model that differs from the one presented here in three as-yet untested ways:

1. Irregularities in the lattice structure of activity nodes. While the model presented here is able to show irregular lattice structures (e.g. heptagons, pentagons), the hexagonal periodicity is hard-wired into the

network topology in the McNaughton et al. (2006) model. Thus, future data substantiating the regular occurrence of lattice irregularities would rule out the McNaughton et al. (2006) model, and, more generally, lend support to the notion, as developed here, that the hexagonal periodicity emerges out of a stability constraint on what patterns are stable within the network, as opposed to a topological constraint that mandates hexagonal periodicity.

2. Fixed orientation of the lattice. In contrast to Hafting et al. (2005), McNaughton et al. (2006) assert that it is not currently known whether grid orientations vary within one animal. That the orientations of grid cell fields vary can be accounted for by differences in the head direction system, since the direction of shift on the grid cell network relative to the direction of travel in space is a free parameter. In the model presented here, the same network is capable of adopting multiple orientations. Thus, this model predicts that it should be possible to see relative changes in the orientations of two simultaneously recorded grid cell fields if they are recorded from different grid cell networks (i.e. by varying the dorsoventral position of the two recording electrodes).

3. Same lattice orientation among all grid networks. The training regimen used by McNaughton et al. (2006) involves afferent hexagonally periodic activity patterns at a fixed orientation. Thus, in contrast to the present model, they predict that the lattice orientation of different grid cell networks will be the same.

Given the significant architectural differences between the two models, future experimental work to study these issues would provide great insight into the neural architectural of grid cell networks.

6.6.3 *dMEC impact on hippocampal remapping*

Redish and Touretzky (1997) argued that navigation was likely accomplished using a set of interdependent representations, two of which were the hippocampal "place" representation, which provided an environment-specific place code, and a universal coordinate map to support path integration based on vestibular and motor information. In their theory, these two representations were tightly coupled: in cue-deficient conditions, the hippocampal place representation could be updated based on idiothetic movement information, whereas in cue-rich conditions, cumulative error in the PI could be corrected by efferent projections from the place code, whose representation is strongly driven by sensory information. Thus, in a novel environment, a bidirectional

mapping is learned between PI coordinates and hippocampal representations of place. When returning to a familiar environment, it follows that the PI system would have to be reset to realign it with the animal?s present location. Failure to reset the system would result in complete hippocampal remapping (Redish and Touretzky, 1998).

Recent data has shed light on the possibility that the hippocampal remapping phenomenon may be mediated by a set of distinct (if interdependent) processes. Specifically, there appears to be evidence that, in novel environments, CA1 and CA3 representations form independently (Leutgeb et al., 2004; Lee et al., 2004a; Wilson et al., 2005). Place fields in CA1 are observed even when the Schaffer collateral projection from CA3 is severed (Brun et al., 2002). This suggests that immediate CA1 remapping is mediated by an extrahippocampal process such as the failure to reset the grid cell networks in dMEC.

In contrast, the gradual remapping in CA1 that manifests over many days of exposure is likely to be intrahippocampally mediated. When visual cues were reshaped (Lever et al., 2002) or repositioned either with respect to each other (Shapiro et al., 1997; Jeffery, 2000) or in conflict with vestibular information (Sharp et al., 1995), the number of CA1 cells that remapped during each successive experimental manipulation progressively increased. Preliminary evidence from recordings in DG shows heightened sensitivity to arena shape changes in DG relative to CA3 (Leutgeb et al., 2005a), suggesting that, with experience, this remapping is propagated along the trisynaptic pathway from DG to CA1. Such DG sensitivity is consistent with behavioral data showing that DG lesions impair the rats' ability to distinguish two locations only if they are nearby (Gilbert et al., 2001).

Together, these data suggest that hippocampal remapping attributable to gross environmental changes (or deficient LTP) may be caused by a failure to reset the dMEC grid cells, whereas subtle changes to spatial relationships between landmarks may be detected within the hippocampus and gradually propagated into CA1 in an experience-dependent manner.

7

CONTEXT AND THE HIPPOCAMPUS

"Context is an elusive concept."
– Kubie and Ranck (1983)

7.1 Introduction

Several theories have posited that the hippocampus is responsible for providing a representation of context, and that this representation underpins an animal's performance of a variety of tasks sensitive to hippocampal lesions (Hirsh, 1974; Nadel and Willner, 1980; Jarrard, 1993; Levy, 1996; Wallenstein et al., 1998; Redish, 1999, 2001; Hasselmo and Eichenbaum, 2005). These tasks include spatial navigation, sequence learning, and hippocampal lesion-sensitive conditioning paradigms such as reversal learning (for review, see Chapter 5). Models have typically focused on one or two of these domains (Levy, 1989; Schmajuk and DiCarlo, 1992; Gluck and Myers, 1993; Levy, 1996; Samsonovich and McNaughton, 1997; Wallenstein and Hasselmo, 1998; Redish and Touretzky, 1998; Doboli et al., 2000; Hasselmo et al., 2002; Hasselmo and Eichenbaum, 2005), developing mechanisms of context learning that are well tailored to the domain of study but do not generalize well across domains.

Moreover, empirical data from both lesion and physiology studies has called into question the viability of existing models. Attractor models of hippocampal place cell remapping (Samsonovich and McNaughton, 1997; Redish and Touretzky, 1998; Tsodyks, 1999; Doboli et al., 2000) cannot account for the gradual separation of maps between similar environments (Jeffery, 2000; Lever et al., 2002). Hippocampal models addressing reversal learning (Schmajuk and DiCarlo, 1992; Gluck and Myers, 1993; Hasselmo et al., 2002) fail to capture the experimental observation that, after a series of discrimination reversals, rats can learn to reverse behavior after a single trial (Buytendijk, 1930; Dufort et al., 1954; Pubols, 1962). Models of sequence learning (Levy, 1989, 1996; Wallenstein and Hasselmo, 1998) are challenged by the failure to find sequence-dependent differences in hippocampal place cell activity in some studies (Lenck-Santini et al., 2001; Hölscher et al., 2004; Bower et al., 2005).

While these previous modeling efforts have offered neural mechanisms

to explain *how* the hippocampus remaps, the present work focuses on a related but distinct question: *why* does the hippocampus remap? More specifically, if remapping serves to create a representation of context, then understanding remapping across problem domains requires a general and concrete definition of what a context is. We propose that context learning may be formalized as decomposing a non-stationary world of hippocampal input patterns (and the sensory input and behaviors they represent) into multiple domains, or contexts, within which the distribution of these input patterns is stationary. What critically defines a context is therefore not a particular class of stimuli (e.g. background cues) or behaviors (e.g. random vs. directed foraging), but a set of time windows within which the statistical structure of sensory experiences and behaviors is stable.

Such a definition of context advances its role in prediction: if recent experiences suggest that the present context is C, then other prior knowledge about context C should also be applicable for the foreseeable future (i.e., until the context changes). The predictive power of a context is determined both by the temporal duration of the context and the (systematic) variability within a context. Contexts that generally last only a short duration fail to be useful in prediction because of the high probability that a different context with different contingencies will soon become active. Contexts that encompass a highly variable set of sensory experiences and behaviors are also poor predictors insofar as they do not delineate among the many possibilities within the context.

As a concrete example, consider the serial reversal learning task in which animals must make a choice between two alternatives (levers, maze arms, etc.). During odd blocks of trials, one choice is rewarded; during even blocks, the other is rewarded. As training progresses, the two reward contingencies may be codified as two different contexts, each with a static reward structure in which only choice one is rewarded (context 1) or only choice two is rewarded (context 2). Thus, when a particular choice is rewarded (or is not rewarded), it suggests which of the two contexts is active, thereby predicting that that choice will be rewarded on subsequent trials as well. If, however, both reward contingencies were grouped together in the same context, the identity of the active context would not be useful in predicting an optimal behavioral strategy. By contrast, in a random foraging paradigm, the random scattering of food within the environment ensures that the reward location is unpredictable. Dissecting a random foraging session into a large number of contexts, each representing a distinct location at which food was found, would result in a complex contextual representation with no predictive value.

Inferring the most informative context model is therefore the central prob-

lem of interest. In this article, we formulate context learning as a model selection problem: into how many contexts should the world be divided? We develop a statistical framework for context learning based on the hypothesis that the degree of remapping in hippocampus reflects two independent factors: the degree of similarity between contexts and the animal's confidence that the two contexts are in fact distinct. The framework shows how contextual representations should evolve over time as the animal progresses through a training regimen, whether that regimen involves different environments, reward contingencies, or sequences. Within the framework, model selection is biased to prefer contexts that are active for longer periods of time, a key constraint that explains the development of hippocampal contextual representations in spatial and reversal learning paradigms and justifies their absence in overlapping sequence learning. The framework also distinguishes between context learning (inferring the context model), which may be a gradual process over many blocks of trials, and context selection (inferring the current context given a context model), which should be a more abrupt phenomenon. This distinction is critical to understanding serial reversal learning: substantial training is required for rats to achieve single-trial reversals, but, once trained, single trial reversals can be realized as a context switch. Several testable predictions are made to motivate future experimental work.

7.2 A Statistical Framework for Context Learning

7.2.1 *Overview*

Hidden Markov models and hippocampal activity patterns. We model context learning as a process in which the hippocampus constructs a generative model of its inputs. Within our statistical framework, which is based on hidden Markov models (HMMs), generative models are composed of states and contexts. An HMM with a given number of states is parameterized by a distribution of expected values for each state and a transition matrix, which expresses the probability of transition from one state to another. Each state represents a conjunctive encoding of the hippocampal input and may be thought of as a particular hippocampal activity pattern. Different states are therefore used to represent different positions within an environment or different stages in a task.

A context is simply a group of states. States are grouped together into contexts such that, while state transitions are frequent, transitions between states in different contexts are rare. Transition probabilities between states

within the same context are not constrained, and each is individually parameterized, but the transition probability between states in different contexts is fixed at a small value. If, for a particular model, context switches occur frequently, the goodness of fit of the model will therefore be judged to be low. Thus, the fixed inter-context transition probability encourages context switches to be rare, or, inversely, context durations to be long.

What is the relationship between states, contexts, and hippocampal activity patterns? In an HMM, states are "identifiable" in the sense that there is an abstract label for each state that is independent of its particular parameters. This permits an HMM to contain multiple distinct states with the same parameters. This state identifiability is a theoretical construct that we view as unreasonable to apply to the hippocampus. Rather, we argue for weaker identifiability constraints:

- States within a context are distinguished only by their expected observations. Thus, within a context, different hippocampal activity patterns can only be observed when different input patterns are presented.
- Contexts are identifiable, i.e. the hippocampus forms a latent representation of context. This implies that the same input pattern presented in two different contexts will result in different hippocampal activity patterns.

These constraints suggest that an observed hippocampal activity pattern (i.e. the currently active state) is determined by two factors: afferent activity and the currently active context. Intuitively, when the same input pattern is represented by multiple hippocampal codes, each code must be part of a different context. These constraints are closely related to the behavior of latent attractors (Doboli et al., 2000), where network activity is determined both by the input pattern driving the network and the network's current attractor basin (which represents the current context).

In order to handle input pattern noise, each state is parameterized by a distribution of input patterns. Thus, determining which state is active is a (minor) statistical inference problem, requiring the determination of the state under which the input pattern is most likely. This is qualitatively similar to pattern completion in associative memory models, where the network activity pattern resulting from a noisy input is made more similar to the best matching previously stored pattern. Several theories have implicated the hippocampus in conjunctive encoding (e.g. Rudy and Sutherland, 1995) and associative memory function (Marr, 1971; O'Reilly and McClelland, 1994; Rolls, 1996), and the present formulation is not incompatible with

these notions. However, the present work focuses not on the utility of the individual states but on their contextual organization.

Adopting new models. In the hippocampus, new experimental conditions are observed to cause the creation of new place cell maps. In the model, a new experimental condition is represented by augmenting the current HMM with new states in a new model context. Given that the input patterns are noisy, under what conditions should a new context be added to the model? A larger, more complex model, one with more contexts, will inevitably provide a better fit of the input patterns. However, one should distinguish the extent to which the model is fitting the statistical structure of the inputs from the extent it is better fitting the noise.

In a Bayesian setting, a model should be adopted when its posterior likelihood is higher than competing models. This posterior probability accounts for the complexity of the model by averaging the goodness of fit of the model over the prior uncertainty in the model's parameterization. Larger models have more parameter values that must be specified, so the prior probability of any particular parameterization is lower. Larger models are therefore implicitly penalized for the size of their parameter space.

In clear cases, such as when entering a completely new and different environment, the sensory input would differ strongly from what is expected by any state under the "current" model. A larger model, augmented with a group of states for the new environment, would therefore be immediately justifiable. In more subtle cases, such as when a new environment is similar to a familiar environment, the posterior likelihood of the larger model would rise more gradually as the animal gains more experience in each environment. This increasing experience offsets the penalty caused by the larger model's added complexity.

A distributed neural representation of context allows multiple models to be expressed simultaneously. Returning to the two-environment example, if a larger model were only weakly favored over a smaller one, then it is possible for some cells to represent the larger model, distinguishing between the two environments, while other cells represent the smaller model, showing the same firing patterns in both. We interpret gradual increases in the number of cells that remap between contexts (gradual remapping) as observed by Lever et al. (2002) and others as a reflection of the relative degree of acceptance of each model: the increasing degree of remapping between two environments reflects the increased statistical likelihood of a model that represents them as two separate contexts.

Independent and dependent contexts. Consider again the case of two similar environments, i.e. where the input patterns at corresponding positions in each environment are similar. These two environments could be represented by a one-context HMM, where each of the p states represents a corresponding position in the two environments. Each state would therefore be optimally parameterized when tuned to the distribution of input patterns observed at that position in either environment. The two environments could also be represented by a two-context HMM, where two different groups of states, each of size p, represent the two environments. Such a two-context model would have double the number of parameters, and the Schwarz criterion (Schwarz, 1978) suggests that this corresponds to a quadratic increase in complexity.

Penalized for such high complexity, preliminary simulations showed that, compared to a one-context model, a two-context model would be considered astronomically improbable until after substantial experience in both environments. This would predict that, without substantial experience, rats should never show any remapping between similar environments, a finding incompatible with the observation of gradual remapping.

Instead of asking whether there is sufficient evidence to justify the complexity of an entirely new context, one might ask whether there is sufficient evidence to justify the complexity required to express how the second environment differs from the first. If the two environments are similar, then it may be simpler to express this difference than to express the second environment independently. This is tantamount to assuming *a priori* that the two contexts are related, but that their differences may nonetheless be valuable to represent in separate contexts.

A two-level contextual hierarchy is therefore considered in which contexts may either be *independent* or *dependent*. At the top level are *independent* contexts, whose states' parameters are not statistically associated with those in any other independent context. The creation of an independent context must be justified with respect to the complexity of its entire set of parameters, which is easily done when the animal enters an environment clearly unrelated to any other. At the second level are *dependent* contexts, each of which is associated with an independent context. Specifically, each state in a dependent context is paired with a corresponding state in an independent context, and states in the dependent context may share parameters with states in that independent context. In addition, the expected hippocampal inputs of states in the dependent contexts are assumed to be similar to those of the states in their respective independent contexts. This reduces the added complexity of the dependent context to more accurately reflect the degree to

which it differs from the independent context.

There is a growing body of evidence to support the notion that similar contexts are not represented independently *in vivo*. Very similar contexts show partially overlapping representations (partial remapping), especially when only a change in task is involved (e.g. Markus et al., 1995; Shapiro et al., 1997; Skaggs and McNaughton, 1998; Wood et al., 2000; Jeffery, 2000). Even between contexts with more pronounced differences (arena wall shape or color changes), the type of remapping observed primarily involves a change in firing rate, but not of the location of the place field (Leutgeb et al., 2005c). We therefore interpret the degree of remapping to be determined, in the asymptote, by the similarity of the contexts. With little experience, the degree of remapping is even less, reflecting the uncertainty that the two contexts are distinct. We do not propose a specific similarity metric – any such metric would vary among subjects – but such a metric should be monotonic: changes that make two contexts more different should not result in less remapping.

Even if the contexts are not independent, why should one be represented as an independent context and the other as a dependent context? (In other words, why not represent them symmetrically as two co-dependent contexts without introducing an explicit hierarchy?) The asymmetry allows the models to be nested, which provides a more direct relationship between model contexts and place cell activity patterns. Consider an experiment involving two similar environmental contexts (E_1 and E_2). Before the start of the experiment, the animal has contextual representations only for its home cage, a transport box, etc.; let this be context model M_0. Upon introduction to E_1, which is nothing like any previously experienced context, a new model is immediately adopted, $M_1 = M_0 \cup \{I_1\}$, i.e. all contexts in M_o and an independent context I_1 for E_1. Subsequent exposure to E_2 leads to consideration of $M_2 = M_1 \cup \left\{D_2^{I_1}\right\} = M_0 \cup \left\{I_1, D_2^{I_1}\right\}$, where $D_2^{I_1}$ is a dependent context associated with I_1 that represents E_2. As more experience is acquired in E_1 and E_2, the relative likelihood of M_2 will gradually increase. Given a nested model structure, the only change from M_1 to M_2 involves the creation of a new context $D_2^{I_1}$ for E_2. The model context I_1 for E_1 exists (has been statistically justified) independently of any experiences in E_2. Thus, the most parsimonious hippocampal representational change would involve the creation of a contextual representation for E_2 without changing the representations of any other contexts.

Gradual remapping appears to show such an asymmetry. Lever et al. (2002) first exposed rats to a cylindrical arena (E_1), then a square arena (E_2). Over the course of many subsequent exposures to both arenas, rats

gradually remapped between them. The observed pattern of remapping suggests that the activity of place cells in the square arena (E_2) diverged from the cylindrical arena E_1; in contrast, fields in the cylindrical arena (E_1) appeared stable over several days.

It should be noted that, whereas the experiments and simulations in this study focus on learning to distinguish two contexts, the statistical framework outlined here can be extended in a straightforward manner to allow an arbitrary number of dependent contexts.

In the following subsections, a detailed description is provided of how simulated hippocampal inputs are constructed, how HMMs with independent and dependent contexts are defined, and how the posterior likelihood of different models is calculated.

7.2.2 *Simulated hippocampal inputs*

While the inputs to the hippocampus are very high dimensional, input patterns to the model, for computational simplicity, were formulated as noisy scalar values. Input patterns were generated based on a simulated "ground truth" of the animal's actual state (location, task stage, etc.) in the world. These environmental states ("e-states") compose the true generative model, not to be confused with states in the hippocampal HMM model. Each e-state was assigned a positive index i, and the hippocampal input value y generated for that e-state was $8\,(i - n/2) + \eta$, where n is the number of e-states in the environment and η is a Gaussian noise term; the y values are therefore roughly centered around 0, as expected under the prior (see below). The standard deviation of the noise term was 0.125, which ensures that different e-states within an environmental context are unambiguously distinct.

To simulate an experiment involving multiple sequences or tasks, a single set of e-states was used; only their order of presentation changed. For example, multiple spatial sequences were simulated using the same set of position e-states, but the order of visited positions was different for each sequence. To simulate small environmental changes, each e-state's input value was mildly perturbed (see specific simulations for details).

Experiments were modeled in discrete time: each discrete time step had an associated e-state. To model a specific experimental paradigm of duration T, a sequence of T e-states was produced, representing the entire course of the experiment over many days. A sequence of hippocampal inputs, $y_{1\ldots T}$, was then generated based on the e-state sequence.

7.2.3 *Hidden Markov models with independent and dependent contexts*

Our simulations used Gaussian HMMs as a model of the hippocampal representation of its input patterns. By using HMMs, we do not intend to suggest that the hippocampus is a finite state machine; rather, HMMs provide a convenient statistical framework in which to represent both a mixture of many different input patterns and their temporal structure (or a Markov approximation thereof).

Formally, a Gaussian HMM is composed of a set of N_S states, each of which is parameterized by a Normal distribution of expected values, $N\left(\mu_s, \sigma_s^2\right)$. A transition probability matrix, A_{ji}, defines the probability of transitioning from state i to state j at each time step. Additionally, a starting state probability p_0 must be specified; in these simulations, the starting state probability was assumed to be uniform over all states.

In the present framework, states are organized into multiple contexts, an organization realized by restrictions placed on transition probabilities. Transition probabilities between states in the same context could vary to fit the observed sequence of transitions in y. However, the sum of transition probabilities to all states in other contexts was fixed at $\gamma = 0.05$. This is the critical parameter that determines how strongly models are biased against context changes.

If the total between-context transition probability is γ, what is the transition probability from some state s to some other state s' in a different context? This is defined based on a hierarchical organization of states and contexts. Consider a model M_k composed of N_S states, $S_1 \ldots S_{N_S}$. These states are organized into N_C contexts, $C_1 \ldots C_{N_C}$. Contexts are organized into N_G context groups, $G_1 \ldots G_{N_G}$, where each context group contains exactly one independent context and any dependent contexts associated with it. For the purpose of defining transition probabilities, the model M_k is therefore organized as a mixture of context groups, each context group a mixture of contexts, each context a mixture of states. For each mixture, the component probabilities are set to be equal. Thus, the probability of transitioning to a particular context group, G_g, is $p_1 = 1/N_G$. The probability of transitioning to a particular context in that group is $p_2 = 1/|G_g|$, the inverse of the number of contexts in group G_g. The probability of transitioning to a particular state in that context is $p_3 = 1/|C_c|$, the inverse of the number of states in context C_c. Thus, the probability of transitioning from state s to state s' is $\gamma p_1 p_2 p_3$. When the values p_1 and p_2 are calculated, the context of s is first excluded, since these transitions are only between states in different contexts.

The hierarchical organization of contexts helps us take into account the

fact that there are other unrelated contexts outside of the experimental apparatuses. A fixed value of N_G is used to reflect that the transition probabilities among unrelated contexts would not change with the addition of one or two more. For the multiple environments simulations, $N_G = 3$, since up to three independent contexts are considered; for the other simulations, $N_G = 1$. (How exposure to just a few or to many completely different environments affects remapping is not known, but an increased willingness to remap after experiencing many previous environments would argue instead that N_G is explicitly represented in the hippocampus.)

We now describe how states in independent and dependent contexts are parameterized. In particular, we describe a model in which a state in a dependent context may share parameters with a state in an independent context, and the prior probabilities of parameters in dependent contexts may depend on corresponding parameters in their respective independent contexts. Qualitatively, the particular scheme for parameter sharing and the priors used here have a singular purpose: to reduce the size of the parameter space (i.e. the complexity) of the dependent context. Other formulations that achieve the same goal would have been equally acceptable, so readers not well-versed in statistics may safely skip the remainder of this section.

To define the parameters of a model, consider a state $\hat{s}$ in a dependent context and its corresponding state s in the associated independent context. The independent state's parameters are defined purely in terms of the state's transition probability vector, $\vec{a}_s$, and mean and variance terms, μ_s and σ_s^2:

$$A_{*s} = \vec{a}_s \tag{7.1}$$

$$p(y|s) \sim N\left(\mu_s, \sigma_s^2\right) \tag{7.2}$$

where A_{*s} denotes a row of the HMM's transition probability matrix and y is the observed input pattern. The dependent state $\hat{s}$ is defined with respect to both the set of parameters for state s and a set of parameters unique to state $\hat{s}$:

$$A_{*\hat{s}} = (1 - z_{\hat{s}})\, \vec{a}_s + z_{\hat{s}}\, \vec{a}_{\hat{s}} \tag{7.3}$$

$$p(y|\hat{s}) \sim (1 - \zeta_{\hat{s}})\, N\left(\mu_s, \sigma_s^2\right) + \zeta_{\hat{s}}\, N\left(\mu_{\hat{s}}, \sigma_{\hat{s}}^2\right) \tag{7.4}$$

The mixing parameters $z_{\hat{s}}$ and $\zeta_{\hat{s}}$ govern the relative contribution of each component. This allows a dependent state to, for example, share the same transition probability vector with its corresponding independent state ($z_{\hat{s}} \approx 0$), but adopt a different distribution of y values ($\zeta_{\hat{s}} \approx 1$). When a mixing parameter is close to zero, the second mixture component plays no role in the likelihood of y under the model, thus reducing the model's effective

HMM Parameters	Prior	Value
Transition probabilities $\vec{a}$	δ_a	0.8
Parameter sharing probabilities $z_{\hat{s}}$ and $\zeta_{\hat{s}}$	δ_1, δ_2	0.1, 0.05
Mean of μ_s	ξ	0
Inverse variance of μ_s	κ_1	0.01
Inverse variance of $\mu_{\hat{s}}$	κ_2	4
Shape parameter of σ_s^{-2}	α_1	2
Shape parameter of $\sigma_{\hat{s}}^{-2}$	α_2	10
Rate parameter of σ_s^{-2}	β_1	0.1
Offset of mean of $\mu_{\hat{s}}$ from μ_s	h	0.4

TABLE 7.1: Hyperparameter values that define the prior distributions over each parameter.

complexity.

The prior over each mixing parameter was a highly sparse beta distribution that mildly favored smaller values: $z_{\hat{s}} \sim B(\delta_1, \delta_2)$ and $\zeta_{\hat{s}} \sim B(\delta_1, \delta_2)$. All fixed hyperparameters are listed in Table 7.1. For the first components' parameters, vague priors were used: $\vec{a}_s \sim \text{Dirichlet}(\delta_a, \delta_a, \ldots, \delta_a)$, $\mu_s \sim N\left(\xi, \kappa_1^{-1}\right)$, $\sigma_s^{-2} \sim \Gamma(\alpha_1, \beta_1)$. The prior over each $\vec{a}_{\hat{s}}$ was the same as $\vec{a}_s$. The priors for $\mu_{\hat{s}}$ and $\sigma_{\hat{s}}$ were biased to be somewhat similar to μ_s and σ_s, further reducing the dependent context's contribution to the complexity penalty: $\mu_{\hat{s}} \sim N\left(\mu_s + h, \kappa_2^{-1}\right)$ and $\sigma_{\hat{s}}^{-2} \sim \Gamma\left(\alpha_2, \alpha_2 \sigma_s^2\right)$. The purpose of h was to encourage model parameterizations in which $\zeta_{\hat{s}} = 0$ instead of $\zeta_{\hat{s}} = 1$ and $\mu_{\hat{s}} = \mu_{s,1}$. A more sophisticated approach might involve splice sampling from the non-conjugate prior $\mu_{\hat{s}} \sim I(\mu_s, r) N\left(\mu_s, \kappa_2^{-1}\right)$, where the indicator function I is zero over some interval around the mean, $[\mu_s - r, \mu_s + r]$; however, the simpler procedure was sufficient for our simulations.

7.2.4 *State inference and model selection*

Let M_k denote a model composed of N_S^k states grouped into N_C^k contexts. Let $\theta_k = \left(\vec{z}, \vec{\zeta}, A, \vec{\mu}, \vec{\sigma}^2\right)$ denote the parameters of the states in M_k. Given a sequence of noisy inputs, $y_{1\ldots T}$, the state inference problem is to infer under which sequence of states the inputs are most probable. For state inference, it is assumed that both M_k and θ_k are known, and that the HMM is the generative model of $y_{1\ldots T}$, i.e. each HMM state corresponds to an e-state in the world. Thus, inferring the HMM state serves as a proxy for inferring the true e-state. This is traditionally done using the Viterbi algorithm (Viterbi, 1967).

The context learning problem can be understood as part of the more

fundamental problem of inferring the generative model of the input patterns: given the sequence $y_{1\ldots T}$, infer the HMM most likely to have generated it. This can be decomposed into two parts: model parameterization and model selection. Model parameterization involves inferring the most likely set of parameters, θ_k, given an HMM M_k with a known structure. Model selection involves inferring which model structure is most likely out of a set of candidates, typically with different numbers of states.

While model selection in a machine learning context typically involves a single batch analysis, the hippocampus provides some representation of context even in completely novel environments. Multiple models with different numbers of contexts were therefore compared based on initial subsequences, $y_{1\ldots\tau}$, $\tau < T$. As τ increased, the models were provided with progressively more input patterns to fit, resulting in changes in their likelihoods.

In a Bayesian setting, models are compared based on relative likelihoods, given the observed data ($y_{1\ldots T}$). Assuming equal prior probability across models, this can be reduced to calculating the odds of $y_{1\ldots T}$ under each model:

$$B_{k\ell} = \frac{P(y_{1\ldots\tau}|M_k)}{P(y_{1\ldots\tau}|M_\ell)} \tag{7.5}$$

This is known as the Bayes factor (for review, see Kass and Raftery, 1995), and all model comparisons calculated will be presented as a log Bayes factor. $p(y_{1\ldots\tau}|M_k)$ is the marginal likelihood of $y_{1\ldots T}$ under M_k, marginalized over the entire parameter space:

$$p(y_{1\ldots\tau}|M_k) = \int_{\theta\in\Theta} p(y_{1\ldots\tau}|\theta_k, M_k)\, p(\theta_k|M_k)\, d\theta_k \tag{7.6}$$

where $p(y_{1\ldots\tau}|\theta_k, M_k)$ is the conditional probability of the inputs given a particular parameterization of the model, and $p(\theta_k|M_k)$ is the prior probability density over the parameter space.

As mentioned above, Equation 7.6 provides an implicit complexity penalty. Importantly, while $p\left(y_{1\ldots\tau}|\theta_k^*, M_k\right)$ depends on $y_{1\ldots\tau}$, $p(\theta_k|M_k)$ does not. Since the difference in conditional probability between models typically grows with the length of $y_{1\ldots\tau}$, the Bayes factor asymptotically determines by how well each model fits $y_{1\ldots\tau}$. However, for smaller input sequences ($\tau \ll T$), the prior uncertainty over the parameter space, $p(\theta_k|M_k)$, can strongly influence the Bayes factor.

The computational challenge in calculating Bayes factors is to accurately calculate $p(y_{1\ldots\tau}|M_k)$, which requires integration over the parameter space

(Equation 7.6). A closed-form solution does not exist, but any accurate numerical approximation method is equally acceptable, and the results presented here do not depend on the particular method employed. We use a Monte Carlo integration technique described in the following section.

7.2.5 *Estimating marginal model probabilities*

For these simulations, Equation 7.6 was numerically approximated by Monte Carlo integration of the parameter samples using an importance distribution. The importance distribution was constructed in an automated fashion using 500 samples from the posterior parameter distribution, $p(\theta_k|y_{1\ldots\tau}, M_k)$, and the Gibbs kernels from the Markov chain Monte Carlo (MCMC) sampler that generated the parameter samples (for details, see Frühwirth-Schnatter, 2004). Monte Carlo integration becomes unstable when the tails of the importance distribution are narrower than the posterior distribution along any dimension of the parameter space. To guard against this, the variances of the μ and σ^{-2} components were doubled, and the variance of the a_s component was increased by multiplying each Dirichlet parameter by 0.75.

The MCMC sampler was constructed based on previous sampling techniques for standard HMMs (Chib, 1996). Briefly, the Markov chain was constructed from Gibbs kernels that sample sequentially from both the parameters of the model and latent indicator vectors, which indicate, for each mixture distribution within the model and each time step t, which mixture component is implicated in the observed value y_t. For the shared parameter HMMs, there are three indicator vectors. The vector $S_{1\ldots T}$ indicates the active HMM state at each time step, and was sampled using a standard HMM Gibbs move:

$$p(S_t = i|\ldots) \propto A_{i,S_{t-1}}\, p\left(y_t|\mu_i, \sigma_i^2\right) A_{S_{t+1},i} \tag{7.7}$$

The vector $u_{1\ldots T-1}$ indicates which component of the transition probability mixture (Equation 7.3) for the current state contributed to each transition. The vector $v_{1\ldots T}$ indicates which component of the hippocampal input value mixture (Equation 7.4) for the current state contributed to each observation. Gibbs moves were as for standard mixtures:

$$p(u_t = \text{Dep}|\ldots) \propto A_{S_{t+1},\hat{s}} / \left(A_{S_{t+1},s} + A_{S_{t+1},\hat{s}}\right) \tag{7.8}$$

$$p(v_t = \text{Dep}|\ldots) \propto p\left(y_t|\mu_{\hat{s}}, \sigma_{\hat{s}}^2\right) / \left(p\left(y_t|\mu_s, \sigma_s^2\right) + p\left(y_t|\mu_{\hat{s}}, \sigma_{\hat{s}}^2\right)\right) \tag{7.9}$$

The values of u_t and v_t are defined only for times when the HMM state indicator vector indicates a state in a dependent context.

Gibbs moves for the HMM parameters were complicated by sharing of

parameters between independent and dependent states. Gibbs moves for the mixing parameters were:

$$p(z_{\hat{s}}|\ldots) \propto \text{Beta}(\delta_1 + n_s^u, \delta_2 + n_{\hat{s}}^u) \tag{7.10}$$

$$p(\zeta_{\hat{s}}|\ldots) \propto \text{Beta}(\delta_1 + n_s^v, \delta_2 + n_{\hat{s}}^v) \tag{7.11}$$

where $n_s^u = \#\left(u_t^{\hat{s}} = \text{Ind}\right)$, $n_{\hat{s}}^u = \#\left(u_t^{\hat{s}} = \text{Dep}\right)$, $n_s^v = \#\left(v_t^{\hat{s}} = \text{Ind}\right)$ and $n_{\hat{s}}^v = \#\left(v_t^{\hat{s}} = \text{Dep}\right)$. The counting function $\#()$ returns the number of occurrences over a time-indexed vector in which the specified condition is satisfied. Gibbs moves for the transition probabilities were:

$$p(\vec{a}_{s_1}|\ldots) \propto \text{Dirichlet}\left(\delta + n_1^{\text{ind}}, \delta + n_2^{\text{ind}}, \ldots\right) \tag{7.12}$$

$$p(\vec{a}_{\hat{s}_1}|\ldots) \propto \text{Dirichlet}\left(\delta + n_1^{\text{dep}}, \delta + n_2^{\text{dep}}, \ldots\right) \tag{7.13}$$

The transition probabilities for a state in an independent context, s_1, were updated based on a transition count, $n_{s_2}^{\text{ind}}$, which counts the number of occurrences of a transition from $s_1$to s_2. If the independent context had a dependent context (with states $\hat{s}_1$ and $\hat{s}_2$ that are paired with states s_1 and s_2), then $n_{s_2}^{\text{ind}}$ also included the number of occurrences of a transition from $\hat{s}_1$ to $\hat{s}_2$ when $u_t = \text{Ind}$. For a state in a dependent context, $\hat{s}_1$, the transition count $n_{\hat{s}_2}^{\text{dep}}$ included transitions from $\hat{s}_1$ to $\hat{s}_2$ when $u_t = \text{Dep}$.

The observation parameters for a state in an independent context were sampled using the following Gibbs moves

$$p(\mu_s|\ldots) \propto N\left(\left(\sigma_s^{-2}\sum_{t\in\mathcal{T}_s} y_t + \kappa_1\xi\right)\sigma_\mu^2,\ \sigma_\mu^2\right) \tag{7.14}$$

$$p\left(\sigma_s^{-2}|\ldots\right) \propto \Gamma\left(\alpha_1 + \frac{1}{2}|\mathcal{T}_s|,\ \beta_1 + \frac{1}{2}\sum_{t\in\mathcal{T}_s}(y_t - \mu_s)\right) \tag{7.15}$$

where $\sigma_\mu^{-2} = \sigma_s^{-2}|\mathcal{T}_s| + \kappa_1$ and $\mathcal{T}_s$ is the set of times for which $S_t = s$ or for which $v_t = \text{Ind}$ and $S_t = \hat{s}$. The observation parameters for a state in a dependent context were sampled using the following Gibbs moves:

$$p(\mu_{\hat{s}}|\ldots) \propto N\left(\left(\sigma_{\hat{s}}^{-2}\sum_{t\in\mathcal{T}_{\hat{s}}} y_t + \kappa_2(\mu_s + h)\right)\sigma_\mu^2,\ \sigma_\mu^2\right) \tag{7.16}$$

$$p\left(\sigma_{\hat{s}}^{-2}|\ldots\right) \propto \Gamma\left(\alpha_2 + \frac{1}{2}|\mathcal{T}_{\hat{s}}|,\ \alpha_2\sigma_s^2 + \frac{1}{2}\sum_{t\in\mathcal{T}_{\hat{s}}}(y_t - \mu_{\hat{s}})\right) \quad (7.17)$$

where $\sigma_{\mu}^{-2} = \sigma_{\hat{s}}^{-2}|\mathcal{T}_{\hat{s}}| + \kappa_2$ and $\mathcal{T}_{\hat{s}}$ is the set of times for which $v_t = \text{Dep}$ and $S_t = \hat{s}$.

Five thousand samples from the importance distribution were used to estimate Equation 7.6. The availability of samples from the posterior distribution permitted evaluation of Equation 7.6 with the computationally more expensive technique of bridge sampling. Unlike traditional importance sampling, an accurate estimate via bridge sampling does not require the importance distribution to be broader than the posterior distribution (Meng and Wong, 1996; Frühwirth-Schnatter, 2004). In our many test cases, we found that the differences based on which integration technique was used were negligible, suggesting that the importance distribution was accurately covering the entire mass of the posterior distribution.

7.3 Simulation Results: Multiple Environments

7.3.1 *Gradual remapping*

Recent data have demonstrated that the development of distinct spatial maps for two environments can be gradual (Tanila et al., 1997; Jeffery, 2000; Lever et al., 2002). Tanila et al. (1997) found that repeatedly rotating two sets of cues in opposite directions engendered an increase in remapping between the cue configurations over time. Jeffery (2000) found that when the same arena was placed in two different room locations, the number of place fields that differentiated between arena room locations gradually increased over several days. Lever et al. (2002) similarly found that, when hunting for food pellets alternately in cylindrical and square arenas, different place cells developed distinct place fields in each arena on different days; two weeks of training were required to achieve complete remapping. In each case, repeated conflicts between cue configurations were gradually resolved by developing separate contexts for each configuration.

Our modeling of gradual remapping is described with reference to the Lever et al. (2002) study, though it could equally well be interpreted with respect to the others. The arena geometry change is interpreted as a small but consistent difference in the afferent input pattern for which a model with a separate context for each arena will, with enough experience, be better suited.

To model gradual remapping, we constructed input sequences, $y_{1\ldots T}$,

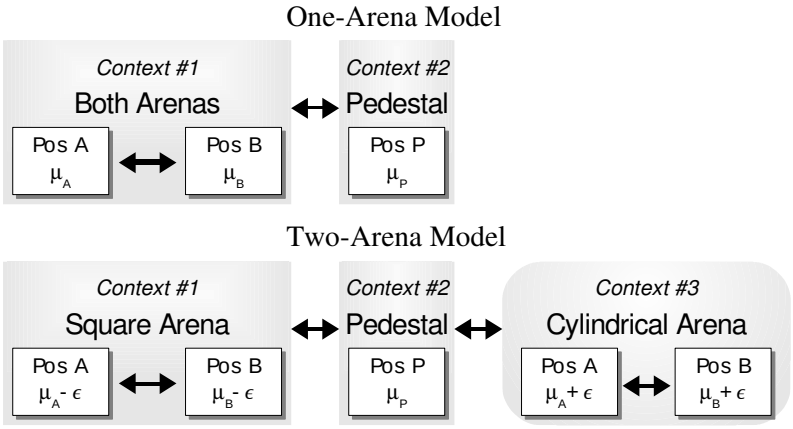

FIGURE 7.1: Diagram of one- and two-arena models used to model gradual remapping. The small white boxes represent states in each model, and the large grey boxes indicate how the states are grouped into contexts. Arrows indicate transition probability parameters that are either sampled (within a context) or fixed (between contexts).

generated from alternating sessions in each arena, interposed by time spent on a holding pedestal (during which an experimenter would swap arenas). For computational tractability, two positions within the environment – A and B – were modeled (e.g. one in each half of the arena). The simulated rat spent 2 time-steps on the pedestal and 10 time-steps in each arena, alternating between the positions in order to remove trajectory differences that might suggest the environments were different. An entire sequence contained 16 environment visits, half in each arena, for a total sequence length of 192 samples.

At each position, the input was perturbed depending on the shape of the arena. For example, at position A, the input had mean $\mu_A - \varepsilon$ in the square arena but $\mu_A + \varepsilon$ in the cylindrical arena. The input standard deviation for all states was $\sigma = 0.125$, and the value of ε used was 0.175, so the input distributions of the two contexts overlapped.

The likelihoods of the one- and two-arena models shown in Figure 7.1 were compared using the model selection framework described above. The "No pretraining" line in Figure 7.2 shows that, as more experience is gained in each environment, the Bayes factor gradually increases toward a decisive positive value. This suggests that the observed gradual remapping reflects an underlying statistical process: an evidence-based transition to a more complex contextual model.

If one were to pretrain rats by exposing them first to one arena for multiple

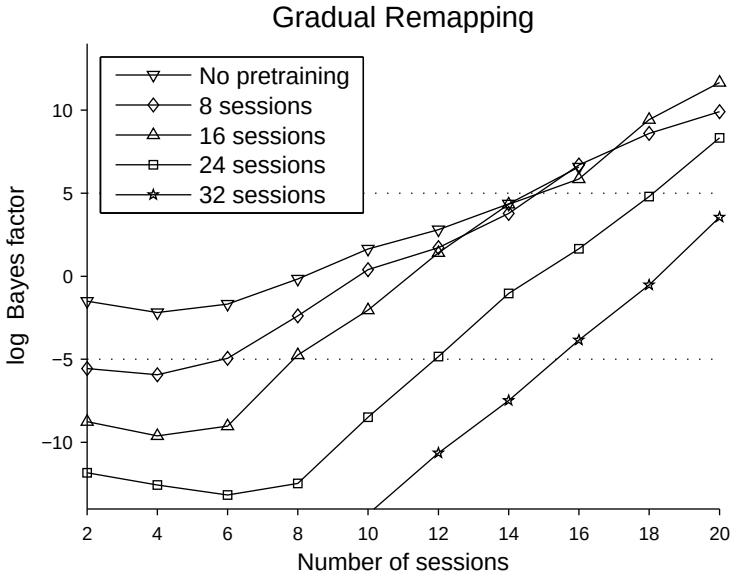

FIGURE 7.2: With more experience in each environment, the relative likelihood of the two-arena model increases toward certainty. As a result of first pretraining in one arena, the two-arena model is initially less likely, but preference for this model then increases more rapidly.

sessions, how should this pretraining experience affect the rate of remapping? The "N sessions" lines in Figure 7.2 reflect a training sequence that starts with N sessions of pretraining in one environment, followed by alternating sessions in the two environments. These simulations predict that pretraining should both delay the onset of gradual remapping and hasten its completion. With few experiences, the second arena looks like a "noisy" version of the first; however, the larger sequence of experiences due to pretraining eventually permits the environments to be distinguished more rapidly.

7.3.2 *Failure to generalize*

In an extension of the Jeffery (2000) study, Hayman et al. (2003) trained rats in a box placed in two different locations in the room. After place cells gradually learned to remap between box positions, the color of the box and floor was changed from white to black. This substantial sensory change resulted in an immediate, complete remapping. Interestingly, despite multiple days of training in a white box in the same two locations, place cells did not immediately discriminate the two locations of the black box.

Their result can be understood as a consequence of the relative differences of the box position and box color manipulations. Specifically, while the

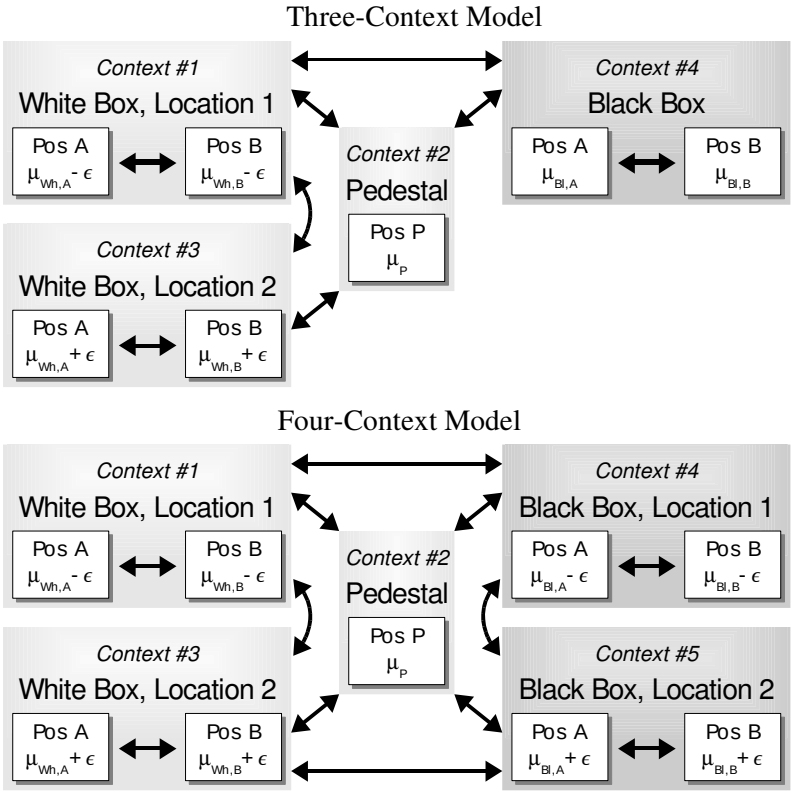

FIGURE 7.3: Diagram of three- and four-context models used to model the Hayman et al. (2003) study. The two white box contexts differ subtly, as do the two black box contexts. However, the white box and black box contexts are highly differentiated. The dotted lines at ± 5 denote the thresholds beyond which there is essentially no statistically uncertainty in which model is preferred.

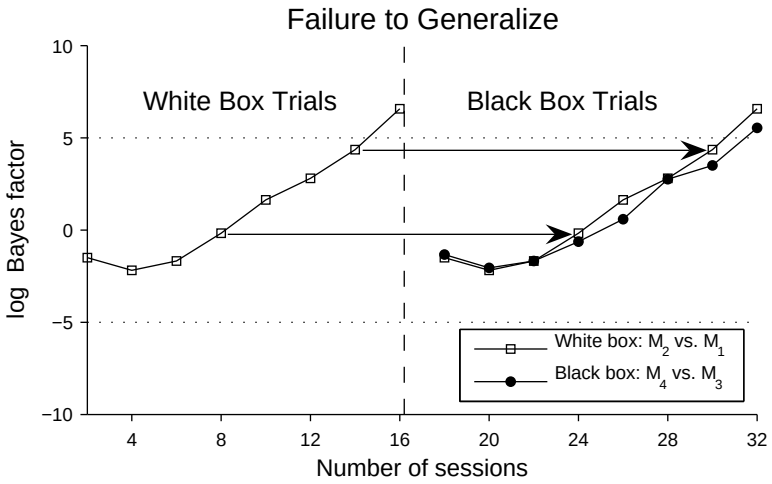

FIGURE 7.4: The likelihood of the four-context model relative to the three-context model (Black Box) increases at the same rate as the two-context model relative to the one-context model (White Box). A copy of the White Box log Bayes factors are superimposed on top of the Black Box log Bayes factors to illustrate their similarity.

position shift created a relatively subtle change, the color change was large. We model their experiment by extending the gradual remapping simulation to include two additional "black box" contexts whose hippocampal inputs are similar to each other but not to states in the original two contexts. Specifically, we constructed input sequences, $y_{1\ldots T}$, generated from alternating sessions in the white box in room locations 1 and 2, followed by alternating sessions in the black box in room locations 1 and 2. Box visits were, as before, interposed by time spent on a holding pedestal. The two white box positions were modeled identically to the square and cylindrical arenas in the previous simulation. The two black box room locations were modeled in an analogous fashion: the hippocampal inputs for corresponding rat positions in the two black boxes differed by $\pm\epsilon$, but their μ values differed substantially from hippocampal inputs in the white box.

Figure 7.3 shows the three- and four-context models evaluated on the sequence of hippocampal inputs. Figure 7.4 shows the evolution of the likelihoods of the three- and four-context models over the course of training in the black box, subsequent to full training in the white box. For comparison, training of one- and two-context models in the white box (structured as in Figure 7.1) is shown as well. Since the white and black boxes differ so strongly, the question of whether the black box locations should be represented as one context or two is statistically unrelated to the representation of the white

box locations. The relative likelihood of the four-context model therefore increases at the same gradual rate, predicting that rats should show the same gradual remapping between room locations in both white and black boxes. Consistent with this prediction, Hayman et al. (2003) found in the one rat whose rate of remapping was fastest that the rat remapped between black box locations at the same rate as between white box locations. (The black box sessions were not continued long enough to assess remapping in the rats who were slower to remap in the white box.)

7.3.3 *Morph environments*

Leutgeb et al. (2005b) trained rats in square and cylindrical environments until the degree of remapping between arenas reached asymptote. Then, they exposed the rats to a sequence of arenas whose wall shape systematically morphed between the square and cylinder. They found that the rats' place cell activity in the morph environments reflected various averages of the two maps.

What if the rats were first trained on the morph environments and then on just the square and cylindrical environments? In the gradual remapping simulations, separate contexts for the square and cylindrical arenas develop because they result in distinct clusters of hippocampal input patterns. However, substantial pretraining in the morph environments would generate one broad cluster of input patterns, predicting that morph training should inhibit remapping during subsequent training in just the square and cylindrical arenas.

To model this hypothesized experiment, five arenas were used, and input patterns were generated that varied linearly from the square arena (arena 1) to the cylindrical arena (arena 5). For example, at position A, the inputs in the five arenas had means $\mu_A - \varepsilon$, $\mu_A - \frac{1}{2}\varepsilon$, μ_A, $\mu_A + \frac{1}{2}\varepsilon$, and $\mu_A + \varepsilon$. We constructed input sequences beginning with sessions in just the morph arenas, selecting arena 3 twice as often as 2 or 4 to reinforce a single cluster distribution of input patterns. These sessions were followed by alternating sessions in the square and cylindrical arenas. All other details were the same as in the gradual remapping simulation described previously.

Simulation results are shown in Figure 7.5. Compared to the baseline condition (no morph training), the 16 morph training sessions roughly doubled the time required to complete remapping during subsequent cylinder / square training. With 16 pretraining sessions, remapping was further delayed. This prediction stands in direct contrast to hippocampal models based on independent component analysis (ICA), which would predict place code differentiation during the morph pretraining, as the arena shape acts as an

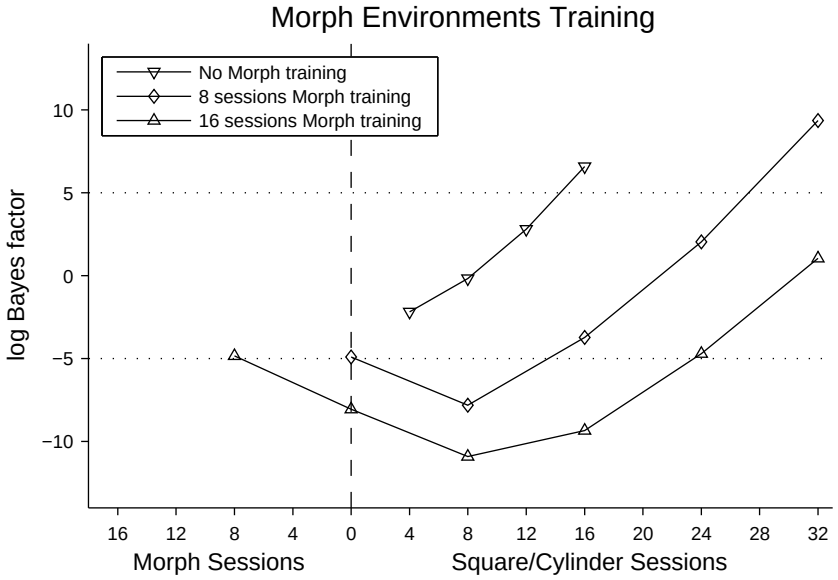

FIGURE 7.5: Morph experiment simulation: pretraining in the morph environments inhibits adoption of a two-context model during later training only in square and cylindrical arenas. Simulation results involving morph training were calculated after each block of 8 sessions, and simulation results not involving morph training are plotted after each block of 4 sessions. All plots begin after completion of the first block of training, i.e., after 4 or 8 sessions.

independent source of variation (Lörincz and Buzsáki, 2000).

The disorientation study by Knierim et al. (1995) invites a similar interpretation. Rats were trained to forage for food in a cylinder with an orienting cue card. They found that when rats were disoriented by carrying them around in a closed box before being placed in the arena, their place fields failed to align with the cue card after several sessions, while non-disoriented rats maintained the alignment. They interpreted their results to suggest that the cue card was perceived to be in a different location each time the rats entered the arena after disorientation. After several sessions, the cue card was perceived as "unstable" and therefore was ignored. Moreover, even after multiple subsequent sessions without disorientation, place fields never developed a consistent alignment with the cue card.

Their experiment likely reflects the influence of the head direction system on the place code more than any computational process within the hippocampus itself. Nonetheless, their findings suggest a similar type of statistical inference elsewhere in the brain: if a cue varies in an apparently random manner, its impact on the overall representation of an environment should

be minimized.

7.4 Simulation Results: Reversal Learning

While many conditioning paradigms are not hippocampus dependent, reversal learning has consistently shown dependence on hippocampal function (Kimble and Kimble, 1965; Silveira and Kimble, 1968; Winocur and Olds, 1978; Berger and Orr, 1983; Neave et al., 1994; McDonald et al., 2002; Ferbinteanu and Shapiro, 2003). Consistent with this interpretation are place cell studies showing context-specific firing patterns during spatial reversal learning tasks (Ferbinteanu and Shapiro, 2003; Smith and Mizumori, 2006).

One of the most interesting aspects of reversal learning is that repeated reward reversals lead to progressively faster behavioral reversal. Two studies have demonstrated that after repeated reward reversals, rats are capable of reversing behavior after a single error trial (Buytendijk, 1930; Dufort et al., 1954). Additionally, Pubols (1962) showed near-perfect reversal performance (< 0.5 errors on average after initial error trial), and Brunswick (1939) showed that, even when single-trial reversal performance is not yet achieved, most of the improvement is observed on the second trial. Thus, rats can be trained to select a different (previously learned) behavioral strategy after a single error trial.

Several studies have also explored the impact of partial reinforcement on reversal learning, considering cases in which the "correct choice" is only rewarded on some percentage of trials, as well as cases in which the "incorrect choice" is also rewarded on some smaller percentage of trials (Brunswick, 1939; Wike, 1953; Grosslight et al., 1954; Elam and Tyler, 1958; Wise, 1962; Pennes and Ison, 1967). The pattern of data across these studies suggests that the more similar the original and reversed contingencies, the more slowly the animal learns the reversal. While intuitive, this suggests that the impact of a trial on a reward association is weighted by how informative the trial is perceived to be. If the expectation of a particular outcome (reward, no reward) is more uncertain, then observing the outcome provides less information about whether the distribution of outcomes has changed.

Previous approaches to modeling reversal learning have posited that the discrimination is relearned during each reversal. Hasselmo et al. (2002) theorized that the hippocampal facilitation of reversal learning was due to quick unlearning and relearning of the association between choice and reward within the hippocampal representation. Learning in their model is unsupervised (Hebbian), and they suggest that reversal learning deficits due to hippocampal or theta modulation impairment are caused by the inability to

separate new learning from past associations. Unfortunately, since this model completely relearns the reward association after each reversal, no savings with repeated reversals is predicted. Also, since the current association is dissociated from the previous association, partial reinforcement would not affect the speed of reversal learning under this model.

Another series of models have proposed that the hippocampus plays a role in stimulus representation (Gluck and Myers, 1993; Myers et al., 1995; Gluck and Myers, 1996), performing both stimulus compression and predictive stimulus differentiation. A similar model has been proposed by Schmajuk and DiCarlo (1991, 1992). Both models propose that the hippocampus is critically involved in learning a conjunctive stimulus layer (hidden layer), though their network topologies differ somewhat. In addition, both models train the hidden layer using variants of the backpropagation learning rule; Schmajuk and DiCarlo propose a more direct, biologically plausible implementation, while Gluck and Myers suggest only that a functionally similar computational process occurs *in vivo*.

Both models demonstrate some savings during repeated retraining (on serial reversals or serial extinctions and renewals). In the Gluck and Myers model, hidden layer discrimination increases gradually, making associations with the output layer simpler to learn. In the Schmajuk and DiCarlo model, the learning rate of the input-to-hidden-layer weights is higher, so the increased weighting of the hidden layer representation at the output layer over the repeated reversals leads to faster relearning. In either case, both models fundamentally rely on some degree of retraining during each reversal. However, the single trial reversals observed experimentally occurred based on an error trial alone (Buytendijk, 1930; Dufort et al., 1954). Even with an arbitrarily high learning rate, it is not clear how one could retrain a backpropagation network without at least one positive trial.

In addition, the partial reinforcement conditions are not well explained by these models, or other delta-rule based models (e.g. Rescorla and Wagner, 1972; Pearce and Hall, 1980), since learning is not adjusted based on the entropy of the trial's outcome. Similarly, none of these models accounts well for the partial reinforcement extinction (PRE) effect, in which extinction training is prolonged following partial reinforcement. With generalized delta rule learning, partial reinforcement results in a weaker association between behavioral choice and outcome. During a reversal (or extinction), this weaker association would be easier to expunge than a stronger association would be. However, the data show the opposite results. A few recent statistical models account well for the basic PRE effect (Gallistel and Gibbon, 2000; Courville, 2006), but they do not address the progressive improvement observed in

reversal learning.

An alternative interpretation of the reversal learning data, as proposed by Hirsh (1974), is that the hippocampus represents each reward condition as a separate context. If the reward contingencies of both contexts are represented simultaneously, then after initial training no retraining should be required after each reversal. Rather, recent trial outcomes can be compared to prior knowledge of each context to infer which context is active. In a full-reinforcement reversal learning paradigm, knowledge of both contexts allows a single error trial to be sufficient to infer a context switch, since the choice during the error trial should yield no reward only in "the other" context. One might think of this process as analogous to the problem of self-localization in spatial navigation, a function that has also been attributed to the hippocampus (Touretzky and Redish, 1996). The present statistical formulation of context learning also correctly quantifies the increased uncertainty in the adoption of a new context under partial reinforcement conditions.

Interestingly, if the reversal is performed in a different environment, the environmental cues substantially improve adaptation to the reversal, even in hippocampal animals (McDonald et al., 2002). This finding further reinforces the notion that what the hippocampus contributes to reversal learning is a contextual cue to separate the two discriminations.

To model reversal learning, we constructed a simple model of discrimination learning (Figure 7.6). The discrimination begins at the start state, denoting the availability of two options (A and B). The start state leads to either of two states reflecting the rat's choice. The choice states in turn lead to reward and no reward states, indicating the outcome of that choice. (Such reward states are justified by studies showing reward-related responses in the hippocampus; Tabuchi et al., 2003; Smith and Mizumori, 2006.)

7.4.1 *Serial reversal learning*

To simulate serial reversal learning training, alternating blocks of 10 trials were generated in which, for odd blocks, choice A was rewarded, and, for even blocks, choice B was rewarded. To simulate behavioral learning, the rat's choice was selected randomly such that, during the first block, the probability of selecting the correct choice exponentially approached 58.5% from 50%; during the second block, the correct choice probability exponentially approached 62% from 41.5% (100% - 58.5%); during the third block, the correct choice probability exponentially approached 65.5% from 38%. This continued for 10 blocks, increasing the final correct choice probability in each successive block by 3.5%. The exponential "decay" of choice probabilities within each trial approximates the trial-by-trial error data of Brunswick (1939).

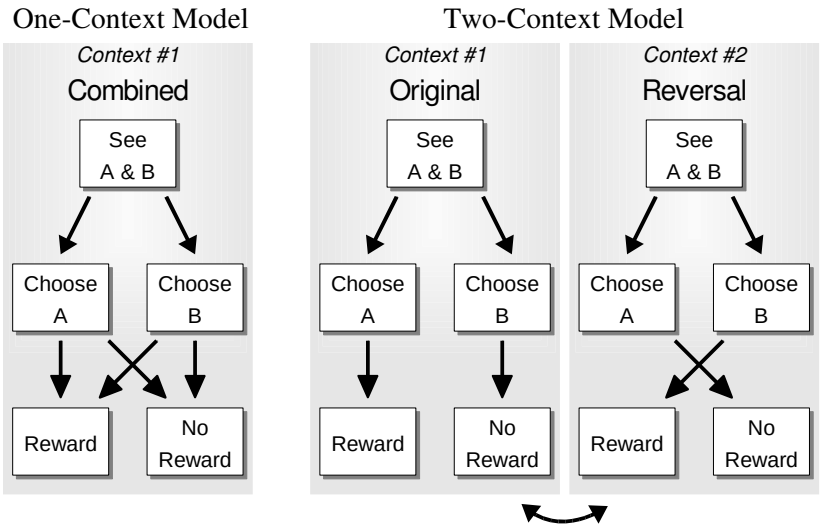

FIGURE 7.6: Diagram of one- and two-context models used to model reversal learning. Arrows within each context indicate transition probabilities that are typically high; however, transitions between any two states are possible.

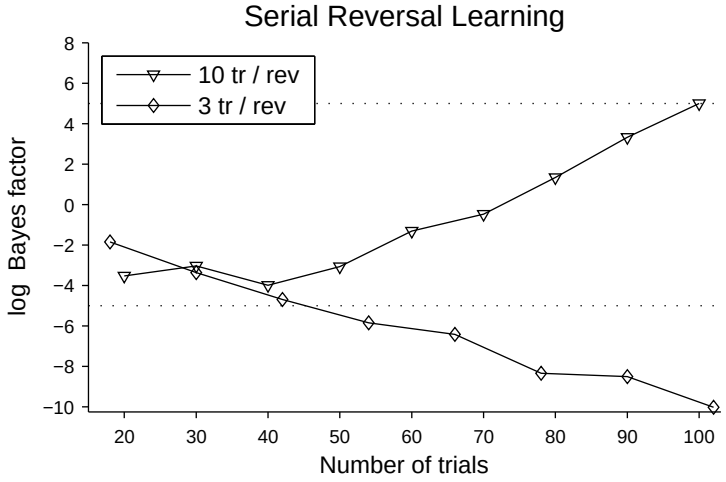

FIGURE 7.7: Serial reversal learning simulation: whereas rapid reversals are indistinguishable from partial reinforcement, less frequent reversals lead to the adoption of a separate context for each reward contingency.

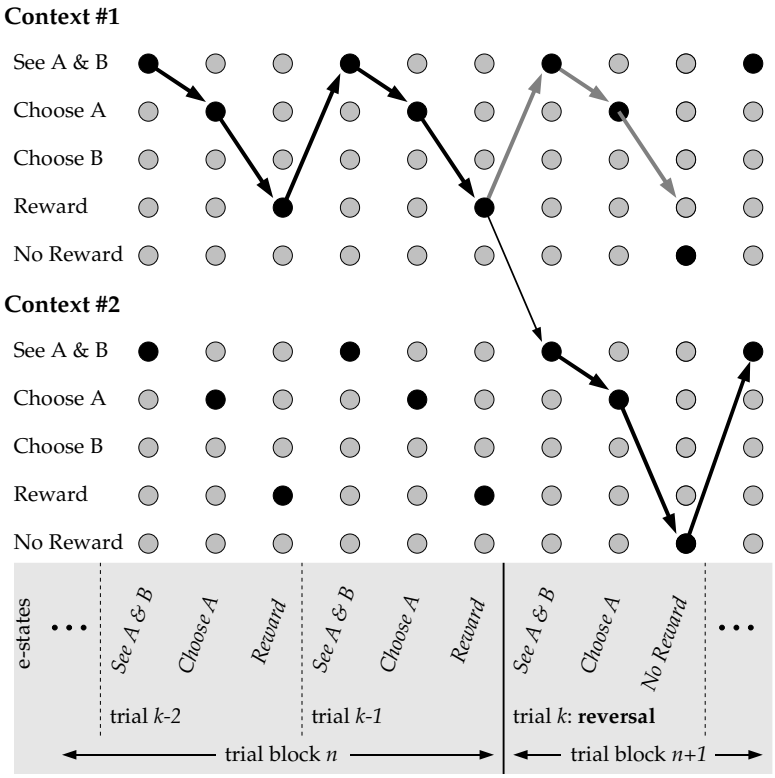

FIGURE 7.8: Context selection in the two-context model. Each circle represents a state in the HMM, and each column represents the full set of states of the two-context model at a particular point in time. The gray box shows e-states (hippocampal inputs) from three trials. Trials $k-2$ and $k-1$ are at the end of a block of trials in which choice A is rewarded. Trial k is the first trial of a new block in which choice B is rewarded. HMM states that are probable given the hippocampal input are darkened. With full reinforcement, only one choice in each context is associated with reward. Therefore, while the gray path (remaining in Context #1) initially appears to be more likely, the lack of reward indicates that trial k is actually part of Context #2. Once the context switch has been inferred, the most likely path through the states will continue within Context #2 until choice B is no longer rewarded.

One- and two-context models, shown in Figure 7.6, were compared as the number of training blocks was increased, and the results are shown in Figure 7.7. With increased training, the likelihood of the two-context model gradually increases, predicting a gradual adoption of context-specific hippocampal firing patterns. Smith and Mizumori (2006) showed that context-specificity developed during reversal learning training, though they do not examine the time course of remapping across training sessions. Figure 7.8 illustrates context selection in the two-context model, demonstrating how a context switch can be inferred after a reversal.

If the number of trials per block is substantially reduced, the reward structure becomes difficult to distinguish from partial reinforcement. When training sequences were generated comprising three trials per block, the two-context model required too many context switches to be justified, since context switches are constrained to be unlikely. This resulted in the progressively decreasing likelihood of the two-context model, as shown in Figure 7.7.

7.4.2 *Partial reinforcement and reversal*

In order not to conflate the effects of partial reinforcement during the original and reversal discriminations, the training paradigm typically applies partial reinforcement to choices during the original discrimination, leaving the reversal condition unambiguous (Wike, 1953; Grosslight et al., 1954; Elam and Tyler, 1958; Wise, 1962; Pennes and Ison, 1967). We consider three cases: full reinforcement (100:0), in which choice A is always rewarded and choice B is never rewarded; partial reinforcement of A (75:0), in which choice A is rewarded only 75% of the time; and partial reinforcement of A and B (75:25), in which choice B is also rewarded 25% of the time. In all cases, only choice B is rewarded during the reversal discrimination (0:100).

Training consisted of a single original and single reversal block, the length of each being equal and set so that the Bayes factor would be roughly 5.0 by the end of training on both blocks. The rat's choice was selected randomly, where the probability of a correct choice exponentially approached 90% from 50% in the original block, and 90% from 10% in the reversal block.

Figure 7.9 shows that, as the original discrimination is made more ambiguous, learning to differentiate it as a separate context from the reversal discrimination requires progressively more trials. We therefore predict that context-specific hippocampal representations would develop more slowly in these partial reinforcement paradigms.

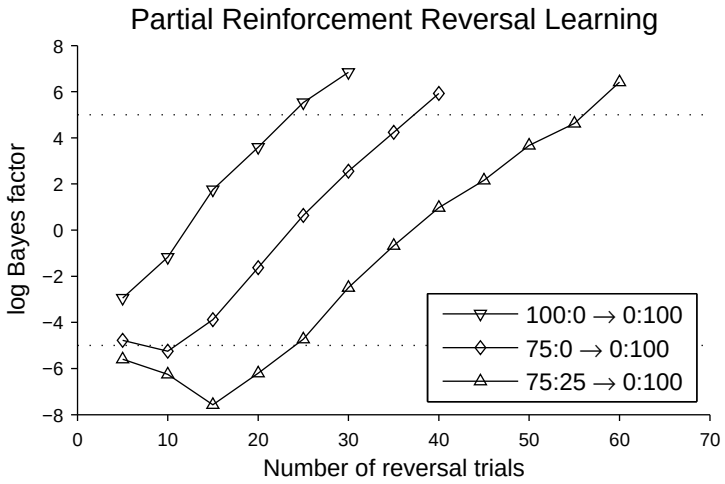

FIGURE 7.9: Partial reinforcement reversal learning simulation results show that a two-context model is adopted more slowly when the two reward contingencies are more similar.

7.5 Simulation Results: Sequence Learning

The hippocampus has been implicated in a variety of tasks involving sequences (Kesner and Novak, 1982; Chiba et al., 1994; Gilbert et al., 2001; Agster et al., 2002; Fortin et al., 2002; Kesner et al., 2002). For example, Fortin et al. (2002) found that, after being presented with a random sequence of odors, hippocampal rats could not choose the earlier of two odors from the sequence. Of particular interest has been the study of overlapping sequences, in which multiple sequences share common middle elements but distinct beginning and ending elements. Agster et al. (2002) found that, while hippocampal rats could disambiguate two partially overlapping olfactory sequences, inter-trial interference or a delay condition could impair their performance substantially.

Several models have proposed that the hippocampus develops separate contextual representations of each sequence that serve to associate the ambiguous middle elements with the rest of the sequence (Levy, 1989, 1996; Wallenstein and Hasselmo, 1998). These models predict that, if a rat were to repeatedly travel down a common maze arm that was part of two different paths (e.g. a continuous figure-8 pattern), place cells on the common arm would fire differently depending upon which path the animal was traveling.

Some studies have confirmed this finding (Frank et al., 2000; Wood et al., 2000) while others found no path-related differences (Lenck-Santini et al., 2001; Hölscher et al., 2004). Intriguingly, Bower et al. (2005) were able to

reproduce both cases by varying the shaping procedure used to train the rats. The studies that found no path-related differences nonetheless reported that rats were able to learn the task, a result consistent with Ainge and Wood (2003) who found that hippocampal lesions did not impair the continuous version of the task. However, when a small delay was added at the start of the common maze arm, Ainge and Wood found that hippocampal rats were impaired. Subsequently, two groups trained unlesioned rats on a figure-8 maze, each finding, paradoxically, that path-specific modulation of place fields on the common maze arm occurred with no delay (when the task was not hippocampus dependent) but largely disappeared when a delay was added (Ainge et al., 2005; Robitsek et al., 2005). Despite this disappearance, rats could still perform the task. As Bower et al. (2005) point out, the sequence dependence of place cell activity is likely attributable to differing input from some extra-hippocampal brain area instead of a contextual representation developed within the hippocampus.

An alternative hypothesis, adopted by Hasselmo and Eichenbaum (2005), is that sequence replay (Foster and Wilson, 2006) in the hippocampus guides behavior independently of whether or not path-specific remapping is seen on the common arm. Specifically, an extra-hippocampal brain area incrementally learns rules for completing each sequence, given its beginning; replay of the beginning of the current sequence is then sufficient to complete the task. (With a higher learning rate, Hasselmo et al. (2002) would likely provide an elegant model of such one-shot learning.) Such a division of labor explains why Agster et al. (2002) found evidence of rats learning overlapping sequences but failing, under some circumstances, to recall which sequence had most recently begun.

The rapid alternation between sequences on a figure-8 maze is incompatible with our constraint that context switches should be rare. To demonstrate this, we constructed a simulation of the figure-8 task. Five positions on the track were modeled, and were traversed in a six-step loop: Start Left, Center, End Right, Start Right, Center, End Left, repeat. A trial constituted a pass through one start arm, the center arm, and the opposing end arm. One- and two-context models (see Figure 7.10) were compared, and the results are shown in Figure 7.11. Even after just 12 trials, the two-context model is astronomically unlikely.

To demonstrate that the context switch penalty in the two-context model is the specific cause of the low Bayes factors, the one-context model was also compared with the generative model (see Figure 7.10). Figure 7.11 shows that decisive preference for the generative model is attained by 30 trials.

Bower et al. (2005) have considered training regimes that promote sequence

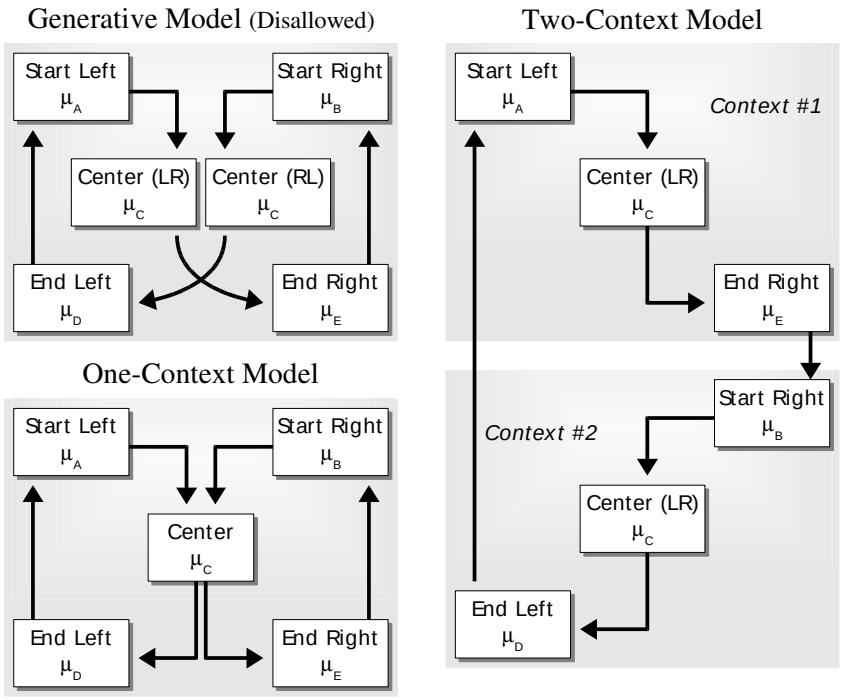

FIGURE 7.10: Diagram of the generative model and one- and two-context models used to model sequence. Note that the generative model includes two states (Center LR and Center RL) in the same context that represent the same distribution of input patterns (identical μ_C values). Our weak-identifiability constraint (see Statistical Framework: Overview) that such states can be represented only when in separate contexts excludes the generative model from the class of possible hippocampal context models. Arrows within each context indicate transition probabilities that are typically high; however, transitions between any two states are possible.

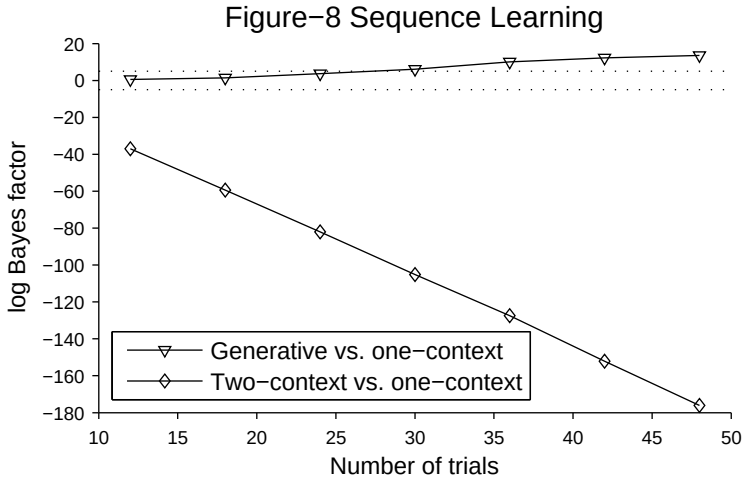

FIGURE 7.11: Sequence learning simulation results show that the one-context model is overwhelmingly more likely than the two-context model, due to the constraint that transitions between contexts are of low probability. If the generative model were admissible, it would be adopted over the one-context model.

disambiguation of the common path, presumably due to afferent input from another brain region. However, were rats to be trained repeatedly on one sequence, then the other, our framework would predict that sequence-specific encodings of the common path would reliably develop due to intra-hippocampal mechanisms sensitive to temporal mismatch.

7.6 DISCUSSION

7.6.1 *Significance features of the theory*

Advances in Bayesian computational statistics techniques over the last decade have opened the door to evaluating the marginal model probabilities of many new classes of models. Using these techniques, insights into animal learning and human reasoning have begun to arise from their formulation as Bayesian inference problems, often over multiple generative models that provide competing explanations of a corpus of data (Tenenbaum and Griffiths, 2001; Courville et al., 2003, 2004; Griffiths and Tenenbaum, 2005; Daw et al., 2005). In a similar vein, the present work advances a theory of context learning in which hippocampal input patterns are grouped together in the same context when they reliably cluster together in time. Choosing between generative models that group the experiences into different numbers of clusters is therefore the fundamental challenge of context learning.

Sequence learning models (Levy, 1989, 1996; Wallenstein and Hasselmo, 1998) provide a rather different notion of context, one oriented toward binding together elements of a temporal sequence. Interestingly, this binding process simultaneously serves to differentiate elements based on their order within the sequence. By contrast, in our theory, context learning groups all sequence elements together without disambiguating multiple occurrences of the same element. Thus, sequence learning models predict that alternating, overlapping sequences should be represented with different contextual bindings, whereas our theory groups them together into one context. The failure for sequence-dependent hippocampal representations to be consistently observed (e.g. Bower et al., 2005) or to serve a behavioral function (Ainge and Wood, 2003) argues against such representations playing a critical role in the disambiguation of alternating, overlapping sequences. Others have argued that such sequential encodings are instead formed in medial temporal lobe areas outside the hippocampus proper such as the entorhinal cortex (Howard et al., 2005).

An important aspect of this theory is that it distinguishes between the inference processes of context learning and context selection: context learning may be gradual over many days, while context selection should be an abrupt process. Multi-basin attractor models (Samsonovich and McNaughton, 1997; Redish and Touretzky, 1998; Tsodyks, 1999; Doboli et al., 2000) have demonstrated how multiple contexts, each a stable basin of attraction, could be simultaneously represented within the same network. In such models, context selection involves restabilization of the network in the most appropriate basin. What these models lack is an explanation of the gradual development of new contextual representations.

By contrast, backpropagation models (Schmajuk and DiCarlo, 1992; Gluck and Myers, 1993) have attempted to address the gradual development of new (hidden layer) representations. However, these networks do not have multi-stable activity patterns; they have no "memory" of the current context across time beyond what is encoded in the weights. Thus, they fail to capture the one-trial context switching behavior observed in reversal learning. In addition, backpropagation and similar delta rule learning models do not properly adapt learning to the information provided by each trial and therefore cannot account for the slower reversal of partial reinforcement reward contingencies.

Finally, while backpropagation is an efficient search technique for learning high dimensional mappings, the complexity of the function represented by the neural network is not explicitly considered. (In a machine learning setting, overcomplexity can result in "overfitting," which is typically detected via

cross-validation of the model on a separate data set.) Similarly, independent component analysis (used in a hippocampal model by Lörincz and Buzsáki, 2000) does not adjust the number of independent components based on the observed data. In contrast, the present framework considers the inherent trade-off between model fit and model complexity that underpins any model selection problem. Our framework can most clearly be dissociated from other models that do not consider this trade-off by determining whether remapping is observed subsequent to training in a set of morph boxes: our simulations predict that remapping should not occur.

7.6.2 *Localizing contextual representations within the hippocampus*

Since our theory concerns contextual representations in the hippocampus, we discuss how it relates to known facts about hippocampal anatomy and physiology. First, we argue that abrupt and gradual remapping are mediated by distinct physiological processes. Specifically, whereas abrupt remapping is caused by a change in the path integrator representation located in dorsal medial entorhinal cortex (dMEC), gradual remapping is caused by experience-dependent representational changes within the DG / CA3 network. Then, we discuss the evidence for pattern separation and completion in the DG / CA3 network and how such mechanisms could underpin context learning. We contrast the role of the DG / CA3 network with the role of CA1 in gating the projection of the DG / CA3 representation to efferent cortical areas.

Abrupt remapping and attractor dynamics. Marr (1971) first proposed that the architecture of the hippocampus is well suited to encode new memory traces using orthogonalized representations, which minimize interference between stored patterns. Recordings from hippocampal pyramidal cells confirmed that sparse, orthogonalized representations are formed to encode different places and other features within an environment (O'Keefe and Dostrovsky, 1971; Wood et al., 1999), as well as different environments as a whole (Muller and Kubie, 1987). Attractor models of hippocampal function (Samsonovich and McNaughton, 1997; Redish and Touretzky, 1998; Doboli et al., 2000), which address the remapping of place fields between environments, were born out of the observation that, when place cells remap, they appear to remap together. For example, Bostock et al. (1991) found that the change of a cue card's color sometimes caused a remapping, and, when observed, all simultaneously recorded cells remapped. When introduced repeatedly to the same environment, rats with deficient LTP sometimes recalled a completely different map (Barnes et al., 1997; Kentros et al., 1998).

One difficulty in interpreting such remapping data is in disambiguating

the role of the path integrator, currently believed to be in dorsal medial entorhinal cortex (dMEC) (Fyhn et al., 2004; Hafting et al., 2005; Fuhs and Touretzky, 2006; McNaughton et al., 2006; see Chapter 6), from the place code. As noted by Touretzky and Redish (1996), failure to reset the path integrator should cause a substantial change in the afferent input to the hippocampus, resulting in abrupt hippocampal remapping independent of any attractor dynamics in the hippocampus. This reset failure likely underpins the results of Leutgeb et al. (2005c), who showed that switching between two rooms causes a different and more radical form of remapping than switching between arena shapes in the same room. While dMEC grid cells change phase when the arena is placed in a novel room (Hafting et al., 2005), preliminary findings by Leutgeb et al. (2006a) show that grid cell phases remain constant when the arena changes shape in the same room, while DG and CA3 undergo rate remapping (see also Quirk et al., 1992). Thus, abrupt remapping, even when delayed from the first exposure to the environment (e.g. Bostock et al., 1991; Brown and Skaggs, 2002), is likely attributable to a reset failure of the path integrator which causes remapping simultaneously in all subfields of the hippocampus. By this interpretation, delayed abrupt remapping reflects a stochastic process where, for those rats that attend to the environmental change, there is a fixed probability of PI reset failure upon each visit to the perturbed environment, which should result in an exponentially distributed time to first remapping.

Gradual remapping, context learning and the DG / CA3 network. In contrast to abrupt remapping, the gradual differentiation of contextual representations should be attributable to circuitry within the hippocampus, specifically DG and CA3. Several authors have suggested that exposure to a novel context results in an orthogonalized representation being formed in DG which is then propagated to CA3 (Marr, 1971; McNaughton and Morris, 1987; Treves and Rolls, 1992, 1994; O'Reilly and McClelland, 1994). Preliminary evidence suggests that the rate remapping between similar contexts observed by Leutgeb et al. (2005c) originates in DG (Leutgeb et al., 2006a). In a familiar context, perforant path input and recurrent collaterals in CA3 guide the recall of the previously learned contextual representation. O'Reilly and McClelland (1994) explored a simplified model of DG and CA3, showing that, while similar patterns could be mapped to an even more similar CA3 representation (pattern completion), more strongly differing input patterns could be separated as well (pattern separation).

The neural mechanisms of pattern separation and completion resemble the statistical process of clustering: map noisy input patterns into more similar

representations to denote their association with the same cluster; map input patterns associated with different clusters into more distinct representations. One might therefore think of a clustering neural network as an extension of the O'Reilly and McClelland (1994) model in which the threshold between separation and completion is not static, but dynamically adjusted based on the distribution of input patterns. While O'Reilly and McClelland only elucidated the benefits of CA3 perforant path plasticity, NMDA-dependent synaptic plasticity is well known to exist within DG and the CA3 recurrent collaterals. In addition, there is intriguing new evidence that mossy fiber synapses show heterosynaptic plasticity, though changes in synaptic efficacy appear to depend on neighboring synapses in stratum radiatum instead of the depotentiation of the post-synaptic cell (Schmitz et al., 2003). We propose that one function of this learning is to adjust the separation-completion threshold, gating when new contextual representations would be propagated from DG to CA3.

In a familiar context, once the path integrator and any other brain state is reset, the activity patterns projected from DG onto CA3 should match the patterns projected from the perforant path and recurrent collaterals, modulo some noise. However, if this familiar context is perturbed into a similar but distinct second context, then there should be some mismatch between the pathways: the representation in DG should more accurately reflect the current input patterns, whereas the perforant path and recurrent circuitry should reinforce a previously learned representation. Critically, our framework suggests that if the differences between recalled and input patterns are small, the impact of the DG representation should be minimal, and, if such differences do not repeat, the impact of DG should remain minimal. However, if the same differences are observed repeatedly, the impact of DG on the CA3 representation should increase, causing remapping in CA3 (Fuhs and Touretzky, 2000). In this way, more complex context models may be adopted.

Small but repeated differences would be expected to cause incremental changes in synaptic connectivity in the DG perforant path and mossy fiber pathway to strengthen the impact of DG on CA3. Interestingly, perforant path plasticity in DG can last for months (Abraham et al., 2002), suggesting that this pathway is capable of accruing changes in synaptic efficacy over many training days. However, granule cell neurogenesis causes cells to be gradually replaced, slowly fading away previous associations (Feng et al., 2001). These opposing forces provide a natural balance for input pattern density estimation. New input patterns can be registered and their impact can be strengthened with repeated exposure; however, this strengthening is

tempered by the replacement of granule cells to clear out old memories. It follows from this proposal that DG principal cells should fire at a lower rate in a novel environment and, with repeated exposure, gradually increase their rates as the environment becomes familiar.

While presenting a neural network model of context learning is beyond the present scope, recent physiology, gene expression, and lesion studies are consistent with the proposal that a neural instantiation of context learning should be localized to the DG / CA3 network, whereas CA1 integrates the experience-dependent DG / CA3 representation with the entorhinal cortical representation. Both pattern separation and pattern completion have been observed in CA3 in response to changes in environmental cues (Vazdarjanova and Guzowski, 2004; Lee et al., 2004b; Leutgeb et al., 2004; Guzowski et al., 2004). Vazdarjanova and Guzowski (2004) found in an immediate-early gene expression study that similar environments were represented with a greater similarity in CA3 than CA1, while two very different environments were represented with less similarity in CA3 than CA1. Both Lee et al. (2004b) and Leutgeb et al. (2004) found that the CA1 representation more directly reflected the current sensory environment, whereas the CA3 representation reflected either pattern completion of a previously learned contextual representation (Lee et al., 2004b) or pattern separation to create a new contextual representation (Leutgeb et al., 2004).

A behavioral role of CA3 pattern completion is suggested by Nakazawa et al. (2002), who found that performance in a cue-degraded version of the Morris water maze is impaired by CA3 NMDA receptor knockout. Lee and Kesner (2002; 2003) showed that delayed non-match to place (DNMP) was impaired in a novel (but not familiar) environment by CA3 NMDA inactivation or DG or CA3 lesion. These deficits may be interpreted as a failure to recall or learn a conjunctive representation of position and target (hidden platform, object, etc.) that could be retrieved via pattern completion using the target as an autoassociative memory cue. (Evidence for such a target representation has been found in prelimbic / infralimbic cortex; Hok et al., 2005.) Consistent with this hypothesis, DG appears necessary to create such conjunctive representations: DG lesions reduce performance to chance on a working-memory water maze task in which the platform is moved to a new location each day (Xavier et al., 1999). In relation to our theory, these data point to DG and CA3 to construct a model-based representation of the animal's experiences, including various conjunctive associations instrumental in solving behavioral tasks.

CA1 appears to relay a composition of the model-based CA3 representation and the context-free entorhinal cortical information to efferent cortical areas.

Hasselmo and colleagues have presented a series of models and pharmacological data supporting the notion that increased cholinergic modulation decreases the contribution of CA3 to the CA1 representation, but increases plasticity of both the CA3 recurrent and Schaffer collaterals (Hasselmo and Schnell, 1994; Hasselmo et al., 1995, 1996). More recently, Yu and Dayan (2005) have proposed a theory of acetylcholine and noradrenaline in which acetylcholine represents expected uncertainty, whether due to the context being new or to known unpredictability within a familiar context. Taken together, these models suggest that CA1 should be influenced by CA3 when the context is informative (low ACh); CA3's influence should be reduced when the context is less informative (high ACh). This gating of the CA3 representation has been confirmed experimentally in novel environments: CA1 shows a stable representation while the CA3 representation evolves over the course of 20 to 30 minutes (Leutgeb et al., 2004). Additionally, several monoaminergic neurotransmitters have been implicated in modulating the balance of CA3 and EC input to CA1 (Otmakhova and Lisman, 1998; Otmakhova et al., 2005).

The complementary roles of the DG / CA3 and CA1 networks provide some insight with which to interpret discrepancies between the double rotation experiments of Shapiro et al. (1997) and Lee et al. (2004b). When local and distal cues were rotated in opposite directions, both studies observed "heterogeneous" or "discordant" responses. However, Shapiro et al., recording mostly from CA1 at the beginning of the experiment, initially observed many more place fields to rotate with the distal cues than the local cues. In fact, the ratio of place fields rotating with each set of cues observed by Shapiro et al. much more closely resembles the ratio of place fields in CA3 tied to each set of cues observed by Lee et al. (2004b), suggesting that, in the Shapiro et al. study, CA3 strongly influenced the representation in CA1. Shapiro et al. repeatedly trained rats on two standard condition sessions and a single double rotation session (local and distal cues rotated 180º apart) each day, as well as various other less frequent cue scrambling and deletion probe trials. They observed that, over time, cells (predominantly in CA3 by the end of the experiment) more strongly remapped between standard and double rotation conditions. This change in the degree of remapping could reflect either the change in cell populations they recorded from or an experience-dependent effect; they do not address this issue statistically. Nonetheless, if we assume that the initial primacy of distal influence on CA1 place cells reflects pattern completion in CA3, then the increase in remapping reflects an experience-dependent transition in CA3 between pattern completion and pattern separation in order to adopt separate contexts for the standard and

double rotation cue configurations. This is consistent with our simulations of gradual remapping, which show that two repeatedly experienced and distinct conditions should be differentiated by context. By contrast, Lee et al. trained rats equally on four rotation angles in addition to the standard condition, a training paradigm more akin to our morph experiment simulations. Though the relatively short duration of training by Lee et al. prevents any decisive conclusions, their observation that CA3 maintained pattern completion throughout the course of their experiment (I. Lee & J. J. Knierim, personal communication) is consistent with our morph experiment simulation results which predict that experience-dependent remapping should not occur in such a case.

8

SPACE AND CONTEXT IN TWO IDENTICAL BOXES: A PLACE CELL STUDY

An earlier version of this chapter appeared in the Journal of Neurophysiology as "Influence of path integration versus environmental orientation on place cell remapping between visually identical environments." by Mark C. Fuhs, Shea R. VanRhoads, Amanda E. Casale, Bruce McNaughton and David S. Touretzky (2005, vol. 94, pp. 2603–2616; reprinted with permission). The experiment was performed at the University of Arizona, Tucson by Shea R. VanRhoads, Amanda E. Casale and Bruce McNaughton. I carried out the analysis of the data.

8.1 Introduction

This chapter presents the work of Fuhs et al. (2005), in which the Skaggs and McNaughton experiment is revisited and extended. Skaggs and McNaughton (1998) measured the effects of sensory-vestibular conflict between two putatively different contexts, two boxes connected by a corridor. Since rats walked between the boxes, vestibular cues alone should clearly differentiate the boxes; however, the boxes were visually identical, so, based on sensory cues alone, the two boxes were the same. What they observed was a partial remapping between the boxes, which they interpreted as a representation of a conflict between a linear path integrator and the sensory cues. In short, the linear path integrator indicated that the boxes occupied two different spaces, but the sensory cues (and behavioral task, random foraging) indicated that they were the same environmental context.

Section 4.1.3 reviewed the effects of sensory-vestibular conflicts caused by translational manipulations. In general, the impact of the linear path integrator was limited. For example, in the linear track contraction experiments (Gothard et al., 1996a, 2001; Redish et al., 2000; Rosenzweig et al., 2003), the place code invariably realigned with the room's distal landmarks after a limited maintenance of place field firing relative to the start box location. By contrast, sensory-vestibular conflicts caused by rotational manipulations have a strong effect on hippocampal place cell firing (for review, see Section 4.1.2), motivating the present experiment.

In the present experiment, rats were retested in a two-box apparatus designed to match that used by Skaggs and McNaughton; this is referred to

as the same-orientation condition, as the two boxes were placed in the same orientation in the room. The experiment was extended with a second condition, the opposite-orientation condition, in which the boxes were abutted entrance-to-entrance, one box rotated 180° with respect to the other (see Figure 8.1). In this condition, there is conflict between the sensory cues and both the linear and angular path integrators.

In contrast to the findings of Skaggs and McNaughton, the present experiment found no remapping in the same-orientation case. However, a complete remapping between the same boxes in the opposite-orientation case was eventually observed. Interestingly, the onset of remapping, though abrupt, varied between rats. The time course of remapping is explored in detail. Also, some preliminary evidence is discussed in the Discussion section that sheds light on a possible source of the discrepancy between the results of the Skaggs and McNaughton study and the present one.

8.2 Methods

8.2.1 *Subjects and apparatus*

Three male FBNF1 rats (Harlan Sprague Dawley), age 9-13 mo, were used in this experiment. The apparatus consisted of two identical boxes similar to those in the Skaggs and McNaughton (1998) experiment. The boxes were 60 $\times$ 60 $\times$ 60 cm in size, with a trapezoidal doorway in one wall whose width progressed from 15 cm at the bottom to 33 cm at the top. (The opening was wider at the top to permit free travel of the recording cable.) In the same-orientation condition, the boxes were placed side by side with identical orientations, such that the prominent features of each box (a light and a doorway) faced the same direction. The doorway panels of each box were connected to one another by a 15-cm-wide hallway, as shown in Fig. 8.1A, same orientation. An attempt was made to keep the rats' visual experiences in the two boxes as nearly identical as possible. The room lights were kept off. Distally, black curtains devoid of cues encircled the boxes. To obscure any overhead distal cues, a single dim light (0.9 W) surrounded by a semitransparent shade was fixed in each box near the top of the inner wall opposite the box doorway. These lights were adjusted so that they were similarly bright in each box and cast similar faint shadows on the box walls and floor. The boxes were placed such that the commutator-to-headstage cable fell in the exact center of the entire two-box/hallway apparatus. To prevent the rat from experiencing disparate angular force on his head and neck from the tug of this cable in the two boxes, a triangular dowel-rod

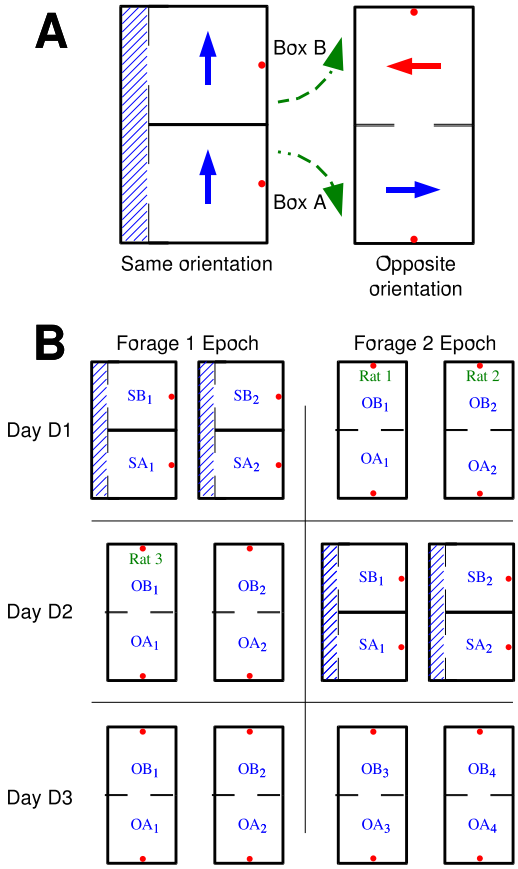

FIGURE 8.1: Summary of experimental manipulations. A) In the same-orientation condition, the boxes were connected by a corridor; in the opposite-orientation condition, the corridor was removed and the boxes were rotated and abutted. Red dots indicate lights mounted on the walls. B) Each recording session contained 2 foraging epochs, consisting of 2 visits to each box, giving a total of 8 visits per day. Order of visits within an epoch was always , A_1, B_1, A_2, B_2. Notation SA_1 describes a visit in the same-orientation configuration, box A, visit 1. Green "Rat n" tags in the top right corner of certain boxes indicate the visit in which a rat first showed remapping. (Reprinted from Fuhs et al., 2005.)

guide track was attached to the top of each box near the doorway. A fresh, large sheet of brown packaging paper was placed under the entire two-box/hallway structure at the beginning of each recording session, to prevent scent markings or other rat-produced floor cues from accumulating. The boxes, hallway, and all exposed nuts and bolts were painted a uniform dark gray to minimize spurious reflections in the tracker data. The positions of the boxes were outlined in tape on the recording room floor to ensure consistent alignment.

8.2.2 *Recording*

Rats were implanted and recordings made using the procedures described in Skaggs and McNaughton (1998). Neuronal spike signals from the dorsal CA1 pyramidal layer were amplified by a factor of 1,000 to 5,000, band-pass–filtered between 600 Hz and 6 kHz, and transmitted to a Cheetah Data Acquisition system (Neuralynx). Signals crossing a minimal threshold, set just above background noise levels, were digitized at 32 kHz and sampled for a duration of 1 ms, beginning 0.25 ms before the spike peak. A cluster of light-emitting diodes was mounted on the headstage to allow position tracking by means of a video camera that was placed directly above the experimental apparatus and recorded with a sampling frequency of 60 Hz.

8.2.3 *Unit isolation*

Putative single neurons were isolated based on the relative action potential amplitudes and principal components from the four tetrode channels (Gray et al., 1995; McNaughton et al., 1983; Wilson and McNaughton, 1994), by means of a semiautomatic clustering algorithm (BBClust, author: P. Lipa). The resulting classification was corrected and refined manually with dedicated software (MClust, author: A. D. Redish), resulting in a spike train time series for each of the discriminated cells. Only single neurons satisfying standard criteria (e.g. Skaggs and McNaughton, 1998) for hippocampal pyramidal cells were included.

8.2.4 *Protocol*

Rats underwent 19 to 26 recording sessions, one per day. Each session was composed of five epochs: Sleep 1 (about 30 min), Forage 1 (16-20 min), Sleep 2 (about 40 min), Forage 2 (16-20 min), and Sleep 3 (about 30 min). On odd-numbered days Box 1 was installed in position A for the first foraging epoch, then moved to position B for the second epoch. On even-numbered days the opposite was done. The swapping of boxes between epochs was done to help distinguish box-specific associations from location-specific associations.

The intervening Sleep 2 epoch allowed the rat to rest from foraging, and provided time for the box locations to be exchanged and other changes to be made to the apparatus, such as installing or removing the connecting corridor. The brown paper covering the floor was also replaced during Sleep 2.

Rats were not intentionally disoriented during this experiment. At the start of each day's session, they were transported from the colony room to the recording room along the most direct route possible. The rats were carried in a flowerpot lined with a towel. On reaching the recording room, they were deposited in another towel-lined flowerpot sitting on a cinderblock 6-12 in. away from the midpoint of the two-box apparatus. While in this flowerpot, the headstage was attached and the Sleep 1 epoch began. At the beginning of a foraging epoch, the rat was gently grasped by the experimenter, lifted from the flower pot, and deposited in box A as close to the doorway as possible, facing the center of the box. At the conclusion of a foraging epoch the rat was gently grasped with one or two hands, removed from box B, and placed back in the flowerpot to begin the next sleep epoch.

The rat's task during the two foraging epochs was always the same: to search for small chocolate pellets, one box at a time. Each epoch consisted of four 4.5-min box visits, starting in box A, moving to box B, then back to box A, and finally back to box B. During each visit, the rat was restricted to that box by a closed door. At the end of 4.5 min, the door would open and the rat would exit the box, run down the hallway, enter the other box, and resume foraging as the door closed behind it. In the opposite-orientation condition the boxes were separated by a single door, which was raised when it was time for the rat to travel to the other box. The door was then reversed before being lowered, so that the same side always faced the rat.

Rats were food deprived to promote foraging. They were offered water to drink from a 1-ml syringe at the beginning of Sleep 2. Small balls of slightly moist rat food were also sprinkled in the boxes in addition to the chocolate pellets.

For the first 16-23 days, the rats became familiar with the two-box environment and experienced a series of manipulations in the same orientation configuration. These included, in order: turning the lights off after the start of a trial (3-4 days), introducing a white cue card into one box to visually differentiate it from the other box (4 days; this proved ineffective), placing a white towel on the floor of one box (4 days), and starting the rat in Box B instead of Box A (at most 1 day). The data from these manipulations are not presented here. Starting on the 17th, 24th, or 18th day (for rats 1, 2, and 3, respectively), place cells were recorded as the rat experienced the 3-day

sequence of conditions shown in Fig. 8.1B. For commonality across rats, this sequence is referred to as days 1-3. Day 1 was a rat's first exposure to the opposite-orientation condition.

8.2.5 *Place field calculation*

For analysis purposes, each of the two foraging epochs was subdivided into four visits: first visit to box A, first visit to box B, second visit to box A, and second visit to box B. This yielded a total of eight box-visits per day, for each of which a 16×16 pixel spatial firing rate map, or place field, was calculated. To facilitate further analyses, each place field was smoothed by fitting it to a two-dimensional cubic B-spline (De Boor, 2001) with five segments along each axis. The spline, which is a nonparametric regression model of the firing rate map, was then used as the estimate of the cell?s true spatial firing rate in that visit. The 16×16 pixel maps for each box visit in Figures 8.3, 8.4, 8.5 and 8.10 are the result of this smoothing process.

8.2.6 *Cell selection*

The analysis began by finding the pyramidal cells with spatially selective firing fields. The inclusion of interneurons, which fire throughout the environment and have no discernible place fields, would have skewed the measure of remapping. For each smoothed place field, of which there were eight per cell (one per box visit), the spatial information content (SIC) was calculated according to the method of Skaggs et al. (1993). To be included in the analysis, a cell's SIC value had to pass two tests in two successive visits to the same box. For example, if the cell passed the tests in visits SA_1 and SA_2, or OB_1 and OB_2 (see Figure 8.1B for visit notation), it would be included.

The first test, called I_{min}, was an absolute measure of spatial information content: the SIC value had to be 0.15 bits/spike. This low threshold excluded only the interneurons. Interneuron firing can have some weak spatial correlation, but because of the high spike rate the number of bits per individual spike is low.

The second test was a visit-specific statistical measure: a place field's SIC value in a given visit was compared with a distribution of SIC values for 250 "randomized" fields computed from the same spike sequence, but with randomized position data. To generate these data, the rat's actual position data were reversed in time and then divided into 10-s windows. These windows were then randomly shuffled, resulting in a new temporal sequence of positions. After each shuffle, a spatial firing rate map was constructed. The SIC values for these 250 randomized maps formed a distribution of expected values under the null hypothesis that the cell was not spatially

Rat	Day	Raw Cells	Satisfied I_{min}	Satisfied $I_{95\%}$	Net Cells
1	1	7	6	7	6
	2	42	39	28	25
	3	16	15	13	11
2	1	55	49	45	41
	2	58	48	51	44
	3	33	22	27	19
3	1	5	4	3	2
	2	10	10	9	8
	3	9	6	7	4
Totals		235	199	190	160

TABLE 8.1: Summary of cells recorded in the experiment.

selective. The critical value $I_{95\%}$ for the visit was set at the 95th percentile of this SIC distribution. A cell passed the test in a given visit if the SIC of its unshuffled firing rate map exceeded this critical $I_{95\%}$ value. This test should classify fields as spatially selective with a type I error of 5%.

Whereas the I_{min} test measures a cell's intensity of firing rate variation across spatial locations, the $I_{95\%}$ test measures a cell's consistency of firing rate within spatial locations. The $I_{95\%}$ test therefore disqualifies two types of cells: those with unreliable fields whose firing rates are not strongly correlated with space, and those with low firing rates, for which the sample of spikes is too small to substantiate consistent spatially selective firing. In the extreme case of a cell with only a few spikes, those spikes could be from a single burst that occurred on one pass through the putative place field. In that case, shuffling the position data in 10-s windows will maintain the high SIC value, causing the cell to fail the $I_{95\%}$ test because shuffled sequences have information content comparable to the original.

Table 8.1 shows that for each epoch, the number of cells that satisfied both the I_{min} and $I_{95\%}$ tests was almost always less than the number that satisfied each test individually. Thus the tests are somewhat independent. If they are effective at selecting legitimate place cells, those cells should show a high degree of correlation between successive visits to the same box, illustrated in Figure 8.2. In Fig. 8.2A, cells that failed the $I_{95\%}$ test (95% shaded area) have substantially lower correlations between SA_1 and SA_2 visits than those that passed the test (5% unshaded area). The correlation between two smoothed place fields were found by calculating the pixelwise correlation coefficient of the fields, excluding undersampled pixels. In Figure 8.2B, excluded cells may occasionally have high correlation values because they are interneurons that

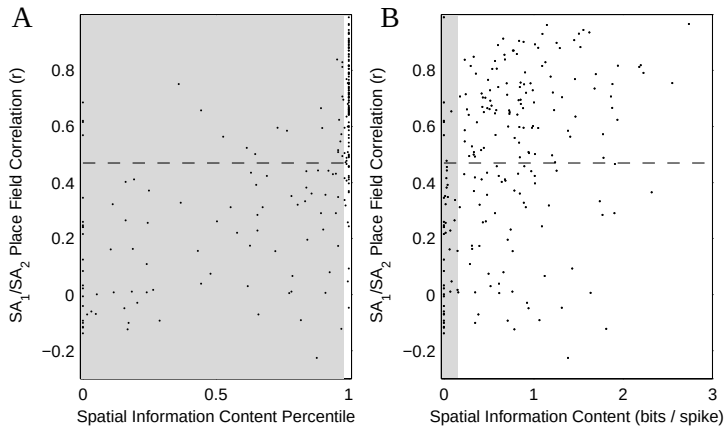

FIGURE 8.2: Two tests for spatial selectivity based on spatial information content (SIC): a 95th percentile ranking, and a fixed minimum value of 0.15 bits/spike. Each test is plotted against another measure of spatial selectivity: the correlation between place fields from successive visits to the same box. Data are from place fields recorded in the first and second visits to box A each day. 95th percentile of the distribution of correlations expected between remapped fields was plotted as a horizontal dashed line in each graph; values above this line indicate that the fields' correlation is unlikely to be coincidental. Gray areas indicate regions in which cells would not qualify for inclusion in the subsequent analysis. A) SIC percentile ranking was strongly indicative of field correlation between successive visits. B) Cells with very low SIC values (0.15 bits/spike) also had low correlations between successive box A visits, and were probably interneurons. (Reprinted from Fuhs et al., 2005.)

fire nearly everywhere, but they fail the I_{min} test because their SIC values are 0.15 bits/spike. The dashed lines in both figures mark the threshold for statistical significance of $SA_1 \times SA_2$ correlations, calculated by comparing the distribution of correlations across two visits to the same box with that for the Complete Remapping distribution, described below.

8.2.7 *ANOVA groups*

Our goal is to assess whether the hippocampus as a whole remapped between various pairs of visits, such as successive visits to the same box, or between box A and box B, or between same-orientation and opposite-orientation conditions. Because the cells' spatial firing patterns are not perfectly correlated even across visits to the same box (note the distribution of points along the x-axes in the plots in Figures 8.6, 8.7, 8.8, and 8.9), and because random changes in individual cells might be observed even when the hippocampus as a whole does not remap, it is necessary to construct statistical tests to formally assess the degree of correlation observed between visits. Given any two visits, one can measure the correlation of place cell firing fields between those visits for all the cells in the population, yielding a distribution of correlation values. Analyses of variance can then test whether this distribution is statistically different from other distributions of interest. A total of eight salient distributions were identified, summarized in Table 8.2.

The *$SA_1 \times SA_2$ No Remapping* distribution consisted of correlations between successive box A visits, with one exception: box B fields had to be used for the first day of rat 1 because of a recording problem during the SA_1 visit. No remapping was expected or qualitatively observed during successive visits to the same box, so this distribution was used as the prototypical No Remapping distribution.

The *$SA \times SB$* distribution consisted of correlations between the first box A visit and the first box B visit, with one exception: the second visits had to be used for the first day of rat 1, resulting from the aforementioned recording problem in SA_1.

The *$OA_1 \times OA_2$* distribution consisted of correlations between successive box A visits in the opposite-orientation configuration.

The *$SA \times OA$* distribution contained correlations between fields from the temporally closest visits to box A in each configuration, i.e., between SA_2 and OA_1 on day 1, and between OA_2 and SA_1 on day 2. Because the boxes in the opposite-orientation configuration were rotated 90° in the room frame with respect to the boxes in the same-orientation configuration, fields from the opposite-orientation configuration were first rotated 90° before correlations were calculated. However, cells recorded during the second day for rat

Distribution	Component Correlations
$SA_1 \times SA_2$ "No Remapping"	Rat 1: $SA_1 \times SA_2$ day 1 (see text)
117 cells	Rats 2, 3: $SA_1 \times SA_2$ day 1
	All: $SA_1 \times SA_2$ day 2
$SA \times SB$	Rat 1: $SA_2 \times SB_2$ day 1
117 cells	Rats 2, 3: $SA_1 \times SB_1$ day 1
	All: $SA_1 \times SB_1$ day 2
$OA_1 \times OA_2$	All: $OA_1 \times OA_2$ day 1
113 cells	All: $OA_1 \times OA_2$ day 1
$SA \times OA$	All: $SA_2 \times OA_1$ day 1
120 cells	All: $OA_2 \times SA_1$ day 2
$OX_2 \times OX_3$ Day 3	All: $OA_2 \times OA_3$ day 3
59 corr. from 35 cells	All: $OB_2 \times OB_3$ day 3
$OA \times OB$ "Before remap"	Rat 2: $OA_2 \times OB_2$ day 2
40 cells	Rat 3: $OA_1 \times OB_1$ day 1
$OA \times OB$ "After remap"	Rat 1: $OA_1 \times OB_1$ day 1
76 cells	Rat 2: $OA_2 \times OB_2$ day 1
	Rat 3: $OA_1 \times OB_1$ day 2
	All: $OA_1 \times OB_1$ day 3
"Complete Remapping"	All: Cell i $OA_1 \times$Cell j OB_1,
5113 cell pairs	where tetrode$(i) \neq$ tetrode(j)

TABLE 8.2: Distributions used in ANOVA tests for remapping. There were no statistical differences in the median correlation among the first six distributions (above dividing line) or the last two (below dividing line). However, the shaded and unshaded distributions differed significantly from each other. See Section 8.3 for details of the statistical analysis.

1 showed a stronger correlation without rotation, and therefore were not rotated before calculating the correlations reported here.

The $OX_2 \times OX_3$ *Day 3* distribution, where X could be either A or B, contained correlations between fields in the same box during the second visit of the first foraging epoch and the first visit of the second foraging epoch on day 3. This distribution measured consistency of fields across the intervening Sleep 2 period.

The $OA \times OB$ *Before remap* distribution contained correlations between boxes A and B in the opposite-orientation configuration before remapping was observed. The onset of remapping varied across rats; because rat 1's fields remapped on first entry to box B in the opposite-orientation configuration (see Section 8.3), it was excluded from this group. Fields in box B were rotated 180° before correlations were calculated.

The $OA \times OB$ *After remap* distribution contained correlations between boxes A and B in the opposite-orientation configuration after remapping had been observed. Fields in box B were rotated by 180° before correlations were calculated for the first 2 days of recording, but were not rotated on the third day.

The *Complete Remapping* distribution modeled the expected distribution of correlations between place fields when the hippocampus remapped completely. To estimate this distribution, cells' box A firing fields were correlated with the box B firing fields of other cells, on different tetrodes, within the same dataset. Because some cells had fields in only one box, this estimate includes cases where a cell remaps by gaining or losing a field between boxes.

Once again, these eight groups are distributions of correlations between two place fields for a cell. When computing these distributions, if neither field was spatially selective (SIC value failed to exceed $I_{95\%}$ and I_{min}), that correlation was excluded from the distribution because the cell was considered uninformative about the similarity of representations in the two environments.

For the bulk of the analysis, correlations were measured by calculating the Pearson r correlation coefficient. This requires first normalizing each place field by its variance. A drawback of this approach is that if a cell remaps by changing only its firing rate and not the shape of its firing field, the correlation between normalized fields will be high and no remapping will be detected. Such "rate remapping" has been reported by Lever et al. (2002) and Hayman et al. (2003). Therefore to test for systematic rate differences across conditions, we also computed a second set of eight distributions using

the following difference metric in place of the Pearson r correlation

$$D = \frac{\sum_x |f_1(x) - f_2(x)|}{\sum_x |f_1(x)| + |f_2(x)|} \tag{8.1}$$

where the variable x ranges over map locations, and $f_1(x)$ and $f_2(x)$ are two place field firing rate maps that have been zero-normalized by subtracting off their respective mean firing rates. Intuitively, the metric calculates the ratio of the difference between rate maps to the difference between each rate map and its mean. The metric equals 0 for identical fields and rises toward a maximum of 1 when fields are either not well spatially aligned or have substantially different firing rates.

8.2.8 *P-HMM reconstruction*

By averaging over several minutes, rate maps provide a robust estimate of place cells' overall spatial firing patterns, which permits the discovery of heterogeneous behavior within the cell population, such as varying degrees of partial remapping. We also performed a complementary analysis, population reconstruction, which assumes homogeneity within the population to reconstruct a variable of interest, such as the animal?s location, from the activity of many cells within a small window of time.

Our population reconstruction analysis used three box visits during which the same cells were recorded: two Reference visits, one from each box, between which remapping was observed, and a Test visit which could be from either box. The Test visit was divided into time windows of 250 ms and, for each window, t, a spike count vector $N_t(1 \ldots C)$ was created containing the number of spikes recorded from each cell, $1 \leqslant c \leqslant C$.. The results of the analysis then indicated whether the spike count vector during each Test time window was more likely to have been generated by the box A or box B map, or whether the cells were inactive.

A Poisson hidden Markov model (P-HMM) was constructed from the rate maps of each pair of Reference visits to provide a model of the expected distribution of spike count vectors in either box. The P-HMM used three states: Box A, Box B, and Quiescent. The state transition probabilities were weighted strongly in favor of remaining in the same state ($p = 0.97$), which was intended to reflect the idea that transitioning between maps occurs much less frequently than traveling within a map. Transitions from a state to either of the other two states had a probability of 0.015. The P-HMMs were defined to output spike count vectors of length C.

It was assumed that spikes from a place cell occurred according to a Poisson

distribution whose mean depended on the rat's location, and that place cells were conditionally independent of each other, given the hippocampal map and the rat's location. These assumptions were used in the Bayesian reconstruction methods in Zhang et al. (1998), the statistical machinery they develop being similar to the present approach.

Based on these assumptions, each element of the P-HMM's output vector was Poisson distributed and independent of the other elements. Specifically, for a given time window t, the distribution of spike counts for element c was modeled as

$$N_t(c) \sim \text{Poisson}(f_c(S_t, x_t)) \tag{8.2}$$

where $\text{Poisson}(\mu)$ denotes a Poisson distribution with mean μ, and $f_c(S_t, x_t)$ is the observed mean firing rate of cell c at position x_t according to the cell c's Reference rate map associated with state S_t.

The Quiescent state was associated with rate maps whose values were all zero, representing a state in which no cells were active. This proved to be an important inclusion in the P-HMMs because during the beginning of a visit cells sometimes showed little activity. When only a handful of cells were recorded, this could be interpreted as representing the activity patterns of the "wrong" box.

The reconstruction process was based on the supposition that a P-HMM was the generative model of the spike count vectors recorded in the Test visit. The Viterbi algorithm Forney (1973), the standard algorithm for inferring an HMM's sequence of "hidden" states from a sequence of its outputs, was used to infer from the observed spike counts the most likely sequence of active hippocampal maps.

8.3 Results

A total of 235 cells were recorded from three rats during the 3-day experimental sequence shown in Fig. 1B. The spatial information content of 160 cells met or exceeded both the I_{min} and $I_{95\%}$ thresholds during two or more consecutive visits to the same box (possibly across epochs), and were included in further analyses (Table 8.1).

Eight distributions of place field correlation coefficients were computed from the 160 place cells as described in ANOVA groups in Section 8.2.7 and Table 8.2. Two of these distributions, the No Remapping and the artificially constructed Complete Remapping distributions, were included as baseline measures of what the distribution of place field correlation coefficients should be in the extreme cases of no or complete remapping. The other

six distributions were compared with each other and with these two baseline distributions to assess the extent to which remapping occurred under various conditions.

8.3.1 *Remapping occurred in the opposite-orientation configuration*

All eight distributions listed in Table 8.2 were analyzed together using a Kruskal-Wallis nonparametric ANOVA, which showed significant differences among them ($\chi^2_7 = 1291.3$; $p \ll 0.001$). Post hoc comparisons revealed no significant differences among the first six ($p > 0.5$), including the *No Remapping* distribution, suggesting that the same hippocampal map was being recalled in each of the conditions measured by these distributions. Post hoc comparisons also revealed significant differences ($p \ll 0.001$) between these first six distributions and the last two (the *OA* × *OB After Remap* distribution and the artificially constructed *Complete Remapping* distribution), There were no significant differences between the *OA* × *OB After Remap* and *Complete Remapping* distributions ($p \approx 0.1$). Thus remapping was observed only in the *OA* × *OB After Remap* distribution, and the degree of remapping was statistically indistinguishable from a complete remapping.

8.3.2 *No remapping detected in the same-orientation configuration*

Although Skaggs and McNaughton (1998) observed partial remapping between the two boxes in the same-orientation configuration, there was no evidence for remapping between boxes in the corresponding condition in this study, including those visits subsequent to opposite-orientation sessions. The median correlation values of the same box ($SA_1 \times SA_2$) and different box ($SA \times SB$) distributions were both 0.71. Post hoc comparisons between same box and different box correlations revealed no significant difference ($p > 0.5$).

A Kruskal-Wallis nonparametric ANOVA revealed no significant differences among the six sets of fields (3 rats × 2 days) that constituted the $SA_1 \times SA_2$ group, our prototypical *No Remapping* group ($\chi^2_5 = 4.8$; $p \approx 0.44$). Similarly, no significant differences among the six sets of fields that constituted the $SA \times SB$ group were observed ($\chi^2_5 = 9.25$; $p \approx 0.1$). A visual comparison of SA fields to SB fields in Figures 8.3, 8.4, and 8.5 confirms that cells behaved similarly in the two boxes. Figure 8.6 compares the correlations in the same box ($SA_1 \times SA_2$) distribution with the different box ($SA \times SB$) distribution. Cells tend to cluster near the 45° line, meaning that their correlations between boxes are similar to their correlations between visits to the same box.

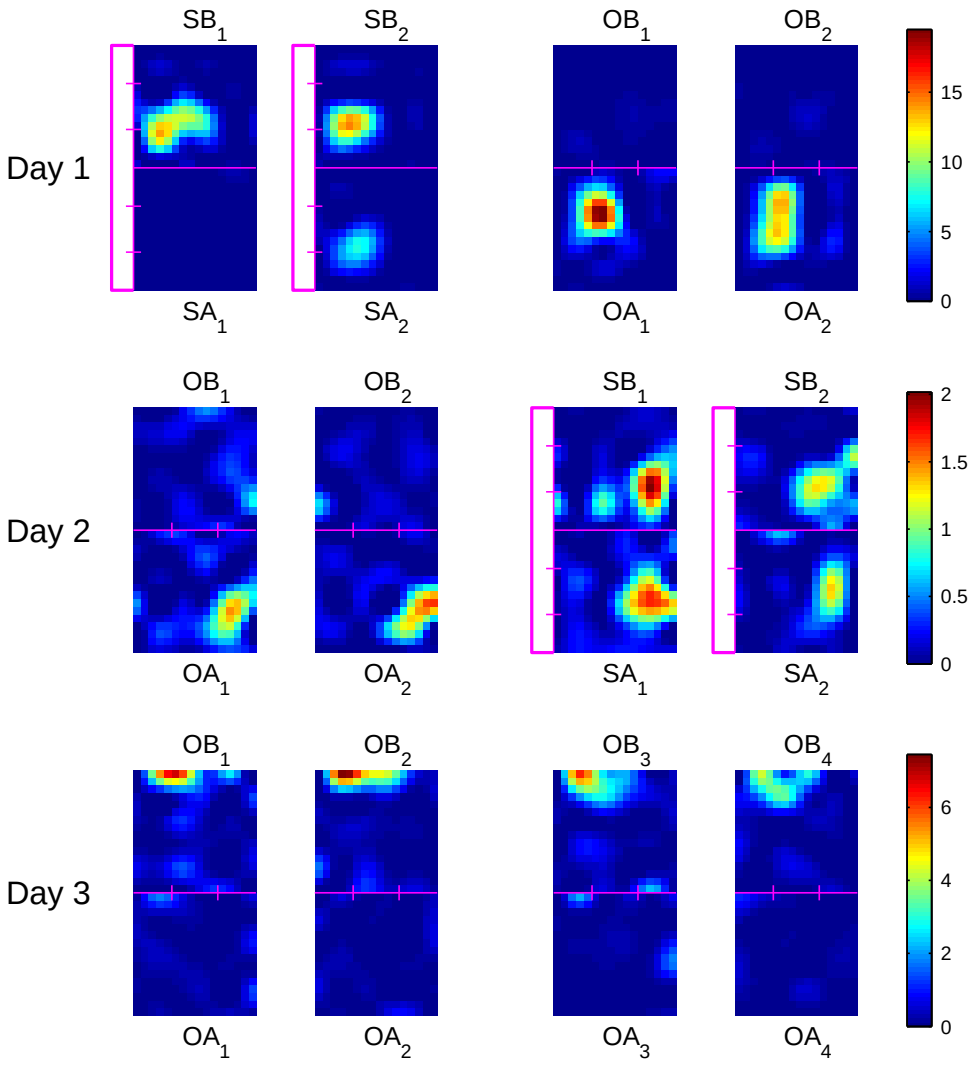

FIGURE 8.3: Place fields recorded in rat 1 show that the first remapping in box B occurred in visit OB_1 on day 1. Surprisingly, but consistent with the bulk of the day 2 place cell population, the fields in OA_1 and OA_2 on day 2 had the strongest correlations with the 0° rotations of the fields in SA_1 and SA_2. On day 1, and on both days for the other 2 rats, the strongest correlations were at 90°. A different cell is shown for each day. Color bars at right show spike rate in Hz. (Reprinted from Fuhs et al., 2005.)

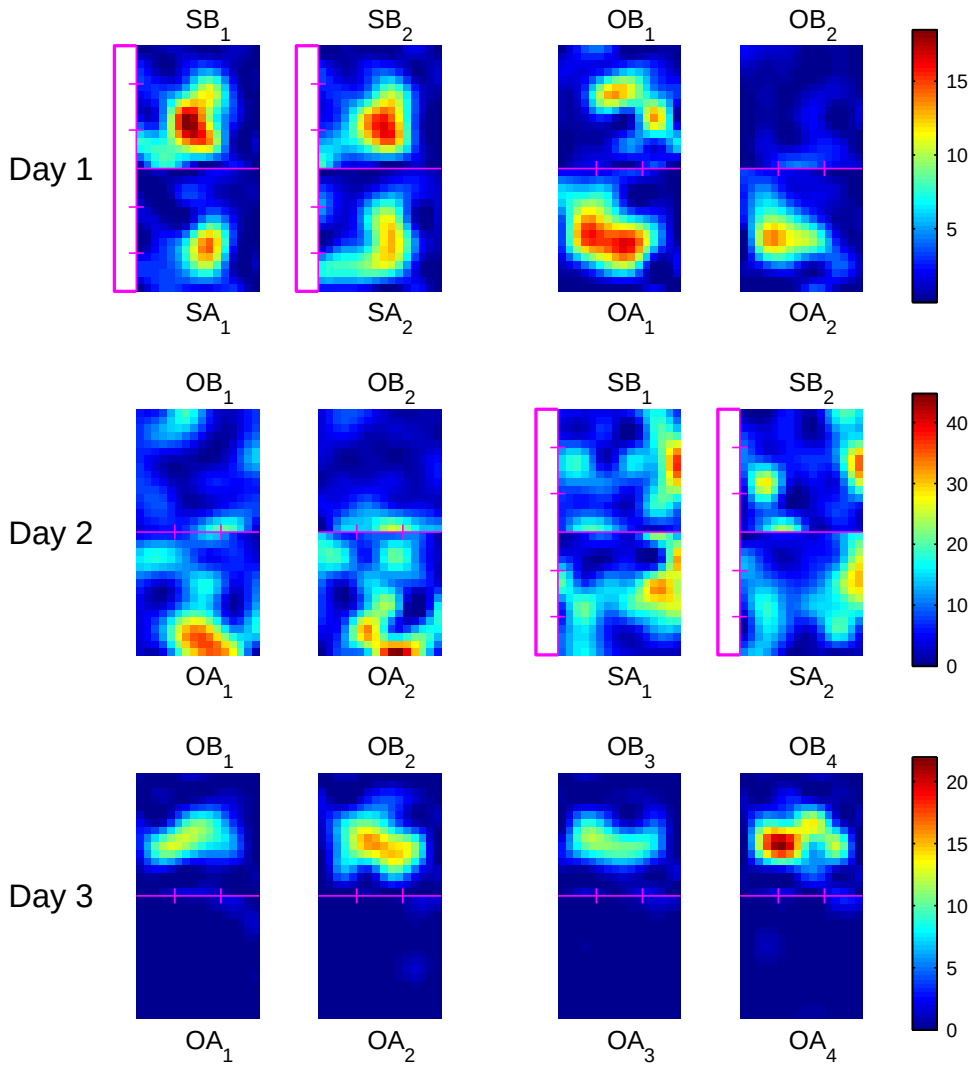

FIGURE 8.4: Place fields recorded in rat 2 show that the first remapping in box B occurred in visit OB_2 on day 1. Note that the field in SA_2 day 1 shows the strongest correlation with the field in OA_1 day 1 at the 90° rotation. A different cell is shown for each day. Color bars at right show spike rate in Hz. (Reprinted from Fuhs et al., 2005.)

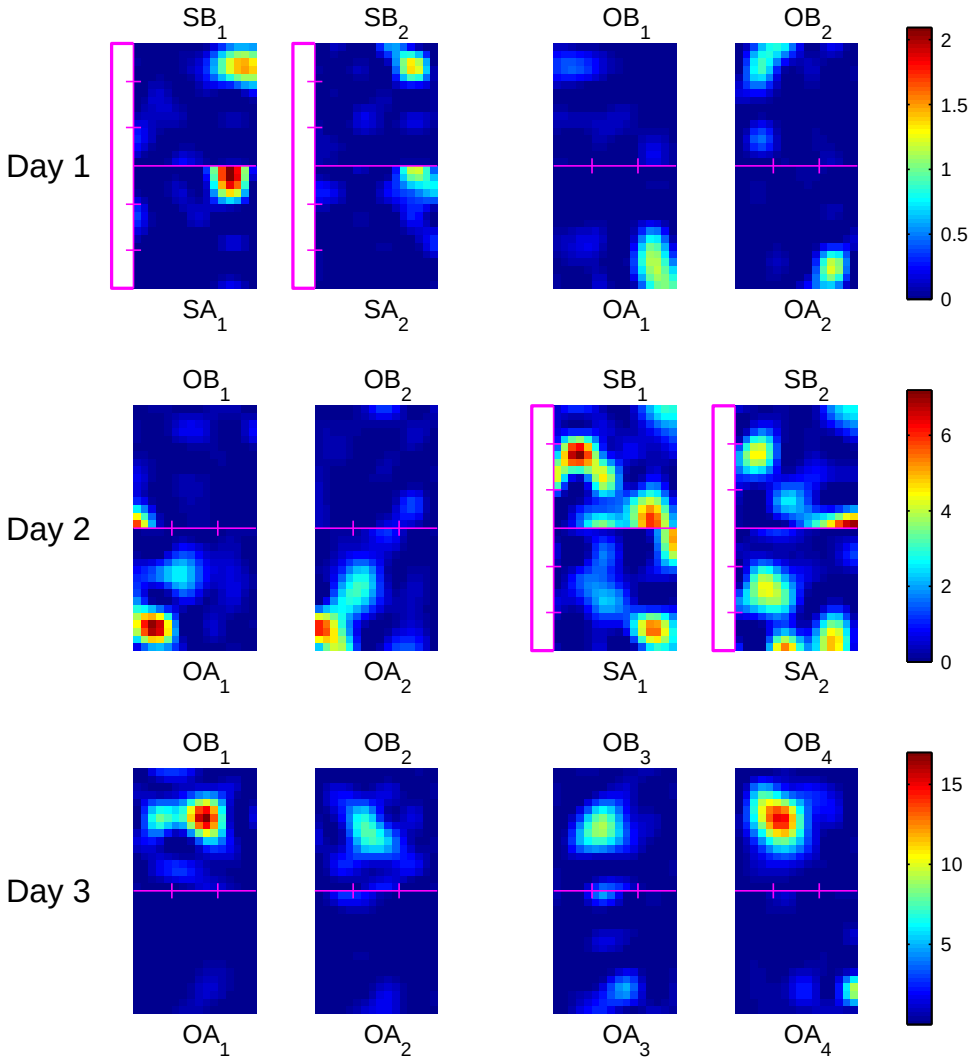

FIGURE 8.5: Place fields recorded in rat 3 show that the first remapping in box B occurred in visit OB_1 on day 2: the field in the bottom left corner of OA_1 is absent from the top right corner of OB_1. A different cell is shown for each day. Color bars at right show spike rate in Hz. (Reprinted from Fuhs et al., 2005.)

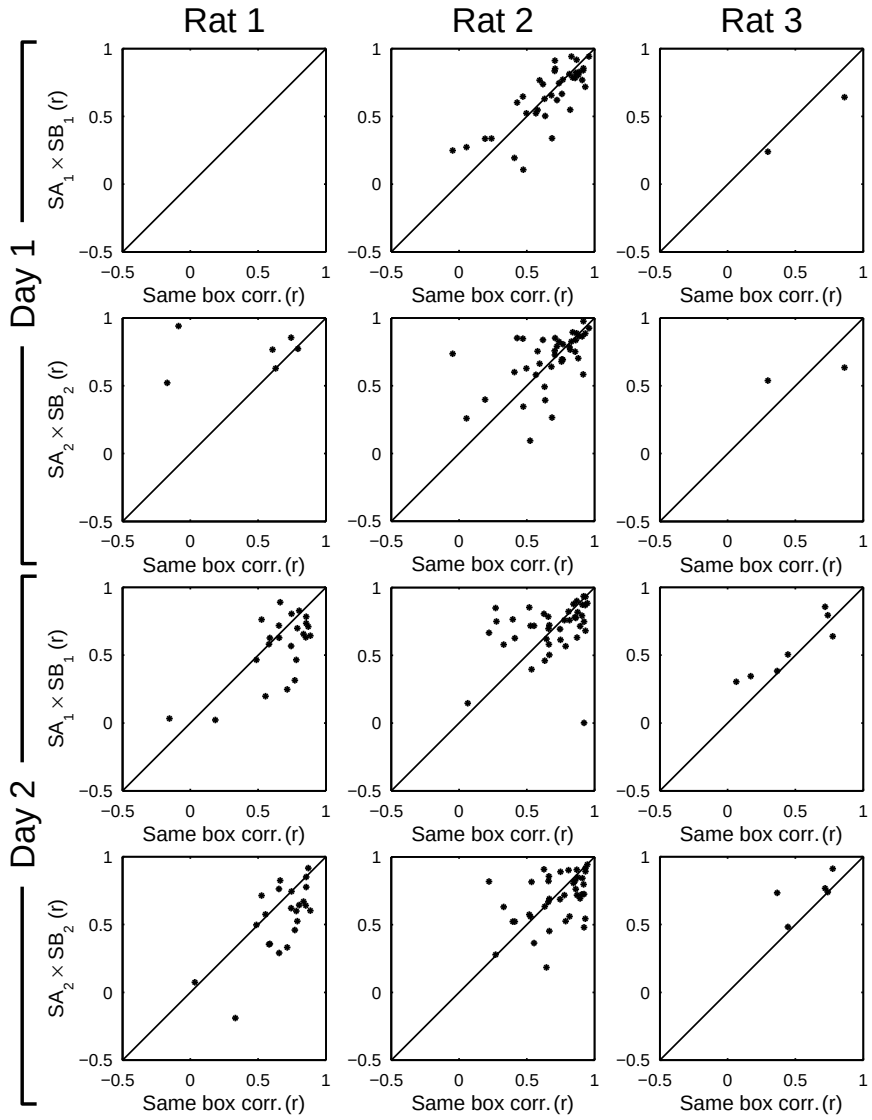

FIGURE 8.6: Rats did not remap between boxes in the same-orientation configuration. Place field correlations between boxes A and B are plotted against correlations between successive visits to the same box (usually box A, but occasionally box B if there was no field in box A). Points lie mostly on the 45° line, suggesting that field correlations are as strong between boxes as within a box. (Reprinted from Fuhs et al., 2005.)

8.3.3 *Same-orientation box A fields were maintained in the opposite-orientation condition*

For all three rats, the box A fields in the opposite-orientation condition were isomorphic to those in the same-orientation condition. For each visit, we found the best fitting rotation from among 0, 90, 180, or 270°, and then computed the correlation. The best-fitting rotation was 90° in all but one case. The median correlation between fields in the temporally closest same-orientation and opposite-orientation box A visits ($SA \times OA$) was 0.61. Post hoc comparisons showed no significant difference between this distribution and the same-box distributions from either configuration ($SA_1 \times SA_2$ and $OA_1 \times OA_2$, $p > 0.5$).

A Kruskal-Wallis nonparametric ANOVA revealed significant differences among the six sets of correlations that constituted the $SA \times OA$ distribution (3 rats $\times$ 2 days; $\chi^2_5 = 16.0$; $p < 0.01$). The rat 1, day 2 set had the lowest median correlation of the six (0.49), and post hoc comparisons showed that the rat 1, day 2 $SA \times OA$ correlations were significantly lower than the rat 2, day 2 set ($p < 0.05$).

To understand to what extent these differences were attributable simply to differences in overall place field stability among recording sessions, each $SA \times OA$ distribution was compared with the corresponding $SA_1 \times SA_2$ distribution using Kolmogorov-Smirnov tests. Only during the second day of rat 1 was there a significant difference ($D_{25,25} = 0.54$, $p < 0.002$), suggesting that variations in $SA \times OA$ correlations were indeed linked to varying place field stability. Even for the rat 1, day 2 set, the $SA \times OA$ group median of 0.49 suggests at most a partial remapping between configurations.

The second day in rat 1 was different from the others in that, despite the rotation of box A by 90° between the opposite-orientation and same-orientation configurations, the place fields did not rotate. Figure 8.7A compares the within-configuration box A correlations to the between-configuration box A correlations for each rat and each day. Except for this second day in rat 1, cells that showed consistent fields in box A between visits in the opposite-orientation configuration also showed consistent fields across configurations when rotated by 90°. Figure 8.7B shows the within-configuration versus between-configuration plots for rat 1, day 2 for the four possible rotations of the fields consistent with the shape of the box walls. Only the 0° rotation shows a strong positive correlation.

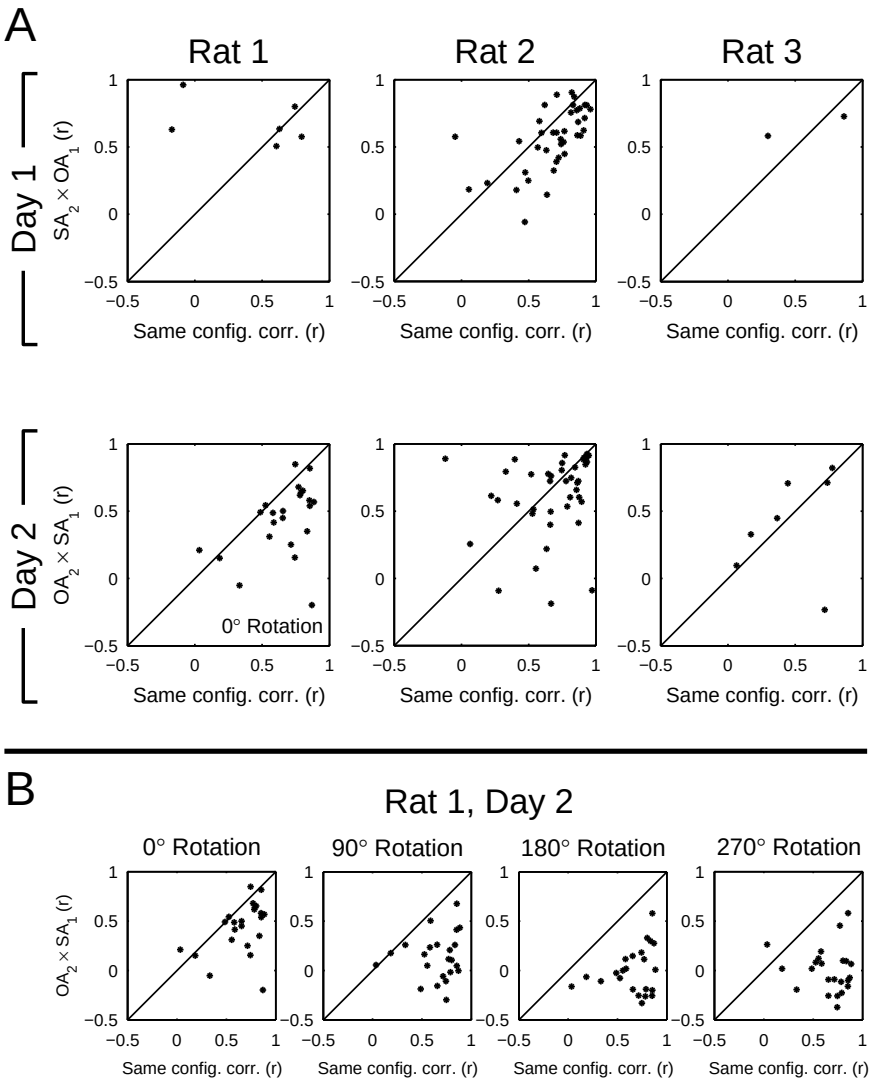

FIGURE 8.7: Box A fields in the opposite-orientation condition were isomorphic to those in the same-orientation condition. Correlations between box A place fields in the 2 conditions are plotted against correlations between successive visits to box A. A) except for the second day of rat 1 (labeled 0° Rotation), all the fields correlated best when rotated by 90°. B) during the second day for rat 1, fields appeared to remain fixed in the room frame rather than rotating with the box. Only the 0° rotation gave correlations significantly centered at the diagonal line. (Reprinted from Fuhs et al., 2005.)

8.3.4 *Opposite-orientation box A fields were stable across visits*

Overall, box A place fields were as stable between visits in the opposite-orientation configuration as in the same-orientation configuration. The $OA_1 \times OA_2$ median correlation was 0.67, which was not significantly different from the $SA \times SB$ distribution ($p > 0.5$). However, a Kruskal-Wallis nonparametric ANOVA revealed significant differences between the six sets of fields that constituted the $OA_1 \times OA_2$ distribution (3 rats $\times$ 2 days; $\chi^2_5 = 13.5$; $p < 0.02$). The rat 1, day 2 group again had the lowest median (0.57), and post hoc comparisons showed it to be significantly lower than the median for the first day for the same rat (P 0.05). No other differences were significant.

8.3.5 *Box B fields eventually differed from box A in the opposite-orientation configuration*

For all three rats, place cells eventually showed completely different fields between boxes A and B in the opposite-orientation configuration. However, the rats differed with respect to the visit in which box B differentiation first occurred. Rat 1's place fields remapped on first visiting box B in the opposite-orientation configuration (Figure 8.3, visit OB_1 day 1). Rat 2's fields remapped on its second visit to box B (Figure 8.4, visit OB_2 day 1). Rat 3's fields remapped on its third visit to box B (Figure 8.5, visit OB_1 day 2).

Figure 8.8 shows how the place fields correlated between boxes as a function of how they correlated between successive visits to box A over the first 2 days. Remapping is indicated by a majority of points falling significantly below the 45° line. There are two interesting features to note. First, scanning down each column, it can be seen that once a rat's place fields remapped between the boxes, they continued to remap in all subsequent visits. (The between-box correlations were calculated after rotating the box B fields by 180°, but once remapping occurred, these correlations were no stronger than unrotated correlations.) Second, as noted earlier, fields in box A in the opposite-orientation configuration were stable across visits and were not disrupted by the remapping in box B; this can be seen from the tendency of points to fall on the right side of each plot.

A Kruskal-Wallis ANOVA was performed on the first five pairs of opposite-orientation between-box visit correlations from each rat (i.e. $OA_1 \times OB_1$ on days 1 and 2; $OA_2 \times OB_2$ on days 1 and 2, and $OA_1 \times OB_1$ on day 3) and the *No Remapping* ($SA_1 \times SA_2$) distributions. Significant differences were observed consistent with the previously described timing of remapping for each rat ($\chi^2_{15} = 138.0$; $p < 0.001$). Post hoc comparisons showed the following:

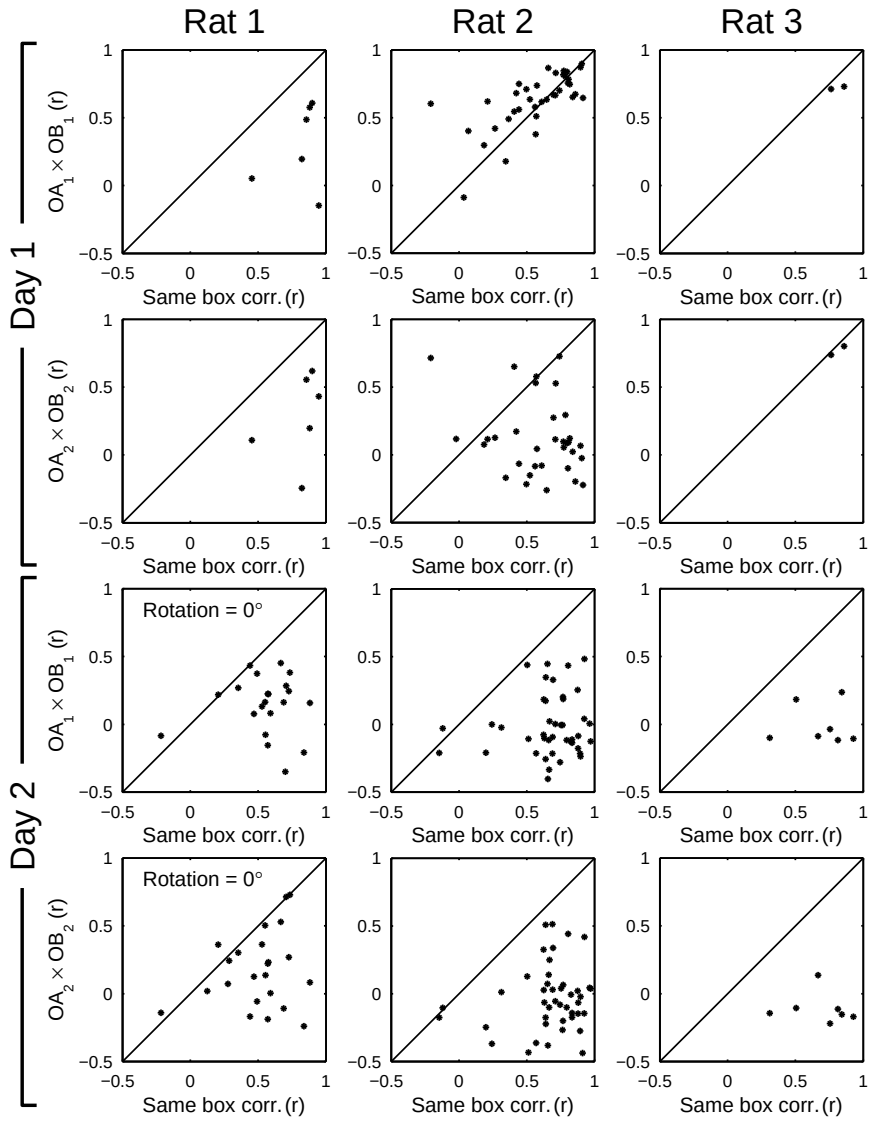

FIGURE 8.8: Rats first showed remapping during their first, second, or third visit to box B in the opposite-orientation configuration. Correlations between box A and box B place fields in the opposite-orientation configuration are plotted against correlations between successive opposite-orientation visits to box A. Box B fields were rotated 90° before calculating correlations, unless otherwise shown. Strong correlations (points on or above the diagonal line) are shown during the first box B visit by rat 2 and the first two visits by rat 3, indicating that remapping had not yet occurred. (Reprinted from Fuhs et al., 2005.)

Rat 1. There were no significant differences among the five visit pairs, which is consistent with the rat remapping on first exposure to the opposite-orientation condition. None of these was significantly different from any of the other "remapped" visit pairs ($p > 0.5$). The second visit pair of the second day yielded the largest number of cells with spatially selective fields in at least one box (24), and the between-box correlations during the second visit pair of the second day were significantly lower than the first visit pair of rat 2, which had not remapped yet ($p < 0.05$).
Rat 2. The first visit pair showed significantly stronger correlations than the following four ($p < 0.005$), which were not significantly different from the other "remapped" visit pairs ($p > 0.5$). This is consistent with the rat remapping on second exposure to the opposite-orientation condition.
Rat 3. Only two cells were recorded during the first two visit pairs (day 1); they both showed strong correlations between boxes. By contrast, in the third and fourth visit pairs (day 2), none of the eight cells showed even a weak correlation between boxes. Correlations in the fifth visit pair (day 3) were similarly low, but the small sample size (four cells) was not sufficient to show a statistical difference between these $OA \times OB$ pairs and the *No Remapping* distribution. Nonetheless, the values for the third and fourth visit pairs were significantly lower than for the first visit pair of rat 2 ($p < 0.025$), consistent with the rat remapping at the beginning of day 2.

The between-box correlations for the visits just before remapping in rats 2 and 3 were combined into an $OA \times OB$ *Before remap* distribution (median correlation 0.67), and the between-box correlations for the visit initially showing remapping in all three rats were combined into an $OA \times OB$ *After remap* distribution (median correlation 0.06). $OA \times OB$ *Before remap* showed high correlation values and was not significantly different from either same-box correlation group ($SA_1 \times SA_2$ or $OA_1 \times OA_2$, $p > 0.5$). By contrast, $OA \times OB$ *After remap* was significantly different from $OA \times OB$ *Before remap* ($p < 0.001$), and not significantly different from the artificially created *Complete Remapping* distribution ($p > 0.5$).

8.3.6 *Place fields on day 3 were stable across epochs*

Figure 8.9 shows, for day 3, how the place fields correlated between boxes as a function of how they correlated between successive visits to box A. The correlations were measured without rotation, because two cells each recorded from rats 1 and 2 appeared to show consistent fields between boxes when not rotated (Figure 8.10). The third rat showed equally low correlations for all four rotations consistent with the box shape (0, 90, 180, and 270°).

The $OX_2 \times OX_3$ *day 3* distribution, constituting box A and B field correla-

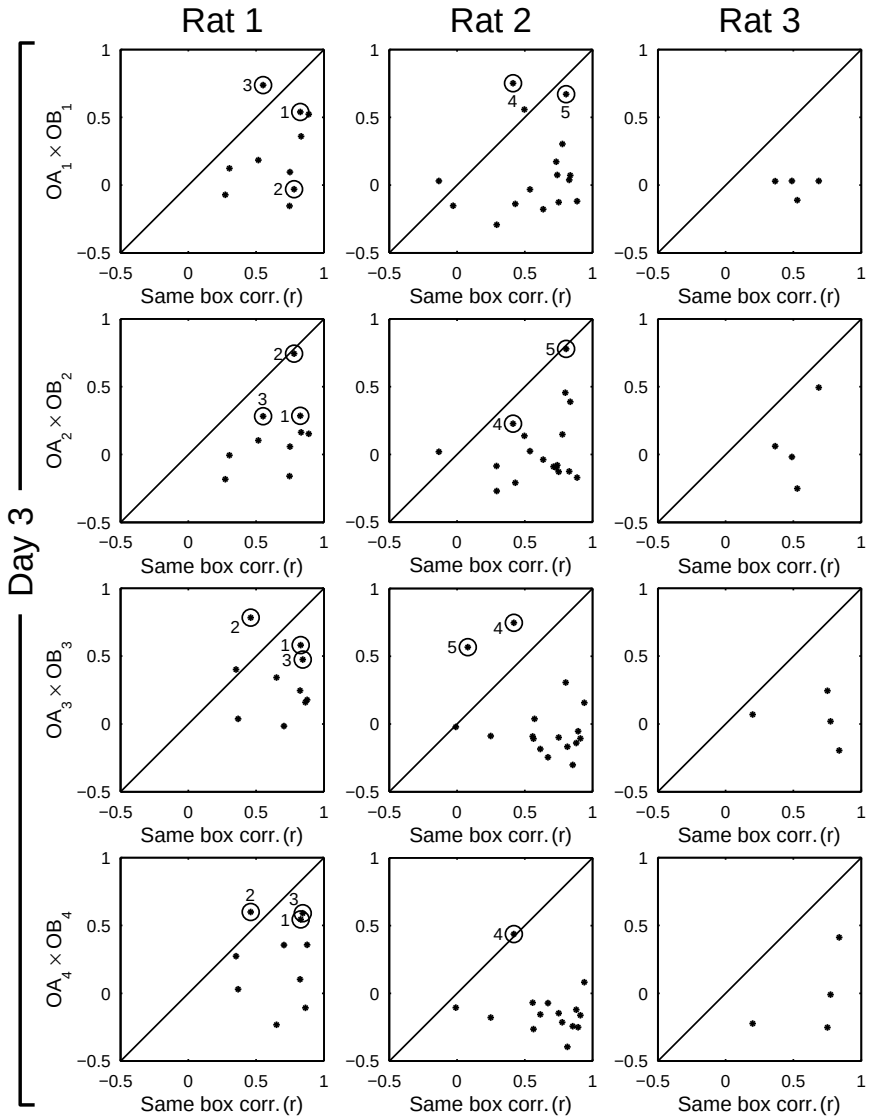

FIGURE 8.9: All three rats showed strong remapping between the two boxes on day 3: most points on the right are well below the diagonal line. Points labeled 1-3 are the cells for rat 1 that exhibited consistent fields across boxes instead of remapping; points labeled 4 and 5 are the 2 cells for rat 2 that did the same. (Reprinted from Fuhs et al., 2005.)

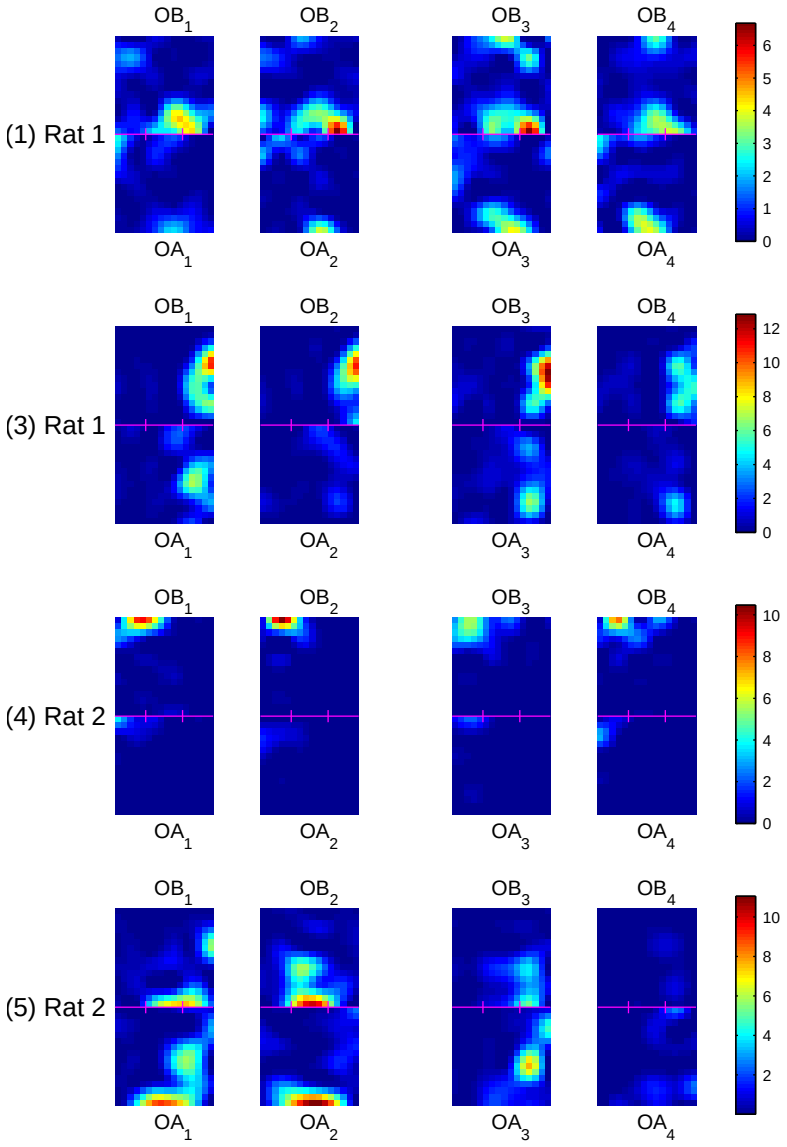

FIGURE 8.10: Four of the 5 place cells (from 2 rats) that showed similar fields between boxes on day 3. Cell numbers (in parentheses) match those in Figure 8.9. Cells 1 and 4 may be examples of rate remapping. Cell 2 from Figure 8.9 is omitted because it had only a very faint field in one box, although the between-box correlation coefficient was high. (Reprinted from Fuhs et al., 2005.)

tions between the second visit of the first foraging epoch and the first visit of the second foraging epoch, was not significantly different from any of the other non-remapping distributions ($p > 0.5$). A Kruskal-Wallis ANOVA comparing the six sets of correlations in this distribution (2 boxes $\times$ 3 rats) showed no significant differences among them (χ^2_5=6.6; $p \approx 0.25$). Thus all rats were remapping between boxes on day 3, and the maps were stable across the two opposite-orientation epochs.

8.3.7 *No evidence of rate remapping in the absence of field remapping*

The main Kruskal-Wallis ANOVA was recalculated using the same distributions in Table 8.2 but using a place field difference metric (Equation 8.1) that is strongly sensitive to absolute rate variations. Again, significant differences were found ($\chi^2_7 = 1268.8$; $p < 0.001$). Post hoc comparisons again found no significant differences between the first six distributions ($p < 0.5$), but significant differences between these six distributions and the *OA* $\times$ *OB After Remap* distribution ($p < 0.001$). The median D values for the *No Remapping* and *OA* $\times$ *OB Before Remap* distributions were 0.37 and 0.42, respectively, whereas the median D value for the *OA* $\times$ *OB After Remap* was 0.74. The median D value of the Complete Remapping distribution (0.81) was found to be even higher than that of the *OA* $\times$ *OB After Remap* distribution ($p < 0.01$), although this difference is likely explained by the artificially constructed nature of the *Complete Remapping* distribution. Absolute firing rates are likely to vary between cells to a greater degree than between conditions for the same cell, and tetrode cell isolation techniques do not generally produce an equal sampling of spikes from all cells.

8.3.8 *Population dynamics within a session*

We analyzed the temporal dynamics of place cells during opposite-orientation box A and B visits on the day in which remapping was first observed. The purpose of the analysis was to determine to what extent place cells showed activity consistent with the same map throughout the visit (see Section 8.2.8) For each rat, a Poisson hidden Markov model (P-HMM) was constructed using firing rate maps from Reference visits OA_2 and OB_2. These box A and box B Reference visits were used as archetypes of the firing patterns expected in each box subsequent to remapping between them. (Place field correlations indicated that place cells completely remapped between these visits.) Spike trains from two Test visits, OA_1 and OB_1, recorded just before the Reference visits, were then analyzed by inferring the sequence of P-HMM states (*Box A*, *Box B*, or *Quiescent*) most likely to have generated each spike train.

The sequences of inferred states for the Test visits of each rat are shown in

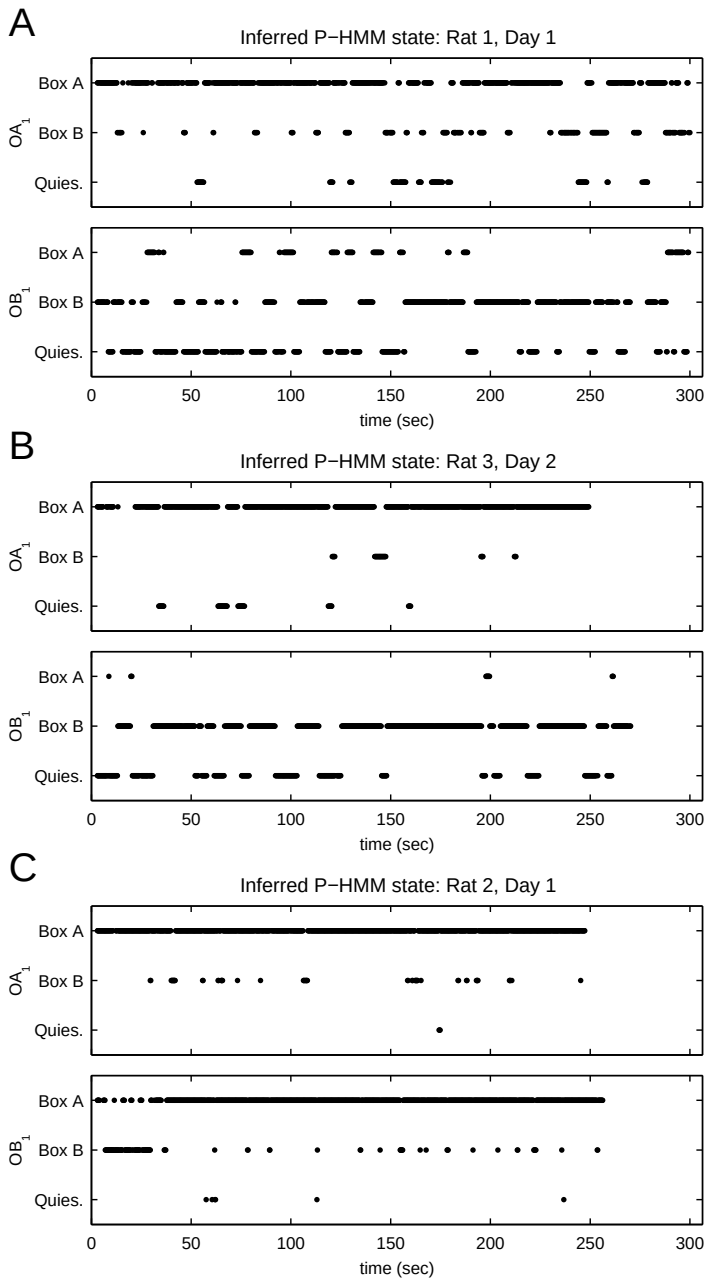

FIGURE 8.11: Hippocampal state (*Box A*, *Box B*, or *Quiescent*) inferred by the P-HMM model during the course of a box visit. A and B) rats 1 and 3 are mostly in state *Box A* during OA visits, and state *Box B* or *Quiescent* during OB visits. C) rat 2 favors state *Box B* in the first 27 s of visit OB_1 on day 1, but then switches to *Box A*. (Reprinted from Fuhs et al., 2005.)

Figure 8.11. The top graph for each rat shows the reconstruction performed on the box A Test visit, providing a baseline indication of how well the reconstruction method was able to infer the identity of the box. The bottom graph for each rat shows the reconstruction performed on the box B Test visit. For rats 1 and 3, the box B reconstructions are consistent with the fields having remapped on entry into the box. Interestingly, the likelihood of the Quiescent state being active decreases with time in both cases (Rat 1: $r = 0.28$, $p < 0.005$; Rat 3: $r = 0.22$; $p < 0.02$), suggesting that cell firing became more robust as the visit progressed. This phenomenon is consistent with previous reports progressively stronger cell firing over the course of a session (Wilson and McNaughton, 1993; Mehta et al., 1997).

Based on firing rate map correlations, Rat 2 was judged to still be using the box A map during the reconstructed box B Test visit (OB_1 day 1) shown in Figure 8.11C. The reconstructed state sequence is mostly consistent with this description. However, during the first 27 s, the place cells appear to adopt the box B map later used in the following box B visit, before switching back to the box A map.

To ascertain how well the place cell firing patterns were fit by the box B state, two portions of the box visit were defined: the box B portion ($t < 27$ s) and the box A portion ($t > 27$ s). Four groups of P-HMM log observation likelihoods were then analyzed: $P(N_t \mid \textit{Box A})$, $0 < t < 27$ s; $P(N_t \mid \textit{Box B})$, $0 < t < 27$ s; $P(N_t \mid \textit{Box A})$, $t > 27$ s; and $P(N_t \mid \textit{Box B})$, $t > 27$ s). A Kruskal-Wallis ANOVA was performed on the four groups and found significant differences ($\chi^2_3 = 226.1$; $p \ll 0.001$). Post hoc comparisons revealed that groups $P(N_t \mid \textit{Box B})$, $0 < t < 27$ s and $P(N_t \mid \textit{Box A})$, $t > 27$ s were not significantly different, yet both groups' median correlations were significantly higher than group $P(N_t \mid \textit{Box B})$, $t > 27$ s), suggesting that the box B state was characterizing the first 27 s of population activity nearly as well as the box A state characterized the remainder of the session.

8.4 Discussion

8.4.1 *Same-orientation case*

Linear path integration could in principle have differentiated the two boxes in the same-orientation configuration, and might account for the partial remapping Skaggs and McNaughton observed (Touretzky, 2005). No partial remapping, not even rate remapping, however, was observed under nearly identical conditions in the present experiment. Skaggs and McNaughton rats were naive to the apparatus, whereas our rats had already undergone 16-23

days of trials in the same-orientation configuration. Therefore to rule out experience-dependent effects, data were examined from the rats? first 2 days of exposure to the apparatus, i.e., days 1 and 2 of the 19- to 26-day recording sequence. In these first 2 days, the between-box correlations were actually slightly stronger than the within-box, between-epoch correlations ($p < 0.05$), indicating that partial remapping did not occur even at the beginning of the experiment.

The present study used Fisher Brown Norway rats, in contrast to the albino rats used by Skaggs and McNaughton. Albino rats have roughly half the visual acuity of non-albinos. Also, the box lights used by Skaggs and McNaughton were dimmer and better shielded than in the present experiment. If the failure to find partial remapping in the present experiment were due to the higher light levels and greater visual acuity, then training in darkness should decrease the place code similarity between boxes. Preliminary data from one rat confirms this finding. When trained initially in the two boxes in darkness, the place codes in the two boxes were completely different. Thus, as the quality of sensory information provided to the rat is decreased from that available in the present experiment, to what was available in the Skaggs and McNaughton, to darkness, the degree of remapping between the boxes increases.

When the rats in this experiment left box A, traveled down the corridor, entered box B, and ended up back on the box A map, did they jump to it abruptly or transition smoothly? To see whether an abrupt change in map, or position within a map, was detectable in our data, we generated sets of ensemble activity patterns for rat 2 on days 1 and 2 and rat 1 on day 2 as they traveled from box A through the corridor and into box B. (Rat 3 had too few cells to be included in this analysis.) The patterns were constructed by dividing each visit into 500-ms bins and computing the average spike rate of each cell within each bin. We then looked at a trajectory from box A through the corridor and into box B, measuring the maximum correlation at each time step between the current activity pattern and the activity pattern at the entrance to box A, which is also the expected pattern in box B because the rats were not remapping between boxes. When the rats entered box B from the corridor, we did not see any evidence that the correlation increase was sudden, as would be expected if the hippocampal representation abruptly changed from that of the corridor to that of the box A map. However, given the paucity of trajectories through the corridor, it is unlikely that such a jump would be detectable in the present data.

If the hippocampus did not abruptly jump between maps in this condition, one possible explanation is that the representation of the apparatus

formed a non-Euclidean map in which the same box A doorway could be smoothly entered from either end of the straight corridor. The inability of the linear path integrator to differentiate two same-orientation boxes, even after place fields were remapping in the opposite-orientation condition, is a striking demonstration of the difference between linear and angular idiothetic information.

8.4.2 *Opposite-orientation case*

The opposite-orientation portion of the experiment tested the influence of both linear and angular path integration on place fields. All three rats eventually exhibited different maps in the two boxes when their orientations were 180° apart. For rats 2 and 3, however, this result did not emerge until the second or third box B visit, respectively.

Several experiments have shown that repeated instances of orientation discordance can weaken visual cues' control of place fields. Knierim et al. (1995) showed that repeatedly disorienting rats before placing them in a cylinder with a white cue card along the wall prevented the card from acquiring directional control of place fields. The disorientation would have caused the card to appear at a different allocentric bearing on each trial. In a subsequent experiment, Knierim et al. (1998) introduced a conflict between visual and vestibular cues by rapidly rotating the cylinder wall and floor through an angle of 135-180° while the rat remained inside. The rat's visual system would indicate no self-motion under these conditions, whereas its vestibular system would sense the rotation. Knierim et al. found that place cells could remap in this situation, and head direction cells could fail to stay aligned with the cue card.

When the rat entered box B in the present experiment, visual landmarks would indicate that the world had rotated by 180°, whereas its vestibular sense would deny that any rotation had occurred – the complement of the Knierim et al. (1998) scenario. However, if either type of discordance between visual landmarks and the rat's internal direction sense weakens the landmarks' control of place fields, it follows that repeated trips between boxes A and B in the opposite-orientation configuration should eventually lead to the abandonment of the box A map in box B.

Rat 1 seemed to be the most sensitive to orientation discordance because its cells remapped immediately on entering box B in the opposite-orientation configuration, in visit OB_1 on day 1. They continued to remap during visit OB_2, and again on day 2 in visits OB_1 and OB_2; but then, after the sleep period between epochs, when the rat was again passively transported to box A – now in the familiar side-by-side configuration – its fields maintained the

same orientation relative to the room as before, rather than rotating by 90° to align with the within-box cues. (See Fig. 3, visit SA_1 on day 2.) In this rat, after experiencing just two opposite-orientation epochs constituting four trips from box A to box B, the visual cues appear to have lost directional control of place cells. Jeffery and O'Keefe previously showed that rats could learn to ignore visual cues and follow idiothetic cues when a prominent visual cue repeatedly shifted location.

Rats 2 and 3 continued to use a 90° rotation of the box A map in the same-orientation configuration. (Compare visits OA_1 and OA_2 with visits SA_1 and SA_2 on day 2, Figures 8.4 and 8.5.) It is significant that the rats entered box A at the start of the SA_1 visit by passive transport. Stackman et al. (2003) showed that rats do not integrate vestibular and optic flow cues accurately when transported passively through a 180° heading change. The 90° rotation in the present experiment may therefore have been barely noticeable. In contrast, the rats always entered box B by active locomotion, so they should have been able to maintain their angular orientation Taube and Burton (1995). The 180° heading change required to realign the box A map with the box B landmarks would have been highly salient.

Another possible response to orientation discordance is for place fields to both remap and dissociate (follow different sets of cues), as seen in double-cue rotation tasks Knierim (2002); Tanila et al. (1997). Dissociation was not observed in the present experiment, possibly because there were no distal cues; the room was kept dark and within-box illumination was faint. So the discordance perceivable to the rats was between visual and vestibular inputs, not between competing sets of visual cues. The outcome might have been different if the rats had access to prominent room cues. What distinguishes our result from previous discordance results, specifically the double-cue rotation experiments of Tanila et al. (1997) and Knierim (2002), is that the remapping was complete, and in two of the three rats, the remapping was delayed rather than immediate (Bostock et al., 1991). Whether the delay is dependent on some representation of the animal's cumulative experience is not known; however, as discussed in the next section, it may simply reflect the increasing cumulative likelihood of a failure to realign the box A map in box B space.

8.4.3 *Remapping versus map extension*

We define two maps as distinct if there is no continuous path from a place on one map to a place on the other. If such a path exists, then the two maps are really just different regions of the same map. In topological terms, a ?place? is a vector of place cell firing rates and a ?map? is a collection of such vectors

forming a two-dimensional manifold. Maps are distinct if the manifolds do not intersect.

The appearance of a separate box B map in the opposite orientation condition could arise in two ways. The rat could be literally remapping, i.e., jumping from the initial manifold to a separate manifold containing the box B map. Or the rat could be using a single manifold for the entire apparatus, but the portion containing the box B activity patterns might initially be unobservable because the rat jumped (or smoothly transitioned) back to the box A portion of the manifold on entering box B. In this situation, the subsequent change in box B patterns we attributed to "remapping" (jumping to a new manifold) would actually be the result of the rat's no longer resetting its position on the current manifold.

The P-HMM reconstruction for rat 2 is consistent with the notion that the rat extended the box A map into box B rather than switching to a new map. Because remapping in box B did not emerge until the second visit on day 1, we can compare firing fields of the same cells before the emergence (visit OB_1) and afterward (visit OB_2). Figure 8.11C shows that during visit OB_1, the rat started out using what would later appear as the box B map. It switched to the rotated box A map after about 27 s. This is consistent with the rat initially maintaining its orientation sense during the first 27 s of the box B visit, which would be natural if the box B map were simply an extension of the existing box A map through the doorway and into additional territory.

The argument would be strengthened if place fields spanning the two boxes could be found. We did see a few such fields, but they appeared to be tied to the doorway because they did not remap when the rat adopted a separate representation for box B.

Although remapping in the opposite-orientation configuration was complete on day 2, there is a hint of partial remapping on day 3 (Figure 8.10). Rat 1 had three cells and rat 2 had two cells that appeared to stay in the box A reference frame, displaying similar fields in box B despite the fact that the cues were 180° opposite. (However, cell 4 and possibly also cell 1 in Figure 8.10 are exhibiting rate remapping.) The remaining cells recorded on day 3 remapped between boxes. This suggests either that the remapping on day 2 was substantial but not actually complete, or else the rats' representation for the task had begun evolving in a new direction by day 3. To date there have been no reports of two completely distinct hippocampal states becoming more similar over time, so the former hypothesis seems more likely.

In conclusion, linear and angular path integration function differently in the rat. Linear path integration appears easily overridden by visual landmarks, whereas angular path integration is more sensitive to cue discordance.

In the present experiment, although repeated visits to the two boxes in the same orientation condition produced no measurable remapping, one to three visits in the opposite orientation condition prompted (nearly) complete remapping. That this remapping was both delayed and abrupt suggests a sudden failure to reset idiothetic representations, rather than a gradual adaptation of the hippocampal code.

9 CONCLUSION

> "Reasoning draws a conclusion, but does not make the conclusion certain, unless the mind discovers it by the path of experience. "
> – Roger Bacon (1214–1294)

9.1 Summary of Research Contributions

The goal of this thesis has been to contribute to an understanding of neural systems within the hippocampal region using advanced techniques from machine learning, computational statistics, and artificial neural networks. I have focused on two anatomical areas, dorsal medial entorhinal cortex (dMEC) and hippocampus, presenting a neural network model of path integration in dMEC and both theoretical and empirical contributions to an understanding of contextual representations in hippocampus.

Chapter 6 presented a neural network model of dMEC that unified recent physiological findings in dMEC with previous theories suggesting the existence of a neural path integration system in rats. The model provides an account of how the hexagonally periodic firing fields of dMEC grid cells can arise from a recurrent neural network with local, radially symmetric connections. In addition, it demonstrates how a network with hexagonally periodic firing fields can be used, with the addition of a velocity signal, to integrate motion across an environment. The differences in field size, spacing and orientation at different dorso-ventral levels of dMEC can be understood to reflect different networks with different of velocity modulation. How a sparse place code could be formed from the conjunctive encoding of multiple grid cell networks, each with highly periodic activity patterns, is also explored.

Chapter 7 presented a Bayesian statistical theory of contextual representations in hippocampus. The principal goal of this work is to advance a single notion of context that can be applied in both spatial and non-spatial domains. This single notion is that a context is a statistically stationary distribution of experiences that occur in temporal proximity to one another. Statistical stationarity permits inference across time: the distribution of future experience in a context should be similar to past experiences in that context.

Thus, context selection (or context recall) should be an informative inference process; knowing the current context should permit expectations within the context that would not in general be reasonable.

Since an animal's world is not a stationary distribution of experiences, the challenge of context learning is to divide the animal's experiences into separate contexts, each of whose distribution of experiences is stationary. The theory shows that this process, context learning, can be understood as a model selection problem and that the gradual development of distinct hippocampal contextual representations can be understood as reflecting the adoption of progressively more complex context models. Because contexts represent a temporal clustering of experiences, the theory explains why contextual representations develop for different reward contingencies in reversal learning but not for overlapping trajectories in alternating sequence tasks. Because context selection and context learning are distinct inference processes within this theory, it is able to capture the gradual adoption of a context model to distinguish between contingencies in serial reversal learning, as well as the single-trial behavioral reversals observed in well-trained animals.

In Chapter 8, I analyzed data from an experiment performed in Bruce McNaughton's lab in which rats foraged for food in two identical boxes with either the same or opposite (180^o) orientation in the room. The analysis showed that, while rats represented the two boxes in the same orientation condition using the same hippocampal code, a second hippocampal code developed for the second box in the opposite orientation case. Thus, linear and angular path integration could both serve to distinguish the two boxes, the hippocampal code differentiated them. When when only linear path integration could serve to differentiate the two boxes, the hippocampal code was shared.

Of particular interest was the time course of differentiation of the two boxes. Each rat began to differentiate the two opposite-orientation boxes after a different number of visits to each box (first, second and third, respectively). As part of the analysis, I developed a novel application of inhomogeneous Poisson HMMs to decoding the rats' hippocampal code. Interestingly, the decoding showed that hippocampal representations need not be constant throughout a visit to an environment; rather, hippocampal representations can abruptly change several tens of seconds after entry.

9.2 Future Experimental Work

9.2.1 *Grid cells*

The model presented in Chapter 6 and the model presented by McNaughton et al. (2006) differ in a number of fundamental ways. There are three experiments that would help to distinguish between them:

1. Irregularities in the lattice structure of activity nodes during exploration in an open field. While the model presented here is able to show irregular lattice structures (e.g. heptagons, pentagons), the hexagonal periodicity is hard-wired into the network topology in the McNaughton et al. (2006) model.

2. A single orientation shared by every grid cell network. Such a result could be tested by recording simultaneously from cells at different dorso-ventral levels within the same animal. In contrast to Hafting et al. (2005), McNaughton et al. (2006) assert that it is not currently known whether grid orientations vary within one animal. The model presented here supports multiple orientations and place code analysis suggests that coding efficiency is substantially improved by grid cell networks with varying activity pattern orientations.

3. A fixed difference in orientations between grid networks. Though the orientation of different networks may vary, there may be a consistent relationship between them across environments. The training regimen used by McNaughton et al. (2006) involves afferent hexagonally periodic activity patterns at a fixed orientation. Thus, in contrast to the present model, they predict that the lattice orientation of different grid cell networks will be the same.

In addition, our dMEC model does not directly address praxic navigation, where a homing vector is constructed to return to the starting point of a previously integrated path. Discovering the locus of such a system within the brain and exploring its physiological properties would provide important insights into the role of dMEC in spatial navigation.

9.2.2 *Hippocampus*

The context learning theory presented in Chapter 7 makes several predictions about future experiments. First, in an extension of the Hayman et al. (2003) study, rats should gradually remap between black box room positions at the same rate as between white box room positions. Second, when rats are

trained initially in a series of morph boxes whose shapes vary from square to cylinder, the hippocampal place code should not distinguish among arena shapes. Third, in a serial reversal learning paradigm, remapping between reward contingencies should develop in a gradual manner, and the rate of remapping should decrease when reward contingencies are made more similar. However, if each reversal block is very small (e.g. 3 trials / reversal), no remapping between contingencies should be observed. Finally, since the theory showed that the alternating T-maze task should not lead to the development of distinct contextual representations to distinguish left-to-right and right-to-left trials, replay of the previous trial or of earlier portions of the current trial should be necessary in cases where the task is hippocampus-dependent (e.g., when a delay is present). In cases where trial-type remapping is observed, trial-type differences should be due to afferent input and observable in superficial EC.

Critical to understanding how new contextual representations are formed is place cell data from DG and CA3, in addition to CA1, in all of the paradigms addressed by the model. Since the most likely neural instantiation of the theory involves the pattern-separation and pattern-completion mechanisms associated with DG and CA3, mossy-fiber inactivation (as in Lassalle et al., 2000) should inhibit gradual remapping between two similar contexts.

The experiment presented in Chapter 8 suggests the following question: why was partial remapping between same-orientation boxes observed by Skaggs and McNaughton (1998)? Preliminary evidence suggests that the salience of visual cues is the significant difference, and further experimental work should be (and is being) carried out to explore the relative contributions of visual cues and linear path integration to the formation of the hippocampal place code.

9.3 Future Theoretical Work

9.3.1 *Space*

In designing the grid cell model, simplicity and accessibility were favored over path integration accuracy. The best approach to improving path integration accuracy is likely by incorporating learning into the model, which naturally invites the question of development: how could such a model be constructed during the development of the pup's neural circuitry? Hahnloser (2003) provides an elegant explanation of the self-organization of the head direction system; however, application of Hahnloser's technique to a path-integration system is not straight-forward. McNaughton et al. (2006)

present a developmental model of their neural architecture, but the details of their developmental model constrain it to have properties that differ critically from our model.

Müller and Wehner (1988) demonstrated a simple model of path integration involving vector arithmetic in polar coordinates whose systematic errors are similar to those observed (Seguinot et al., 1993). However, I am not aware of any attractor-based models that show such systematic errors (nor am I aware that any attempts have been made to show such a bias). Showing how these systematic errors arise from a developmental model of the dMEC circuitry would be an important contribution to understanding praxic navigation in rodents.

While grid cells are known to be sensitive to sensory cues, the dMEC model currently does not consider how associations should be formed between sensory cues and grid network phases. It has been argued that the hippocampus is critical for learning these associations (Redish and Touretzky, 1997, 1998), even if they are eventually consolidated to an extra-hippocampal area. Concurrent simulations of dMEC and hippocampus could explore how rats solve the simultaneous localization and mapping problem (Smith et al., 1990; Montemerlo et al., 2002): learning associations between PI coordinates and sensory cues in a novel environment.

The dMEC model is a firing rate model, and currently does not consider precise temporal pattern of spikes. However, the timing of spikes of hippocampal place cells is known to vary within the theta cycle depending on the rat's position relative to the center of the cell's field; preliminary evidence suggests this is true of layer II grid cells as well. Though the septum is critical for theta modulation, the origin of phase precession is not known, nor is it well understood how the animal's environment and behavior influence the phase precession effect. Moreover, it is not clear how phase precession in dMEC might cause phase precession in the hippocampus, where place cell firing is believed to depend on multiple simultaneously active neurons. Further research to explore these issues will be necessary.

9.3.2 *Context*

The context learning theory presented in Chapter 7 includes a series of simplifying assumptions in order to make model selection more computationally tractable. By using HMMs as a generative model of the animal's experiences, time is discretely modeled, and time dependency can only be modeled across a single time step. Daw et al. (2006) have suggested that a more sophisticated semi-Markov process better explains physiology studies of the dopamine system. In a semi-Markov process, an explicit "dwell time" is incorporated

into the model to reflect an expectation about the duration of time the animal expects to spend in each state or context. Context learning *in vivo* may similarly involve a more explicit inference about the duration of time an animal expects to be in each context. In addition, transition probabilities between contexts could be defined as parameters that are adapted based on the observed sequence of context transitions.

Skaggs and McNaughton (1998) started each session of their two-box experiment in box A. On the final day, rats were instead introduced into box B, but recalled the box A map. Since the boxes could only be distinguished based on a hallway connecting them, this recall was presumably due to previous days training where they became aware during travel between boxes that box A was their starting box.

The theory also assumes a perfect memory of past experiences, an assumption certain to be untrue. It would be interesting to explore to what extent memory limitations, perhaps imposed by the representation of the full history of experiences by a set of (in)sufficient statistics, would impact the predictions made by the present framework.

While HMMs are composed of discrete states, the distribution of hippocampal input patterns is likely better represented in a continuous fashion. To deal with more physiologically realistic input patterns, the generative model should be adapted to model regions of continuously varying input patterns. This leads to the question of what sort of topology one might expect to see both in the hippocampal input patterns and in the hippocampal representation itself. The development and application of machine learning tools to analyze physiological data sets to explore this and similar questions is an important avenue for future research.

On a more theoretical note, a more sophisticated notion of context might be considered in which contexts were not exclusive of one another. In the theory presented, contexts represent clusters of experiences where each experience is associated with one context. (Technically, each experience is probabilistically associated with every context; however, this uncertainty is intended to reflect uncertainty in context membership, not membership in multiple contexts.) This theory could be generalized to consider non-exclusive contexts, where multiple contexts may be simultaneously active. For example, m behavioral contexts and n environmental contexts could either be represented conjunctively via $m \times n$ exclusive contexts, or by $m + n$ non-exclusive contexts. In the latter case, recall of the correct behavioral context could occur independently of recall of the correct environmental context, permitting recall of a behavioral context in a new environment (or vice versa). This is related to the idea of maplets (Touretzky and Muller, 2006),

in which the hippocampal CA3 circuitry forms many small sub-networks, each with multiple attractor basins. Maplets have been shown to better account for discordant hippocampal representations than single attractor networks. In addition, Chapter 6 explores how multiple networks in dMEC might support such discordant activity in hippocampus.

Finally, while CA1 has historically been by far the most common recording site within hippocampus, explorations of CA3 and DG during subtle contextual manipulations are beginning to be performed. These new data will permit realistic exploration of detailed neural network models of context learning. In particular, how are new contextual representations gradually formed in an experience-dependent manner? An early study exploring potential learning rules to separate maps was offered by Fuhs and Touretzky (2000). However, that model lacked a principled basis on which to form new maps. With a theoretical foundation for context learning having been presented here, neural network instantiations that embody this foundation are ripe for exploration.

BIBLIOGRAPHY

Abraham, W. C., Logan, B., Greenwood, J. M., and Dragunow, M. (2002). Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *Journal of Neuroscience*, 22(21):9626–9634.

Ackil, J. E., Mellgren, R. L., Halgren, C., and Frommer, G. P. (1969). Effects of CS preexposures on avoidance learning in rats with hippocampal lesions. *J Comp Physiol Psychol*, 69(4):739–747.

Aggleton, J. P., Neave, N., Nagle, S., and Sahgal, A. (1995). A comparison of the effects of medial prefrontal, cingulate cortex, and cingulum bundle lesions on tests of spatial memory: evidence of a double dissociation between frontal and cingulum bundle contributions. *Journal of Neuroscience*, 15(11):7270–7281.

Agnihotri, N. T., Hawkins, R. D., Kandel, E. R., and Kentros, C. (2004). The long-term stability of new hippocampal place fields requires new protein synthesis. *Proc Natl Acad Sci U S A*, 101(10):3656–3661.

Agster, K. L., Fortin, N. J., and Eichenbaum, H. (2002). The hippocampus and disambiguation of overlapping sequences. *Journal of Neuroscience*, 22(13):5760–5768.

Ainge, J. A., van der Meer, M. A. A., and Wood, E. R. (2005). Disparity between sequence-dependent hippocampal activity and hippocampal lesion effects on a continuous t-maze task. Program no. 776.13. In *Society for Neuroscience Abstracts*, Washington, DC. Society for Neuroscience.

Ainge, J. A. and Wood, E. (2003). Excitotoxic lesions of the hippocampus impair performance on a continuous alternation t-maze task with short delays but not with no delay. Program no. 91.1. In *Society for Neuroscience Abstracts*, Washington, DC. Society for Neuroscience.

Amaral, D. G., Ishizuka, N., and Claiborne, B. (1990). Neurons, numbers and the hippocampal network. *Prog Brain Res*, 83:1–11.

Awh, E., Jonides, J., and Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *J Exp Psychol Hum Percept Perform*, 24(3):780–790.

Baeg, E. H., Kim, Y. B., Huh, K., Mook-Jung, I., Kim, H. T., and Jung, M. W. (2003). Dynamics of population code for working memory in the prefrontal cortex. *Neuron*, 40(1):177–188.

Bangasser, D. A., Waxler, D. E., Santollo, J., and Shors, T. J. (2006). Trace conditioning and the hippocampus: the importance of contiguity. *J Neurosci*, 26(34):8702–8706.

Bannerman, D. M., Good, M. A., Butcher, S. P., Ramsay, M., and Morris, R. G. M. (1995). Distinct components of spatial learning revealed by prior training and NMDA receptor blockade. *Nature*, 378:182–186.

Barnes, C. A., Suster, M. S., Shen, J., and McNaughton, B. L. (1997). Multistability of cognitive maps in the hippocampus of old rats. *Nature*, 388(6639):272–275.

Barrientos, R. M., O'Reilly, R. C., and Rudy, J. W. (2002). Memory for context is impaired by injecting anisomycin into dorsal hippocampus following context exploration. *Behav Brain Res*, 134(1-2):299–306.

Berger, T. W. and Orr, W. B. (1983). Hippocampectomy selectively disrupts discrimination reversal conditioning of the rabbit nictitating membrane response. *Behav Brain Res*, 8(1):49–68.

Beylin, A. V., Gandhi, C. C., Wood, G. E., Talk, A. C., Matzel, L. D., and Shors, T. J. (2001). The role of the hippocampus in trace conditioning: temporal discontinuity or task difficulty? *Neurobiol Learn Mem*, 76(3):447–461.

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2:32–48.

Blair, H. T., Lipscomb, B. W., and Sharp, P. E. (1997). Anticipatory time intervals of head-direction cells in the anterior thalamus of the rat, implications for path integration in the head-direction circuit. *Journal of Neurophysiology*, 78(1):145–159.

Blair, H. T. and Sharp, P. E. (1995). Anticipatory head direction signals in anterior thalamus: Evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. *Journal of Neuroscience*, 15(9):6260–6270.

Blair, H. T. and Sharp, P. E. (1998). Angular velocity-modulated head direction cells are segregated by hemisphere in the lateral mammillary nucleus of the rat. *Neuron*, 21(6):1387–1397.

Bostock, E., Muller, R. U., and Kubie, J. L. (1991). Experience-dependent modifications of hippocampal place cell firing. *Hippocampus*, 1(2):193–206.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol Bull*, 114(1):80–99.

Bouton, M. E. and King, D. A. (1983). Contextual control of the extinction of conditioned fear: tests for the associative value of the context. *J Exp Psychol Anim Behav Process*, 9(3):248–265.

Bouton, M. E., Westbrook, R. F., Corcoran, K. A., and Maren, S. (2006). Contextual and Temporal Modulation of Extinction: Behavioral and Biological Mechanisms. *Biol Psychiatry*. in press.

Bower, M. R., Euston, D. R., and McNaughton, B. L. (2005). Sequential-context-dependent hippocampal activity is not necessary to learn sequences with repeated elements. *J Neurosci*, 25(6):1313–1323.

Brown, J. E. and Skaggs, W. E. (2002). Discordant coding of spatial locations in the rat hippocampus. *Journal of Neurophysiology*, 88:1605–1613.

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

Brunswick, E. (1939). Probability as a determiner of rat behavior. *J Exp Psychol*, 25:175–197.

Burwell, R. D. (2000). The parahippocampal region: corticocortical connectivity. *Ann N Y Acad Sci*, 911:25–42.

Buytendijk, F. J. J. (1930). Über das umlernen. *Arch. nëerl. Physiol.*, 15:283–310.

Calton, J. I., Stackman, R. W., Goodridge, J. P., Archey, W. B., Dudchenko, P. A., and Taube, J. S. (2003). Hippocampal place cell instability after lesions of the head direction cell network. *Journal of Neuroscience*, 23(30):9719–9731.

Chan, K. H., Jarrard, L. E., and Davidson, T. L. (2003). The effects of selective ibotenate lesions of the hippocampus on conditioned inhibition and extinction. *Cogn Affect Behav Neurosci*, 3(2):111–119.

Chang, J. Y., Chen, L., Luo, F., Shi, L. H., and Woodward, D. J. (2002). Neuronal responses in the frontal cortico-basal ganglia system during delayed matching-to-sample task: ensemble recording in freely moving rats. *Experimental Brain Research*, 142:67–80.

Chen, L. L., Lin, L. H., Barnes, C. A., and McNaughton, B. L. (1994a). Head-direction cells in the rat posterior cortex: II. Contributions of visual and ideothetic information to the directional firing. *Experimental Brain Research*, 101:24–34.

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

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, 75:79–97.

Chiba, A. A., Kesner, R. P., and Reynolds, A. M. (1994). Memory for spatial location as a function of temporal lag in rats: role of hippocampus and medial prefrontal cortex. *Behavioral and Neural Biology*, 61:123–131.

Clark, R. E., West, A. N., Zola, S. M., and Squire, L. R. (2001). Rats with lesions of the hippocampus are impaired on the delayed nonmatching-to-sample task. *Hippocampus*, 11:176–186.

Clark, R. E., Zola, S. M., and Squire, L. R. (2000). Impaired recognition memory in rats after damage to the hippocampus. *Journal of Neuroscience*, 20:8853–8860.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., and Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386(6625):604–608.

Collett, T., Cartwright, B. A., and Smith, B. A. (1986). Landmark learning and visuo-spatial memories in gerbils. *Journal of Comparative Physiology A*, 158:835–851.

Conklin, J. and Eliasmith, C. (2005). A controlled attractor network model of path integration in the rat. *Journal of Computational Neuroscience*, 18:183–203.

Corcoran, K. A., Desmond, T. J., Frey, K. A., and Maren, S. (2005). Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction. *J Neurosci*, 25(39):8978–8987.

Corcoran, K. A. and Maren, S. (2001). Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *J Neurosci*, 21(5):1720–1726.

Corcoran, K. A. and Maren, S. (2004). Factors regulating the effects of hippocampal inactivation on renewal of conditional fear after extinction. *Learn Mem*, 11(5):598–603.

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., and Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, 279(5355):1347–1351.

Courville, A. C. (2006). A latent cause theory of classical conditioning. Technical Report CMU-RI-TR-06-29, Doctoral Dissertation, Robotics Institute, Carnegie Mellon University.

Courville, A. C., Daw, N. D., Gordon, G. J., and Touretzky, D. S. (2003). Model uncertainty in classical conditioning. In Thrun, S., Saul, L., and Schölkopf, B., editors, *Advances in Neural Information Processing Systems*, volume 16, pages 977–984, Cambridge, MA. MIT Press.

Courville, A. C., Daw, N. D., and Touretzky, D. S. (2004). Similarity and discrimination in classical conditioning: a latent variable account. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems*, volume 17, pages 313–320, Cambridge, MA. MIT Press.

Cressant, A., Muller, R. U., and Poucet, B. (1997). Failure of centrally placed objects to control firing fields of hippocampal place cells. *Journal of Neuroscience*, 17(7):2531–2542.

Darwin, C. A. (1873a). Origin of certain instincts. *Nature*, 7:417–418.

Darwin, C. A. (1873b). Perception in the Lower Animals. *Nature*, 7:360.

Daw, N. D., Courville, A. C., and Touretzky, D. S. (2006). Representation and timing in theories of the dopamine system. *Neural Comput*, 18(7):1637–1677.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*, 8(12):1704–1711.

Day, L. B., Weisend, M., Sutherland, R. J., and Schallert, T. (1999). The hippocampus is not necessary for a place response but may be necessary for pliancy. *Behavioral Neuroscience*, 113(5):914–924.

Day, M., Langston, R., and Morris, R. G. M. (2003). Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature*, 424:205–209.

De Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, rev. edition.

DeCoteau, W. E. and Kesner, R. P. (1998). Effects of hippocampal and parietal cortex lesions on the processing of multiple-object scenes. *Behavioral Neuroscience*, 112(1):68–82.

DeCoteau, W. E. and Kesner, R. P. (2000). A double dissociation between the rat hippocampus and medial caudoputamen in processing two forms of knowledge. *Behavioral Neuroscience*, 114(6):1096–1108.

Dennett, D. C. (1994). Cognitive science as reverse engineering: Several meanings of "top-down" and "bottom-up". In Prawitz, D., Skyrms, B., and Westerstahl, D., editors, *Proceedings of the 9th International Congress of Logical, Methodology and Philosophy of Science*, Amsterdam. North-holland.

Doboli, S., Minai, A. A., and Best, P. J. (2000). Latent Attractors: A Model for Context-Dependent Place Representations in the Hippocampus. *Neural Computation*, 12(5):1003–1037.

Duffy, S. N., Craddock, K. J., Abel, T., and Nguyen, P. V. (2001). Environmental enrichment modifies the PKA-dependence of hippocampal LTP and improves hippocampus-dependent memory. *Learning & Memory*, 8:26–34.

Dufort, R. H., Guttman, N., and Kimble, G. A. (1954). One-trial discrimination reversal in the white rat. *J Comp Physiol Psychol*, 47(3):248–249.

Dusek, J. A. and Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Science USA*, 94:7109–7114.

Duva, C. A., Floresco, S. B., Wunderlich, G. R., Lao, T. L., Pinel, J. P. J., and Phillips, A. G. (1997). Disruption of spatial but not object-recognition memory by neurotoxic lesions of the dorsal hippocampus in rats. *Behavioral Neuroscience*, 111(6):1184–1196.

Eilam, D. and Golani, I. (1989). Home base behavior of rats (*rattus norvegicus*) exploring a novel environment. *Behavioral Brian Research*, 34:199–211.

Elam, C. B. and Tyler, D. W. (1958). Reversal-learning following partial reinforcement. *Am J Psychol*, 71(3):583–586.

Ermentrout, B. (1994). Reduction of Conductance-Based Models with Slow Synapses to Neural Nets. *Neural Computation*, 6:679–695.

Etienne, A. S. and Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, 14:180–192.

Etienne, A. S., Maurer, R., and Saucy, F. (1988). Limitations in the assessment of path dependent information. *Behavior*, 106:81–111.

Etienne, A. S., Maurer, R., Saucy, F., and Teroni, E. (1986). Short-distance homing in the golden hamster after a passive outward journey. *Animal Behavior*, 34:696–715.

Etienne, A. S., Maurer, R., and Seguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *Journal of Experimental Biology*, 199(1):201–209.

Fagan, A. M. and Olton, D. S. (1986). Learning sets, discrimination reversal, and hippocampal function. *Behav Brain Res*, 21(1):13–20.

Fanselow, M. S. (1990). Factors governing one trial contextual conditioning. *Animal Learning & Behavior*, 18:264–270.

Feng, R., Rampon, C., Tang, Y., Shrom, D., Jin, J., Kyin, M., Sopher, B., Martin, G. M., Kim, S., Langdon, R. B., Sisodia, S., and Tsien, J. (2001). Deficient neurogenesis in forebrain-specific *presenilin-1* knockout mice is associated with reduced clearance of hippocampal memory traces. *Neuron*, 32:911–926.

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

Forney, G. D. J. (1973). The viterbi algorithm. *Proc IEEE*, 61:266–278.

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

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

Frank, L. M., Brown, E. N., and Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27:169–178.

Frohardt, R. J., Guarraci, F. A., and Bouton, M. E. (2000). The effects of neurotoxic hippocampal lesions on two effects of context after fear extinction. *Behav Neurosci*, 114(2):227–240.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques. *Econometrics Journal*, 7:143–167.

Fuhs, M. C. and Touretzky, D. S. (2000). Synaptic learning models of map separation in hippocampus. *Neurocomputing*, 32:379–384.

Fuhs, M. C. and Touretzky, D. S. (2006). A spin glass model of path integration in rat medial entorhinal cortex. *J Neurosci*, 26(16):4266–4276.

Fuhs, M. C., VanRhoads, S. R., Casale, A. E., McNaughton, B., and Touretzky, D. S. (2005). Influence of path integration versus environmental orientation on place cell remapping between visually identical environments. *J Neurophysiol*, 94(4):2603–2616.

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., and Moser, E. I. (2005). Preserved spatial and temporal firing structure in entorthinal grid cells during remapping in the hippocampus. Program no. 198.6. In *Society for Neuroscience Abstracts*, Washington, DC. Society for Neuroscience.

Fyhn, M., Molden, S., Hollup, S., Moser, M.-B., and Moser, E. I. (2002). Hippocampal neurons responding to first-time dislocation of a target object. *Neuron*, 35:555–566.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., and Moser, M.-B. (2004). Spatial representation in the entorhinal cortex. *Science*, 305:1258–1264.

Gallistel, C. R. and Gibbon, J. (2000). Time, rate, and conditioning. *Psychol Rev*, 107(2):289–344.

Gatling, F. (1952). The effect of repeated simtulus reversals on learning in the rat. *Journal of Comparative and Physiological Psychology*, 45:347–351.

Gerrard, J. L., Bower, M. R., Insel, N., Lipa, P., Barnes, C. A., and McNaughton, B. L. (2001). A long day's journey into night. program number 643.12. *Soc Neurosci Abstr*, 27:643.12.

Gilbert, P. E. and Kesner, R. P. (2003). Localization of function within the dorsal hippocampus: the role of the CA3 subregion in paired-associate learning. *Behavioral Neuroscience*, 117(6):1385–1394.

Gilbert, P. E. and Kesner, R. P. (2006). The role of the dorsal CA3 hippocampal subregion in spatial working memory and pattern separation. *Behav Brain Res*, 169(1):142–149.

Gilbert, P. E., Kesner, R. P., and DeCoteau, W. E. (1998). Memory for spatial location: Role of the hippocampus in mediating spatial pattern separation. *Journal of Neuroscience*, 18(2):804–810.

Gilbert, P. E., Kesner, R. P., and Lee, I. (2001). Dissociating hippocampal subregions: a double dissociation between dentate gyrus and CA1. *Hippocampus*, 11:626–636.

Gluck, M. A. and Myers, C. E. (1993). Hippocampal Mediation of Stimulus Representation: A computational Theory. *Hippocampus*, 3(4):491–516.

Gluck, M. A. and Myers, C. E. (1996). Integrating behavioral and physiological models of hippocampal function. *Hippocampus*, 6(6):643–653.

Golani, I., Benjamini, Y., and Eilam, D. (1993). Stopping behavior: Constraints on exploration in rats (*rattus norvegicus*). *Behavioral Brain Research*, 53(1-2):21–33.

Gold, A. E. and Kesner, R. P. (2005). The role of the CA3 subregion of the dorsal hippocampus in spatial pattern completion in the rat. *Hippocampus*, 15(6):808–814.

Goodridge, J. P., Dudchenko, P. A., Worboys, K. A., Golob, E. J., and Taube, J. S. (1998). Cue control and head direction cells. *Behavioral Neuroscience*, 112(4):1–13.

Goodridge, J. P. and Taube, J. S. (1994). The Effect of lesions of the postsubiculum on head direction cell firing in the anterior thalamic nuclei. *Society for Neuroscience Abstracts*, 20(1):805.

Goodridge, J. P. and Taube, J. S. (1995). Preferential use of the landmark navigational system by head direction cells in rats. *Behavioral Neuroscience*, 109(1):49–61.

Goodridge, J. P. and Taube, J. S. (1997). Interaction between the postsubiculum and anterior thalamus in the generation of head direction cell activity. *Journal of Neuroscience*, 17:9315–9330.

Goodridge, J. P. and Touretzky, D. S. (2000). Modeling attractor deformation in the rodent head-direction system. *J Neurophysiol*, 83(6):3402–3410.

Gothard, K. M., Hoffman, K. L., Battaglia, F. P., and McNaughton, B. L. (2001). Dentate Gyrus and CA1 ensemble activity during spatial reference frame shifts in the presence and absense of visual input. *Journal of Neuroscience*, 21(18):7284–7292.

Gothard, K. M., Skaggs, W. E., and McNaughton, B. L. (1996a). Dynamics of mismatch correction in the hippocampal ensemble code for space: Interaction between path integration and environmental cues. *Journal of Neuroscience*, 16(24):8027–8040.

Gothard, K. M., Skaggs, W. E., Moore, K. M., and McNaughton, B. L. (1996b). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *Journal of Neuroscience*, 16(2):823–835.

Gray, C. M., Maldonado, P. E., Wilson, M., and McNaughton, B. (1995). Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods*, 63(1-2):43–54.

Griffiths, T. L. and Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognit Psychol*, 51(4):334–384.

Grosslight, J. H., Hall, J. F., and Scott, W. (1954). Reinforcement schedules in habit reversal; a confirmation. *J Exp Psychol*, 48(3):173–174.

Grünbaum, B. and Shephard, G. C. (1987). *Tilings and Patterns*. Freeman, New York.

Guzowski, J. F., Knierim, J. J., and Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron*, 44(4):581–584.

Hafting, T., Fyhn, M. H., Moser, M.-B., and Moser, E. I. (2004). Mnemonic properties of position-modulated neurons in dorsocaudal medial entorhinal cortex – or the entorhinal cortex as a cognitive map. Program no. 330.7. In *Society for Neuroscience Abstracts*, Washington, DC. Society for Neuroscience.

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

Hahnloser, R. H. R. (2003). Emergence of neural integration in the head-direction system by visual supervision. *Neuroscience*, 120:877–891.

Hampson, R. E., Heyser, C. J., and Deadwyler, S. A. (1993). Hippocampal Cell Firing Correlates of Delayed-Match-to-Sample Performance in the Rat. *Behavioral Neuroscience*, 107(5):715–739.

Hargreaves, E. L., Rao, G., Lee, I., and Knierim, J. J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science*, 308(5729):1792–1794.

Hasselmo, M. E., Bodelón, C., and Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Comput*, 14(4):793–817.

Hasselmo, M. E. and Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Netw*, 18(9):1172–1190.

Hasselmo, M. E. and Schnell, E. (1994). Laminar Selectivity of the Cholinergic Suppression of Synaptic Transmission in Rat Hippocampal Region CA1: Computational Modeling and Brain Slice Physiology. *Journal of Neuroscience*, 14(6):3898–3914.

Hasselmo, M. E., Schnell, E., and Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *J Neurosci*, 15(7 Pt 2):5249–5262.

Hasselmo, M. E., Wyble, B. P., and Wallenstein, G. V. (1996). Retrieval of episodic memories: Role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus*, 6(6):693–708.

Hayman, R. M. A., Chakraborty, S., Anderson, M. I., and Jeffery, K. J. (2003). Context-specific acquisition of location discrimination by hippocampal place cells. *European Journal of Neuroscience*, 18:1–10.

Hebb, D. O. (1949). *The Organization of Behavior*. Wiley, New York.

Hetherington, P. A. and Shapiro, M. L. (1997). Hippocampal Place Fields Are Altered by the Removal of Single Visual Cues in a Distance-Dependent Manner. *Behavioral Neuroscience*, 111(1):20–34.

Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: A theory. *Behavioral Biology*, 12:421–444.

Hobin, J. A., Ji, J., and Maren, S. (2006). Ventral hippocampal muscimol disrupts context-specific fear memory retrieval after extinction in rats. *Hippocampus*, 16(2):174–182.

Hoh, T., Beiko, J., Boon, F., Weiss, S., and Cain, D. P. (1999). Complex behavioral strategy and reversal learning in the water maze without NMDA receptor-dependent long-term potentiation. *J Neurosci*, 19(10):RC2.

Hok, V., Save, E., Lenck-Santini, P. P., and Poucet, B. (2005). Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proc Natl Acad Sci U S A*, 102(12):4602–4607.

Hollup, S. A., Kjelstrup, K. G., Hoff, J., Moser, M.-B., and Moser, E. I. (2001a). Impaired recognition of the goal location during spatial navigation in rats with hippocampal lesions. *Journal of Neuroscience*, 21(12):4505–4513.

Hollup, S. A., Molden, S., Donnet, J. G., Moser, M.-B., and Moser, E. I. (2001b). Accumulation of hippocampal place fields at the goal location in annular watermaze task. *Journal of Neuroscience*, 21(6):1635–1644.

Hölscher, C., Jacob, W., and Mallot, H. A. (2004). Learned association of allocentric and egocentric information in the hippocampus. *Exp Brain Res*, 158(2):233–240.

Holt, W. and Maren, S. (1999). Muscimol inactivation of the dorsal hippocampus impairs contextual retrieval of fear memory. *J Neurosci*, 19(20):9054–9062.

Honey, R. C. and Good, M. (1993). Selective hippocampal lesions abolish the contextual specificity of latent inhibition and of conditioning. *Behavioral Neuroscience*, 107:23–33.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective ocmputational abilities. *Proceedings of the National Academy of Sciences, USA*, 79:2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, USA*, 81:3088–3092.

Howard, M. W., Fotedar, M. S., Datey, A. V., and Hasselmo, M. E. (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(1):75–116.

Huerta, P. T., Sun, L. D., Wilson, M. A., and Tonegawa, S. (2000). Formation of temporal memory requires NMDA receptors within CA1 pyramidal neurons. *Neuron*, 25:473–480.

Jarrard, L. E. (1993). On the role of the hippocampus in learning and memory in the rat. *Behavioral Neural Biology*, 60(1):9–26.

Jarrard, L. E., Feldon, J., Rawlins, J. N., Sinden, J. D., and Gray, J. A. (1986). The effects of intrahippocampal ibotenate on resistance to extinction after continuous or partial reinforcement. *Exp Brain Res*, 61(3):519–530.

Jeffery, K. J. (2000). Plasticity of the hippocampal cellular representation of space. In Holscher, C., editor, *Neuronal Mechanisms of Memory Formation*, pages 100–124. Cambridge University Press, Cambridge, UK.

Jeffery, K. J. and O'Keefe, J. M. (1999). Learned interaction of visual and idiothetic cues in the control of place field orientation. *Experimental Brain Research*, 127:151–161.

Ji, J. and Maren, S. (2005). Electrolytic lesions of the dorsal hippocampus disrupt renewal of conditional fear after extinction. *Learn Mem*, 12(3):270–276.

Jung, M. W. and McNaughton, B. L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, 3(2):165–182.

Jung, M. W., Qin, Y., Lee, D., and Mook-Jung, I. (2000). Relationship among discharges of neighboring neurons in the rat prefrontal cortex during spatial working memory tasks. *J Neurosci*, 20(16):6166–6172.

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

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773–795.

Kentros, C., Hargreaves, E., Hawkins, R. D., Kandel, E. R., Shapiro, M., and Muller, R. U. (1998). Abolition of Long-Term Stability of New Hippocampal Place Cell Maps by NMDA Receptor Blockade. *Science*, 280:2121–2126.

Kentros, C. G., Agnihotri, N. T., Streater, S., Hawkins, R. D., and Kandel, E. R. (2004). Increased attention to spatial context increases both place field stability and spatial memory. *Neuron*, 42:283–295.

Kesner, R. P., Gilbert, P. E., and Barua, L. A. (2002). The role of the hippocampus in memory for the temporal order of a sequence of odors. *Behavioral Neuroscience*, 116(2):286–290.

Kesner, R. P., Lee, I., and Gilbert, P. (2004). A behavioral assessment of hippocampal function based on a subregional analysis. *Reviews in the Neurosciences*, 15:333–351.

Kesner, R. P. and Novak, J. M. (1982). Serial position curve in rats: role of the dorsal hippocampus. *Science*, 218(4568):173–175.

Kimble, D. P. (1968). Hippocampus and internal inhibition. *Psychological Bulletin*, 70(5):285–295.

Kimble, D. P. and Kimble, R. J. (1965). Hippocampectomy and response perseveration in the rat. *J Comp Physiol Psychol*, 60(3):474–476.

Kleene, S. C. (1943). Recursive predicates and quantifiers. *Transactions of the American Mathematical Society*, 53:41–73.

Knierim, J. J. (2002). Dynamic interactions between local surface cues, distal landmarks, and intrinsic circuitry in hippocampal place cells. *Journal of Neuroscience*, 22:6254–6264.

Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1995). Place cells, head direction cells, and the learning of landmark stability. *Journal of Neuroscience*, 15:1648–1659.

Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1998). Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *Journal of Neurphysiology*, 80:425–446.

Knierim, J. J. and Rao, G. (2003). Distal landmarks and hippocampal place cells: effects of relative translation verses rotation. *Hippocampus*, 13:604–617.

Kubie, J. L. and Ranck, Jr., J. B. (1983). Sensory-Behavioral Correlates in Individual Hippocampus Neurons in Three Situations: Space and Context. In Seifert, W., editor, *Neurobiology of the Hippocampus*, pages 433–447. Academic Press Inc., New York.

Káli, S. and Dayan, P. (2000). The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model. *J Neurosci*, 20(19):7463–7477.

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

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

Lee, I., Hunsaker, M. R., and Kesner, R. P. (2005). The role of hippocampal subregions in detecting spatial novelty. *Behav Neurosci*, 119(1):145–153.

Lee, I. and Kesner, R. P. (2002). Differential contribution of NMDA receptors in hippocampal subregions to spatial working memory. *Nature Neuroscience*, 5(2):162–168.

Lee, I. and Kesner, R. P. (2003). Differential roles of dorsal hippocampal subregions in spatial working memory with short verses intermediate delay. *Behavioral Neuroscience*, 117(5):1044–1053.

Lee, I., Rao, G., and Knierim, J. J. (2004a). A double dissociation between hippocampal subfields: Different time course of CA3 and CA1 place cells for processing changed environments. *Neuron*, 42:803–815.

Lee, I., Yoganarasimha, D., Rao, G., and Knierim, J. J. (2004b). Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. *Nature*, 430:456–459.

Lenck-Santini, P.-P., Muller, R. U., Save, E., and Poucet, B. (2002). Relationships between place cell firing fields and navigational decisions by rats. *J Neurosci*, 22(20):9035–9047.

Lenck-Santini, P.-P., Rivard, B., Muller, R. U., and Poucet, B. (2005). Study of CA1 place cell activity and exploratory behavior following spatial and nonspatial changes in the environment. *Hippocampus*, 15(3):356–369.

Lenck-Santini, P. P., Save, E., and Poucet, B. (2001). Place-cell firing does not depend on the direction of turn in a Y-maze alternation task. *Eur J Neurosci*, 13(5):1055–1058.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., and Moser, E. I. (2005a). Pattern separation in the dentate gyrus during morphing of two environments. Program no. 198.5. In *Society for Neuroscience Abstracts*, Washington, DC. Society for Neuroscience.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., and Moser, E. I. (2006a). Expansion recoding is expressed in ca3 but not in the dentate gyrus. program no. 68.3. In *Society for Neuroscience Abstracts*, Washington, DC. Society for Neuroscience.

Leutgeb, J. K., Leutgeb, S., Treves, A., Meyer, R., Barnes, C. A., McNaughton, B. L., Moser, M.-B., and Moser, E. I. (2005b). Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron*, 48(2):345–358.

Leutgeb, S., Jezek, K., Colgin, L. L., Leutgeb, J. K., McNaughton, B. L., Moser, E. I., and Moser, M. B. (2006b). Continuous and discontinuous representations in ensembles of hippocampal place cells. program no. 68.5. In *Society for Neuroscience Abstracts*.

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

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B., and Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, 305(5688):1295–1298.

Lever, C., Wills, T., Cacucci, F., Burgess, N., and O'Keefe, J. (2002). Long-term placticity in hippocampal place-cell representation of environmental geometry. *Nature*, 416:90–94.

Levy, W. B. (1989). A computational approach to hippocampal function. In Hawkins, R. D. and Bower, G. H., editors, *Computational Models of Learning in Simple Neural Systems*, volume 23 of *The Psychology of Learning and Motivation*, pages 243–305. Academic Press, San Diego CA.

Levy, W. B. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, 6(6):579–591.

Long, J. M. and Kesner, R. P. (1996). The effects of dorsal versus ventral hippocampal, total hippocampal, and parietal cortex lesions on memory for allocentric distance in rats. *Behavioral Neuroscience*, 110(5):922–932.

Long, J. M. and Kesner, R. P. (1998). Effects of hippocampal and parietal cortex lesions on memory for egocentric distance and spatial location information in rats. *Behavioral Neuroscience*, 112(3):480–495.

Lörincz, A. and Buzsáki, G. (2000). Two-phase computational model training long-term memories in the entorhinal-hippocampal region. *Ann N Y Acad Sci*, 911:83–111.

Louie, K. and Wilson, M. A. (2001). Temporally structured replay of awake hippcampal ensemble activity during rapid eye movement sleep. *Neuron*, 29:145–156.

Mackintosh, N. J. (1962). The effects of overtraining on a reversal and a nonreversal shift. *Journal of Comparative and Physiological Psychology*, 55(4):555–559.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annu Rev Neurosci*, 24:897–931.

Maren, S., Aharonov, G., and Fanselow, M. S. (1997). Neurotoxic lesions of the dorsal hippocampus and Pavlovian fear conditioning in rats. *Behav Brain Res*, 88(2):261–274.

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

Marr, D. (1971). Simple memory: A theory of archicortex. *Philosophical Transactions of the Royal Society of London*, 262(841):23–81.

Masters, J. and Skaggs, W. (2001). Effects of hippocampal remapping on a goal directed navigational task. *Society for Neuroscience Abstracts*, 27:643.9.

Matus-Amat, P., Higgins, E. A., Barrientos, R. M., and Rudy, J. W. (2004). The role of the dorsal hippocampus in the acquisition and retrieval of context memory representations. *J Neurosci*, 24(10):2431–2439.

Maurer, A. P., Vanrhoads, S. R., Sutherland, G. R., Lipa, P., and McNaughton, B. L. (2005). Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. *Hippocampus*, 15(7):841–852.

McClelland, J. L. and Goddard, N. H. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus*, 6(6):654–665.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457.

McDonald, R. J., King, A. L., and Hong, N. S. (2001). Context-specific interference on reversal learning of a stimulus-response habit. *Behav Brain Res*, 121(1-2):149–165.

McDonald, R. J., Ko, C. H., and Hong, N. S. (2002). Attenuation of context-specific inhibition on reversal learning of a stimulus-response task in rats with neurotoxic hippocampal damage. *Behav Brain Res*, 136(1):113–126.

McDonald, R. J. and White, N. M. (1993). A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. *Behav Neurosci*, 107(1):3–22.

McGregor, A., Hayward, A. J., Pearce, J. M., and Good, M. A. (2004). Hippocampal lesions disrupt navigation based on the shape of the environment. *Behavioral Neuroscience*, 118(5):1011–1021.

McHugh, T. J., Blum, K. I., Tsien, J. Z., Tonegawa, S., and Wilson, M. A. (1996). Impaired hippocampal represntation of space in CA1-specific NMDAR1 knockout mice. *Cell*, 87:1339–1349.

McNaughton, B. L. (1989). Neuronal Mechanisms for Spatial Computation and Information Storage. In Nadel, L., Cooper, L., Culicover, P., and Harnish, R. M., editors, *Neural Connections, Mental Computation*, pages 285–350. MIT Press, Cambridge, MA.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., Kudrimoti, H., Qin, Y., Skaggs, W. E., Suster, M., and Weaver, K. L. (1996). Deciphering the hippocampal polyglot: The hippocampus as a path integration system. *Journal of Experimental Biology*, 199(1):173–186.

McNaughton, B. L., Barnes, C. A., Meltzer, J., and Sutherland, R. J. (1989). Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. *Experimental Brain Research*, 76:485–496.

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

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

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

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

McNaughton, B. L., O'Keefe, J., and Barnes, C. A. (1983). The stereotrode: A new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *Journal of Neuroscience Methods*, 8:391–397.

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

Mehta, M. R., Quirk, M. C., and Wilson, M. A. (2000). Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron*, 25(3):707–715.

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 6:831–860.

Micheau, J., Riedel, G., v. L. Roloff, E., Inglis, J., and Morris, R. G. M. (2004). Reversible hippocampal inactivation partially dissociates how and where to search in the water maze. *Behavioral Neuroscience*, 118(5):1022–1032.

Miller, K. D., Keller, J. B., and Stryker, M. P. (1989). Ocular dominance column development: analysis and simulation. *Science*, 245(4918):605–615.

Mittelstaedt, H. (1962). Control systems of orientation in insects. *Annual Review of Entemology*, 7:177–198. Cited in (maurer95).

Mittelstaedt, M. L. and Mittelstaedt, H. (1980). Homing by path integration in a mammal. *Naturwissenschaften*, 67:566–567.

Mochida, H., Sato, K., Sasaki, S., Yazawa, I., Kamino, K., and Momose-Sato, Y. (2001). Effects of anisomycin on LTP in the hippocampal CA1: long-term analysis using optical recording. *Neuroreport*, 12(5):987–991.

Montemerlo, M., Thrun, S., Koller, D., and Wegbreit, B. (2002). FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI-02)*, pages 593–598, Menlo Park, CA. AAAI.

Morris, R. G. M. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, 12:239–260.

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

Morris, R. G. M., Schenk, F., Tweedie, F., and Jarrard, L. E. (1990). Ibotenate Lesions of Hippocampus and/or Subiculum: Dissociating Components of Allocentric Spatial Learning. *European Journal of Neuroscience*, 2:1016–1028.

Moser, E. I. (2006). Spatial maps in hippocampal and parahippocampal cortices. Program No. 466. Lecture delivered at the Society for Neuroscience Annual Meeting, Washington, DC.

Moser, M. B., Moser, E. I., Forrest, E., Anderson, P., and Morris, R. G. M. (1995). Spatial learning with a minislab in the dorsal hippocampus. *Proceedings of the National Academy of Sciences, USA*, 92:9697–9701.

Mulder, A. B., Tabuchi, E., and Wiener, S. I. (2004). Neurons in hippocampal afferent zones of rat striatum parse routes into multi-pace segments during maze navigation. *Eur J Neurosci*, 19(7):1923–1932.

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

Muller, R. U. and Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, 7:1951–1968.

Muller, R. U., Kubie, J. L., Bostock, E. M., Taube, J. S., and Quirk, G. J. (1991). Spatial firing correlates of neurons in the hippocampal formation of freely moving rats. In Paillard, J., editor, *Brain and Space*, pages 296–333. Oxford University Press, New York.

Myers, C. E., Gluck, M. A., and Granger, R. (1995). Dissociation of hippocampal and entorhinal function in associative learning: A computational approach. *Psychobiology*, 23:116–138.

Nadel, L. and Willner, J. (1980). Context and conditioning: A place for space. *Physiological Psychology*, 8(2):218–228.

Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., Kato, A., Carr, C. A., Johnston, D., Wilson, M. A., and Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science*, 297:211–218.

Nakazawa, K., Sun, L. D., Quirk, M. C., Rondi-Reig, L., Wilson, M. A., and Tonegawa, S. (2003). Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. *Neuron*, 38(2):305–315.

Neave, N., Lloyd, S., Sahgal, A., and Aggleton, J. P. (1994). Lack of effect of lesions in the anterior cingulate cortex and retrosplenial cortex on certain tests of spatial memory in the rat. *Behavioural Brain Research*, 65:89–101.

O'Keefe, J. (1989). Computations the hippocampus might perform. In Nadel, L., Cooper, L., Culicover, P., and Harnish, R. M., editors, *Neural Connections, Mental Computation*, pages 225–284. MIT Press, Cambridge, MA.

O'Keefe, J. and Conway, D. H. (1978). Hippocampal Place Units in the Freely Moving Rat: Why They Fire Where They Fire. *Experimental Brain Research*, 31:573–590.

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

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

Oler, J. A. and Markus, E. J. (2000). Age-related deficits in the ability to encode contextual change: a place cell analysis. *Hippocampus*, 10:338–350.

Olton, D. S. (1972). Discrimination reversal performance after hippocampal lesions: an enduring failure of reinforcement and non-reinforcement to direct behavior. *Physiol Behav*, 9(3):353–356.

O'Reilly, R. C. and McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, 4(6):661–682.

Otmakhova, N. A., Lewey, J., Asrican, B., and Lisman, J. E. (2005). Inhibition of perforant path input to the CA1 region by serotonin and noradrenaline. *J Neurophysiol*, 94(2):1413–1422.

Otmakhova, N. A. and Lisman, J. E. (1998). Dopamine selectively inhibits the direct cortical pathway to the CA1 hippocampal region. *J Neurosci*, 19(4):1437–1445.

Otnaess, M. K., Brun, V. H., Moser, M.-B., and Moser, E. I. (1999). Pretraining prevents spatial learning impairment after saturation of hippocampal long-term potentiation. *The Journal of Neuroscience*, 19:RC49 (1–5).

Otto, T. and Eichenbaum, H. (1992). Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: Evidence for hippocampal processing in recognition memory. *Hippocampus*, 2(3):323–334.

Oxford English Dictionary (1989). Oxford University Press. http://www.oed.com.

Pearce, J. M. and Hall, G. (1980). A model for pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev*, 87(6):532–552.

Pennes, E. S. and Ison, J. R. (1967). Effects of partial reinforcement on discrimination learning and subsequent reversal or extinction. *J Exp Psychol*, 74(2):219–224.

Phillips, R. G. and Le Doux, J. E. (1992). Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behavioral Neuroscience*, 106(2):274–285.

Postle, B. R., Awh, E., Jonides, J., Smith, E. E., and D'Esposito, M. (2004). The where and how of attention-based rehearsal in spatial working memory. *Brain Res Cogn Brain Res*, 20(2):194–205.

Pubols, B. H. (1962). Serial reversal learning as a function of the number of trails per reversal. *Journal of Comparative and Physiological Psychology*, 55(1):66–68.

Purves, D., Bonardi, C., and Hall, G. (1995). Enhancement of latent inhibition in rats with electrolytic lesions of the hippocampus. *Behav Neurosci*, 109(2):366–370.

Pylyshyn, Z. (1989). Computing in cognitive science. In *Foundations of Cognitive Science*. MIT Press, Cambridge.

Quirk, G. J., Muller, R. U., and Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience*, 10(6):2008–2017.

Quirk, G. J., Muller, R. U., Kubie, J. L., and Ranck, Jr., J. B. (1992). The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *Journal of Neuroscience*, 12(5):1945–1963.

Ragozzino, M. E., Detrick, S., and Kesner, R. P. (1999). Involvement of the prelimbic-infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. *Journal of Neuroscience*, 19(11):4585–4594.

Rampon, C., Tang, Y.-P., Goodhouse, J., Shimizu, E., Kyin, M., and Tsien, J. Z. (2000). Enrichment induces structural changes and recovery from nonspatial memory deficits in CA1 NMDAR1-knockout mice. *Nature*, 3(3):238–244.

Rawlins, J. N., Feldon, J., Tonkiss, J., and Coffey, P. J. (1989). The role of subicular outputs in the development of the partial reinforcement extinction effect. *Exp Brain Res*, 77(1):153–160.

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

Redish, A. D. (2001). The hippocampal debate: are we asking the right questions? *Behavioral Brain Research*, 127:81–98.

Redish, A. D., Battagli, F. P., Chawla, M. K., Ekstrom, A. D., Gerrard, J. L., Lipa, P., Rosenzweig, E. S., Worley, P. F., Guzowski, J. F., McNaughton, B. L., and Barnes, C. A. (2001). Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *Journal of Neuroscience*, 21:RC134 (1–6).

Redish, A. D., Elga, A. N., and Touretzky, D. S. (1996). A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems*, 7(4):671–685.

Redish, A. D., McNaughton, B. L., and Barnes, C. A. (1998). Reconciling Barnes et al. (1997) and Tanila et al. (1997a,b). *Hippocampus*, 8(5):438–443.

Redish, A. D., Rosenzweig, E. S., Bohanick, J. D., McNaughton, B. L., and Barnes, C. A. (2000). Dynamics of hippocampal ensemble activity realignment: time versus space. *Journal of Neuroscience*, 20(24):9298–9309.

Redish, A. D. and Touretzky, D. S. (1997). Cognitive maps beyond the hippocampus. *Hippocampus*, 7(1):15–35.

Redish, A. D. and Touretzky, D. S. (1998). The role of the hippocampus in solving the Morris Water Maze. *Neural Computation*, 10(1):73–111.

Reid, L. S. (1953). The development of noncontinuity behavior through continuity learning. *Journal of Experimental Psychology*, 46(2):107–112.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditiong: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. and Prokesy, W. F., editors, *Classical Conditioning II: Current Research and Theory*, pages 64–99. Appelton Century Crofts, New York.

Richmond, M. A., Yee, B. K., Pouzet, B., Veenman, L., Rawlins, J. N., Feldon, J., and Bannerman, D. M. (1999). Dissociating context and space within the hippocampus: effects of complete, dorsal, and ventral excitotoxic hippocampal lesions on conditioned freezing and spatial learning. *Behav Neurosci*, 113(6):1189–1203.

Rivard, B., Li, Y., Lenck-Santini, P.-P., Poucet, B., and Muller, R. U. (2004). Representation of objects in space by two classes of hippocampal pyramidal cells. *J Gen Physiol*, 124(1):9–25.

Robitsek, R. J., Fortin, N., and Eichenbaum, H. (2005). Hippocampal unit activity during continuous and delayed t-maze spatial alternation. Program no. 776.7. In *Society for Neuroscience Abstracts*, Washington, DC. Society for Neuroscience.

Rolls, E. T. (1996). A theory of hippocampal function in memory. *Hippocampus*, 6:601–620.

Rosenzweig, E. S., Redish, A. D., McNaughton, B. L., and Barnes, C. A. (2003). Hippocampal map realignment and spatial learning. *Nature Neuroscience*, 6(6):609–615.

Rotenberg, A., Mayford, M., Hawkins, R. D., Kandel, E. R., and Muller, R. U. (1996). Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. *Cell*, 87:1351–1361.

Rudy, J. W. and Matus-Amat, P. (2005). The ventral hippocampus supports a memory representation of context and contextual fear conditioning: implications for a unitary function of the hippocampus. *Behav Neurosci*, 119(1):154–163.

Rudy, J. W. and Sutherland, R. J. (1995). Configural association theory and the hippocampal formation: An appraisal and reconfiguration. *Hippocampus*, 5:375–389.

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

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

Saucier, D. and Cain, D. P. (1995). Spatial learning without NMDA receptor-dependent long-term potentiation. *Nature*, 378:186–189.

Save, E., Nerad, L., and Poucet, B. (2000). Contribution of multiple sensory information to place field stability in hippocampal place cells. *Hippocampus*, 10(1):64–76.

Schmajuk, N. A. and DiCarlo, J. J. (1991). A neural network approach to hippocampal function in classical conditioning. *Behav Neurosci*, 105(1):82–110.

Schmajuk, N. A. and DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. *Psychol Rev*, 99(2):268–305.

Schmitz, D., Mellor, J., Breustedt, J., and Nicoll, R. A. (2003). Presynaptic kainate receptors impart an associative property to hippocampal mossy fiber long-term potentiation. *Nature Neuroscience*, 6(10):1058–1063.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

Searle, J. R. (1980). Minds, brains and programs. *The Behavioral and Brain Sciences*, 3:417–424.

Seguinot, V., Maurer, R., and Etienne, A. S. (1993). Dead reckoning in a small mammal: The evaluation of distance. *Journal of Comparative Physiology A*, 173:103–113.

Selden, N. R., Everitt, B. J., Jarrard, L. E., and Robbins, T. W. (1991). Complementary roles for the amygdala and hippocampus in aversive conditioning to explicit and contextual cues. *Neuroscience*, 42(2):335–350.

Shapiro, M. L., Tanila, H., and Eichenbaum, H. (1997). Cues the hippocampal place cells encode: Dynamic and hierarchcal representation of local and distal stimuli. *Hippocampus*, 7(6):624–642.

Sharp, P. E., Blair, H. T., and Cho, J. (2001). The anatomical and computational basis of the rat head-direction cell signal. *Trends Neurosci*, 24(5):289–294.

Sharp, P. E., Blair, H. T., Etkin, D., and Tzanetos, D. B. (1995). Influences of vestibular and visual motion information on the spatial firing patterns of hippocampal place cells. *Journal of Neuroscience*, 15(1):173–189.

Shatz, C. J. (1996). Emergence of order in visual system development. *Proc Natl Acad Sci U S A*, 93(2):602–608.

Shen, J., Kudrimoti, H. S., McNaughton, B. L., and Barnes, C. A. (1998). Reactivation of neuronal ensembles in hippocampal dentate gyrus during sleep after spatial experience. *J Sleep Res*, 7 Suppl 1:6–16.

Shimizu, E., Tang, Y., Rampon, C., and Tsien, J. Z. (2000). NMDA receptor-dependent synaptic reinforcement as a crucial process for memory consolidation. *Science*, 290:1170–1174.

Shohamy, D., Allen, M. T., and Gluck, M. A. (2000). Dissociating entorhinal and hippocampal involvement in latent inhibition. *Behav Neurosci*, 114(5):867–874.

Silva, A. J., Stevens, C. F., Tonegawa, S., and Wang, Y. (1992). Deficient hippocampal long-term potentiation in alpha-calcium-calmodulin kinase II mutant mice. *Science*, 257(5067):201–206.

Silveira, J. M. and Kimble, D. P. (1968). Brightness discrimination and reversal in hippocampally-lesioned rats. *Physiology and Behavior*, 3:625–630.

Sinden, J. D., Jarrard, L. E., and Gray, J. A. (1988). The effects of intrasubicular ibotenate on resistance to extinction after continuous or partial reinforcement. *Exp Brain Res*, 73(2):315–319.

Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1995). A Model of the Neural Basis of the Rat's Sense of Direction. In Tesauro, G., Touretzky, D. S., and Leen, T. K., editors, *Advances in Neural Information Processing Systems 7*, pages 173–180. MIT Press.

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.

Skaggs, W. E. and McNaughton, B. L. (1998). Spatial firing properties of hippcampal CA1 populations in an environment containing two visually identical regions. *Journal of Neuroscience*, 18(20):8455–8466.

Skaggs, W. E., McNaughton, B. L., Gothard, K. M., and Markus, E. J. (1993). An information theoretic approach to deciphering the hippocampal code. In Hanson, S., Cowan, J., and Giles, L., editors, *Advances in Neural Information Processing Systems 5*, pages 1030–1037. Morgan Kaufmann, San Mateo, CA.

Smith, D. M. and Mizumori, S. J. Y. (2006). Learning-related development of context-specific neuronal responses to places and events: the hippocampal role in context processing. *J Neurosci*, 26(12):3154–3163.

Smith, R., Self, M., and Cheeseman, P. (1990). Estimating uncertain spatial relationships in robotics. In Cox, I. and Wilfong, G., editors, *Autonomous Robot Vehicles*, New York. Springer.

Solomon, P. R. and Moore, J. W. (1975). Latent inhibition and stimulus generalization of the classically conditioned nictitating membrane response in rabbits (Oryctolagus cuniculus) following dorsal hippocampal ablation. *J Comp Physiol Psychol*, 89(10):1192–1203.

Solomon, P. R., Schaaf, E. R. V., Thompson, R. F., and Weisz, D. J. (1986). Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behav Neurosci*, 100(5):729–744.

Song, P. and Wang, X. (2005). Angular path integration by moving 'hill of activity': A spiking neuron model without recurrent excitation of the head direction system. *Journal of Neuroscience*, 25(4):1002–1014.

Stackman, R. W., Clark, A. S., and Taube, J. S. (2002). Hippocampal spatial representations require vestibular input. *Hippocampus*, 12:291–303.

Stackman, R. W., Golob, E. J., Gassett, J. P., and Taube, J. S. (2003). Passive transport disrupts directional path integration by rat head direction cells. *Journal of Neurophysiology*, 90:2862–2874.

Stackman, R. W. and Taube, J. S. (1997). Firing properties of head direction cells in the rat anterior thalamic nucleus: Dependence upon vestibular input. *Journal of Neuroscience*, 17(11):4349–4358.

Stackman, R. W. and Taube, J. S. (1998). Firing properties of rat lateral mammillary single units: Head direction, head pitch, and angular head velocity. *Journal of Neuroscience*, 18(21):9020–9037.

Steele, R. J. and Morris, R. G. (1999). Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP5. *Hippocampus*, 9(2):118–136.

Tabuchi, E., Mulder, A. B., and Wiener, S. I. (2003). Reward value invariant place responses and reward site associated activity in hippocampal neurons of behaving rats. *Hippocampus*, 13(1):117–132.

Tai, C. T., Clark, A. J., Feldon, J., and Rawlins, J. N. (1991). Electrolytic lesions of the nucleus accumbens in rats which abolish the PREE enhance the locomotor response to amphetamine. *Exp Brain Res*, 86(2):333–340.

Tang, Y.-P., Wang, H., Feng, R., Kyin, M., and Tsien, J. Z. (2001). Differential effects of enrichment on learning and memory function in NR2B transgenic mice. *Neuropharmacology*, 41:779–790.

Tanila, H., Shapiro, M. L., and Eichenbaum, H. (1997). Discordance of spatial representation in ensmbles of hippocampal place cells. *Hippocampus*, 7(6):613–623.

Taube, J. S. (1995). Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *Journal of Neuroscience*, 15(1):1953–1971.

Taube, J. S. and Burton, H. L. (1995). Head direction cell activity monitored in a novel environment and during a cue conflict situation. *Journal of Neurophysiology*, 74(5):1953–1971.

Taube, J. S., Klesslak, J. P., and Cotman, C. W. (1992). Lesions of the rat postsubiculum impair performance on spatial tasks. *Behavioral and Neural Biology*, 5:131–143.

Taube, J. S., Muller, R. I., and Ranck, Jr., J. B. (1990a). Head direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience*, 10:420–435.

Taube, J. S., Muller, R. I., and Ranck, Jr., J. B. (1990b). Head direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *Journal of Neuroscience*, 10:436–447.

Taube, J. S. and Muller, R. U. (1998). Comparisons of head direction cell activity in the postsubiculum and anterior thalamus of freely moving rats. *Hippocampus*, 8(2):87–108.

Tenenbaum, J. B. and Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behav Brain Sci*, 24(4):629–40; discussion 652–791.

Thompson, L. T. and Best, P. J. (1989). Place cells and silent cells in the hippocampus of freely-behaving rats. *Journal of Neuroscience*, 9(7):2382–2390.

Thompson, L. T. and Best, P. J. (1990). Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats. *Brain Research*, 509(2):299–308.

Thue, A. (1892). Om nogle geometrisk taltheoretiske theoremer. *Forandlingerneved de Skandinaviske Naturforskeres*, 14:352–353.

Thue, A. (1910). Über die dichteste zasammenstellung von kongruenten kreisen in der ebene. *Christinia Vid Selk Skr*, 1:1–9.

Tolman, E. C. (1948). Cognitive Maps in Rats and Men. *Psychological Review*, 55:189–208.

Toni, N., Buchs, P. A., Nikonenko, I., Bron, C. R., and Muller, D. (1999). LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature*, 402(6760):421–425.

Touretzky, D. S. (2005). Path integrator contributions to hippocampal map formation. *Neurocomputing*, 65–66:291–296.

Touretzky, D. S. and Muller, R. U. (2006). Place field dissociation and multiple maps in hippocampus. *Neurocomputing*, 69:1260–1263.

Touretzky, D. S. and Redish, A. D. (1996). A theory of rodent navigation based on interacting representations of space. *Hippocampus*, 6(3):247–270.

Touretzky, D. S., Redish, A. D., and Wan, H. S. (1993). Neural representation of space using sinusoidal arrays. *Neural Computation*, 5(6):869–884.

Touretzky, D. S., Weisman, W. E., Fuhs, M. C., Skaggs, W. E., Fenton, A. A., and Muller, R. U. (2005). Deforming the hippocampal map. *Hippocampus*, 15(1):41–55.

Treves, A. and Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, 2:189–199.

Treves, A. and Rolls, E. T. (1994). Computational analysis of the role of hippocampus in memory. *Hippocampus*, 4(3):374–391.

Tsien, J. Z., Huerta, P. T., and Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor dependent synaptic plasticity in spatial memory. *Cell*, 87:1327–1338.

Tsodyks, M. (1999). Attractor Neural Network Models of Spatial Maps in Hippocampus. *Hippocampus*, 9:481–489.

Van Elzakker, M., O'Reilly, R. C., and Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. I. An empirical analysis. *Hippocampus*, 13(3):334–340.

Vazdarjanova, A. and Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *J Neurosci*, 24(29):6489–6496.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13:260–269.

Wallace, A. R. (1873a). Inherited Feeling. *Nature*, 7:303.

Wallace, A. R. (1873b). Perception and instinct in the lower animals. *Nature*, 8:65–66.

Wallenstein, G. V., Eichenbaum, H., and Hasselmo, M. E. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neuroscience*, 21(8):317–323.

Wallenstein, G. V. and Hasselmo, M. E. (1998). GABAergic modulation of hippocampal population activity: sequence learning, place field development, and the phase precession effect. *Journal of Neurophysiology*, 78:393–408.

Webster, D. B. and Voneida, T. J. (1964). Learning deficits following hippocampal lesions in split-brain cats. *Exp Neurol*, 10:170–182.

Whishaw, I. Q., Cassel, J.-C., and Jarrard, L. E. (1995). Rats with Fimbria-Fornix lesions display a place response in a swimming pool: A dissociation between getting there and knowing where. *Journal of Neuroscience*, 15(8):5779–5788.

Whishaw, I. Q. and Jarrard, L. E. (1996). Evidence for extrahippocampal involvement in place learning and hippocampal involvement in path integration. *Hippocampus*, 6:513–524.

Wiener, S. I. (1993). Spatial and behavioral correlates of striatal neurons in rats performing a self-initiated navigation task. *Journal of Neuroscience*, 13(9):3802–3817.

Wiener, S. I. and Taube, J. S. (2005). *Head direction cells and the neural basis of spatial orientation*. MIT, Cambridge, MA.

Wike, E. L. (1953). Extinction of a partially and continuously reinforced response with and without a rewarded alternative. *J Exp Psychol*, 46(4):255–260.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N., and O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308:873–876.

Wilson, A., Brooks, D. C., and Bouton, M. E. (1995). The role of the rat hippocampal system in several effects of context in extinction. *Behav Neurosci*, 109(5):828–836.

Wilson, I. A., Ikonen, S., Gallagher, M., Eichenbaum, H., and Tanila, H. (2005). Age-associated alterations of hippocampal place cells are subregion specific. *J Neurosci*, 25(29):6877–6886.

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

Wilson, M. A. and McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265:676–679.

Winocur, G. and Olds, J. (1978). Effects of context manipulation on memory and reversal learning in rats with hippocampal lesions. *J Comp Physiol Psychol*, 92(2):312–321.

Wise, L. M. (1962). Supplementary report: the Weinstock partial reinforcement effect and habit reversal. *J Exp Psychol*, 64:647–648.

Wittenberg, G. M., Sullivan, M. R., and Tsien, J. Z. (2002). Synaptic reentry reinforcement based network model for long-term memory consolidation. *Hippocampus*, 12(5):637–647.

Witter, M. P. and Amaral, D. G. (2004). Hippocampal Formation. In Paxinos, G. T., editor, *The Rat Nervous System*, chapter 21, pages 635–704. Academic Press, San Diego, CA, 3 edition.

Witter, M. P., Wouterlood, F. G., Naber, P. A., and Haeften, T. V. (2000). Anatomical organization of the parahippocampal-hippocampal network. *Annals of the New York Academy of Sciences*, 911:1–24.

Wong, R. O. (1999). Retinal waves and visual system development. *Annu Rev Neurosci*, 22:29–47.

Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, 397:613–616.

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., and Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27:623–633.

Xavier, G. F., Oliveira-Filho, F. J. B., and Santos, A. M. G. (1999). Dentate gyrus-selective colchicine lesion and disruption of performance in spatial tasks: difficulties in 'place strategy' because of a lack of flexibility in the use of environmental cues. *Hippocampus*, 9:668–681.

Young, S. L., Bohenek, D. L., and Fanselow, M. S. (1994). NMDA processes mediate anterograde amnesia of contextual fear conditioning induced by hippocampal damage: immunization against amnesia by context preexposure. *Behav Neurosci*, 108(1):19–29.

Yu, A. J. and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4):681–692.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6):2112–2126.

Zhang, K., Ginzburg, I., McNaughton, B. L., and Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J Neurophysiol*, 79(2):1017–1044.

Zinyuk, L., Kubik, S., Kaminsky, Y., Fenton, A. A., and Bures, J. (2000). Understanding hippocampal activity by using purposeful behavior: Place navigation induces place cell discharge in both task-relevant and task-irrelevant spatial reference frames. *Proceedings of the National Academy of Sciences, USA*, 97(7):3771–3776.

COLOPHON

This thesis was typeset with LaTeX using the Lyx graphical front-end. The thesis appearance is based on André Miede's elegant Classic Thesis style, heavily adapted to be compatible with Lyx and to suit scientific publication standards. The document text was typeset in Hermann Zapf's Palatino and Euler type faces.